Research article

Gauging queryable iterative estimator uncorrelatedness from incompatibilistic propagational Poissionian noise in eigen-normalized non-negativity constraints employing analogs of the Pythagorean theorem and parallelogram laws in sub-meter resolution pseudo-Euclidean space in C++ for semi-parametrically prognosticating synergistic semi-logarithmic *Aedes aegypti* non-ordinate axis-scaled landscape weightage covariances of episodical sylvatic yellow fever case distributions for an agro-irrigated riceland village ecosystem in Gulu, Uganda

BENJAMIN G. JACOB and ROBERT J. NOVAK

1- Department of Global Health, College of Public Health, University of South Florida, Tampa, FL.

Corresponding Author *and co-authors information:
E-mail: bjacob1@health.usf.edu

RUNNING HEAD: C++ models for predicting geolocations of Yellow Fever

Abstract: C++ is a programming language which has imperative, object-oriented and generic programming features and facilities for low-level memory manipulation. C++ has also been found useful in many other contexts, with key strengths being software infrastructure and resource-constrained applications, including desktop applications, servers (e.g., e-commerce, web search or SQL servers), and performance-critical applications (e.g., space probes). A C++ model was constructed to forecast, yellow fever (YF), case distribution in an agro-irrigation,

riceland, village ecosystem, eco-epidemiological, study site in Gulu, Uganda. Initially, a panchromatic, grid-stratified, QuickBird 0.61m, spatial resolution, uncoalesced endmember dataset of transmittance, land use land cover, (LULC), spectral signatures of the study site was differentitated by visible and near-infrared, (NIR) wavelength, irradiance frequencies. Next, an object-based classification algorithm employed a divergence measure to match mixed pixels ("mixels") to the endmember, LULC, derivative spectra. The spectrum of a selected mixel was calculated as a linear combination of the endmember LULC spectra weighted by the area coverage of each endmember within the mixel if the scattering and absorption of electromagnetic radiation was dominated by a single component on the capture point, habitat surface. A mosaiciked dataset was then employed to manage, display, serve, and distribute, raster, time series, intermittent, YF virus, mosquito, vector, *Aedes aegypti,* oviposition, capture point, immature, habitat geolocations along the riceland, peripheral, forest-canopied, interface corridor. A new mosaiciked dataset was created as an empty container in an ArcGIS geodatabase cyberenvironment with default properties which was subsequently added to the raster dataset. Thereafter, the YF virus data was cartographically illustratable as non Fractional abundances of non-mixelated, endmember riceland, agro-village, non-homogeneous, LULC, geoclassifiable eco-zones (e.g., sparsely shaded, pre-flooded, densely canopied, post-tillering) within a supervised classification matrix. Non-normalized, geometric, seasonal, YF, case distribution data was articulated from various Bayesian perpsectives. The models took inputs from table file in the geodatabase, which subsequently imput information of yearly population and YF cases from 1990 to 2012. Log-likelihood functions were also generated in PROC REG for each geosampled, YF virus, mosquito vector (i.e., *Ae.aegypti*) endmember, LULC-specified, visible and NIR, wavelength regressor employing $\log(f(y_i|\mu,\sigma)) = \log(\phi(y_i;\mu,\text{var}=\sigma^2))$ whilst the joint log-likelihood function was exponentially logrithmically quantitated by $\log(f(y|\mu,\sigma))$.

Consequently, the model output rendered $\sum_i \log(\phi(y_i;\mu,\text{var}=\sigma^2))$ and

$$\lim_{n\to\infty}\ln p_n \prod_{k=1}^{n}\frac{1}{1+\frac{1}{p_k}} = \lim_{n\to\infty}\ln p_n \prod_{k=1}^{n}\frac{\frac{1}{1-\frac{1}{p_k}}}{\frac{1}{1-\frac{1}{p_k}}} = \frac{\prod_{k=1}^{\infty}\frac{1}{1-\frac{1}{p_k^2}}}{\lim_{n\to\infty}\frac{1}{\ln p_n}\prod_{k=1}^{t}\frac{1}{1-\frac{1}{p_k}}} = \frac{\zeta(2)}{e^\gamma} = \frac{\pi^2}{6e^\gamma}$$

$=.0915$. The forecasts, (ε_t), were positively correlated with a relative change in YF cases, where a regression line was denoted as $\varepsilon_t = 1.85\log\left(\frac{y_t}{y_{t-1}}\right)+0.22$, which revealed a pseudo R^2 of 0.93. To uncorrelate, noisy, LULC, oviposition residuals for deriving meaningful, unbiased, non-linear, computable forces in the YF model, a sequential iterative algorithm was employed in AUTOREG. A second-order, pertubation algorithm investigated the error and sensivity of the forecasts. Specified orthogonal matrices updated a subset of endmember synthetic functions and their renderings. A log-determinant Jacobian term orthogonally eigen-decomposed oviposition, LULC gridded data which ensured that the integral of the model's likelihood function equaled 1 when integrated over the densities of all unknown, endmember, signature, frequency estimators. The model assumption that the post-unspecified, marginally monotone, semi-parameterically, supplemental, sub-meter resolution, uncoalesced, wavelength, irradiance distributions were multivariate Gaussian was violated. Generalized autoregressive moving average (GARMA) models were extended to autoregressive, integrated, moving average, (SARIMA) models in PROC ARIMA for optimally conducting, observation-driven, sub-mixel, signature modelling of the non-Gaussian, non-stationary, YF-related,time series, LULC count variables. A negative binomial, mean model fit

[i.e., $\text{GSARIMA}(3',1,0) \times (1,1,0)_{12}$] was compared to that of a Bayesian Gaussian fit [i.e, $\text{SARIMA}(3',1,0) \times (1,1,0)_{12,}$] employing Box-Cox transformed YF data. For the Gaussian [i.e., $\text{SARIMA}(3',1,0) \times (1,1,0)_{12}$], the posterior distributions appeared to be platykurtic. The logarithm of the posterior density was calculable as follows $\log(p(\theta|y))$ which was equivalent to $\log(\pi(\theta)) + \sum_{i=1}^{n} \log(f(y_i|\theta))$ where θ was a vector of the geometric, seasonal, endogenous regressors. Geospatial outliers were teased out in the residual plots. The normalized modes were validated employing a Monte Carlo simulation in PROC MCMC. Subsequently, the time series YF data was eco-epidemiologically forecasted in C++ employing a stochastic-dynamic, random, weighted matrix. Homoscedastic residuals were cross-tabulated via various hierarchies of continuous or categorical, finite, seasonal, transitional, rice agro-irriagted LULC endmembers. The model included, meteorological, values. The final model tabulation revealed prognosticated population growth and number of case occurrences to 2020 in C++ for the Gulu study site. C ++ language may have implementations for constructing real-time, geometric, endmember, YF virus, forecasting cyber-platforms including the Free Software Foundation Low Level Virtual Machine Microsoft, Intel and IBM. In so doing, cartographic and geostatistical descriptors of geo-spectrotemporally, fractionalized, seasonal, hyperproductive, sylvatic, *Ae aegypti*, ,oviposition, sub-meter resolution, newly transitioned inhomogeneously forest, canopied, to rice-agro-iririgated, LULC abudance may be elucidated for implementing larval control strategies in expanding agro-village, African, complex due to anthropogenic pressure for parsimonously forecasting YF case distributions.

Keywords: Yellow fever virus, ArcGIS, time series, C++, Gulu, Uganda.

Introduction

Due to increasing anthropogenic population growth in African, riceland, agro-village complexes, seasonal, transitional, landscape shifts from inhomogeneous, forest-canopied, land use land cover (LULC) to agro-irrigation LULCs along the periphery of these ecosystems is common. Increased interface between anthropogenic population and intermittent, forest-canopied, discontinuous LULCs may facilitate sylvatic, yellow fever (YF), endemic transmission by increasing exposure of farm workers and their families to prolific, (i.e., seasonal hyperproductive), larval, mosquito, vector habitats of *Aedes aegypti*, (Linnaeus). Yellow fever is a viral hemorrhagic fever transmitted by mosquitoes infected with the YF virus (YFV) whose case fatality rates in severe cases can exceed 50% (www.who.gov). *Aedes aegypti* is the primary vector of YF (http://entnemdept.ufl.edu/). The virus could hence be transmitted via *Ae. aegypti* to humans whenst farmers and their families encroach into forest-canopied LULCs along African riceland, agro-village, peripheral, discontinuous corridors during occupational or recreational activities *Ae. aegypti* is the most efficient vector for arboviruses because it is highly anthropophilic, frequently bites, and thrives in close proximity to humans [WHO, 2009]. An intermediate (savannah) cycle exists that involves transmission of YF virus from mosquitoes to humans living or working in jungle border areas (http://www.cdc.gov).

Literature has revealed a lithany of contributions of anthropogenic-induced, seasonal, rice-cycle, LULC changes (e.g., pre-flooded to post-tillering) in expanding African, agro-irrigated ecosystems contributing to aquatic, vector, larval habitats of malaria, mosquito, *Anopheles arabiensis* (Mwangangi et al. 2008, Muturi et al. 2007, Jacob et al. 2007). However, there are substantially less contributions to the potential risks associated with seasonal, endemic, YF, transmission on immature, sylvatic, *Ae aegypti,* hyperproductive foci along sparsely, forest-canopied, agro-irrigated, African, riceland, capture point, peripheral, LULC, oviposition sites.

Amongst the limited contribution, Ricardo Lourenço-de-Oliveira (2002) examined the oral susceptibility to YFV of 23 Brazilian samples of *Ae. aegypti* which was evaluated on areas of LULC change induced by anthropogenic pressure. For comparative purposes, six additional samples from Southeast Asia, West Africa and North and South America were tested. Mosquito samples were collected by employing ovitraps settled around human dwellings, from March to December 2001, except for the West Palm Beach, Florida, Ho Chi Minh Vietnam, Phnom Penh, Cambodia and Boulbinet, Guinea samples which were collected either as larvae or pupae. Field collected mosquitoes (F_0) raised to adult in the laboratory were morphologically identified to species and fed on guinea pigs to obtain eggs. The procedures to maintain and infect mosquitoes followed those described by Vazeille-Falcoz et al. (1999).

A total of 2,487 *Ae. aegypti* from Brazil were tested. Except for *Ae. aegypti* collected in Milhã, where 148 tested females in areas of geocalssified LULCs were negative, all other evaluated samples from Brazil were orally susceptible to YFV. However, the infection rates were extremely heterogeneous when all Brazilian samples were considered ($P < 0.05$). In the great majority of the Brazilian *Ae. aegypti* samples (14 out of 23) infection rates were low, (i.e., under 13%). This was the case of most of *Ae. aegypti*, immature, habitat samples from Northeast (Pacujá, Quixeramobim, Salvador, Feira de Santana) and Southeast regions (Cariacica, Leandro Ferreira and five samples from the State of Rio de Janeiro). When only samples from the endemic area of sylvatic, YFV, transmission were taken into account at the eco-epidemiological study sites, the infection rates were higher, although heterogeneous ($P < 0.05$), ranging from 11.1% in Rio Branco to 46.4% in Ananindeua in the LULC change sites. The highest infection rate for *Ae. aegypti* was observed for Foz do Iguaçu (48.6%), which was geographically located (henceforth geolocated) in an LULC transition area of sylvatic, endemic, YFV transmission in Brazil. The Asian samples were more susceptible (47%-64.4%) than those from Brazil (0-48.6%), Venezuela (13.6%) and USA (24.8%). Although the authors were able to regress some discrete, integer, count values representing geosampled, clinical, diagnostic, field-operational data on remote-specified, oviposition, LULC data and generate some eco-epidemiological statistics, the research failed to optimally quantitate circumstancial cases where the pathogen may not have reached all susceptible population within gridded, geoclassified, stratified, LULC areas at the eco-epidemiological study sites; thus, assuming geomorphological homogeneity (e.g., continuous canopy cover). In has been recognized through multiple contributions to literature (Mwagangi et al. 2008, Jacob et al. 2007, Muturi et al. 2007, Gu and Novak 2005), that grid-stratifiable, vector, arthropod, eco-epidemiological, immature, LULC, capture point, habitat data display extreme, geospatial, endogeneity and topological heterogeneity (eco-georeferenceable, geosampled, seasonal, oviposition, capture point, depth levels), especially whenst regressively, forecasting, sub-meter resolution, geoclassifiable, geo-spectrotemporal, oviposition, sub-meter resolution, LULCs based on semi-parametric, iterable, interpolative,

eigen-decomposable, endmember, signature covariates. Hence, the data collections were an incomplete reflection of eco-regions where newly transitioned LULCs were optimally suitable for *Ae. aegypti* and host presence.

Spatial heterogeneity in an entomological, eco-epidemiological, vector, arthropod, endmember, time series, grid-stratifiable, sub-meter resolution, LULC model can refer to the variation or instability in observational units (e.g., gridded, eco-georeferenecable discontinuously canopied, forestland, sylvatic, *Ae aegypti,* oviposition, capture points) across an eco-geographic region (agro-irrigated African floodplain) which may imply that the functional forms and/or behavioral parameters vary by geolocation. The adaptation for oviposition preference may have been part of the overall evolution of domesticity that likely occurred in North Africa when ancestral sylvan *Ae. aegypti*. became isolated from sub-Saharan Africa due to the Sahara Desert (Tabachnick 1991). In general oviposition choice in mosquitoes is largely due to volatiles produced by the microorganisms in the larval (Yelfwagash et al. 2017,). Thus, as long as appropriate volatiles are produced by a standing forested, sparasely canopied pool of water, resding next to an expandinf African agro-irrigated riceland corridor an opportunistic species like *Ae. aegypti* may oviposit there.This is supported by situations where this domestic form outside Africa has reverted to developing in natural water. This has occurred mostly on islands or other isolated sites. Chadee et al. (1998) report 12 types of natural habitats where *Ae.aegypti* can be found in Jamaica, Puerto Rico and Trinidad including rock holes, tree holes, leaf axils, bamboo joints and coconut shells. Larvae developing in rock holes has been documented on the east coast of Africa (Trpis 1972) and in Anguilla (Wallis & Tabachnick 1990). *Ae.aegypti* has been observed ovipositing in tree holes in New Orleans [cited in Wallis and Tabachnick (1990)].

Geographically weighted regression (GWR) may be employable to characterize *Ae. aegypti*, oviposition, sub-meter resolution, geoclassifiable, immature habitat variations, by approximatimg parameter estimators for each LULC site or observational unit (e.g., forest-canopied, capture point along an expanding African, riceland corridor) based on all observations within a neighborhood (optimally pre-determined using cross validation), as described in Fotheringham (2003). Ghosh et al. (2008) analyzed impervious (or heavily developed) LULC shares via a continuous-response GWR model, using data points across Minnesota's Twin Cities metro area. Páez (2006) calibrated a binomial probit GWR model with heteroscedastic error terms to characterize the development of 324 vacant 1-hectare grid cells near California's Bay Area Rapid Transit lines. Applications of multinomial GWR models for land use patterns can be found in McMillan and McDonald (1999) and Wang et al. (2011). The second type of spatial effect, spatial autocorrelation spillover may arise primarily due to imperfect information on observational, inhomogeneous, forest-canopied, *Ae. aegypti*, oviposition, capture point, eco-epidemiological, grid-stratified, sub-meter resolution, peripheral African riceland, LULC units and measurement errors. For example, information on endmember, LULC variables like soil types, may be regularly absent from *Ae. aegypti* models of African agro-irrigated land use and change, resulting in correlations across nearby discontinuous, forest-canopied sites' along with their geoclassified LULC error terms. Moreover, geo-aggregated spatial entomological, data values (such as geo-spectrotemporally geosampled, LULC, ovipoition, capture point, count density values) and arbitrary spatial boundaries (such as Census tract designations) can introduce

forms of spatial autocorrelation into an agro-irrigated, African, riceland, sub-meter reosolution eco-epidemiological, oviposition, forecast, vulnerability, study site model especially whenst quantitating probability of endemic, YF transmission based on seasonal LULC change due to anthropogenic pressure. A common treatment for this effect may be to directly specify a spatial structure, such as a spatial autoregressive (SAR) or spatial moving average (SMA) models (see Anselin 1988, Anselin and Hudak 1992, LeSage and Pace 2009). Work on discrete states of land use change with such specifications can be found in Chakir and Parent's (2009) spatial multinomial probit model (for cross-sectional data), Munroe et al.'s (2002) series of binary probit and random effects probit models (using panel techniques), and Wang and Kockelman's (2009a, 2009b, 2009c) dynamic spatial ordered probit model with a temporal component

In statistics, a probit model is a type of regression where the dependent variable can only take two values. The purpose of the model is to estimate the probability that an observation (e.g., geo-spectrotemporally geosampled, eco-georeferenceable, seasonal hyperproductive, *Ae. aegypti* capture point,) with particular characteristics (e.g., inhomogeneous forest-canopied LULC) will fall into a specific regression category. (e.g., a binary classification model).A probit model is a popular specification for an ordinal or a binary response model (Bliss 1935);

In time series, vector, arthropod-related, eco-epidemiological, oviposition, capture point, signature LULC, endmember, analyses, binomial regression is a technique in which the response (often referred to as Y) is the result of a series of Bernoulli trials, or a series of one of two possible disjoint outcomes (traditionally denoted "success" or 1, and "failure" or 0) (Jacob et al. 2009, Griffith 2005). In the theory of probability and statistics, a Bernoulli trial (or binomial trial) is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success is the same every time the experiment is conducted (Papoulis 1984). In binomial regression, the probability of a success is related to explanatory variables: the corresponding concept in ordinary regression is to relate the mean value of the unobserved response to explanatory variables (Hosmer and Lemeshew 2002). Binomial regression models in vector arthropod entomological science are essentially the same as binary choice models, one type of discrete choice model.

Discrete choice models, or qualitative choice models can describe, explain, or predict choices between two or more discrete alternatives, such as probability of endemic YF transmission if a farm worker residing an expanding, African, agro-ecosystem, eco-epidemiological, irrigated, study site enters or not enters an inhomogeneous, forest-canopied, peripheral, sub-meter resolution, geoclassifiable LULC. Such choices contrast with standard consumption models in econometrics in which the quantity of each good consumed is assumed to be a continuous variable. In the continuous case, calculus methods (e.g., first-order LULC conditions) may be optimally employable to determine the probability of an endemic, YF, transmission based on the times an individual enters a "hot spot" zones (area of positively autocorrelated, *Ae. aegypti,* forest, discontinuous, capture points along a riceland perphery) which may be modeled empirically using regression analysis. Alternatively, a discrete choice, sub-meter resolution, geometric, grid-stratifiable, sub-meter resolution, endmember, LULC analysis may examine situations in which the potential outcomes are discrete, such that the optimum is not characterized by standard first-order conditions. Thus, instead of examining "how much the probability of endemic YF transmission" could occur in an expanding African

floodplain due to anthropogenic pressure employing continuous choice LULC variables, discrete choice analysis would examine which behavior response of the individual residing in the agro-ecosystem may cause a higher probability of an endemic YF transmission event to occur. However, a discrete choice geo-spectrotemporal endmember, sub-meter resolution, eco-epidemiological, LULC analysis can also be used to examine the chosen quantity when only a few distinct quantities, such as the number of seasonal, hyperproductive, oviposition, sub-meter resolution, grid-stratifiable, LULC sites along inhomogeneous, forest-canopied. peripherial riceland corridors of an expanding African, agro-ecosystem are avialable. Techniques such as logistic regression and probit regression can be employed for endmember LULC empirical analysis of discrete choice (Griffith 2003). Discrete choice models specify the probability that an individual chooses an option among a set of alternatives (Hosmer and Lemeshew 2000). The endmember LULC description of discrete choice behavior in an *Ae. egypti*, forecast vulnerability model may thus be reflected by the individual behavior of a farm worker or other African riceland resident to enter or not enter into a discontinuous, forest-canopied, grid-straifiable geolocation along the peripherial corridor of the agro-irrigated eco-georferenecable ecosystem which may be quantiatable as intrisically probabilistic event for endemic YF transmission.

In practice, no one individual can know all co-factors affecting endemic, YF, transmission in an African expanding riceland, agro-ecoystem nor their determinants which may be partially observed or imperfectly measured. Therefore, entomological time series, discrete choice models for prognosticating YF endemic transmission in farm workers and other residents in expanding, African, riceland, agro-irrigated environments may rely on stochastic assumptions and specifications to account for unobserved LULC factors related to a) choice alternatives, b) taste variation over people (interpersonal heterogeneity) and over time (intra-individual choice dynamics), and c) heterogeneous choice sets. The different formulations may be been summarized and classified into groups for modeling endemic probability of YF transmission.

Regardless, the primary difference in any geo-spectrotemporal or geo-spatiotemporal, forecast, capture point, *Ae aegypti* oviposition, capture point, eco-epidemiological, endmember endemic, transmission-oriented, risk model is in the theoretical motivation: Discrete LULC choice models are motivated using utility theory so as to handle various types of correlated and uncorrelated seasonal, time series, covariate coefficient choices, while binomial regression models are generally described in terms of the generalized linear model.

In statistics, the generalized linear model (GLM) is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement (e.g., eco-georeferenced, seasonal, hypeproductive, *Ae. aegypti*, oviposition site on a sub-meter resolution, grid-stratifiable, inhomogeneous forest-canopied LULC) to be a function of its prognosticated value .Generalized linear models in vector entomological, time series endmember, oviposition, eco-epidemiological forecast, vulnerability models were formulated as a way of unifying various other statistical models,

including linear regression, logistic regression and Poisson regression (Jacob et al. 2005, Griffith 2005) These authors propose an iteratively reweighted least squares method for maximum likelihood estimation of field remote and clinically specified diagnostic LULC, vector arthropod-related model parameters. Maximum-likelihood estimation remains popular in vector ecology and is the default method on many statistical computing packages. These sub-modeling endmember, techniques extends beyond the exponential-family-type generalized linear models to other distributions, to non-linear parameterizations, and to dependent observations. Various criteria for estimation other than maximum likelihood, including resistant alternatives, may be used. The algorithms are generally numerically stable, easily programmed withouthe aid of packages, and highly suited to interactive computation. Other approaches, including iterative Bayesian approaches and least squares fits to variance stabilized responses, have been developed for various vector arthropods ovipsoition, eco-epidemiolgical, predictive risk modeling

Henceforth, geo-spectrotemporally geosampled entomological eco-epidemiological, LULC, capture point, *Ae aegypti,*sub-meter resolution, grid-stratifiable discrete choice, forecast, vulnerability models may be described primarily with a latent variable indicating the "utility" of making a choice, and with randomness introduced through an error geosampled habitat variable distributed according to a specific probability distribution. statistics, latent variables as opposed to observable variables), are variables that are not directly observed but are rather inferred (through a mathematical model) from other variables that are observed (directly measured).

Endogeneity is often described in geo-spatiotemporal, vector arthropod, eco-epidemiological, geo-spectrotemporal, geosampled, endmember, geoclassifiable, LULC, forecast, vulnerability, eco-georerenecable, capture point, signature model as having three sources: 1)omitted density, count variables, 2) habitat wavelength, frequency dimension measurement error;and,3) simultaneity (Jacob et al. 2012, Jacob et al. 2009) Though it is often helpful to mention these "sources" separately, confusion sometimes arises because they are not truly distinct. Imagine a regression forecasting the effect of an empirical, geo-spectrotemporally geosampled, sub-meter resolution, grid-stratified, capture point, oviposition LULC dataset of uncoalesced, sub-meter resolution, 3-dimensional, immature, explicative, African , riceland, agro-irrigated, *Ae, aegypti*, catchment watershed, endmember, covariates based on seasonal, immature, habitat, larval, density count. Perhaps the measure of eco-georeferenceable, forecast, vulnerability, YF, endmember, LULC, 3-Dimensional (D) model, slope coefficients based or the number of geo-spectrotemporally geosampled, seasonal hyperproductive, *Ae, aegypti*, immature,capture points in an agro-irrigated, African floodplain may be pertinent, diagnostic, parameterizable,sub-meter resolution, grid-stratifiable covariate estimators regardless of the time of the rice-cycle. Additionally, if an arbovirologist, medical entomologist or YF experimenter has an idea of what type of contour lines may create discontinuous, floodplain, surface, runoff statistics, then he or she may be able to optimally tabulate a propogagtional measurement error in the eco-epidemiological, capture point, oviposition, *Ae aegypti*, immature habitat, unmixed, explanatory, diagnostic, sub-mixel, regression-related, capture point, wavelength, transmittance, LULC frequencies.

Alternatively, a arbrovirologist, medical entomologist or a YF experimenter could describe the situation as an omitted variables problem for approximating uncertaintyiteratively interpolative , endmember,sub-meter resolution, wavelength, frequency, signature, LULC estimates. If monthly precipitation evapotranspiration. and elevation coefficients are selected as explicative diagnostic, regressors in the geo-spectrotemporal, geosampled, YF, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, endmember, LULC, grid-stratified, model and are measured at the same time this would be example of simultaneity, but it too, might be reframed in terms of omitted variables.

Potential implications of precipitation evapotranspiration. and elevation variability along with LULC on stream flow have been assessed by Akopoti et al. (2016). in the Black Volta basin employing the Soil and Water Assessment Tool (SWAT) model. The spatio-temporal variability of rainfall over the Black Volta was assessed using the Mann-Kendall monotonic trend test and the Sen's slope for the period 1976–2011. The statistics of the trend test showed that 61.4% of the rain gauges presented an increased precipitation trend whereas the rest of the stations showed a decreased trend. However, the test performed at the 95% confidence interval level revealed that the detected trends in the rainfall data were not statistically significant. Land use trends between the year 2000 and 2013 revealed that within thirteen years, LULC classes like bare land, urban areas, water bodies, agricultural lands, deciduous forests and evergreen forests had increased respectively by 67.06%, 33.22%, 7.62%, 29.66%, 60.18%, and 38.38%. Only grass land decreased by 44.54% within this period. Changes in seasonal stream flow due to LULC were assessed by defining dry and wet seasons. The results showed that from year 2000 to year 2013, the dry season discharge has increased by 6% whereas the discharge of wet season has increased by 1%. The changes in stream flows components such us surface run-off lateral flow and ground water (GW) contribution to stream flow and also on evapotranspiration changes due to LULC was evaluated. The results showed that between the year 2000 and 2013, and have respectively increased by 27% and 19% while GW flow decreased by 6% while ET has increased by 4.59%. The resultant effects are that the water yield to stream flow has increased by 4%.

The climate and geophysical characteristics of the basin (such as topography, soil, vegetation and vegetation) as well as anthropogenic activities are the main factors that influence the ecohydrological biophysical processes in African, riceland, agro-irrigated, ecosystem basins. Thus, e LULC processes that influence the evolution of the surface of these basins (and hence any *Ae. aegypti*, immature, habitat, oviposition, capture point geo-spectrotemporally geosampled at the basin), will depend on the geo-spatiotemporal patterns of the precipitation and evaporation rates. The distribution of rainfall is one of the favorable factors that can influence the number of *Aedes aegypti* mosquito Amongst the remotely sensed, sub-meter resolution, LULC, eco-epidemiological models, geophysically based distributed models are well established models for analyzing the impact of land management practices on water, sediment, and agricultural chemical yields in large complex watersheds. Unfortuantely, the SWAT is a comprehensive, semi-distributed river basin model that requires a large number of input parameters, which can complicate model parameterization and calibration for optimally quantitating abundance of *Ae. aegypti* foci on grid-stratifiable, sub-meter resolution, expanding, endmember, geoclassified, LULC plots in African ricefields due to anthropogenic pressure.

A simulation-based approach may be employable for attributing the observed flooded LULC trends to the postulated drivers of seasonal, hyperproductive, *Ae. aegypti*, oviposition foci in African riceland environments. Aich et al. (2015) contributed to the ongoing discussion on whether LULC or climate trends have influence on the observed increase of flood magnitudes in the Sahel. An ecohydrological model Soil and Water Integrated Model (SWIM) with a new, dynamic LULC module was set up for the Sahelian part of the Niger River until Niamey, including the main tributaries Sirba and Goroul. The model was driven with observed, reanalyzed climate and LULC data for the years 1950–2009. In order to quantify the shares of influence, one simulation was carried out with constant land cover as of 1950, and one including LULC. As quantitative measure, the gradients of the simulated trends were compared to the observed trend. The modeling studies showed that for the Sirba River only the simulation which included LULC was able to reproduce the observed trend. The simulation without LULC showed a positive trend for flood magnitudes, but underestimated the trend significantly. For the Goroul River and the local flood of the Niger River at Niamey, the simulations were only partly able to reproduce the observed trend. The new endmember LULC module enabled some first quantitative insights into the relative influence of LULC and climatic changes. For the Sirba catchment, the results implied that LULC and climatic changes contributes in roughly equal shares to the observed increase in flooding.

Unfortunately for optimally prognosticating, seasonal, *Ae. aegypti* abundance and distribution in expanding African agro-irrigated environments parts of the subcatchment must be included,which SWIM cannot provide for any ecohydrological geosampled body. The SWIM model can only simulate ecohydrological LULC processes, vegetation, erosion and nutrient cycles at the catchment scale. A subcatchment is composed of hydrotopes, which are sets of elementary units with homogeneous soil and land use types(Jensen 2005). In each hydrotope, nutrient transport and transformation are simulated to model processes from the hydrological system to the river network. A subcatchment is used to model the runoff from a given area of land(www.esri.com). Optimally each subcatchment LULC should generate a runoff hydrograph, that typically routes into a downstream reach or iririgation pond in an African, riceland agro-village complex for quantitating *Ae. aegypti,* oviposition sites on geoclassifiable LULCs. The SWIM model results may be able to distinguish that climatic changes and LULC are drivers for the flood increase in African ricelands ; however their shares cannot be quantified for quantitating *Ae. aegypti* eco-epidemiological, capture points on newly transitioned inhomogeneous, forest-canopied LULC to agro-irrigation LULC along the periphery of these ecosystems.

The SWAT model is a complex semi-distributed process-based model. It was developed by the Agricultural Research Service of the United States Department of Agriculture and can model changes in hydrology processes, vegetation, erosion, and nutrient loadings at the sub-catchment scale. It divides the catchment into subcatchments and subsequently into Hydrologic Response Units (HRUs). Different combinations of land use, soil types and slope in each subcatchment can be represented by the HRUs. The processes related to water, sediment and nutrient transport for a geosampeld, geo-spectrotemporal, endmember, sub-meter resolution, endmember LULC, capture points may be modeled at the HRU scale. The hydrological processes may then be distributed in five compartments: the stream, the soil surface, the soil water layers, the shallow unconfined aquifer, and the deep confined aquifer. Up to ten soil layers

can be divided in SWAT(www.esri.com). Surface runoff can be calculated either by Soil Conservation Service-Curve Number (SCS-CN) method or Green and Ampt method. Erosion of the oviposition site may be estimated with the Modified Universal Soil Loss Equation (MUSLE). With the daily time step, SWAT simulates nutrient transports and transformations in soil profiles, river network, various water bodies (e.g., pond, lakes, and wetland), and the interaction processes between different systems. The SWAT model may also differentiate between nutrient fluxes from different ecohydrological, geoclassifiable, sub-meter resolution, LULC sources in expanding African riceland agro-village complexes that may influence *Ae. aegypti* abundance and distribution in these irrigated ecosystems

Althoug SWAT distributed models are capable of capturing the spatial distribution of input forecast, vulnerability, geo-spectrotemporally geosampled, *Ae aegypti* sub-meter resolution, endmember, LULC variables including metrological conditions (rainfll, temperature etc) and physical parameters (land use, soil, elevation etc). distributed models are data intensive they need quality data, hard to configure and they require greater simulation and calibration time(Jensen 2005). The performance of these models is quite low in hydrological remote area (data scare regions). Semi-distributed models, lump metrological variables and biophysical parameters in sub-basins,. Hence, parameter calibration of the geosampled, *Ae aegypti*, oviposition, LULC etsimators may not reveal all capture point heterogeneous variables in the model.

Unobserved heterogeneity is simply variation/differences (e.g.,seasonal, endmember, LULC, African, riceland, *Ae aegypti*, oviposition sites along discontinuous forest canopied, peripheral ecosystem geoclocations) amongst cases which are not measured (Cressie 1993). If an arbovirologist, medical entomologist or YF experimenter understands endogeneity such a sub-meter resolution endmember, LULC, eco-epidemiological, forecast, vulnerability model, the implications of unobserved heterogeneity in a regression context may be definable. For example, let a be an empirical geo-spectrotemporal, endmember geosampled dataset of immature habitat returns for a seasonal, hypeproductive, geosampled, *Ae aegypti*, African, oviposition, LULC, time series be defined as $\beta+b_i\beta+b_i$ with $E(b_i)=0 E(b_i)=0$ in PROC REG. In so doing $y_i=x_i(\beta+b_i)+w'_i\gamma+\epsilon_i, y_i=x_i(\beta+b_i)+w_i'\gamma+\epsilon_i$ may optimally quantitate seasonal, hypeproductive, capture points along the ecosystem peripheral, geoclassified LULCs whenst $y_i y_i$ is a typically, log- transformed density count rate, $x_i x_i$ is the sample frame time and $w_i w_i$ is a set of other controls. An example of endogeneity then may be quantitated along with noisy LULC variables when $x_i x_i$ is correlated with $\epsilon_i \epsilon_i$ (e.g., precipitation diffusion floodplain rates is correlated with number of agro-irrigation ditches, which is not amongst the other prognosticators). If an arbovirologist, medical entomologist or YF experimenter estimates a single geosampled, *Ae aegypti,* endmember, sub-meter resolution, oviposition, LULC coefficient $y_i=x_i\beta+w'_i\gamma+(\epsilon_i+b x_i)=x_i\beta+w'_i\gamma+\epsilon\sim_i y_i=x_i\beta+w_i'\gamma+(\epsilon_i+b x_i)=x_i\beta+w_i'\gamma+\epsilon\sim I$, then the included variable $x_i x_i$ may be correlated with the error term $\epsilon\sim_i \epsilon\sim_i$, inducing the same problems as the case of endogeneity in the sub-mter resolution, forecast, vulnerability, eco-epidemiological, YF, geo-spectrotemporal, signature model.

Unfortunately, the authors of Ricardo Lourenço-de-Oliveira (2002) also cateogorized vectorial competence of these samples high, even though YFV infected mosquito LULC data had not been tested for violations of assumptions (i.e., heteroskedascity) in the experimental,

regressively, geo-spectrotemporally or geo-spatiotemporally geosampled, diagnostic, time series, endmember, residual, LULC prognosticators. Many articles in the literature on stochastic, medical entomological, vector, arthropod-related, geoclassified, sub-meter resolution, LULC, immature habitat, diagnostic, riceland, parameter estimators (e.g., meteorological, or Euclidean distance measurements from a capture point to an eco-georferenecable, agro-village, grid-stratified, orthogonal centroid) have provided arguments for optimally quantitating the nonconstancy of coefficients across multiple, bio-geophysical, eco-georeferenceable, eigen function, orthogonally decomposable, synthetic, spatial filter, LULC endmember observations (Jacob et.al. 2012). Suffice it to note that if parameterizing sylvatic, endmember, YFV, geometric, eigen-decomposable, seasonal, geo-spectrotemporally geosampled, regression coefficients is to be regarded as a vital modeling methodology for retrieving true partial derivatives from an empirical optimizable, *Ae. aegypti*, oviposition, endmember, LULC dataset of uncoalesced, sub-meter resolution, quantitative, iterative, qualitatively, interpolative, geoclassifiable, wavelength, forest-canopied or African, riceland, agro-irrgated, village, stochastic sub-mixwel signature explanators with respect to the diagnostic, endogeneity, it is improbable that these partial derivatives are identical for any two different, geosampled, capture point, hyperproductive, foci grid-stratifiable observations. Panchromatic, geometric, geoclassifiable time series, dependent, frequentistic, LULC data is optimal for geostatistically, optimally targeting, vector, arthropod-related, endemic, transmission-oriented, eco-georeferenceable, seasonal, hyperproductive, capture point, oviposition foci on geo-spectrotemporal, geoclassified ,grid-stratifiable, sub-meter, resolution, LULC**s** (Jacob and Novak 2014).

Frequentist probability or frequentism is an interpretation of probability; it defines an event's probability (e.g., an *Ae. aegypti,* sub-meter resolution, orthogonal, endmember, LULC model forecasts of capture point, seasonal hyperproductive foci) as the limit of its relative frequency in a large number of trials(Hosmer and Lemeshew 2002). There are two broad categories of explanative, probability signature LULC interpretations in time series, vector eco-epidemiological entomological, oviposition, remotely sensed data which can be called "geophysical" and "evidential" probabilities(Jacob et al. 2009). Geohysical explicative probabilities, (i.e., objective or frequency probabilities), can be associated with vector arthropods in random, geophysical ecosystem, geomorphological, terrain-related, grid-stratifiable, LULC transitions grid-stratifiable geolocations such as an those commonly seen in agro-irrigated, African, rice-cycle, LULC changes (e.g., discontinuous, peripheral, forest-canopy to riceland, *Ae aegypti, ,* immature, seasonal, hyperproductive, oviposition habitats). In such agro-irrigated LULC ecosystems, a given type of event (e.g., flooding event) may yield a lower, *Ae. aegypti,* immature habitat, eco-epidemiological, capture point, seasonal, count density at a persistent rate, or "relative frequency", in a long run of trials. Geophysical endmember probabilities may be usable to either explain, or may be invoked to explain, these stable frequencies in these entomological models.

All interpretations of probability in a seasonal, geo-spectrotemporal, vector, entomological, sub-meter, resolution, geosampled, signature endmember, LULC model are associated with approaches to statistical inference, including theories of estimation and hypothesis testing. The physical interpretation, of frequentist statistical methods may be robustly employable for remotely targeting, seasonal, hyperproductive, *Ae. aegypti,* immature habitat, capture points in expanding African, riceland, agro-irrigated ecosystems due to expanding anthropogenic populations, but iterative, interpolative explicative

forecasts must be field validated ("ground truthed"). Prior to implementation of a LULC , entomological, LULC model the existence and importance of propogagtional error probabilities, must be considered in the entomological, forecasts, so that the calculation of evidential probabilities may be statistically valid (Jacob et al. 2005).

There are four principal assumptions which justify the usage of a time series, eco-epidemiological, linear, vulnerability, probabilistic, remotely sensed, explanative, predictive, YF, geometric, diagnostic, endmember, orthogonal, LULC, signature, orthogonal paradigms for purposes of generating inference of specific, seasonal, transitional, inhomogeneous, peripherally expanding, agro-irrigated, eco-georeferenecable, African ricelands endmeber estimators into discontinuous, LULC, canopied corridors where, flooded, oviposition, hyperproductive, seasonal, sylvatic, *Ae. aegypti,* immature habitats may occur. Rice plants along gap edges forage for light by occupying both horizontal and vertical gap space and this morphological flexibility has implications for individual plant success, as well as forest structure and dynamics (Ackerly and Bazzaz 1995). To determine if contrasts in light availability lead to the development of canopy displacement, Jacob et al. (2015) investigated the responses of tree canopies to the heterogeneous light environments at the edges of multiple experimental canopy gaps to distinguish hyperproductive, eco-epidemiological, capture point onchocerciasis (river blindness"), black fly vector *Similium damnosum* s.l. in a riverine tributary in northern Togo. Canopies and trunks of gap edge trees were mapped, and their spatial distributions were analyzed using sub-meter resolution, endmember, LULC data. The authors found that tree canopies were displaced towards LULC gap centers where the eco-epidemiological oviposition foci was geolocated. The magnitude and precision of canopy displacement were greater for subcanopy trees than for canopy trees. The magnitude and precision of canopy displacement were generally greater for earlier successional trees than for later successional trees surrounding a seasonal, hyperproductive, *S. damnosum* s.l., trailing vegetation, capture point. Canopy depth was significantly greater on gap-facing sides of trees than on forest-facing sides of trees. Thus, the authors of Jacob et al. (2015) determined that trees along gap edges forage for light by occupying both horizontal and vertical gap space especially where seasonal, hyperproductive, immature habitats of *S. damnosum* s.l. occur in African meandering seasonal, flood-prone, agro-irrigated tributaries. This morphological flexibility may have implications for individual rice plant success, as well as canopy gaps in forest structures and dynamics for prognosticating seasonal hypeproductive capture point, *Ae. aegypti* foci in expanding African, Riceland villages due to anthropogenic pressure.

The first assumption is linearity and additivity of the relationship between   an, expositorial, intuitive, prognosticative, diagnostic, entomological, geoclassified, orthogonal, signature, LULC variable [e.g., annual, YF, prevalence, incidence rates] and a series of independent, geo-spectrotemporal or geo-spatiotemporal, geosampled, endmember, continuous and categorical, sub-meter resolution, time series, geoclassifiable, vulnerability variables such as weekly, meteorological, endmember, LULC estimators or log-transformed, surface and subsurface, geomorphic structures [e.g., catchment three (3)-dimensional, agro-irrigated, watershed covariates]. The linearity assumption can best be tested with scatterplots (Hosmer and Lemeshew 2002). These plots may depict a curvilinear and a hetroscedastic relationship in an empiricial optimizable, uncoalesced, LULC, endmember, signature dataset of uncoalesced, sub-

meter resolution, remotely sensed, *Ae. aegypti,* oviposition, capture point seasonal, hyperproductive habitats in an expanding, African, agro-irrigated riceland.

The following statements in the CORR prodcedure was employed by Jacob and Novak (2014) for requesting an endmember, iteratively interpolative, time series, malarial, mosquito, *Anopheline arabiensis,* seasonal hyperproductive, oviposition, sub-meter resolution, geoclassifiable, uncoalesced, endmember, LULC, capture point, signature, correlation analysis within a scatter plot matrix. The CORR procedure is a statistical procedure for numeric random variables that computes Pearson correlation coefficients, nonparametric measures of association, and the probabilities associated with these statistics. The correlation statistics include; Pearson product-moment and weighted product-moment correlation, Spearman rank-order correlation, Kendall's tau-b, Hoeffding's measure of dependence, D andPearson, Spearman, and Kendall partial correlation. PROC CORR also computes Cronbach's coefficient alpha for estimating reliability (www.sas,edu).The explicative, default, endmember, LULC correlation analysis includes descriptive statistics, Pearson correlation statistics, and probabilities for each analysis exploratory geo-spectrotemporal, geosampled, LULC variable. It may be possible for an arbovirologist, medical entomologist or other researcher to save the correlation statistics regressively derived from an eco-epidemiological, eco-georeferenceable, sub-meter resolution, LULC geo-spectrotemporal, *Ae. aegypti,* geoclassified, geosampled, African agro-irrigated, riceland, forecast, vulnerability, signature,oviposition, LULC model in a SAS dataset for use with other endmember statistical and reporting procedures.

Simple Correlation Analysis for a Fitness Study using PROC CORR is the simplest form of PROC CORR output (www.sas.edu). Pearson correlation statistics may be computed for all numeric geo-spectrotemporal, geosampled, African, riceland, ago-irrigated and discontinuous, forest-canopied, sub-meter resolution, geoclassifiable, *Ae. aegypti,* empirical oviposition variables for investigating the effect of various LULC, uncoalesced, signature covariates on immature habitat fitness in these peripheral ecosystem capture points The statement in Jacob and Novak (2014) that would produces the SAS outputwas :

```
options YF mosquito habitat size=60;
proc corr data=fitness;
run;
```

A sub-meter resolution, grid-stratified,eco-epidemiological, eco-georeferenceable, LULC orthogonal, signature dataset containing multiple, empirical, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, geosampled, clinical, field and remote, optimally specified, YF prognosticators may be employable to unbiasedly, regressively quantitate, agro-irrigated, hyperproductive, seasonal, immature habitat, LULC estimators geosampled in an African agro-irrigated, riceland environment, This dataset may contian multiple bio-geophysical, geo-spectrotemporal, geosampled, endmember, capture point, seasonal, eco-georeferenecable, sub-meter resolution, endmember, LULC observations, one of which may contain a missing value for the immature density variable (e.g., larval *Weight3*) for example. Initially, the capture point, immature habitats external, topological, geo-morphological, LULC feature attributes may be examined employing:

```
ods graphics on;
title 'Ae. aggypti Larval Measurement Data';
proc corr data=mosq1 nomiss plots=matrix(histogram);
  var Height Width Length3 Weight3;
 run;
ods graphics off;
```

The "Simple Statistics" table in Figure 1 display univariate descriptive statistics for the hypothetical, *Ae.aegypti,* larval habitat, endmember, sub –meter resolution, LULC, endemic ,vulnerability, forecast dataset

Figure 1: The CORR Procedure variable output for 4 African riceland and forest discontinuously canopied ovispoition, capture point, topological regressors

Simple Statistics						
Variable	**N**	**Mean**	**Std Dev**	**Sum**	**Minimum**	**Maximum**
Height	34	15.22057	1.98159	517.49950	11.52000	18.95700
Width	34	5.43805	0.72967	184.89370	4.02000	6.74970
Length3	34	38.38529	4.21628	1305	30.00000	46.50000
Weight3	34	8.44751	0.97574	287.21524	6.23168	10.00000

An arbovirologist, medical entomologist or YF experimenter may specify the NOMISS option for quantiating the same empirical dataset of oviposition, eigen-decomposable, *Ae. aegypti* capture point, riceland, agro-village, geosampled observations to computate the correlation of each pair of parameterizable, hyperproductive, seasonal geosampled, immature habitat   estimators. A Pearsons Correlation matrix may be generated.The NOMISS option can exclude geoclassified endmember, LULC observations with missing values of the VAR statement variables from the LULC, forecast, vulnerability analysis. The OUTP= option can create an e output LULC *Ae. aegypti* capture point dataset (i..e, *CorrOutp*) that may contain the Pearson correlation statistics.

In statistics, the Pearson correlation coefficient [PCC], d also referred to as the Pearson's *r,* Pearson product-moment correlation coefficient (PPMCC) or bivariate correlation] is a measure of the linear correlation between two explanative endmember variables(e.g., geosampled, YF-related, geo-spectrotemporal, African, riceland, partially, forest-canopied, sub-meter resolution, endmember, LULC, geoclassified, *Ae. aegypti,* agro-irrigated, seasonal, hyperproductive foci) *X* and *Y.* Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations(Hosmer and Lemeshew 2002). The form of the definition involves a "product moment", that is, the mean (the first moment about the origin) of the product of the mean-adjusted random variables, hence the modifier *product-moment* in the name. It has a value between +1 and −1, where 1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation. Pearson's

correlation coefficient is the covariance of the two variables divided by the product of their standard deviations (Cressie 1993). The absolute values of both the sample and population Pearson correlation coefficients are less than or equal to 1. Correlations equal to 1 or −1 correspond to data points (geosampled, African, agro-irrigated, riceland, *Ae aegypti,* eco-epidemiological, capture points along peripheral, discontinuous, canopy, forested corridors) lying exactly on a line (in the case the sample correlation), or to a bivariate distribution entirely supported on a line (in the case of a capture point , immature habitat, population correlation).

Hence, an arbovirologist, medical entomologist or YF experimenter could transform X to $a + bX$ and transform Y to $c + dY$, where a, b, c, and d are constants with $b, d > 0$ in an eco-epidemiological, time series, siganature, YF forecast, sub-meter resolution, time series, LULC equation without changing the diagnostic, normalized, correlation coefficient. The Pearson correlation coefficient is symmetric: $\text{corr}(X,Y) = \text{corr}(Y,X)$. A key mathematical property of the Pearson correlation coefficient such that it is invariant under separate changes in location and scale in any two variables(Hosmer and Lemeshew 2002). Employing the hypothetical, entomological signature LULC dataset, an uncertainty-related, Pearson's correlation test was constructed (see Figure 2).

Figure 2:Pearson's Correlation Coefficients, Prob>IrI under HO: Rho=0 for a hypothetical dataset of forecasted, *Ae,*aegypti riceland, agro-village, LULC , ovispoition regressors

	Height	Width	Length3	Weight3
Height	1.00000	0.92632	0.95492	0.96261
		<.0001	<.0001	<.0001
Width	0.92632	1.00000	0.92171	0.92789
	<.0001		<.0001	<.0001
Length3	0.95492	0.92171	1.00000	0.96523
	<.0001	<.0001		<.0001
Weight3	0.96261	0.92789	0.96523	1.00000
	<.0001	<.0001	<.0001	

In our example provided the geo-spectrotemporally geosampled newly transitioned, African, riceland, agro-irrigated, sub-meter resolution, geoclassifiable, endmember, LULC variables and forest-discontinuous, canopy, iterative interpolative signature dataset were highly correlated. For example, the correlation between Height and Width of an eco-epidemiological,

capture point, eco-georferenecable, *Ae. aegyti* capture point, hyperproductive, seasonal, forest-canopied foci was 0.92632.

To test the performance of a selected *Ae. aegyti* capture point, LULC variable selection method in linear regression with normal errors using simulated African riceland agro-irrigated sub-meter resolution data, a model :$y=X\beta+\epsilon, y=X\beta+\epsilon$, may be employable where yy is n×1, β is p×1n>pn>p, $\epsilon j \sim ind.N(0,\sigma^2)\epsilon j \sim ind.N(0,\sigma^2)$, j=1,…,nj=1,…,n., .Another way way would be to simulate all $x1,x2,...,xpx1,x2,...,xp$ together and assigning each geo-spectrotemporal, geosampled , geoclassifiable, endmember, explantory LULC variable a coefficient where the endogeneity may be quantitated by the error term ϵ. In such circumstances the explanative, YF, dependent variable would be the sum of the X′β and ϵ. Many statistical packages have functions that can specify the correlation between the x , geosampled African rice-land agro-irrigated and discontinuous, forest canopied, sub-meter resolution, geoclassifiable LULC , oviposition, endmember variables,. In Stata, for instance, that could be achieved with the corr2datacommand.,

```
// set a certain number of African riceland Ae. aegypti observations
set obs 1000

// generate the explanatory variables (here we simulate 2 variables
from a normal distribution)
gen x1 = rnormal(5,3)
gen x2 = rnormal(9,1)

// generate the error term (here is the simple case where the error is
distributed as N(0,1) - for other distributions use the according
sampling technique)
gen e = rnormal(0,1)

// generate the dependent variable and assign coefficients to the
explanatory variables (0.5 for x1 and 0.9 for x2, for instance)
gen y = 0.5*x1 + 0.9*x2 + e

// run the linear YF regression model using y on x1 and x2
reg y x1 x2
```

The corr2data command can render more options to specify *Ae. aegypti*, African, riceland, sub-meter resolution, orthogonal, endmember, LULC correlations between the variables. So if the signature model has high collinearity between x1 and x2, an arbovirologist, medical entomologist or YF experimenter may simulate measurment error correlations in the entomological, signature forecast, vulnerability, LULC model It can also be used to generate a heteroscedastic relationship between one or more of the explanatory habitat variables with the error. Below is an example on how to add superfluous *Ae.aegypti*, sub-mixel, timem series, LULC variables. In order to facilitate this quantiation ,an arbovirologist, medical entomologist

or YF experimenter may need to specify a correlation matrix before generating the African riceland LULC , oviposition data, for example:

```
  |    x1     x2     x3      e
----+------------------------------
 x1 |  1.0000
 x2 |  0.3000  1.0000
 x3 |  0.0100 -0.0000  1.0000
  e |  0.0000 -0.0000 -0.0000  1.0000
```

which may be achieved via

```
mat C = (1, 0.3, 0.01, 0 \ 0.3, 1, 0, 0 \ 0.01, 0, 1 , 0 \ 0, 0, 0, 1)
corr2data x1 x2 x3 e, n(1000) means(5 7 13 0) sds(3, 1, 2, 1) corr(C)
corr
```

where x3 can be made "superfluous" by simply not including it in the construction of y. If the geo-spectrotemporally geosampled, explanatory, YF forecastable variable is correlated with x1, it may still be usable through the correlation matrix C which may also reveal how superfluous x3 actually is. y may be generated in the entomological, vector arthropod, forecast, vulnerability, geo-spectrotemporal, signature model in Stata employing:

```
// generate the dependent variable
gen y = 0.5*x1 + 0.9*x2 + e

// run the regression with the useless variable
reg y x1 x2 x3
```

But this would be the generic set-up to generate such African agro-irrigated, riceland, sub-meter resolution, LULC uncoalesced signature data which should work in any other statistical package in the same way. Presumably other packages have additional/different functions that use but the steps done here are a basic way to achieve this.

Jacob and Novak (2014) generalized a simulation malarial, mosquito (*An. arabeinsis*) ovispoition, capture point, eco-epidemiological, forecast, vulnerability model employing PROC MCMC in SAS employing the method proposed by Gelman and Rubin (1992a) for monitoring the convergence of iterative, sub-meter resolution, LULC simulations by comparing between and within variances of multiple chains, in order to obtain a family of tests for convergence in an oviposition,eco-epidemiological,capture point, forecast, vulnerability model.The authors reviewed methods of inference from multiple endmember LULC simulations in order to develop convergence-monitoring summaries that were relevant for the purposes of optimally targeting, seasonal, eco-georeferenceable, hyperproductive, grid-stratifiable, immature, riceland habitats. An iterative method was given for selecting the biasing parameter, k, in the regression equation. The method rendered a distribution of squared errors for the geo-sampled, *An. arabiensis,* geoclassifiable, LULC regression coefficients that had a smaller mean and a smaller variance

$$k_a = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

than least squares or the single iteration estimate   The authors recommend applying a

battery of endmember tests for unmixing entomological vectoir arthropod , seasonal hypeproductive foci based on the comparison of inferences from individual iteratively interpolated, LULC signature sequences and from the mixture of geosampled parameterizable estimator sequences. The authors also suggest multivariate analogues, for assessing convergence of several, African, agro-irrigation, geosampled, riceland, endmember, sub-meter resolution, geoclassifiable LULC observations simultaneously.

The form of the definition in an eco-epidemiological, YF, forecast, vulnerability, time series, African, riceland, discontinuous, forest-canopy, sub-meter resolution, agro-irrigation, riceland, eco-georeferenceable, LULC model would involve a "product moment"(Jacob et al. 2013) that is, the mean (the first moment about the origin) of the product of the mean-adjusted, oviposition, sylvatic, capture point, seasonal, endmember, hyperproductive, *Ae. aegypti*, random variables; hence the modifier product-moment in the name.The product moment correlation coefficient is a measuremnet of the degree of scatter(Cressie 1993). It is usually denoted by r and r can be any value between -1 and 1. It is defined as follows:$r = \underline{s_{xy}} \sqrt{} (s_x s_y$ where s_{xy} is the covariance of x and y, $s_x = \frac{1}{n} \sum (x - \overline{x})$ .The product moment correlation coefficient (pmcc) can be used to tell us how strong the correlation between two variables is. A positive value indicates a positive correlation and the higher the value, the stronger the correlation (Griffith 2003). Similarly, a negative value indicates a negative correlation and the lower the value the stronger the correlation. If there is a perfect positive correlation (in other words the points all lie on a straight line that goes up from left to right), in an *Ae aegypti,* eco-epidemiological, geo-spectrotemporal, forecast, vulnerability endmeber, oviposition, LULC model then r = 1.If there is a perfect negative correlation, then r = -1.If there is no correlation, then r = 0. r would also be equal to zero if the geosampled capture point variables were related in a non-linear way (they might lie on a quadratic curve rather than a straight line, for example).

The PLOTS=MATRIX (HISTOGRAM) option requests a scatter plot matrix for the VAR statement variables(www.sas.edu). With VAR statement input, the rows of the contingency table may correspond to the observations (geosampled uncoalesced, *Ae. aegypti,* African, riceland LULC regression estimators) of the input data set, and the columns correspond to the VAR statement variables. In so doing, the values of the habitat variables would then be definable employing the table frequencies.

PROC FREQ statements also request frequency plots for the crosstabulation tables. PROC FREQ produces these plots by using ODS Graphics to create graphs as part of the procedure output. ODS Graphics must be enabled before producing plots. The PLOTS(ONLY)=FREQPLOT option requests frequency plots. The TWOWAY=CLUSTER *plot-option* specifies a cluster layout for the two-way frequency plots.

```
ods graphics on;
proc freq data=SummerSchool;
   tables Gender*Internship*Enrollment /
          chisq cmh plots(only)=freqplot(twoway=cluster);
   weight Count;
```

```
run;
ods graphics off;
```

In an sub-meter resolution, ricland agro-irrigated, African, *Ae. aegypti,* LULC, eco-epidemiological, geo-spectrotemporal, forecast vulnerability model negative values may be treated as missing. In so doing, the observation may be excluded from the capture point, endmember, oviposition analysis. The values are not required to be integers. Row labels for the table may be specified in SAS with an ID variable. Column labels may be constructed from the variable name or variable label (*Ae aegypti* , post-harvesting, capture point). An arboviroloist, medical entomologist or YF experimenter may specify multiple correspondence analysis (MCA). In so doing, the row and column labels will be the same and may be constructed in SAS from the variable names or labels, so as to include an ID statement. With MCA, the VAR statement must list the variables in the order in which the rows occur(www.sas.edu).

A Scatter Plot Matrix may also be generated in Mathlab based on the of the regression-related, entomological, LULC, sub-meter resolution uncoalesced, capture point, seasonal hyperproductive, oviposition prognosticators. Scatter plot is a useful exploratory tool for multivariate, endmember, LULC data analysis and is one of the most commonly used statistical graphics(https://www.mathworks.com). Scatterplot matrices are good for determining rough linear correlations of metadata that contain continuous entomological, LULC variables (Jacob at al. 2007) an arbovirologist, medical entomologist or YF experimenter may create a scatter plot matrix with one or two Matrix Inputs employing randomized, African, grid-stratifiable, orthogonally eigen-decomposable, sub-meter resolution, LULC data. The subplot in the ith row, jth column of the matrix is a scatter plot of the ith column of X against the jth column of X(Griffith 2003). In so doing, the diagonal will be histogram plots of each column of X which may aid in specifying the LULC marker type and the color for the time series, seasonal, hyperproductive. *Ae aegypti,* eco-epidemiological, capture point, scatter plots in Mathlab emplying

```
X = randn(50,3);
plotmatrix(X,'*r'
```

Figure 3: A hypothetical scattergraph employing multiple sub-meter resolution LULC Cartesian coordinates to display regressed oviposition, *An. egypti* seasonal geosampled sub-mter resolution, endmember values
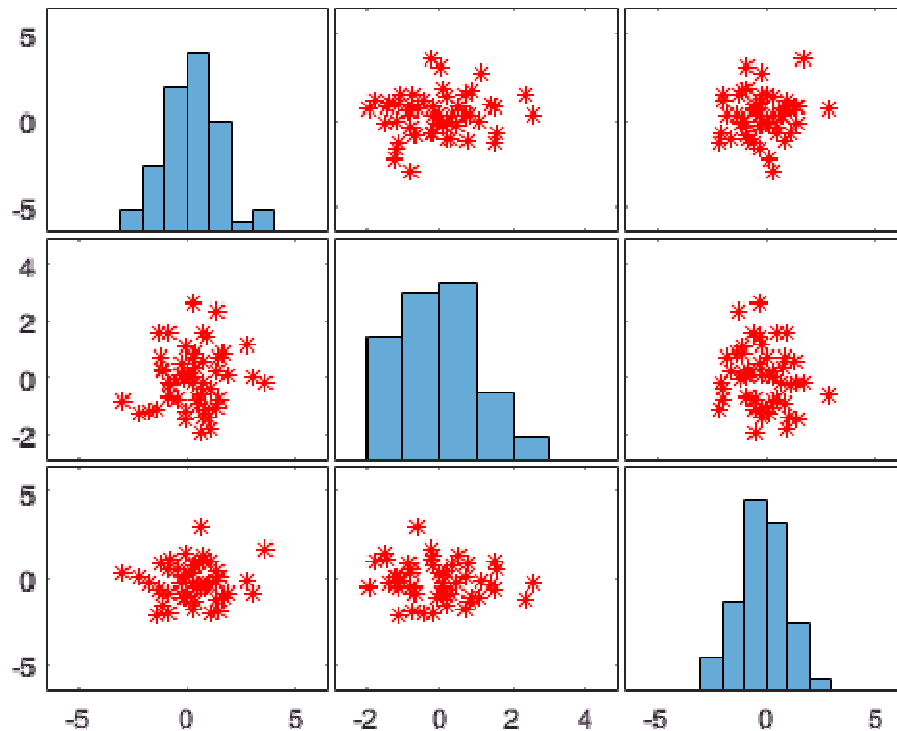
In traditional SAS/GRAPH®, to create a quality time series, eco-epidemiological, YF, regression-related, sub-meter resolution scatter matrix cooperation of the GPLOT procedure, several SYMBOL statements and GOPTION statements would be necessitated. New with SAS® 9.2, the SGSCATTER procedure can produce a variety of sylvatic, YF, *Ae.aegypti,* oviposition, sub-meter resolution, LULC, capture point, scatter plots and put them into panels with different grid-stratified ,orthogonal layouts within just a few lines of code. Creating different types of riceland, African, agro-village LULC, scatter plots with PROC SGSCATTER and using ODS GRAPHICS and ODS styles may enhance cartographically and geostatistically illustrating seasonal, hyperproductive*, sylvatic, Ae .aegypti,* oviposition, sub-meter resolution, riceland,capture point, geoclassified , grid-stratified, LULCs spilling into inhomogeneous ,forest canopy, in these expanding irrigated ecosystems due to anthropogenic pressures.

Traditional SAS/GRAPH® procedures create graphs that are saved in SAS® catalogs and can be displayed and edited in the GRAPH window. GOPTIONS, SYMBOL, AXIS and other SAS/GRAPH® statement control the appearance of the graph. On the other hand, the new statistical graphics procedures, such as PROC SGSCATTER can create and display graphs of seasonal, hyperproductive, capture point, *Ae. aegypti*, oviposition, foci on newly transitioned

discontinuous, forest-canopied, sub-meter resolution, geoclassified, LULC in expanding, African, agro-irrigated, riceland village in standard image formats, such as BMP and PNG,.This could be facilitated using the Output Delivery System (ODS) directly. Instead of GOPTION and SYMBOL statements, ODS statement, especially ODS graphics, may be usable in controlling the appearance of endemic, YF, transmission-oriented, time series, LULC graphs produced by PROC SGSCATTER. The new SGSCATTER procedure can provide an exciting method to produce paneled, forecast, YF, vulnerability, time series, explanative, scatter plots. Its simple and natural syntax and seamless cooperation with ODS GRAPHICS may make PROC SGSCATTER a powerful tool for sylvatic, *Ae.aegytpti* geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, sub-meter resolution, LULC, data visualization.

In order to create this display, Jacob and Novak (2014) specified the ODS GRAPHICS ON statement in addition to the PLOTS= option for geo-spectrotemporally and geo-spatiotemporally mapping , seasonal, hyperproductive, *An. arabiensis*, ovispoition, sub-meter resolution, geoclassifiable, eco-georeferenceable, capture points. To explore the correlation between Height and Width predictor, LULC malaria, capture point, seasonal hyperproductive, oviposition, sub-meter resolution, variables the following statements were employed. .

```
ods graphics on;
proc corr data= malmosq nomiss
      plots=scatter(nvar=2 alpha=.20 .30);
  var Height Width Length3 Weight3;
 run;
ods graphics off;
```

The NOMISS option was specified with the original VAR statement to ensure that the same set of oviposition wavelength, frequency, sub-meter rersolution, grid-straified, *An. arabiensis* LULC observations was used for the analysis. The PLOTS=SCATTER(NVAR=2) option requested a scatter plot for the first two variables in the VAR list. The ALPHA=.20 .30 suboption requested 92% signature paradigm prediction ellipses, respectively.

Prediction ellipse is a region for predicting a new observation from the population, assuming bivariate normality (Hosmer and Lemeshew 2002). In Jacob and Novak (2014) this forecast vulnerability, signature, endmember map approximated an eco-epidemiological, capture point region (positively autoccorelated) containing a specified percentage of the immature geosampled population(i.e., 'hot spot). The displayed prediction ellipse was centered at the means $(\bar{x}, \bar{y})$. Note that the following statements displayed a scatter plot for the *An. arabiensis* hyperproductive, seasonal, foci estimators *Height* and *Width*:

```
ods graphics on;
proc corr data=Mal mosq1
      plots=scatter(alpha=.20 .30);
  var Height Width;
 run;
ods graphics off;
```

For any eco-epidemiological, geo-spectiotemporal or geo-spatiotemporal, sylvatic, *Ae aegypti*,YF, eco-georeferenceable, sub-meter resolution LULC, forecast, vulnerability model the multiple linear regression, analysis would require that the error between observed and predicted, immature, habitat density values (i.e., the residuals of the regression) should be normally distributed. This assumption can best be checked by plotting residual values on a histogram with a fitted normal curve or by reviewing a Q-Q-Plot of the eco-epidemiological, time series, YF, model, diagnostic output.

Normality in a forecast ,vulnerability, eco-epidemiological, sylvatic, YF, *Ae. aegypti* , oviposition, African, riceland, agro-ecosystem, endmember, LULC model can be checked with a goodness of fit test (e.g., the Kolmogorov-Smirnov test), though this test must be conducted on the residuals themselves. When the LULC, orthogonal, grid-stratified, time series, eco-epidemiological, YF data is not normally distributed, a non-linear transformation (e.g., log-transformation) might correct this issue in SAS if one or more of the individual explanatory, predictor variables are to blame, though this may not directly respond to the normality of the residuals.

Further, a multiple, linear, YF, sylvatic, *Ae aegypti*, oviposition, capture point, forecast, vulnerability, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, African, riceland, agro-irrigated, regression-related, prognosticative model will assume that there is little or no multicollinearity in the geosampled, LULC, parameter estimators. Multicollinearity occurs when the independent variables are not independent from each other (Rao 1972). Multicollinearity in such a paradigm may be checked against 4 key criteria:

1) Correlation matrix – When computing the matrix of Pearson's Bivariate Correlation amongst all sylvatic, *Ae. aegypti*, potential, African, riceland or forest-canopied, explicative, independent variables (i.e., optimally the correlation coefficients shoud be smaller than .08)(see Jacob et al. 2009).

2) Tolerance – The tolerance measures the influence of one independent variable on all other independent explanatory variables; the tolerance is calculated with an initial linear regression analysis (Hosmer and Lemeshew 2002). Tolerance in an eco-epidemiological, YF, sub-meter resolution, geo-spectrotemporal or geo-spatiotemporal, LULC, endmember, signature, vulnerability model may be defined as $T = 1 – R^2$ for conducting a vigrious first step regression analysis(see Jacob et al. 2012). With $T < 0.2$ there might be multicollinearity in the uncoalesced, LULC endmember, sub-meter resolution signature and with $T < 0.01$ there certainly would be.

3) Variance Inflation Factor (VIF) – The variance inflation factor of for a medical entomological,forecast, vulnerability, sub-meter resolution, eco-epidemiological, LULC model linear regression may be defined as $VIF = 1/T$. Similarly with $VIF > 10$ there is an indication for multicollinearity to be present.

4) Condition Index – The condition index may be calculable for an oviposition, sub-meter resolution, *Ae aegypti*, YF, eco-epidemiological, eco-georeferenceable, sylvatic, grid-stratifiable, LULC representing inhomogeneous forecast canopy and or riceland African

ecosystems using a factor analysis on the independent geosampled variables. Values of 10-30 may indicate a mediocre multicollinearity in the regression variables, whereas values > 30 may indicate strong multicollinearity.

If multicollinearity is found in an empiricial, orthogonal, dataset of geo-spectrotemporal or geo-spatiotemporal, geosampled, eco-epidemiological, African, agro-ecosystem, geoclassified, YF, oviposition, sylvatic, *Ae. aegypti,* sub-meter resolution, data frequency, signature,sub-meter resolution, estimators, one remedy might be centering the data employing a grid-stratified LULC. To center the endemic, transmission LULC data, an arbovirologist, medical entomologist or other experimenter may simply deduct the mean score from each geosampled, riceland, agro-irrigation, or forest-canopied LULC geosampled observation. Other alternatives to tackle the problem of multicollinearity in a multivariate,endemic, YF, time series, LULC regression is to conduct a factor analysis before the linear analysis and to rotate the factors to insure independence of the factors in the analysis.

An endemic, transmission-oriented, geo-spectrotemporal or geo-spatiotemporal, sylvatic, YF, forecast, vulnerability, eco-epidemiological, eco-georeferenceable, multiple, linear regression, LULC analysis would require that there is little or no autocorrelation in the geo-spectrotemporal or geo-spatiotemporal, geosampled African, riceland, LULC data. Autocorrelation occurs when the residuals are not independent from each other. In other words when the value of y(x+1) is not independent from the value of y(x).

The Goldfeld-Quandt Test may test for heteroscedasticity in an uncoalesced dataset of sub-meter resolution, uncoalesced, eco-epidemiological, inhomogeneous, forest-canopied, riceland, oviposition, capture point, transitional, LULC, wavelength, signature frequencies for determining abundance and distribution of oviposition, *Ae.aegypti,* seasonal, hyperproductive foci in African agro-irrigation ecosystems. In statistics, the Goldfeld–Quandt test checks for homoscedasticity in regression analyses (Hosmer and Lemeshew 2002) . It does this by dividing a dataset into two parts or groups, and hence the test is sometimes called a two-group test. The test would split the multiple linearly regressed, oviposition, *Ae.aegypti* seasonal, hyperproductive foci data into high and low density value to see if the samples are significantly different. If homoscedasticity is present in any eco-epidemiological, orthogonal, YF, predictive risk-related, multiple, linear regression model estimator, a non-linear correction might fix the diagnostic, residiual without allowing multicollinearity into the model.

Expanding agro-irrigation, African, riceland, ecosystem, flood plains facilitate high geohydrologic connectivity (e.g., water flux between the channel and flood plain) marked by complex seasonal patterns of inundation, extensive penetration of channel water laterally into the alluvial aquifer, and springbrooks formed by ground water erupting onto the agro-village flood-plain surfaces. After cartographically optimally delineating and geostatistically classifying sub-meter resolution, flood-plain, geoclassifiable, LULC elements (e.g., tillered vegetation patches and agro-irrigation channel reaches) in an African riceland flood plains in central Kenya, Jacob et al. (2007) remotely regressively quantitated expanding, malarial, mosquito, vector, *An. arabiensis*, seasonal, hyperproductive, eco-georeferenceable, capture points in peripheral corridors of the agro-village complex. Amongst the geoclassified, LULC, regression-related, paramterizable.field-operational covariates selected to delineate the prolific, capture point,

oviposition foci was contour survey data retrieved from a sub-meter resolution, 3 –dimensional (D) Digital Elevation Model (DEM) which aided in identifying expected relationships amongst flood-plain, LULC, element types, surface scour frequency, and flood-plainslope coefficients. Data analyses of the eco-georeferenced, eco-epidemiological, seasonal, hyperproductive, sylvatic, *Ae. aegypti* may reveal that scour frequency is inversely proportional to the elevation of the flood plain after the tillering rice-cycle stage, except when localized geomorphic controls such as natural levees prevent normal high flows from inundating and scouring relatively low, flood-plain, LULC elements within the periphery of an African, riceland, agro-irrigated ecosystem. Further, while different flood-plain, sub-meter resolution, geoclassifiable, rice-cycle, LULC, element types may occupy distinct elevation zones on these flood plains, the elevation of each zone above the agro-irrigation channel may differentiate with localized channel entrenchment along the peripheral, inhomogeneus, partially, forest-canopied, peripheral corridors of these ecosystems. It may be that topographic variation amongst flood-plain, LULC elements in an African, expanding, agro-village, ricland complex is greater than the variation within elements during post-harvesting, suggesting that sub-meter resolution, scaled, flood-plain topography may be optimally usable for remotely characterizing, seasonal, hypeproductive, sylvatic, *Ae aegypti*, oviposition,signature, capture points within geoclassifiable, forest-canopied, flood-plain discontinuous peripheral elements.

Field data literature suggest strong associations between specific geo-spectrotemrpoal or LULC classes of flood-plain elements and preferential ground-water flow paths in the upper alluvial aquifer (Brooker 2002). Combined with preexisting geo-spatiotemporal or geo-spatiotemporal, geoclassified satellite surveys, eco-georeferenecable, oviposition, seasonal, LULC,capture point,grid-stratified, *Ae aegypti*, hyperproductive foci, geoclassifiable, data may reveal a sinuous lattice of preferential, ground-water, flow paths (e.g., buried abandoned streambeds) in the upper alluvial aquifer along the peripheral edges of the expanding, African, riceland environment at approximately the same elevation as the main channel's streambed where the geoclassifiable, forest-canopied, discontinuous LULCs are being converted to riceland oviposition, seasonal, hyperproductive,capture point, sylvatic, *Ae. aegypti,* immature habitats. Employing geostatistical interpretations of remotely discernible, orthogonal, grid-stratifiable, sub-meter resolution, regressively quantizable relationships amongst geoclassifiable, time series, multivariate. LULC, element types, (e.g., elevation,), preferential ground-water flow paths may be optimally characterized employing 3-D, surface and subsurface, geomorphological, terrain-related, caovariate, parameter estimation across an entire riceland, agro-village, flood plain. In so doing, a heuristic, eco-epidemiological, probabilistic, YF model effort may be utilized for optimally investigating how much expanding African, agro-irrigated, riceland ecosystems due to anthropogenic pressures are geospatially spilling into forest-canopied, discontinuous LULCs. Potential, hyperendemic, geostatistical,"hot spots."(e.g., positively autocorrelated eco-georeferenceable, *Ae.aegypti* oviposition, eco-epidemiological, capture points) may be identified in SAS/GIS.

The expected value of a geometric probabilistically regeressable, endmember, time series, geoclassifiable, grid-stratifiable, *Ae aegypti,* oviposition, geomorphological, terrain-related, eco-georenceable, explanatorial, LULC specified, response variable quantitated in SAS/GIS whilst holding the other independent variables fixed would be a straight-line function of each geostatistically significant, forecastable variable within a 95% confidence interval.

The slope coefficient of that line would not be dependent on the values of any time series, diagnostic, clinical, field or remote, time series, specified, sylvatic, YFV-related, empirical, regressively optimizable endmember, oviposition, geoclassifiable, grid-stratifiable, sylvatic, *Ae aegypti,* capture point, or any of the oviposition, LULC, spectral, wavelength frequencies. The effects of different empirical, geometric frequency, endmember, independent variables on the expected value of a geo-spectrotemporal or geo-spatiotemporal, uncoalesced, wavelength, frequency,endmember, LULC, signature, dependent variable (e.g., case distribution data) would then be additive.

Another assumption that would have to be non-violated, in an experimental, prognosticative, endmember, time series, field-operationizable, regressable, geometric, endmember,orthogonal, grid-stratifiable, dataset of uncoalesced, vulnerability, geometric, sub-resolution, eco-epidemiological, YFV, signature paradigm, remotely sensed, paramerizable estimators for optimally targeting seasonal, hyperproductive, eco-georeferenceable, eco-epidemiological, *Ae aegypti*, oviposition, capture points on sub-meter resolution, sparsely forest-canopied, geo-spatiotemporal or geo-spectrotemporal, geoclassifiable LULCs along expanding, African, riceland peripherial corridors   would be statistical independence of the sampling errors. In particular, no correlation should exist between consecutive, iterative, quantitative, inconspicuous, propogagtional regression error [e.g., aspatial heteroskedascity amongst LULC oviposition observations] especially in the case of any uncoalesced, YF-related, sub-meter resolution, geosampled, eigen-decomposable, explanatory, agro-irrigated, African, riceland, sylvatic, *Ae. aegypti,* elucidative wavelength, frequency, diagnostic, orthogonal variables with discontinuous, forest-canopied, synthetic, feature attributes. If two events are independent, statistically independent, or stochastically independent then the occurrence of one does not affect the probability of the other (Hazewinkle 2001).

Similarly, two optimally unbiased, expositiorial, diagnostic, YFV–related, explanatory, clinical, field or remote specified, random, African, agro-village, riceland, LULC variables or inhomogeneous, forest-canopied, LULC variables would be independent if the realization of one on a sub-meter resolution, discontinuous, geoclassifiable, seasonal, hyperproductive, sylvatic, capture point along an expanding   agro-irrigated, ecosystem, geoclassifiable LULC does not affect the probability distribution of the other. The concept of independence may extend to dealing with collections of more than two events whenst, eco-georeferenceable, agro-irrigated, riceland, explicative, geo-spectrotemporal or geo-spatiotemporal, YFV-related, randomized probability regression variables are geosampled in which case the events would be pairwise independent, if each pair are independent of each other. Further, the transitional, triggering, riceland, agro-village, LULC events (i.e., *Ae. aegypti* inhomogenous forest-canopied, newly transitioned agro-irrigated, rice cycles) would be mutually independent if each sample event is independent of each other based on a combination of statistically quantitated events.

Violations of homoscedasticity (i.e., common variance) in an entomological, sub-meter resolution, geo-spectrotemporal or geo-spatiotemporal, diagnostic, sub-meter resolution, [e.g., QuickBird visible and near infra-red (NIR) at 0.61m], geoclassifiable LULC, explanatory, forecast, probabilistic, vulnerability, YFV paradigm (e.g., "heteroscedastic " non-robust parameters) may make it difficult to gauge the true standard deviation of optimal, iteratively, quantitative, explicative, illuminative, seasonal prognostications (e.g., remotely sensed,

optimally targeted, hyperproductive, sub-meter resolution, sylvatic, oviposition, eco-georeferenecable, *Ae. aegypti,* geoclassifiable, grid-stratfied, sub-meter resolution, eco-georeferenceable, sites) along expanding riceland corridors and their unquantiated, propagational regression, probabilistic, uncertainty, erroroneous, LULC estimators. In such vulnerability paradigms the optimally targeted, eco-georeferenceable, hyperproductive, vector arthropod, immature, seasonal, eco-epidemiological, diagnostic, explanatory, capture point, time series, probabilistic variable output would render confidence intervals that are too wide or too narrow. Confidence intervals may provide the likely range of a sample proportion or sample mean from the true proportion mean found in immature, seasonal, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiologically geosampled, *Aedes* populations on partially discontinuous, forest canopy, transitioning geoclassifiable, sub-meter resolution, LULCs to agro-irrigated, riceland LULCs in an expanding, African, agro-village ecosystem which may enable optimally estimating the precision of results rendered from a sample frame for prognosticating the true population of prolific, oviposition, immature, capture point, *Ae aegypti,* seasonal, hyperproductive riceland geolocations. In so doing, a intervention may be developed for implementing a targeted larval control strategy based on hyperproductive, seasonal, eco-georeferenceable, georeferenceable, YF, oviposition sites (e.g., see Gu and Novak 2005) employing algorithms from geostatistical geodatabase cyberenvironments (e.g., ArcGIS) (see Jacob and Novak 2014).

Results for both individual studies and meta-analyses of an orthogonally eigen-decomposable, eco-georeferenceable, autoregressively optimizable, eco-epidemiological, sylvatic,*Ae. aegypti,* hyperproductive, capture point, remotely specifies, oviposition, geoclassifiable, frequency, LULC, sub-meter resolution, uncoalesced dataset of irradiance, frequency, wavelength measurements may be tabulated in SAS/GIS, R or other statistical packages(GeoDa). In so doing, empirically, optimally forecastable, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological sub-meter resolution, endmember, LULC, signature datasets of diagnostic, seasonal, explanative, agro-irrigated, African, riceland, eigen-decomposable, explanatory, empirical estimators may be reported along with an eco-epidemiological, capture point, probabilistically, geoclassifiable, regresseable estimates and an associated confidence interval. For example, the odds ratio of a a vulnerability, sub-meter resolution, geoclassifiable, eco-epidemiological, sylvatic, robustifiable, YFV-related, eco-georeferenceable, explanatorial, diagnostic, regression estimate may be optimally calculable as 0.75 with a 95% with confidence interval of 0.70 to 0.80. The iteratively, quantitative, seasonal, eco-georeferenceable, hyperproductive, capture point, eigen-decomposable, interpolative, explicative estimate (0.75) may then be the best guess of the magnitude and direction of the experimental, oviposition, LULC, control, intervention's effect compared with the YFV control intervention. The confidence interval may optimally describe the propagational uncertainty inherent in this quantitative, geoclassified, seasonal, LULC, oviposition, regression, explicative estimate, whilst simultaneously describing a range of prolific, eco-georeferenceable, optimally regresseable, immature, capture point,uncoalesced, habitat count, density values which an arbovirologist, medical entomologist or YFV experimenter could use to statistically validate the LULC effect actual effects on an expanding, African, riceland, agro-village complex, eco-epidemiological, study site. By employing unbiased, sylvatic, seasonal, *Ae. aegypti,* oviposition, sub-meter resolution, LULC, explanatory, African, riceland agro-irrigation, parameterizable,

covariate, frequency geo-spectrotemporal and or geospatiotemporal, unbiased, wavelength estimators geosampled along an interfacing agro-irrigated and discontinuous, forest-canopied corridor, hypeproductive foci may be optimally remotely identified and henceforth optimally seasonally forecasted employing within a YF, signature paradigm.

If the confidence interval is relatively narrow (e.g., 0.70 to 0.80) in the eco-epidemiological, eco-georeferenecable, seasonal, explanatory, forecast, vulnerability, sub-meter resolution, LULC remotely sensed, sylvatic, YFV, geo-spatiotemporal or geo-spectrotemporal, sub-meter resolution, regression model, the effect size may be known precisely. If the interval is wider (e.g. 0.60 to 0.93), the probabilistic, propogagational uncertainties in the diagnostic, explanatory, geoclassifiable, endogeneous regressors would be greater, although there may still be enough precision to make decisions about the utility of the *Ae. aegypti* control intervention variables to optimally target, eco-georeferenecable, hyperproductive, seasonal, eco-epidemiological, oviposition foci along an expanding agro-irrigated, riceland, African, village complex parsimoniously. Intervals that are very wide (e.g. 0.50 to 1.10) in a partially discontinuous, diagnostic,explanatory, forecast, vulnerability, probabilistic, sub-meter, resolution, riceland, sylvatic, *Ae. aegypti,* oviposition, LULC, parameterizable estimator signature dataset would indicate little knowledge about the effect, and hence further information would be warranted prior to implementation of an effective control intervention,[e.g., an Integrated Vector Management program (see Gu and Novak 2005)] for optimally, precisely, remotely targeting, prolific, eco-georeferenceable, immature habitat, seasonal, hyperproductive, eco-epidemiological, capture points.

A 95% confidence interval is often interpreted as indicating a range within which can be 95% certain that the true effect lies (Fotheringham 2002). This statement may be a loose interpretation in an explicative, geoclassifiable, eco-epidemiological, YF-related, predictive, sub-meter resolution, LULC, probabilistic, *Ae egypti*, oviposition, African, agro-irrigated, riceland, risk model, diagnostic outout. It may be however useful as a rough guide for measuring robustness of a real-time, explanative, optimally forecastable, eco-georeferenceable, geometric, endmember, orthogonal, grid-stratified, time series, regression, probabilistic, regressable LULC estimate. The strictly-correct vulnerability, frequency, wavelength, sub-meter resolution, uncoalesced, endemic, seasonal, geo-spectrotemrpoal or geo-spatiotemporal, geosampled diagnostic interpretation of a confidence interval in such a probabilistic, sylvatic, oviposition, *Ae. aegypti*, sub-meter resolution, signature paradigm would then be based on the notion that the iteratively quantitatively interpolative, parameterizable, LULC covariates results could be extracted if the eco-epidemiological, geoclassifiable, sylvatic, YF study were repeated multiple times at the agro-irrigated, riceland, African, agro-village complex. If the model were repeated often, a 95% confidence interval may be calculable in SAS (e.g., PROC REG). Subsequently, 95% of these intervals could optimally contain the true effect in an empirical,optimizable, uncoalesced signature, uncoalesced LULC dataset of expositorial, residual, vulnerability, eco-georeferenceable, geometric, endmember, explanative forecasts (e.g., targeted, hyperproductive, oviposition, *Ae. aegypti,* LULC sites on sub-meter resolution imaged, geo-spatiotemporal or geospectrotemporal, geoclassifiable, agro-irrigated, oviposition, eco-georeferenceable, discontinuous, partially, forest-canopied, capture point, seasonal foci).

Further, the parsimonious width of the eco-epidemiological, sylvatic, YFV-related, geo-spectrotemporal or geo-spatiotemporal, geometric, endmember, sub-meter resolution, LULC, endmember datset of frequency, uncoalesced, prognosticative, geo-spectrotemporal or geo-spatiotemporal, wavelength, frequency, confidence intervals for remotely optimally capturing an individual geoclassifiable, orthogonal, agro-irrigated, agro-ecosystem, riceland, expanding African, eco-epidemiological, forecast, vulnerability model unbiased, eco-georeferenecable estimator would depend to a large extent on the sample size of the, heuristically optimizable, time series, explicative, clinical, field or remote, endogenous, multivariate observables. Larger, elucidatively log-normalized, geo-spectrotemporal or geo-spatiotemporal, geoclassifiable, diagnostic, time series, uncoalesced, wavelength, frequency, signature, transmittance, emissivity, regressors orthogonally synthesized from eigen-decomposable, grid-stratifiable, sylvatic, capture point, *Ae. aegypti,* oviposition, LULC, wavelength, frequency, spectral signatures and subsequently qualitatively iteratively, interpolated over an expanding, agro-irrigated, African, riceland, sub-meter resolution polygon employing an eigen-function decomposition, algorithm in AUTOREG may render endmember, estimates of hyperproductive, immature habitat effects on discontinuous, forest-canopied LULCs.

In a seasonal, iterative, Bayesian, YFV, explanative, endmember, time series, geo-spectrotemporal or geospatiotemporal, wavelength, frequency, geodatabase cyberplatform, continuous, explicative, latent, outcome precision would depend on the variability in the outcome measurement LULC dataset (e.g., the standard deviations of individual,geosampled, temporal, eco-georeferenceable, sylvatic, oviposition, *Ae. aegypti,* geoclassifiable, capture point, seasonal, immature, forest-canopied, endemic foci) in peripheral eco-zones of agro-irrigated expanding, African ecosystems in geographic space. For dichotomous outcomes it would depend on the risk of the sampling event, and for time-to-event outcomes it would depend on the number of riceland, oviposition, capture point, eco-epidemiological, capture point, immature, habitat events observed (see Jacob et al. 2009). In so doing, optimally log-transformed, uncoalesced, seasonal, covariance, endmember weightages may be optimally employable in sub-meter resolution, eco-epidemiological, forecasting, YV-oriented, regression-related, vulnerability LULC signature, frequency, sub-meter resolution paradigms. Robust computation of expositorily quantizable, standard errors of the probabilistically, regressable,sub-mter resolution, LULC wavelength frequency effects from which the confidence interval determination based on the target (i.e., hyperproductive, seasonal, eco-georferenceable, discontinuous, forest-canopied, oviposition, *Ae. aegypti,* eco-epidemiological, geoclassified foci) may be subsequently optimally orthogonally derivable.

The problem of optimally estimating a covariance, orthogonal, spatial filter, grid-stratifiable, algorithmic, weighted matrix in small, sylvatic, *Ae. aegypti* oviposition, sub-meter resolution, geometric endmembers along African, riceland corridors expanding into discontinuous forest-canopied, peripheral, LULC samples has not been considered in literature. In probability theory and statistics, a covariance matrix (also known as dispersion matrix or variance–covariance matrix) is a matrix whose element in the i, j position is the covariance between the ith and j^{th} elements of a random vector. A random vector is a random variable with multiple dimensions (Hazewinkle 2001).This problem may be especially important in sub-meter

resolution, diagnostic, hyperproductive, seasonally, intermittent, discontinuous, forest-canopied, geo-spatiotemporal or geo-spectrotemporal, grid-stratifiable, geoclassifiable, sylvatic, oviposition, *Ae. aegypti*, sub-meter resolution, LULC, eco-georeferenceable, eco-epidemiological, capture point, orthogonal, autoregressive signature, endmember paradigms where the standard errors of fixed and seasonal, oviposition, parameter estimator, random effects would depend on estimation of the covariance matrix which would be based on the distribution of the random effects. A set of hierarchical priors (HPs) for the covariance matrix that produces posterior shrinkage toward a specified, geoclassifiable, agro-irrigated, LULC, African, riceland, transitioning, orthogonally, time series, regressable eigen-structure in an eco-georeferenceable, sylvatic, *Ae aegypti*, inhomogeneous, ovipostion, sub-meter resolution, grid-stratified, vulnerability, forecast model may quantitate geoclassifiable, partially, forest-canopied,hyperproductive, seasonal, oviposition sites through examination of shrinkage toward diagonality.

Standardizable, sub-meter resolution, diagnostic, uncoalesced, endemic, transmittance, frequency, wavelength, sub-meter resolution, signature, LULC estimators like the unstructured maximum likelihood (ML) estimator or restricted maximum likelihood (REML) estimator may be very unstable in a geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, *Ae aegypti,* probabilistic, interpolative, geometric, LULC, endmember, African, riceland, oviposition,probabilistic paradigm with the smallest orthogonally eigen-decomposable, estimated, time series, eigenvalues being too small and the largest too big. A standard approach to more stably estimating the orthogonal gridded matrix in small geoclassifiable, regressable, sylvatic, time series, eco-epidemiological, uncoalesced, orthogonal, endmember empirical datasets of geosampled, *Ae. aegypti,* oviposition, geo-spatiotemporal or geo-spectrotemporal, LULC, wavelength, frequency, estimators in SAS or R may be to compute the ML or REML estimator under some simple structure that involves estimation of fewer, riceland, agro-irrigated, agro-village, explanatory, eco-georeferenceable, LULC parameters, such as compound symmetry or independence. However, these geoclassifiable, oviposition, LULC, agro-village, riceland, endemic foci and their, time series, clincial, field or remote, optimally specified, endmember oviposition, YF estimators may not be consistent unless the hypothesized structure in the paradigm is correct. If an arbrovirologist, medical entomologist or other experimenter interest focuses on geometric eigen-deomposable, wavelength, frequency, endmember, LULC,orthogonal, estimation of eco-epidemiological, time series, empirically geo-spatiotemporal or geo-spatiotemporal, geosampled, regression coefficients associated with seasonal, African, riceland, agro-irrigated, sylvatic, YFV transmission, then correlated longitudinal data (i.e., geocoordinates of prolific, discontinuous, partially, forest-canopied, sylvatic, *Ae. aegypti*, oviposition, geoclassified, sandwich estimators) must be accounted for within a covariance matrix which may be alternatively useable to provide efficient standard errors for the estimated coefficients.

The sandwich estimator, often known as the robust, covariance, matrix estimator or the empirical covariance matrix estimator, has achieved increasing use with the growing popularity of generalized estimating equations in entomology. Its virtue is that it renders consistent estimates of the covariance matrix for optimally parameterizing regressable, endmember covariate, eigen-decomposable LULC estimates (geo-spectrotemporally or geo-spatiotemporally geosampled, uncoalesced, African, riceland, eco-epidemiological, oviposition, seasonal,

hyperproductive, eco-georeferenced, sub-meter resolution, signature, wavelength frequencies) even when a parametric model fails to hold, or is not even specifed. Surprisingly though, there has been little discussion of the properties of the sandwich method other than consistency in any eco-epidemiological, YFV-related, geo-spectrotemporal or geo-spatiotemporal, geometric, time series, endmember, prognosticative, LULC, oviposition, signature, risk model. The sandwich estimator in quasilikelihood, riceland, sylvatic, *Ae. aegypti*, eco-epidemiological, geometric, endmember, sub-meter resolution, signature LULC models may be asymptotically, feasible for optimally discerning expanding agro-irrigated, geoclassifiable, sub-meter resolution, eigen-decomposable, orthogonal, geoclassifiable, geospatial or geo-spectral, endmember, LULC, components onto submeter resolution, discontinuous, partially, forest-canopied LULCs and in the linear case analytically.

In analytic geometry, an asymptote of a curve is a line such that the distance between the curve and the line approaches zero as they tend to infinity (Hosmer and Lemeshew 2002). Some sources include the requirement that the curve may not cross the line infinitely often, but this is unusual for modern authors). It may be shown that when the quasilikelihood, entomological, endmember, forecast, vulnerability, YF-related, sylvatic, capture point, *Ae aegypti*, oviposition, eco-epidemiological, LULC model is correct, the sandwich covariance matrix estimate is far more variable than the usual parametric variance, orthogonal, endmember, regressable estimate, and its coverage probabilities may be abysmal. The increased variance may be a fixed featureof the method, and the price an arbovirologist, medical entomologist or other YF experimenter may have to pay in terms of quantitating parametric model robustness to obtain consistency. Modifying the method in SAS or R may improve eco-epidemiological, eco-georeferenceable, *Ae aegypti*, capture point, LULC, oviposition, coverage probabilities.

Interpretability is not the only advantage of the linear, probability, explanatory LULC model for fitting an uncoalesced endmember dataset of *Ae. aegypti*, oviposition, seasonal, hypeproductive, capture point, sub-meter resolution eco-cartographically illustraable, inhomogeneous forest canopy and agro-irrigated, immature habitats. Another advantage is computing speed. Fitting a logistic, dichotomous, riceland, African, oviposition, eco-epidemiological, forecast, vulnerability, geometric, endmember, geo-spectrotemporal or geo-spatiotemporal, sylvatic, *Ae. aegypti* forest-canopy, risk model may be inherently slower as the model may be diagnostically fit only by an iterative process of the ML. The slowness of logistic regression is not noticeable if fitting a simple model to a small or moderate-sized entomological, vector, arthropod, time series, empirical geo-spattiotemrpoal or geo-spactrotemporal eco-epidemiological, geosampled dataset. But if an arbovirologist, medical entomologist or YF experimenter desires fitting a very complicated endmember, oviposition, eco-georeferenecable, geometric, sub-meter resolution, LULC model or a very large eco-epidemiological, grid-stratified, eigen-function, eigen-decomposable, *Ae. aegypti,* time series, African riceland agro-village, uncoalesced dataset, logistic regression can be frustratingly misspecified (see Jacob et al. 2005).

Linear, regression, model misspecifications (e.g., false positive, seasonal, hyperproductive, *Ae. aegypti,* oviposition geolocations) forecasted on a sub-meter resolution, grid-stratified, geosampled, African, expanding, riceland corridor adjacent geospatially to inhomogeneous, forest-canopy LULCs may be inconspicuous unless the arbovirologist, medical

entomologist or YF experimenter deduces the non-normalities (e.g., lepokurotic distributions) in the model summary statement. In many instances in entomological forecast vulnerability, time series, endmember, sub-meter resolution, LULC models, erroneous propogational outputs may not be gauged. If errors are non-normally distriuted in a YF, seasonal, hypeproductive, sub-meter resolution, endmember signature, capture point, non-continuous, African riceland, eco-epidemiological, signature model, then the forecasts rendered would be non-robust (e.g.,collinear LULC, oviposition,capture point, estimates) In cicumstances whenst geo-spectrotemrpoal or geo-spatiotemporal, eco-georeferenceable, YF, eco-epidemiological, seasonal, prognosticative, LULC signature models are constructed from sub-meter resolution endmember, wavelengths, residual, parameter estimator uncorrelatedness may be detected by usage of object-based geoclassifiaction algorithms (e.g., Spectral Angle Mapper in ENVI). In a geo-spatiotemporal or, geo-spectrotemporal, eco-epidemiolgoical, predictive, *Ae. aegypti,* oviposition, capture point, eco-eeidmiological, eco-georeferenceable, vulnerability , probabilistic, regression paradigm constructed in SAS (PROC REG) a stepwise backward elimination may determine independence amongst the YF, LULC parameter estimators.

Stepwise regression is a method of fitting models in which the choice of predictor variables (e.g., uncoalesced, *Ae . aegypti,* oviposition, YF , sub-meter resolution, African riceland and or inhomogeneous forest-canopied LULC discrete integer values) is carried out by an automatic process in SAS or R. In each step, an eco-epidemiological, YF seasonal geosampled, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological LULC variables would be considered for addition to or subtraction from the empirical dataset of eigen-decomposable, grid-stratified, sub-meter resolution, wavelength frequencies or other clinical diagnostic grid-stratifiable regressors based on some pre-specified criterion. Usually this would take place in the form of a sequences (e.g., F test or T Test). Other test are also possible for detecting unquantitated propogational, inconspicuous, endmembers and or parameter estimator deviations. The frequent practice of fitting the final selected LULC model followed by reporting estimates and confidence intervals without adjusting them to take the model building process into account has led to using stepwise model building (Hosmer and Lemeshew 2002) Other salient sequence distributions that may be applicable for constructing a robust. optimizable, YF eco-epidemiological, forecast, vulnerability , signature, sub-meter resolution, LULC model would be the adjusted R^2, Akaike information criterion, Bayesian information criterion, Mallows's C_p, PRESS, or false discovery rate

The linear probability model is more accurate by comparison as it can be estimated non-iteratively using ordinary least squares (OLS). This model would ignore the fact that the linear probability, riceland, African agro-village, expanding, vulnerability, eco-epidemiological, riceland, oviposition, discontinuous, forest-canopy, prognosticative model is heteroskedastic with residual variance $p(1-p)$, but the heteroscedasticity is minor if p is between .20 and .80, which is the situation when employing linear probability models at all. OLS estimates of eco-georeferenecable, eco-epidemiological, remote targets of seasonal, hyperproductive, *Ae. aegypti*, oviposition sites on geoclassifiable, sub-meter resolution LULCs can be improved by using heteroscedasticity-consistent standard errors or weighted least squares.

Heteroscedasticity-consistent standard errors may be usable for optimally regressively quantitating diagnostic fit of an agro-village, riceland, African, agro-ecosystem, sub-meter,

resolution, discontinuously canopied, geometeric, endmember, LULC, capture point, vulnerabiltiy model estimators that does contain heteroscedastic residuals. The first such approach was proposed by Huber (1967), and further improved procedures have been produced since for cross-sectional data, time-series data (e.g., geosampled, geo-spectrotemporal or geo-spatiotemporal, uncoalesced, sub-meter resolution, *Ae. aegypti,* capture point , YF African, riceland. Discontinuous forest-canopy model for targeting newly transition discontinuous, forest-canopied LULCs) and GARCH estimation.

Autoregressive conditional heteroskedasticity (ARCH) may be optimally interpreted as the condition when there are one or more seasonal, eco-georeferenceable, sub-meter resolution, goclassifiable, LULC, eco-epidemiological, capture points (e.g., geosampled, discontinuous, forest-canopied, and riceland agroecosystem, *Ae aegypti* oviposition, geoclassifiable stratified, orthogonal grids) in a series for which the variance of the current error term exists. Alternatively, ARCH innovation may be a function of the actual sizes of the previous geosampled time periods' parameterized error terms. In such circumstances in a YF eco-epidemiological, African ricland, vulnerability signature, LULC sub-meter resolution, immature, capture point,habitat model the variance may be related to the squares of the previous innovations in the endmember, time series forecasts. In remotely sensed LULC, vector arthropod, entomological bio-geophysical, endmember ARCH models, time series may be characterized (Griffith 2005). A variety of other acronyms are applicable for particular structures that have a similar basis ARCH model assumptions which may be robustly parsimoniously usable in time series, forecast, vulnerability, eco-epidemiological, endmember, YF signature modeling that exhibit time-varying volatility clustering, (i.e., periods of seasonal wavelengths in an oviposition, flooded, African, riceland agro-ecosystem geo-spectrotemporal, sub-meter resolution, geclassifiable, discontinuous, forest-canopied, LULC paradigm interspersed with periods of relative calm). ARCH-type models are sometimes considered to be in the family of stochastic volatility models, although this is strictly incorrect since at time t the volatility is completely pre-determined (deterministic) given previous values (Box and Jenkins 1985).

An eco-epidemiological, time series, weighted, ARCH, malarial, sub-meter resolution, eigen-decomposable, LULC, orthogonal, grid-stratified, risk map was generated in Jacob et al. (2013) using prevalence as a response variable for devising whether aggregate, eco-epidemiological, capture point, immature habitat counts digitally overlaid over an eco-geographical region subdivided by administrative boundaries (e.g., districts) could help prioritize areas for implementing an IVM control strategy in Uganda. Initially, univariate statistics and regression models were generated from empirical geosampled data to determine seasonal covariates (e.g., rainfall) related to monthly prevalence rates. Specific, eco-epidemiological, district-level,time series, prevalence measures were forecasted employing autoregressive LULC specifications and geo-spatiotemporal data collections for targeting district-evel, eco-georeferenceable geolocations that had higher prevalence rates. Case, as counts, were employed as a response variable in a Poisson probability model framework for quantitating the empirical datasets of district-level, explanatorial, ecogeoreferenceable LULC covariates (i.e., meteorological data, densities and distribution of health centers, etc.) sampled from 2006 to 2010 in Uganda. Results from both a Poissonian and a negative binomial (i.e., a Poisson random variable with a gamma distrusted mean) revealed that the geoclassified LULC covariates

rendered from the model were significant, but furnished virtually no predictive power. Inclusion of indicator variables denoting the time sequence and the district geolocation spatial structure was then articulated with Thiessen polygons which also failed to reveal meaningful paramterizable covariates.

Thereafter, in Jacob et al. (2013) an ARIMA model employing the empirical geosampled, district-level, datasets was constructed which revealed a conspicuous but not very prominent, first-order, temporal, autoregressive, endmember, LULC structure in the individual, district-level, time-series, dependent regressors. A random effects term was subsequently specified using monthly time-series dependent explanators. This specification included a district-specific intercept term that was a random deviation from the overall intercept term which was based on a draw from a normal frequency distribution. The random effects specification revealed a non-constant mean across the districts. This random intercept represented the combined effect of all omitted, endmember, LULC covariates that caused districts to be more prone to the malaria prevalence than other districts.

Additionally, in Jacob et al. (2013) inclusion of a random intercept assumed random heterogeneity in the districts' propensity or, underlying risk of malaria prevalence which persisted throughout the entire duration of the time sequence under study. This random effects term displayed no spatial autocorrelation, and failed to closely conform to a bell-shaped curve. The model's variance, however, implied a substantial variability in the prevalence of malaria across districts. The estimated model contained considerable overdispersion (i.e., excess Poisson variability): quasi-likelihood scale = 76.565. The following equation was then employed to forecast the expected value of the prevalence of malaria at the district-level: prevalence = exp[-3.1876 + (random effect)$_i$] . Compilation of accurate explanative, geo-spatiotemporal or geo-spectrotemporal, endemic, YF geoclassifiable, sub-meter resolution, LULC, African, riceland oviposition data along forest-canopied corridors in agro-irrigated grid-stratfied ecosystems may allow continual updating of the random effects term estimates allowing research intervention teams to bolster the quality of the LULC forecasts for district-level, *Ae aegypti*, ovisposition African riceland agro-irrigation, eco-epidemiological risk modelling efforts.

Robustness in a time series, elucidative, eco-epidemiological, YF, riceland, capture point, sylvatic, *Ae aegypti*, forecast, vulnerability model for optimally targeting, newly trasnitioned, oviposition, intermittent, forest-canopied, LULC, eco-georeferenceable seasonal, into hyperproductive habitat sites would be definable in SAS or R if the ability of the newly, transitioned, regressively, paramterizable, coavariate, geometric, endmember, time series, LULC effects remains consistent under misspecifications of the covariance structure. With large matrices, the inefficiency of the sandwich estimator may become worrisome for constructing, optimal, endemic, YF, geo-spatiotemporal or geo-spectrotemporal, geosampled, endmember, sub-meter resolution, geoclassifiable, LULC datasets of unbiased, forecast-oriented, seasonal, hyperproductive, oviposition, sylvatic, *Ae. aegypti* ,exploratory, uncoalesced, sub-meter resolution, iteratively, quantitatively interpolative, eco-epidemiological, vulnerability models,

An arbobiologist, medical entomologist or YFV experimenter may consider two general shrinkage approaches for optimally estimating the covariance matrix in SAS for optimally

regressively, targeting, hyperproductive, seasonal, eco-epidemiological, immature, *Ae aegypti*, eco-georerenceable, explanative, sub-meter resolution, LULC, capture points along a riceland, grid-stratifiable, partially, forest-canopied, discontinuous interface in an expanding, African, agro-irrigated, agro-ecosystem complex. The first would involve shrinking the eigenvalues of the unstructured ML or REML estimator. The second would involve shrinking an unstructured, riceland, LULC, orthogonal prognosticative estimator toward a structured estimator. For bothcases, the geosampled, geo-spectrotemporal or geospatiotemporal, uncoalesced endmember, LULC-specified, regression variables would determine the amount of shrinkage. These uncoalesced, oviposition, sub-meter resolution, wavelength, frequency, entomological signature estimators may be consistent and render asymptotically efficient estimates for optimally aggregating an eco-epidemiological, sylvatic, *Ae. aegypti* LULC, oviposition, endmember dataset of probabilistic, iteratively interpolative, time series, regression-related, eco-georerenceable, eigen-decomposable, coefficients parsimoniously. Simulations may demonstrate improved operating characteristics of time series, geoclassifiable shrinkage, sub-meter resolution, LULC, wavelength frequency , endmember ovisposition riceland estimators of the covariance matrix and their weighted, regressable, diagnostic, YF, time series, parameterizable, covariate coefficients in finite samples. The final, sub-meter resolution, diagnostic, geometric, endmember, LULC, oviposition, signature paradigm itartive interpolative, dependent variable as selected in PROC MIXED may include a combination of both approaches by shrinking the reference eigen-decomposed eigenvalues of the signature toward the endmember LULC structure where a hyperproductive discontinuous, forest-canopied, sylvatic, *Ae. aegypti,* seasonal, hyperproductive, capture point, inmature habitat occurs along an expanding, African, riceland, agro-ecosystem complex. In so doing, an arbovirologist, medical entomologist or YF experimenter may optimally remotely differentiate between structured and unstructured, eco-georerenceable, explanatorial, eco-epidemiological, oviposition, seasonal, hypeproductive, endmember, LULC, explanatory, orthogonally sub-meter resolution, grid-stratifiable, YFV, oviposition, geo-spectrotemrpoal or geo-spatiotemrpoal, frequency, wavelength, eigenvector, habitat, decomposed, frequency estimators.

PROC MIXED implements two likelihood-based methods: ML and REML (www.sas.edu). A favorable theoretical property of ML and REML is that they accommodate data that are missing at random (see Rubin 1976; Little 1995). PROC MIXED may optimally construct an objective function associated with an explantory, ML or REML, in an expanding, African, riceland, agro-village, sylvatic, *Ae. aegypti,* forecast, vulnerability, eco-epidemiological, oviposition, geometric, endmember, sub-meter resolution, LULC, signature model and maximize it over all unknown geo-spatiotemporal or geo-spectrotemporal, geosampled, geoclassifiable, optimizable, sub-meter resolution, LULC, forecast, canopy, discontinuous parameters.

Employing calculus, it is possible to reduce maximization problems applying one over an empriical dataset of only uncoalesced, eco-epidemiological, *Ae. aegypti*, time series, diagnostic, sub-meter resolution, YFV-related, eigen-decomposable, parameterizable, LULC endmember, signatureizable, explanatory estimators in $\mathbf{G}$ and $\mathbf{R}$ employing the corresponding log-likelihood, elucidative, functions $$\mathrm{ML}: \quad l(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\mathbf{r}'\mathbf{V}^{-1}\mathbf{r} - \frac{n}{2}\log(2\pi)$$ and

REML : $l_R(\mathbf{G},\mathbf{R}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{r}'\mathbf{V}^{-1}\mathbf{r} - \frac{n-p}{2}\log(2\pi)\}$ where the equation $\mathbf{r} - \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and p is the rank of $\mathbf{X}$. PROC MIXED actually minimizes -2 times these functions by employing a ridge-stabilized Newton-Raphson algorithm. Lindstrom and Bates (1988) provide reasons for preferring Newton-Raphson to the Expectation-Maximum (EM) algorithm described in Dempster, Laird, and Rubin (1977) and Laird, Lange, and Stram (1987), as well as eigen-analytical, orthogonally decomposable, endmember, LULC, sub-meter resolution, grid-stratifiable, details for implementing an algorithmic, probabilistic approach to the problem of geometerically, remotely identifying discontinuous, forest-canopied, sub-meter, resolution, geoclassifiable, LULC, capture point, oviposition, YF, eco-epidemiological sites in expanding riceland agro-village, African, agro-irrigated ecosystems. Wolfinger, Tobias, and Sall (1994) present the sweep-based algorithms that are also implementable in PROC MIXED for conducting the same task.

One advantage of employing the Newton-Raphson algorithm for seasonally probabilistically optimally quantitating eco-epidemiological, eco-georeferenecable, eigen-decomposable, orthogonal, seasonal, hyperproductive, capture point, *Ae. aegypti,* eco-georferenceable, oviposition foci on time series, sub-meter resolution, satellite images of expanding African, riceland, agro-ecosystem environments is that the second derivative matrix of the objective function evaluated at the optimal is available upon completion in SAS. This software provides a powerful programming language with components called procedures. Procedures can enable an arbovirologist, medical entomologist or YF researcher to perform many different types of endmember, sub-meter, eco-epidemiological, orthogonal, endmember, risk analysis and LULC data, management functions. Procedures also may render many different types of text-based and graphical seasonal, eco-georeferenceable, African, riceland, agro-irrigated, ecosystem, capture point, hyperproductive habitats on discontinuous, forest-canopied, sub-meter resolution, gridded, LULC, endmember outputs. Combined with other seasonal, eco-epidemiological, eco-georeferenecable, geo-spatiotemporal or geo-spectrotemporal, YFV-related, African, riceland, agro-village complex, endmember, *Ae. aegypti,* uncoalesced, oviposition, sub-meter resolution, LULC-related, attribute features, the SAS language and its procedures may render an immense variety of regressively applicable platforms for remotely, optimally, targeting, seasonal, hyperproductive, immature,eco-georeferenceable, capture point, immature habitats including the following examples:1)Access raw data files and sub-meter resolution, LULC data in external databases and cyberdatabase management systems, 2)Manage data using tools for data entry, editing, retrieval, formatting, and endmember conversion of geoclassifiable, African, riceland, oviposition, LULCs into discontinuous forecast canopy,3) Analyze geoclassifiable, geo-spatiotemporal or geo-spectrotemporal, sylvatic, *Ae. aegypti,* geomorphological, terrain-related and meteorological data employing descriptive statistics and multivariate techniques for optimally constructing optimal, forecast, vulnerability, sub-pixel, eco-epidemiological, signature models employing linear programming. SAS is also portable across computing environments. The applications would function the same and produce the same results regardless of the operating environment on which an arbovirologist, medical entomologist or YF experimenter is running SAS/GIS to process the sub-meter resolution, geoclassifiable LULC, riceland, African, ecosystem data.

Whilst denoting a robust, regression, grid-stratified, orthogonal matrix $\mathbf{H}$, in a oviposition, forecast, vulnerability, sub-meter resolution, African agro-village ecosystem, probabilistic, time series, YFV, endmember, signature paradigm reflecting transitioning riceland, *Ae. aegypti,* immature, capture point, immature habitats into inhomogeenous, forest canopy, seasonal, hyperproductive, geoclassifiable LULC, seasonal sites, the asymptotic theory of ML may reveal that $2\mathbf{H}^{-1}$ is an asymptotic variance-covariance matrix of the probabilistically estimated, YF gesampled, parameters of $\mathbf{G}$ and $\mathbf{R}$. Hence, tests and confidence intervals based on asymptotic normality in an optimally regressable, eco-epidemiological, orthogonal, grid-stratifiable, endmember, eigen-decomposable, eco-georeferenceable dataset of sub-meter resolution, oviposition, wavelength, frequency, geoclassifiable, LULC, sylvatic, *Ae. aegypti*, geo-spatiotemporal or geo-spectrotemporal, geomorphological, geoclassifiable, forecast, vulerability models, uncoalesced, wavelength, frequency, parameterizable estimators can be quantitatively obtained. However, these probabilistic, time series dependent, eco-epidemiological *Ae. aegypti*, oviposition, geoclassifiable, LULC regressors may be unreliable in small geosampled, forecast, vulnerability, endmember, signature datasets especially for revealing partially discontinuous, geoclassifiable, forest-canopied, geometric endmember, LULC parameters along an eco-epidmiological, expanding African, agro-irrigated, riceland, agro-village complex, such as variance components that have sampling distributions that tend to be skewed to the right.

If a residual variance σ^2 is a part of an mixed oviposition, eco-epidemiological, geo-spatiotemporal or geo-spectrotemporal, vulnerability, geometric, endmember, LULC, endemic, YFV, probability, oviposition, signature model, it can usually be profiled out of the likelihood. This may require solving analytically for the optimal σ^2 and plugging this expression back into the likelihood formula in PROC REG. This would reduce the number of optimization, eco-georeferenceable, sylvatic, *Ae. aegypti,* sub-meter,resolution, geoclassfied, LULC parameters by one which may improve convergence properties. PROC MIXED may also profile the forecasting vulnerability, time series, endmember, YFV model, sub-meter resolution, signature, residual variance out of the log-likelihood whenever it appears reasonable to do so. This includes the case when $\mathbf{R}$ equals $\sigma^2 \mathbf{I}$ and when it has blocks with a compound symmetry, time series, or spatial structure. PROC MIXED would not profile the log-likelihood whenst $\mathbf{R}$ has unstructured blocks, whenst an arbovirologist, medical epidemiologist or YFV researchist employs the HOLD= or NOITER option in the PARMS statement, or whenst he or she uses the NOPROFILE option in the PROC MIXED statement.

Instead of ML or REML, an arbovirologist, medical epidemiologist or YFV researchist can employ the non-iterative MIVQUE0 method to estimate $\mathbf{G}$ and $\mathbf{R}$ in an probabilistic, sylvatic, *Ae. aegypti,* oviposition, eco-epidemiological, forecast, vulnerability, LULC, geo-spectrotemporal or geo-spatiotemporal, endmember, signature paradigm. In so doing, PROC MIXED would use MIVQUE0 estimates as starting values for the ML and REML procedures for optimally regressively quantitating YFV-related, geo-spatiotemporally or geopectrotemporally, geosampled, geoclassifiable, explanatory, endmemberm LULC variables. For computating geometric, endmember, variance component, vulnerability, in the diagnostic, eco-epidemiological, YFV models, another estimation method in SAS may be performed that involves equating Type 1, 2, or 3 as expected mean squares to the observed, forest-canopy, and

expanding, agro-village, African, riceland expanding, seasonal, geosampled, LULC values and solving the resulting system. However, Swallow and Monahan (1984) present simulation evidence favoring REML and ML over MIVQUE0 and other method-of-moment estimators.

An arbobiologist, medical entomologist or YFV researchist may address the computational difficulties raised by incorporating priors, and nonconjugate priors into hierarchical, African, riceland, agro-village, forest-canopy, discontinuous, *Ae. aegypti,* oviposition, eco-epidemiological, LULC models. A combination of approximation, Gibbs sampling (possibly with a Metropolis step), and importance reweighting may be optimally employable to fit the sylvatic, oviposition, endmember, discontinuous, forest-canopy, eco-epidmiological, wavelength, frequency, signature, LULC models and compare them with a realization rendered from a hybrid approach to alternative Markov Chain Monte Carlo (MCMC) methods.

The Wishart distribution is frequently employable as the prior on the precision matrix parameter of a multivariate, normalized, geometric, endmember, LULC, oviposition, sylvatic, *Ae aegypti,* seasonal distribution. Because the gamma distribution is the conjugate prior for the precision parameter of a univariate normal distribution(Hosmer and Lemeshew 2002), the Wishart distribution (as its multivariate generalization) would extend conjugacy to ta multivariate, forecast, normalized distribution as rendered from an YFV, eco-epidemiological, eco-georeferenecable, forecast, vulnerability, Bayesian model. In statistics, the inverse Wishart distribution, also called the inverted Wishart distribution, is a probability distribution defined on real-valued positive-definite matrices[2]. In statistics, the inverse Wishart distribution, also called the inverted Wishart distribution, is a probability distribution defined on real-valued positive-definite matrices.

Wishart distribution $W(\Sigma, d, n)$ is a probability distribution of random nonnegative-definite $d \times d$ matrices that is used to model random covariance matrices. The parameter n is the number of degrees of freedom, and Σ is a nonnegative-definite symmetric $d \times d$ matrix that is called the scale matrix. By definition $W \approx W(\Sigma, d, n) \approx Xn$ i=1 $XiX0$ i , $Xi \approx N(0, \Sigma)$ (1.1) so that $W \approx W(\Sigma, d, n)$ is the distribution of a sum of n rank-one matrices defined by independent normal $Xi \in Rd$ with $E(X) = 0$ and $Cov(X) = \Sigma$ according to Jacob et al. (2013) for identifying captuire points on geoclassifiable, grid-startifiable vector entomological, forecast vulnerability, eco-epidemiological, remotely sensed, sub-meter resolution data In particular $E(W) = nE(XiX0$ i) = n $Cov(Xi) = n\Sigma$ (1.2) may be applicable for remotely targeting seasonal hypeproductive foci. In general, any $X \approx N(\mu, \Sigma)$ can be represented $X = \mu + AZ$, $Z \approx N(0, Id)$, so that $\Sigma = Cov(X) = A$ $Cov(Z)A$ $0 = AA0$ .The easiest way to find A in terms of Σ in a YF, eco-epidemiological, *Ae. aegypti*, oviposition, African, riceland, agro-ecosystem, forecast vulnerability, endmember LULC model for targeting prolific, capture points on geoclassifiable, sub-meter resolution, forest-canopied, discontinuous LULCs is based on the prcise eigen-decomposition of the signature estimators which may find a unique lower diagonal matrix A with Aii ≥ 0 such that AA0 = Σ. Then by (1.1) and (1.2) with $\mu = 0$ $W(\Sigma, d, n) \approx Xn$ i=1 $(AZi)(AZi)$ $0 \approx A$ μXn i=1 ZiZ 0 i ¶ A 0 , $Zi \approx N(0, Id) \approx A$ $W(d, n)$ A 0 where $W(d, n) = W(Id, d, n)$ (1.3) In particular, $W(\Sigma, d, n)$ can be easily represented in an eco-epidemiological, YF forecast, vulnerability geo-spectrotemporal or geo-spatiotemporal endmember, sub-meter resolution, forecast, vulnerability

model in terms of W(d, n) = W(Id, d, n). Assume in the model that n > d and Σ is invertible. Then the density of the random d × d matrix W in (1.1) can be written in SAS as f(w, n, Σ) = |w| (n−d−1)/2 exp¡ −(1/2) tr(wΣ −1) ¢ 2 dn/2π d(d−1)/4|Σ| n/2 Qd i=1 Γ ¡ (n + 1 − i)/2 ¢ (1.4) where |w| = det(w), |Σ| = det(Σ), and f(w, n, Σ) = 0 unless w is symmetric and positive definite .

In Bayesian inferencial statistics an inverse wishart distribution may be employable as the conjugate prior for the covariance matrix of a multivariate normal distribution. Hence, spatial Bayesian statistics may be usable for remotely targeting geo-specrotemporal or geo-spatiotemproal, geoclassifiable, hyperproductive, oviposition, *Ae. aegypti,* geometric, endmember, seasonal, LULC sites on grid-stratifiable, sub-meter resolution, satellite imagery employing the conjugate prior for the covariance matrix of a multivariate normal distribution.

A flexible class of prior distributions may be proposed for optimally remotely quantitating sub-meter resolution, discontinuous, forest-canopied, expanding, riceland oviposition, geoclassifiable LULCs employing a covariance matrix in PRINCOMP. The PRINCOMP procedure may perform a principal component analyses employing an empirical, endmember, eco-epidemiological, dataset of uncoalesced, orthogonal, African, riceland, oviposition, geoclassifiable, eco-georeferenceable, sub-meter resolution LULC, geometric endmembers. As input, time series, eco-epidemiological, eigen-decmposable, *Ae. aegypti,* uncoalesced, sub-meter resolution, LULC variables, an arbovirologist, medical entomologist or YF experimenter may quantitate geosampled, riceland agro-irrigation.oviposition, geoclassifiable, iterative, interpolative, endmember,LULC, raw data using a correlation matrix, a covariance matrix, or a sum of squres and cross-products (SSCP) matrix. He or she may create an output, explanatory, endmember, forecast, vulnerability, eco-georeferenecable, discontinuous, forest-canopied, African, riceland, parameter estimator *Ae .aegypti* ovispoition dataset of newly transitioned partially shaded, orthogonal, eigen-decomposed, LULC, eigenvectors into riceland, agro-irrigated, orthogonal eigenvalues employing standardized or unstandardized principal components scores.

Principal component, explanatory, endmember sub-meter resolution grid-stratifiable, orthogonal eigen-analyses may be an optimal methodology for quantitating multivariate, sub-meter, resolution, endmember effects   amongst geo-spatiotemporal or geo-spectrotemporal, iteratively interpolative, quantitative, geo-spectrotemrpoal or geo-spatiotemporal, LULC variables. The choice between employing factor analyses and principal component analyses for constructing a sub-meter resolution, African, riceland, eco-epidemiological, YF , forecast, vulnerability model may depend on the depth of the frequentistic responses required for extrapolating robustifiable, eco-epidemiological, eco-georferenceable, forecasts of seasonal, hyperproductive, sylvatic, capture points, *Ae aegypti*, oviposition foci. Principal component analyses was created for non-high performing computer. Conversely factor analyses may be optimally employable to describe variability amongst observed, correlated, eco-georeferenceable, sub-meter resolution, agro-irrigated African, riceland, endmember, LULC variables in terms of a potentially lower number of unobserved variables (i.e, factors**)**. For example, it may be possible that variations in six observed, discontinuous, forest-canopied, seasonal, grid-stratified, endmember, LULC variables mainly reflect the variations in two unobserved (underlying), African, riceland, eco-epidemiological, eco-georeferenceable LULC inhomogeneous, forest-canopied variables. Factor analysis would search for such joint variations

39

in response to unobserved LULC, explanatory, non-zero autocorrelated, latent variables in the geosampled dataset. The observed variables would be modelled as linear combinations of the potential factors, plus "error" terms. Factor analysis would then find independent latent, explanatory, agro-irrigated, time series, African riceland, agro-village, *Ae. aegypti*, oviposition, sub-meter resolution, endmember, LULC variables that are geostatistically related to discontinuously canopied, forest, periphery LULCs.

PRINCOMP may summarize sylvatic, YF, eco-epidemiological, endmember, LULC data for detecting quantitative, grid-stratifiable, linear relationships between geosampled, geo-spatiotemporal, or geo-spectrotemporal, eco-georeferenceable, orthogonal, *Ae. aegypti*, riceland and forecast-canopied, optimally regressively parameterizable covariates. The application of principal components is discussed by Rao (1964), Cooley and Lohnes (1971), and Gnanadesikan (1977). Excellent statistical treatments of principal components are found in Kshirsagar (1972), Morrison (1976) and Mardia, Kent, and Bibby (1979).

A likelihood approximation may be parsimomnioulsy obtainable employing a matrix logrithm of the covariance matrix via Bellman's iterative solution may be employable for optimally targeting seasonal hypeproductioev Ae aegypti, ovispoition, capture points on sub-meter resolution, grid-stratifiable, African riceland agro-village on newly transitioned inhomogeneous, forest canopied habitats.The Bellman–Ford algorithm is an algorithm that computes shortest paths from a single source vertex to all of the other vertices in a weighted digraph. It is slower than Dijkstra's algorithm for the same problem, but more versatile, as it is capable of handling YF grid-stratfied graphs in which some of the edge weights (discontinuous, forest-canopied, geoclassifiable LULC, discrete integers) are negative numbers. Bellman–Ford is based on the principle of relaxation, in which an approximation to the correct distance (e.g., Euclidean distance from a discontinuous canopy, forestland, *Ae aegypti,* capture point to a African, riceland, orthogonal centroid) is gradually replaced by more accurate values until eventually reaching the optimum solution. In both algorithms, the approximate distance to each vertex is always an overestimate of the true distance, and is replacable by the minimum of its old value with the length of a newly found path. However, Dijkstra's algorithm uses a priority queue to greedily select the closest vertex that has not yet been processed, and performs this relaxation process on all of its outgoing edges; by contrast, the Bellman–Ford algorithm simply relaxes all the edges, and does this times, where is the number of vertices in the graph. In each of these repetitions, the number of vertices in an eco-epidemiological, sub-meter resolution, grid-stratifiable, orthogonal, YF-related, forecast, vulnerability, endmember, model with correctly calculated grid-stratified distances grows, from which it follows that eventually all vertices will have their correct distances. This method allows the Bellman–Ford algorithm to be applied to a wider class of inputs than Dijkstra for targeting, Bayesianized, seasonal, hyperproductive, *Ae. aegypti*, African riceland, agro-village, eco-epidemiological, capture points on newly transitioned forest-canopy, peripheral corridors

In Bayesian probability theory, if the posterior distributions $p(\theta|x)$ are in the same family as the prior probability distribution $p(\theta)$, the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood function. For example, the Gaussian family is conjugate to itself (or self-conjugate) with respect to a Gaussian likelihood function: if the likelihood function is Gaussian, choosing a Gaussian prior over the

mean will ensure that the posterior distribution is also Gaussian. This means that the Gaussian distribution in an   African geoclassified, LULC, riceland, YF, sub-resolution, forecast, vulnerability endmember model for targeting seasonal hyperproductive, capture point, endemic, *Ae. aegypti.* oviposition foci would be a conjugate prior for the likelihood that is also Gaussian.

Let the likelihood function be considered fixed; the likelihood function is usually well-determined from a statement of the data-generating process in a forecast, oriented, risk model. It is clear that different choices of the prior distribution $p(\theta)$ may make the integral more or less difficult to calculate, and the product $p(x|\theta) \times p(\theta)$ may take one algebraic form or another inan African, rice-agro-ecosystem, seasonal, *Ae. aegypti*, eco-epidemiological, forecast, vulnerability, LULC model. For certain choices of the prior in the eco-epidemiological, YF model there would be the posterior which may have the same algebraic form as the prior (generally with different parameter values). Such a choice would be termed a conjugate prior. A conjugate prior is an algebraic convenience, giving a closed-form expression for the posterior; otherwise numerical integration may be necessary (Gelman 1995). Further, conjugate priors for a seasonal, YF, eco-epidmiological, forecast, vulnerability, endmember, sub-meter resolution model for targeting, endemic, oviposition, riceland and or inhomogeneous, forest-canopy,inhomogeneous, LULC sites may render robust intuitions (seasonal, hyperproductive,agro-irrigation, hyperproductive geolocations) by more transparently showing how a likelihood function updates a prior distribution in the model. All members of the exponential family have conjugate priors (Cressie 1993).

The width of a confidence interval for conducting a robust, geo-spectrotemporal, or geo-spatiotemporal, endemic, geosampled, YF meta-analysis in SAS for optimally targeting seasonal, hyperproductive, sylvatic, *Ae. aegypti,* geoclassifiable, geometric, endmember, oviposition, LULC sites in an eco-georeferenceable, ArcGIS-derived, sub-meter resolution, seasonal, grid-stratifiable, orthogonal, time series, geospatial cluster of partial, discontinuous, sub-meter resolution, forest-canopied, LULCs along agro-irrigated, African, expanding, riceland agro-ecosystems would depend on the precision of the individual geosampled, eco-epidemiological, capture point, count density, regressable estimates. Commonly this data would be overlaid digitally onto the uncoalesced, descriptive, satellite, geoclassified, eco-epidemiological, grid-stratified, endmember, LULC data. In addition, for parsimonious, robust, endmember optimal quantitation of random-effects LULC realizations from a forecast, vulnerability, geo-spectrotemporal or geo-spatiotemporal, sylvatic, *Ae. aegypti,* oviposition, geometric, frequency endmember, LULC model, precision,empirical, YF-related, optimizable, dataset of parameterizable, wavelength, frequency, sub-resolution covariates will decrease with increasing heterogeneity and confidence intervals will widen correspondingly based on the geomorphological variation and other geosampled eco-georeferenceable variables (e.g., weekly meteorological unmixed components). Conversely, as temporality in added to the SAS, endmember, YF, forecast, vulnerability, LULC, meta-analysis, the width of the confidence interval could decrease. However, if the additional, sub-meter resolution, wavelength, frequency, irradiance, sub-resolution LULC data increases the heterogeneity in the meta-analysis , a random-effects, time series, eco-epidemiological, geo-spectrotemporal or geo-spatiotemporal,YFV model may be employed to exploit the geosampled, agro-irrigated, African, riceland, discontinuous, forest-canopied covariates which may make it possible to increase confidence interval width.

Confidence intervals and point estimates may have different interpretations in fixed-effect and random-effects time series, eco-epidemiological, wavelength, frequency, remotely sensed, YF, sub-meter resolution, seasonal, transmittance, *Ae. aegypti,* oviposition, LULC, discontinuous, sylvatic models. Whilst the fixed-effect, regression, optimizable explanatory estimate in the forecast-oriented, vulnerability, probabilistic signature paradigm and its confidence interval may address the question 'what is the best (single), sub-meter resolution, and/or estimate for optimal targeting of eco-georeferenceable, sparse, forest-canopied, prolific, seasonal, hyperproductive, sylvatic, *Ae. aegypti,* endmember, oviposition LULC sites along a agro-irrigated, African, riceland corridor, the random-effects estimate would assume there to be a distribution of regression effects, and hence the estimate and its confidence interval may determine the estimate of the effect in the eco-georeferenceable, residual, diagnostic, forecast dataset of targeted, hyperproductive, seasonal, capture point, tabulated, immature productivity.

A confidence interval may be reportable for any level of confidence (although they are most commonly reported for 95%, and sometimes 90% or 99%) for determining robustness of an eco-epidemiological, YF, forecasting, optimizable, remotely sensed, sub-meter resolution, vulnerability, explanatory, diagnostic, optimizable LULC model and its parameterizable, geometric, endmember, covariate, geoclassifiable, paramterizable, estimator datasets. For example, the odds ratio of 0.80 could be reported with an 80% confidence interval of 0.73 to 0.88; a 90% interval of 0.72 to 0.89; and a 95% interval of 0.70 to 0.92 for targeting, hyperproductive, *Ae. aegypti,* oviposition sites on geoclassifiable, partially, discontinuous, forest-canopy, sub-meter resolution, endmember, LULC, risk maps. As the confidence level increases, the confidence interval widens (Rao 1965).

There may be logical correspondence between the confidence interval and the P value in an unbiased, eco-georeferenceable,optimizable, sub-meter resolution, uncoalesced dataset of agro-irrigated, African, expanding, sub-meter, spatial resolution, forecast-oriented, vulnerability, eco-epidemiological, parameterizable, frequency, wavelength, YF estimators for geospatially quantitating levels of African, riceland, oviposition sites on geoclassified discontinuous, partially geoclassified, forest-canopied LULCs. The 95% confidence interval for an effect may exclude the null value (such as an odds ratio of 1.0 or a risk difference of 0), if and only if the test of significance yields a P value of less than 0.05. If the P value is exactly 0.05 in the explanatorial, diagnostic, time series, orthogonal, residual, eco-georeferenceable, eco-epidemiological, grid-stratified, LULC, sub-meter resolution, oviposition forecasts, then either the upper or lower limit of the 95% confidence interval will be at the null value in the entomological model. Similarly, the 99% confidence interval will exclude the null if and only if the test of significance yields a P value of less than 0.01 in the forecast, vulnerability, eco-georeferenceable, endemic, forecast, model output (e.g., hyperproductive, foci, of seasonally, *Ae. aegypti*, geo-spatiotemporal or geo-spectrotemporal, geoclassifiable, endmember, geometric, oviposition, LULC, capture points) in an irrigated, African, riceland, expanding, agro-village complex.

In SAS, the regression, explanative, oviposition, eco-epidemiological, seasonal, hyperproductive, capture point, regression estimate and confidence interval may provide information to assess time series, eco-epidemiological, diagnostic, clinical, field or remote-specified, sub-meter resolution, LULC, signature, parameterizable signature covariates

usefulness in an sylvatic, *Ae. aegypti,* oviposition, geo-spatiotemporal or geo-spectrotemporal, geoclassifiable, ArcGIS interventional, forecast, vulnerability, eco-epidemiological model. For example, suppose that an arbovirologist, medical entomologist or other experimenter is evaluating a treatment that reduces the risk of sylvatic YF by remotely targeting an eco-georeferenceable, sub-meter resolution, explanative, geospatial cluster of seasonal, hyperproductive, *Ae. aegypti,* oviposition, LULC sites at the peripheral of an agro-irrigated, African, discontinuously canopied, riceland agro-ecosystem. In order for the probabilistic, time series, regresseable grid-stratifiable, LULC data to be useful, the quantitated, YF-related, geo-spectrotemporal or geo-spatiotemporal, endmember, sub-meter resolution, diagnostic, geoclassified, geosampled variables would have to help reduce the risk of an endemic transmission event from 30% by at least 5 percentage points to 25% (these values will depend on the specific clinical scenario and diagnostic outcome), for example.

If the sylvatic, LULC, oviposition, sub-meter resolution,*Ae. aegypti,* YF-specified, eco-epidemiological, meta-analysis yields an effect estimate of a reduction of 10 percentage points with a tight 95% confidence interval, say, from 7% to 13% for an expanding, agro-irrigated, African, riceland agro-environment, forecast, vulnerability, endmember, seasonal, LULC model, an arbovirologist, medical entomologist or another experimenter may be able to conclude that the control stragey was useful since both the point estimate and the entire range of the interval exceeded a criterion of a reduction of 5% for diagnostic usefulness. However, if the YF meta-analysis reports the same risk reduction of 10% but with a wider interval, say, from 2% to 18%, in prevalence for the riceland eco-epidemiological, study site the optimal estimate of the effect of treatment, would not be so confident as there would be no exclusion of the possibility that the effect could be between 2% and 5%. If the confidence interval was wider in the explicative, optimizable, residual, eco-epidemiological, eco-georeferenecable, forecasts (e.g., remotely targeted, oviposition, seasonal, *Ae. aegypti,* hyperproductive foci on a sparsely shaded, sub-meter resolution, geoclassified, endmember, LULC strip along a ricefield African corridor) and include the null value of a difference of 0%, the arbovirologist, medical entomologist nor any YF experimenter could exclude the possibility that the control treatment has any effect whatsoever. Confidence intervals with different levels of confidence can demonstrate that there is differential evidence for different degrees of benefit or harm (Rao 1972).

In particular, if the variance of the errors is increasing over time, confidence intervals in a, sub-meter resolution, YF, vulnerability, probabilistic, eco-epidemiological, orthogonal LULC, paradigm constructed for quantitating agro-irrigation, oviposition covariates on discontinuous, forest-canopied, peripheral, African, riceland regions for out-of-sample predictions will tend to be unrealistically narrow. Heteroscedasticity in such a geometric, probabilistic paradigm may have the effect of giving too much weight to a small subset of geosampled, diagnostic, sylvatic, YF, explanatorial, clinical, field or remote, geo-spatiotemporal, geo-spectrotemporal, LULC, frequency, orthogonal, wavelength,uncoalesced, grid-stratifiable, sub-meter reoslution data (namely the subset where the error variance is the largest) whenst optimally estimating the sub-meter resolution, iterative interpolative, explanative, endmember, geometric, LULC coefficients. It might be possible to report the same analysis results (i) with 95% confidence that the intervention does not cause harm; (ii) with 90% confidence that it has some effect; and (iii) with 80% confidence that it has an important endemic transmission benefit. These elements may

suggest both usefulness of the intervention and the need for additional, YF research in expanding, agro-irrigated, African, riceland, agro-village environment due to anthropogenic pressure.

Diagnosing, non-explanatorial, time series, heteroskedastic, orthogonal, wavelength, frequency parameters in an eco-epidemiological, geo-spectrotemporal, sub-meter resolution, geo-spatiotemporal, explicative, forecast, vulnerability dataset of eco-georefernceable, YF, model, unbiased estimators would require devising a plot of residually predicted, optimally fractional, geometric, explicative, endmember, time series, LULC values. Importantly evidence of propagational residuals that grow larger either as a function of time or as a function of the prognosticative, eco-epidemiological, time series, uncoalesced, endmember, orthogonal, signature, regressed explanators can cause frequency mispecifications in a target dataset of seasonally explanative, hyperproductive,uncoalesced, sylvatic, capture point, *Ae. aegypti*, oviposition, sub-meter resolution, geoclassifiable explanative, clinical, field or remote, geo-spectrotemporal or geo-spatiotemporal, geosampled, predictively mapped, sub-meter resolution, LULC sites. To be really geostatistically thorough, however, an arbobiologist, medical entomologist or YF researcher could also generate experimental, explanative, time series, eco-epidemiological, eigen-decomposable, orthogonal, synthetic, spatial filter, residual, explanatory, diagnostic plots of residuals versus independent variables to optimally quantitate consistency in the endmember model dataset. Because of imprecision in uncoalesced, biased, diagnostic, sub-meter resolution, LULC, covariate coefficient, endmember, seasonal, wavelength, YF-related, frequency, estimator , sub-meter resolution datasets, may tend to be slightly larger for diagnostic, geo-spectrotemporal or geo-spatiotemporal, model outputs associated with eco-epidemiological forecasts (e.g., regressed values of independent explanatorial variables that are extreme in both directions) although the effect may not be too dramatic. If the explanatory, dependent variable (e.g., prevalance of YF in a riceland agro-village, environment expanding complex is strictly positive, and if the residual-versus, predicted, diagnostic, explanatorial plot shows that the size of the non-Gaussian, probabilistic, uncertainities is proportional to the size of the YF-related, sub-meter resolution, optimally derived, elucidative, LULC, hyperproductive, eco-epidemiological, capture point, seasonal, vulnerability predictions (i.e., if the regression errors seem consistent in percentage rather than absolute terms), a log transformation in SAS applied to the dependent variable may be appropriate.

The BOXCOX in PROC TRANSREG may perform a transformation of a cature point, seasonal, hyperproductive, oviposition, African riceland, *Ae. aegypti*, eco-georeferenceable, geoclassifiable, LULC geosampled, forest-canopied, discontinuous, sub-meter resolution, orthogonal, time series, diagnostic, LULC-oriented, response, regressed variable x. This variable may be then indexed by λ, and optimally defined thereafter as $x'_\lambda = \dfrac{x^\lambda - 1}{\lambda}$. (Equation 1.1) At first glance, although the formula in Equation (1.1) is a scaled version of the Tukey transformation x^λ, this transformation does not appear to be the same as the Tukey formula **:**

$$HSD = \frac{M_1 - M_2}{\sqrt{MS_w \left[\frac{1}{n}\right]}}$$

where HSD may be the significant, unbiased, explanative, parameter estimator in a vulnerability, eco-epidemiological, *Ae. aegypti,* oviposition, forest-canopied, discontinuous,

orthogonal, eigen-decomposable, time series, African, agro-irrigated, riceland, forecasting, vulnerability, endmember,sub-meter resolution, LULC, wavelength, frequency, uncoalesced, signature, probabilistic paradigm.

Freeman-Tukey (FT) transforms seek to adjust data to make the distribution more similar to a Normal distribution. The basic square root transform is a form of power transform, but the FT variant (see Freeman and Tukey, 1950) was specifically designed for Poisson-like data, especially with a mean value >1. The FT angular or arcsine transform was developed for Binomial-like data, (e.g., presence or absence of YF in an, African, riceland ,expanding, agro-irrigated ecosystem); in particular, empirical endmember, orthogonal, sylvatic, *Ae. aegypti*, paramter estimator dataset (e.g., geo-spatiotemporal or geo-spectrotemporal geosampled, sub-meter resolution, geoclassified, LULCs) representing proportions or percentages.

The angular or arcsin transform applies to source data in the range [-1,1] or more commonly in the range [0,1] and is designed to spread the set of values near the end of the range. Frequently data in the [0,1] range will represent proportions (Cressie 1993).However, these should be genuine measurements and not derived values that are essentially nominal, (e.g. presence/absence data). Once transformed data (e.g., orthogonally eigen decomposed sub-meter resolution, African, riceland, geo-spectrotemporal or geo-spatiotemporal, sub-meter resolution, geoclassified, inhomogeneous, forest canopy, *Ae. aegypti,* oviposition LULCs) are scaled to the range $[-\pi/2,+\pi/2]$. k is typically 0.5 and thus can provide the square root in an eco-epidemiological, YF, forecast, vulnerability, rice-agro-ecosystem, eco-georeferenceable, LULC model. This transform is often used to correct S-shaped relationships between response and explanatory variables.

Tukey suggests exploring simple relationships such as $y = b_0 + b_1 X^{\lambda}$ or $y^{\lambda} = b_0 + b_1 X$ where λ is a parameter (e.g., African, riceland, agro-irrigated, empirical, orthogonal, endmember, sylvatic, *Ae. aegypti,* oviposition, geosampled, LULC variable) chosen to make the relationship as close to a straight line as possible (Abramowitz and Stegun, 1972). If a transformation of the type x^{λ} or y^{λ} work as in Equation (1.1) whenst for robustly optimally regressively quantitating, immature, *Aedes* productivity on geoclassifiable, SAS/GIS-derived, sub-meter resolution, forest-canopied, discontinuous, riceland, geoclassfied LULCs along an expanding, irrigated, riceland African agro-ecosystem, then an arbobiologist, medical entomologist or YFV researcher may consider changing measurement scales during the targeting of the seasonal, hyperproductive, eco-georeferenceable, geo-spectrotemporal or geo-spatiotemporal, geoclassified, endmember eco-epidemiologically mapped oviposition, sites.

In the iteratively, explanatively, qualitatively interpolatable, vulnerability, geometric, endmember, LULC, sub-resolution, orthogonal, forecast, eco-epidemiological YF, *Ae. aegypti*, African, riceland, forest, canopy, discontinuous, sub-meter, resolution model renderings when λ < 0, both x_{λ} and x'_{λ} ,the sign of x^{λ} may change as to preserve the ordering of the geo-spectrotemporal or geo-spatiotemporal, geosampled, geoclassifiable LULCs. Of more interest would be the fact that when $\lambda = 0$, then the Box-Cox, sylvatic, capture point, *Ae. aegypti*, eco-georeferenceable, oviposition, sub-meter resolution, LULC, explanatory, predictor variable would be the indeterminate form 0/0. Rewriting the Box-Cox formula

as $x'_\lambda = \dfrac{e^{\lambda \log(x)} - 1}{\lambda} \approx \dfrac{\left(1 + \lambda \log(x) + \frac{1}{2}\lambda^2 \log(x)^2 + \cdots\right) - 1}{\lambda} \to \log(x)$.[Equation 1.3] would render $\lambda \to 0$ This same results may also be obtainable employing l'Hôpital's rule. Let lim stand for the limit $\lim_{x \to c}$, $\lim_{x \to c^-}$, $\lim_{x \to c^+}$, $\lim_{x \to \infty}$, or $\lim_{x \to -\infty}$ in a YF eco-epidemiolgical, sub-meter resolution, African riceland, LULC model. In such circumstances , a, elucidative, robust, YF, riceland, eco-epidemiological, geo-spectrotemporal or geo-spatiotemporal, endemic, transmission-oriented, sub-meter resolution, eco-epidemiological, predictive, risk model may reveal, seasonal, hyperproductive, eco-georeferenceable, *Ae egypti*, oviposition, on sub-meter resolution, geoclassified LULCs in expanding African, irrigated agro-ecosytems due to anthropogenic pressure.

Suppose that lim $f(x)$ and lim $g(x)$ are both zero or are both in an oviposition, eco-epidemiological, forecast, vulnerability, oviposition, eco-epidemiological, capture point, *Ae. egypti*, eco-georeferenceable, remotely sensed, sub-meter resolution, LULC model. If $\lim \dfrac{f'(x)}{g'(x)}$ has a finite value or if the limit is $\pm\infty$, then $\lim \dfrac{f(x)}{g(x)} = \lim \dfrac{f'(x)}{g'(x)}$. may be calculable for quantitating statistical significance of the geosampled, oviposition, sub-meter resolution, predictive, YF, eco-epidemiological, risk model estimators. A rigorous explanation for Tukey's suggestions in the geometric, endmember, LULC, signature, probabilistic paradigm may reveal that the log transformation (which is not an example of a polynomial transformation) may be inserted at the value $\lambda = 0$ in the wavelength, frequency, *Ae egypti*, oviposition, LULC model for optimally targeting seasonal, hyperproductive, capture points.

Further, x'_λ x = 1 may map to the African, riceland, agro-irrigated, LULC , *Ae egypti*, oviposition, capture point x'_λ = 0 from all regressively, iteratively, quantitatively, interpolatable geo-spectrotemporal, or geospatiotemrpoal, geosampled, eigen-decomposable, grid-stratifiable, orthogonal, eco-georeferenceable, sylvatic, hyperproductive, wavelength, frequency, sylvatic, *Ae. aegypti,* oviposition, partially discontinuous, forest-canopied, sub-meter resolution, geoclassifiable, LULC, site values of λ. The choice $\lambda = 1$ may simply shift x to the value x−1, which could be optimistically quantitated as a straight line in AUTOREG. In so doing seasonal hyperproductive, *Ae aegypti*, capture points , may be discerned cartographically along an inhomogeneous, forest, canopied peripheral corridor.

In the bottom row on a semi-logarithmic scale, the choice $\lambda = 0$ may alternatively correspond to a logarithmic transformation in AUTOREG which may reveal a straight line in an *Ae. aegypti,* immature habitat, capture point, geosampled in an African, riceland, agro-village ecosystem. By superimposing a larger collection of transformations on a semi-logarithmic scale in an empirical, orthogonal, sub-meter resolution, grid-stratified, LULC dataset in AUTOREG of orthogonal, geometrical, endmember, sub-meter resolution, geometric diagnostic, residual, model prognosticators (e.g., eco-georeferenceable, seasonally targeting, hypeproductive, sylvatic, *Ae.aegypti,* oviposition, LULC sites on geo-spectrotemporal or geo-spatiotemporal optimally remotely geoclassifiable, sub-meter resolution, uncoalesced, endmember agro-iririgated, African, riceland, agro-village complex, geosampled, geospectrotemporal, uncoalesced, iteratively interpolative, sub-meter, resolution, uncoalesced, wavelength

frequencies),propogagtional empirical, vulnerability, uncertainty oviposition LULC data may be treated for correlation biases (e.g., multicollinearity).

The power transformation in AUTOREG is optimally geostatistically defineable as a continuously varying function, with respect to the power parameter λ, in a piece-wise function form that makes it continuous at the point of singularity ($\lambda = 0$). Unfortunately, in time series, explanatively dependent, geo-spectrotemporal or geo-spatiotemporal, uncoalesced, sub-meter resolution, geometric, endmember, geoclassifiable, orthogonally eigen-decomposable, LULC, sub-meter resolution, optimizable, regression model prognosticators, heteroscedasticity often arises due to the effects of inflation and/or real compound growth. Some combination of logging and/or deflating may stabilize the variance for optimally robustly quantitating uncoalesced, sub-meter resolution, immature, *Ae.egypti,* oviposition, capture point, productivity on geoclassifiable, ArcGIS geoclassified, forest-canopied, discontinuous LULCs along an expanding agro-irrigated, riceland, African ecosystem.

Logarithmic transformations may be optimally employable to localize potentially nonlinear explanatory transformations of dependent, sub-meter resolution, immature, *Ae.egypti,* oviposition, capture point, African, riceland, agro-village, geo-spectrotemporal or geo-spatiotemporal, LULC variables. The Box-Cox transformation has the form

$$\begin{array}{ll} (y^{\lambda} - 1)/\lambda & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{array}$$

., hence the transformations of the positive explanatory, eco-georeferenceable, dependent variable y is controllable by the parameter λ. Transformations linearly-related to square root, inverse, quadratic, cubic, may be usable to aid in targeting, seasonal, hypeproductive, eco-georeferenceable, eco-epidemiological, grid-startieid, orthogonal, LULC, forecast, vulnerability model The limit as λ approaches 0 is the log transformation. More generally, Box-Cox transformations of the following geo-spatiotemporal, and or geospectrotemporal, *Ae.egypti,* oviposition, endmember, LULC, riceland, forest canopy interface

$$\begin{array}{ll} ((y + c)^{\lambda} - 1)/(\lambda g) & \lambda \neq 0 \\ \log(y + c)/g & \lambda = 0 \end{array}$$

corridor form may fit: By default, $c = 0$ may be revealed in the seasonal, targeted, dataset of hyperproductive, *Ae.egypti,* larval sites on newly transitioned, forest-canopy, eco-epidemiological, capture points to riceland agro-village, LULCs. The parameter may be usable to rescale the diagnostic, orthogonal, forecasted, clinical, field or remote –specified, time series, explanatory, YFV specified, ecogeoreferenceable, LULC, orthogonalized, parameter estimators y so that it is strictly positive. By default, $g = 1$ may also occur in the YF model, diagnostic, summary statements. Alternatively, the residuals may be quantitable by g which may be $y^{\lambda-1}$, whenst y is the geometric mean of y in the model.

The BOXCOX transformation in PROC TRANSREG may be usable to perform a Box-Cox transformation of the dependent variable in an exploratory, forecast, vulnerability, sylvatic, sub-meter resolution, *Ae. egypti*, LULC, oviposition, eco-epidemiological, orthogonal, risk model. An arbovirologist, medical entomologist or YFV researchist may specify a list of power parameters by using the LAMBDA= *t-option*. By default, LAMBDA=–3 to 3 by 0.25. The procedure chooses the optimal power parameter by using a maximum likelihood criterion (see Draper and Smith 1981, pp. 225–226). The PARAMETER=c transformation option may be

specified when the objective is to shift the geoclassified, *Ae. aegypti*, oviposition, sub-meter resolution, geoclassified LULC, YF-related, seasonal geo-spactrotemporal or geo-spatiotemporal, geosampled, quantitative, LULC values to avoid negatives in an eco-epidemiological, diagnostic, orthogonal, forecast, vulnerability model. To divide by $y^{\lambda-1}$, an arbovirologist , medical entomologist or YF reseachist may specify the GEOMETRICMEAN *t-option* for optimally, remotely, targeting, newly transitioned, inhomogeneous, forest-canopied, endmember LULCs into African, riceland, agro-irrigated, seasonal, hyperproductive, capture point, immature habitats.

Here are three examples of using the LAMBDA= *t-option* for constructing an eco-georeferenceable,eco-epidemiological, seasonal, hyperproductive, geo-spectrotemporal or geo-spatiotemporal, autoregressive, probabilistic, forest-canopy and or riceland corridor, *Ae eypti*, forecast, vulnerability, endmember,sub-meter resolution, LULC, risk model BoxCox(y / lambda=0) = identity(x1-x5);

    model BoxCox(y / lambda=-2 to 2 by 0.1) = identity(x1-x5);
    model BoxCox(y) = identity(x1-x5);

Here is the first example:

    model BoxCox(y / lambda=0) = identity(x1-x5);

LAMBDA=0 may specify a Box-Cox transformation with a power parameter of 0 in an oviposition, YF, seasonal, hyperproductive, geo-spectrotemporal or geo-spatiotemporal, forest-canopy and or riceland corridor, *Ae. aegypti*, forecast, vulnerability, endmember, sylvatic, LULC, eco-epidemiological, risk model. Since a single value of 0 may be specified for LAMBDA=, there may be no difference between the following eco-georeferenceable,eco-epidemiological, models:

    model BoxCox(y / lambda=0) = identity(x1-x5);
    model log(y) = identity(x1-x5);

Here is the second example:

    model BoxCox(y / lambda=-2 to 2 by 0.1) = identity(x1-x5);

LAMBDA may specify a list of eco-epidemiology, forecast, vulnerability, endmember, sylvatic, *Ae aeypti*, LULC, forest canopy, inhomogeneous, agro-riceland, eco-georferenceable, corridor model, unbiased, explicative, sub-meter resolution, parameter estimators. PROC TRANSREG may try each power parameter in the list and pick the best transformation. A ML approach [see Neter et al. (1990)] may employ the empirical, geo-spectrotemporal or geospatiotemporal, eco-georeferenceable, heuristically optimizable, sylvatic, *Ae egypti*, endmember, signature parameterizable, sub-pixel, uncoalesced, LULC covariates. With Box-Cox transformations, PROC TRANSREG would find the transformation before the usual iterations

begin in the eco-georeferenceable, LULC, oviposition model. Note that this is quite different from PROC TRANSREG's usual approach of iteratively finding optimal transformations with ordinary and alternating least squares. The method would be analogous to SMOOTH and PBSPLINE, which may optimally quantitate LULC transformations before the iterations begin based on a criterion other than least squares in the forecast, vulnerability, LULC, YF, eco-epidemiological model. Here is the third example: model BoxCox(y) = identity(x1-x5).

The above procedure would be printable employing the optimal power forecast, LULC, sub-meter resolution, sylvatic, oviposition, *Ae.   aegypti*,   orthogonal, risk model, eco-epidemiological, grid-stratified,   vulnerability parameter, a confidence interval on the power parameter (based on the ALPHA= *t-option*), a "convenient" power parameter (selected from the CLL= *t-option* list), and the log likelihood for each power parameter tried.

To illustrate how Box-Cox transformations would work for optimally regressively quantitating oviposition, eco-georeferenceable, geo-spatiotemporal or geo-spectrotemporal, forecast, vulnerability, sylvatic, *Ae aeypti*, LULC,African, riceland and forest, discontinuous, canopy, eco-epidemiological, risk YF model, unbiased estimators $y = e^{x+\varepsilon}$ may be constructed where $\varepsilon \sim N(0, 1)$. The transformed data may be fit with a linear model $\log(y) = x + \varepsilon$.

The following statements would produce

```
  title 'Basic Box-Cox YF Example';

  LULC data x;
    do x = 1 to 8 by 0.025;
      y = exp(x + normal(7));
      output;
    end;
  run;

  ods graphics on;
```

The next example shows how to find a Box-Cox transformation without an autoregressive, sylvatic, YFV-related, geosampled, eco-georeferenceable,eco-epidemiological, seasonal, hypeproductive, geoclassifiable, geo-spectrotemporal or geo-spatiotemporal, autoregressive, sub-meter resolution, forest-canopy and or riceland corridor, *Ae. aegypti*, forecast, vulnerability, endmember, LULC, risk model, independent variable. This methodology seeks to normalize the univariate histogram. This example would generate a minumum of 500 random, seasonal, capture point, eco-georeferenceable, forecast, vulnerability, oviposition, sylvatic, *Ae. aegypti*, risk model, endmember, eco-epidemiological, eco-georeferenceable, time series, LULC observations from a lognormal distribution. In addition, a constant variable *z* would be created that could render zero. This is because PROC TRANSREG requires some independent explanatory variable to be specified in any predictive, grid-stratified, orthogonal paradigm, even if it is constant.

Two options may be optimally specified in the PROC TRANSREG statement for targeting, capture point, hyperproductive, seasonal, *Ae. aegypti,* African, riceland habitats on newly transitioned inhomogeneous, forest canopied , sub-meter resolution, optimally geoclassifiable, endmember, LULC sites. MAXITER=0 may be specified as the Box-Cox transformation is performed before any iterations are begun whenst optimally constructing an eco-epidemiological, forecast, vulnerability, endmember, optimizable, geometrical, African, riceland model. In so doing, all eco-georeferenceable, discontinuous, forest-canopied, oviposition, sub-meter resolution, orthogonal, grid-stratified, African riceland agro-ecosystem, LULC, unbiased, capture points may be quantitated No iterations are needed since no other work is required once the model is built. The NOZEROCONSTANT *a-option* (which can be abbreviated NOZ) may be optimally thereafter specified so that PROC TRANSREG does not print any warnings when it encounters a constant, African, agro-irrigated, riceland or forest-canopied, LULC oviposition, sub-meter resolution, explanatory, time series, independent, elucidative, diagnostic, eco-epidemiological variable. The MODEL statement may ask for a Box-Cox transformation of *y* and an IDENTITY transformation (which does nothing) of the constant endmember, sub-meter resolution, LULC variable *z*. Therefater, PROC UNIVARIATE may be run to determine a histogram of the original geosampled geo-spatiotemporal or geo-spectrotemporal, agro-village, riceland, sylvatic, *Ae. aegypti*, oviposition, geometric, endmember, sub-meter resolution, LULC, uncoalesced, wavelength, signature YFV, explanatory, eco-epidemiological, frequentistic, predictor *y*, and the Box-Cox transformation, *Ty*. The following statements may fit a univariate eco-epidemiological, endmember, geo-spectrotemporal or geo-spatiotemporal, seasonal, African, riceland, YFV, Box-Cox model for remotely optimally targeting hyperproductive, *Ae. aegypti,* capture points.

```
title 'Univariate Box-Cox';

data x;
   call streaminit(17);
   z = 0;
   do i = 1 to 500;
      y = rand('lognormal');
      output;
   end;
run;

proc transreg maxiter=0 nozeroconstant;
   model BoxCox(y) = identity(z);
   output;
run;

proc univariate noprint;
   histogram y ty;
run;

ods graphics off;
```

The PROC TRANSREG results can show that zero is chosen for lambda in the riceland probabilistic, orthogonal paradigm so a log transformation may be optimally chosen. The first histogram may reveal that the original oviposition, sylvatic, *Ae. aegypti,* LULC, sub-meter resolution, optimally regressed, YF, geometric, wavelength, frequency, orthogonal, endmember dataset is skewed, but fortunately a log-transformation may make the entomological regression data appear much more normal.

Logging a time series, eco-epidemiological, seasonal, hyperproductive, oviposition, capture point, *Ae. aegypti,* eco-georeferenceable, geo-spectrotemporal or geo-spatiotemporal, endmember dataset of fractionalized, sub-meter resolution, geometric, frequency, wavelength, African, riceland, forest, inhomogeneously canopied, LULC observables may have an effect very similar to deflating: it may straighten out exponential erroneously quantitatable LULC growth patterns and reduce heteroscedasticity (i.e., stabilize variance) in the diagnostic, forecasted, endemic, YF estimators. Logging is not exactly the same as deflating--it does not eliminate an upward trend in the data (e.g., agro-irrigated, African, riceland, discontinuous, forest canopied sub-meter resolution LULCs) --but it can straighten the trend out so that it can be better fitted by a linear model. Deflation by itself will not straighten out an exponential growth curve reflecting the immature, sylvatic, oviposition, *Ae.aegypti* seasonal, hyperproductivity in a geoclassifiable, ArcGIS-derived, sub-meter resolution, forest-canopied, discontinuous LULC along an expanding, agro-irrigated, riceland, African agro-ecosystem especially if the growth is partly real and only partly due to inflation. If the series reflects inflationary growth, then deflation will help to account for the growth pattern and reduce spatial and aspatial heteroscedasticity in the diagnostic, epidemiological, explicative, orthogonal, eigen-decomposable, residual, explantory eco-georeferenecable, forecasts (e.g., targeted, seasonally, prolific, sylvatic, hyperproductive, *Ae. egypti,* ovipositions geolocations).

An arbovirologist, medical entomologist or YF experimenter may quantitate aspatial heteroskedascity in an aagro-irrigated, African, agro-ecosystem, oviposition, sub-meter resolution, geoclassifiable, Lsub-meter resolution, ULC, endmember diffusion model for optimally revealing level of geospatial spillover into forest-canopied, grid-stratified, sub-meter resolution, discontinuous LULCs along the periphery of the riceland complex by employing a negative binomial regression with a non-homogenous gamma distributed mean. In this model the primary assumption is that the standard deviation is the mean value (see Haight 1967). Overdispersed, Poissionian, probability, riceland, expanding,geo-spectrotemporal, or geo-spatiotemporal, eco-epidemiological, oviposition, *Ae. aegypti,* regressed, LULC data (i.e., outliers) due to violations of the assumption that the variance is equivalent to the mean in the paradigm may be rectified employing a negative binomial regression distribution (Jacob et al. 2005). Hence biased, endmember geometric, grid-stratified, endmember, LULC estimators rendered from an eco-epidemiological, forecast, vulnerability, linear, endmember, sub-meter resolution, YF model may be rectified by using a negative binomial gamma distribution as post-treatment for a Poissonian regression.

For quantitating aspatial uncommon variance (e.g., non-eco-georeferenceable, heteroskedastic, residuals) with latent autocorrelation coefficients, an arbovirologist, medical entomologist or YF experimenter may consider a covariance matrix in AUTOREG for robustly quantitating noisy, geosampled, African riceland, agro-village, ecosystem, oviposition, sub-

meter resolution, geoclassifiable, endmember LULCs that are geospatially spilling over onto discontinuous, forest-canopied, remotely geoclassified, sub-meter resolution, grid-stratified LULCs along the external corridors of the agro-irrigated, African eco-epidemiological, study site. In so doing, he or she may be able to optimally regressively quantitate the frequency, wavelength, spatial, heteroskedastic, interaction, covariate terms in the regression matrix. This endmember paramterizable quantitation of propagational uncertainty disturbances in such a probabilistic, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, eco-georeferenecable, diagnostic, forecast, vulnerability model may reveal moment conditions.

This paper considers spatial heteroskedasticity and autocorrelation consistent (spatial HAC) estimation of covariance matrices of multiple, sub-meter resolution, African, riceland, agro-irrigated *Ae. aegypti,* oviposition, geoclassifiable. LULCs transitioned from inhomogeneous, forest-canopied, grid-stratified, orthogonal, parameterizable, covariate estimators. We generalize the spatial HAC estimators introduced by Neter et al. (1990) to apply to linear and nonlinear, LULC, diagnostic, eco-epidemiological, YF, signature models with moment conditions for optimally regressively quantitating discontinuous, forest-canopied, riceland, African, *Ae. aegypti*, oviposition, eco-epidemiological, forecast, vulnerability, eco-georeferenceable endmember estimators. We establish a dataset of YF model parameter estimator based on regressive consistency, rate of convergence and asymptotic truncated mean squared error (MSE). Based on the asymptotic truncated MSE criterion, we derive the optimal bandwidth parameter and suggest the geo-spctrotemporal and geo-spatiotemporal, eco-epidemiological, LULC data, dependent, estimation procedure using a parametric plug-in method.The finite sample performances of the spatial HAC estimator are evaluated via Monte Carlo simulation employing a data of eco-georeferenceable,eco-epidemiological, seasonal, hyperproductive, LULC, geo-spectrotemporal or geo-spatiotemporal, autoregressive, sub-meter resolution, forest-canopy and or riceland corridor, sylvatic, orthogonal, grid-stratified, *Ae. aegypti*, forecast, vulnerability, endmember, risk-related, endmember model, parameter estimators.

Additionally we deflated the past eco-epidemiological, geo-spectrotemporal or geo-spatiotemrpoal, eco-georeferenceable, sub-meter resolution, YF data and reinflate the long-term, diagnostic, forecasts at a constant assumed rate, whilst simultaneously deflating the past oviposition time series, *Ae. aegypti,* African, riceland agro-irrigated, sub-meter resolution, grid-stratified, LULC data by an index and then "manually" reinflate the long-term forecasts employing an autocorrelation index. In Excel, a column was created of formulas which was used to divide the original geo-spectrotemporal and geospatial LULC values by the appropriate factors. For example, if the geosampled, seasonal, hyperproductive, *Ae. aegypti,* immature productivity, oviposition LULC data deflated at a rate of 5 % annually during a monthly sample frame, we divided by a factor of $(1.05)^{\wedge}(k/30)$ where k was the row index (observation number) in a weighted autogressive, estimation matrix for quantitating, African, riceland, agro-irrigated LULC, seasonal, growth, trend patterns in a stimulated, hierarchical model for capturing emerging females from a newly transitioned, inhomogeneous, oviposition, eco-epidemiological, capture point. Our assumption was that if this route is employable for quantitating LULC sub-meter resolution data at the agro-village field-operation level, inflation rates and other regressable estimates may be useable to prognosticate more than one geosampled sampled period ahead. Further, we assumed that the deflated eco-epidemiological, vulnerability, eco-

georeferenceable rates may forecast confidence limits to a data spreadsheet, then generate and save an eco-epidemiological, vulnerability, forecast-based, grid-stratified map residually potraying geometric, sub-meter resolution, endmember, LULC indices where hyperproductive, *Ae. aegypti*, oviposition capture points seasonally occur. In so doing, the appropriate columns together with other independent, riceland geo-spectrotemporal or geo-spatiotemporal, endmember, predictor variables (e.g., meteorological or partially canopied, eco-georeferenceable, eco-epidemiological, seasonal, hypeproductive, geo-spectrotemporal or geo-spatiotemporal, autoregressive, sub-resolution, forest-canopy and or riceland corridor, sylvatic, *Ae. aeypti*, forecast, vulnerability, endmember, orthogonal, LULC, risk, model covariates).

The statisticians Box and Cox developed a procedure to identify an appropriate exponent (Lambda = l) to use to transform LULC data into a "normal shape." Here the Lambda value indicated the power to which all the geo-spectrotemporal or geo-spatiotemporal, eco-epidemiologically geosampled, sub-meter resolution, African, riceland, agro-irrigated, grid-stratified, LULC data should be raised. In order to do this, the Box-Cox power transformation searches from Lambda = -5 to Lamba = +5 until the best value is found. We employed some common Box-Cox algorithmic transformations, where *Y'* was the transformation of the original geosampled data *Y.* The oviposition, sylvatic, time series geosampled, eco-georeferenceable, eco-epidemiological, *Ae. aegypti,* seasonal, hyperproductive, geo-spectrotemporal or geo-spatiotemporal, autoregressive, sub-resolution, forest-canopy and or riceland corridor, forecast, vulnerability, endmember, LULC, eco-epidemiological, risk model where Lambda = 0 and transformation is not Y^0 but instead the logarithm of *Y.* Table 1 reveals the Box-Cox transformations used to target, seasonal, hyperproductive, eco-georeferenceable, eco-epidemiological, *Ae. aegypti* oviposition, eco-epidemiological, sub-meter resolution, endmember, LULC sites

Table 1: Common Box-Cox Transformations	
l	**Y'**
-2	$Y^{-2} = 1/Y^2$
-1	$Y^{-1} = 1/Y^1$
-0.5	$Y^{-0.5} = 1/(Sqrt(Y))$
0	log(Y)
0.5	$Y^{0.5} = Sqrt(Y)$
1	$Y^1 = Y$
2	Y^2

A Box-Cox power transformation was employed for defining the explanative, diagnostic, eco-georeferenceable, eco-epidemiologist, explanatory, oviposition, LULC, orthogonal, dependent variable which was useful method to alleviate spatial and aspatial heteroscedasticity in the geo-spectrotemporal, geo-spatiotemporal, eco-epidemiological, geosampled, sub-meter resolution, sylvatic, *Ae. aegypti,* multivariate, eigen-decomposeable, orthogonal LULC sites

since the distribution of the explanative, dependent variable was not known in the Ugandan argo-village complex. For situations in which the dependent variable Y may be positive, the following transformation can be used:

$$y_i^{(\lambda)} = \begin{cases} \frac{(y_i^\lambda - 1)}{\lambda} & \text{when } \lambda \neq 0 \\ log(y_i) & \text{when } \lambda = 0 \end{cases}$$

(Cressei 1993). Diagnostic, orthogonal, eigen-endmember, LULC, frequency, wavelength, decomposable, sub-meter resolution, geo-spectrotemporal or geo-spatiotemporal, vulnerability, forecast eco-epidemiological, wavelength, frequency analysis of the probabilistically regressed, immature, *Ae.aegypti* oviposition productivity, geosampled on geoclassifiable, forest-canopied, discontinuous, sub-meter resolutionLULCs along an expanding African, riceland, agro-ecosystem revealed heterogeneous, Gaussian residuals.

Gaussian Process (GP) regression models typically assume that residuals are Gaussian and have the same variance for all observations (e.g., eco-georeferenceable,eco-epidemiological, seasonal, hyperproductive, geo-spectrotemporal or geo-spatiotemporal, autoregressive, sub-resolution, forest-canopy and or riceland corridor, forecast, vulnerability, endmember, LULC, risk model,eco-epidemiological,unbiased estimators) (Hosmer and Lemeshew 2002). However, applications with input-dependent noise (e.g., probabilistically regressed immature, agro-village, sylvatic, *Ae. aegypti* productivity geosampled on geoclassifiable, sub-meter resolution, partially, forest canopied, inhomogeneous, LULC, heteroscedastic, endmember, residuals) frequently arise in practice, as do applications in which the residuals do not have a Gaussian distribution. Jacob and Novak (2014) propose a GP endmember, regression model with a latent oviposition geoclassified, LULC variable that served as an additional unobserved, expanding African, agro-riceland, frequency, wavelength, orthogonal, grid-stratified, LULC, uncoalesced, endmember, covariate into the discontinuous, eco-epidemiological, forecast, vulnerability regression. This GPLC model allowed for heteroscedasticity since it allowed the function to have a changing partial derivative with respect to this unobserved, explanative, endmember, LULC covariate. With a suitable covariance function, a GPLC model may handle (a) Gaussian, YFV, endmember, regression residuals with input-dependent variance, or (b) non-Gaussian, YFV residuals with input-explanatory, dependent variance, or (c) Gaussian residuals with constant variance (Griffith 2003). Jacob and Novak (2014) compared an eco-epidemiological, capture point, seasonal, geosampled, hyperproductive, *An. arabiensis* aquatic larval habitat, eco-georeferenceable, sub-meter resolution, LULC,orthogonal, grid-stratified, model, using synthetic endmember eigen-decomposed datasets, which the authors referred to as GPLV, which only deals with a standard endmember GP model. Markov Chain Monte Carlo methods were developed. Experiments revealed that when the malaria mosquito LULC oviposition, time series data was heteroscedastic, both GPLC and GPLV rendered optimal results (smaller eco-epidemiological, forecast, vulnerability, mean squared error and negative log-probability density) than standard GP regression of the *An. arabiensis,* aquatic, larval habitats. In addition, when the residuals were Gaussian, the GPLC model was as good as GPLV for targeting seasonal hypeproductive capture points, while when the residuals were non-Gaussian, the GPLC model was better than GPLV. Here we assumed that since an arbovirologist, medical entomologist or YFV experimenter could choose from many covariance functions to achieve different degrees of smoothness or different degrees of additive structure, in synthetic regression parameters a quantitable covariance function may automatically determine fgeo-spectrotemporal, sub-meter resolution, eco-

epidemiological, eco-georeferenceable, endmember, LULC data revealing hypeproductive, *Ae. aegypti,* immature habitats employing a GP model.

However, standard GP regression models typically assume that the residuals have independent and identically distributed (i.i.d.) Gaussian distributions that do not depend on the input endmember covariates, though in many applications, the variances of the residuals do depend on the inputs (e.g., non-Gaussian, selected, frequency, wavelength,orthogonally eigen-decomposed, sub-meter resolution, geoclassifiable, *Ae. aegypti* grid-stratified, distributions of the residuals). A GP regression model may deal with input-dependent explanatory, YFV, agro-village, African, riceland, grid-stratifiable, residuals bordering, forest canopy, discontinuous LULCs This model we assumed. could include a latent, explanatory, diagnostic, geo-spectrotemporal or geo-spatiotemporal eco-epidemiological, sub-meter resolution endmember, LULC variable with a fixed distribution as an unobserved explanatory, endmember, time series, input covariate (e.g., eco-georeferenceable,eco-epidemiological, seasonal, hyperproductive, autoregressive, forest-canopy and or African, agro-riceland, discontinuous, forest-canopied, sub-meter resolution, LULC corridor, *Ae aegypti*, forecast, vulnerability, endmember, LULC, risk model, unbiased estimator). When the partial derivative of the response with respect to this unobserved, orthogonal, endmember, LULC covariate changes across geosampled, YFV, explanative observations, the variance of the residuals may be assumed to change. When the latent explanative variable is transformed non-linearly, the time series, *Ae. aegypti*, riceland and inhomogeneous, forest canopied, LULC residuals may be non-Gaussian [i.e., endmember Gaussian, agro-village riceland, process with a partially canopied, discontinuous, latent covariate (GPLC) regression model].

Standard parametric confidence intervals may provide a measure of significance for regression coefficients in an eco-epidemiological, oviposition, sylvatic, YF, *Ae aegypti*, LULC, sub-meter resolution, forecasting, probabilistic, endmember paradigm. The model may require acceptance of Gaussian assumptions regarding estimates of endmember coefficients for their validity. Diagnostic analysis does not support regression assumptions, especially when only limited data is available to estimate the variability from the multitude of sources (Hosmer and Lemeshew 2002).

Alternatives to the standard parametric confidence intervals are the semiparametric or nonparametric methods employing bootstrap estimates of the variability of the coefficient estimates (Cressie 1993). These methods may enable remotely targeting, seasonal, hyperproductive, sylvatic, *Ae. aegypti,* eco-epidemiological, capture point, oviposition, in an expanding, African, riceland environment due to anthropogenic pressure. In statistics, bootstrapping can refer to any test or metric that relies on random sampling with replacement. Bootstrapping may allow assigning orthogonal, endmember, eigen-decomposed, grid-stratifiable measures of accuracy to define terms of bias, variance, confidence intervals, prediction error or some other such measure for optimal, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, *Ae. aegypti*, immature habitat sampling for regressing sylvatic, YF, oviposition, LULC estimates(e.g., eco-georeferenceable, seasonal, hyperproductive, autoregressive, sub-resolution, forest-canopy and or riceland corridor, forecast, vulnerability, endmember, LULC, sub-meter, resolution, risk model covariates). This technique may allow estimation of the sampling distribution of almost any statistic using random sampling method in an expanding

interpolative, sub-meter resolution, LULC map of an agro-irrigated, African, riceland agro-ecosystem into discontinuous, forest-canopied, LULCs. Bootstrapping is the practice of estimating properties of an estimator (such as variance of an *Ae. egypti,* riceland, hyperproductive, seasonal foci) by measuring those properties when sampling from an approximating distribution (Hosmer and Lemeshew 2002). One standard choice for an approximating distribution is the empirical distribution, explanatory orthogonal, grid-stratifiable, sub-meter resolution, LULC function of the observed data (e,g., eco-georeferenceable,eco-epidemiological, seasonal, hypeproductive, geo-spectrotemporal or geo-spatiotemporal, autoregressive, forest-canopy and or riceland corridor, *Ae. aegypti*, forecast, vulnerability, endmember, LULC, risk model, descriptive oviposition covariates). In the case where a set of sub-meter resolution, diagnostic observations can be assumed to be rendered from an YF-related, eco-epidemiological, i.i.d.population, this dataset can be implemented for control strategies (IVM) by constructing a number of resamples with replacement, of the observed, grid-stratifiable, orthogonal, dataset immature estimators (and of equal size to the observed dataset).It may also be usable for constructing hypothesis tests. It is often used as an alternative to statistical inference based on the assumption of a parametric model when that assumption is in doubt, or where parametric inference is impossible or requires complicated formulas for the calculation of standard errors. Although bootstrapping is (under some conditions) asymptotically consistent, it may not provide general finite-sample guarantees in a regression, forecasting, vulnerability, probabilistic paradigm for optimally mapping immature, oviposition, *Ae.aegypti* productivity geosampled on geoclassifiable, ArcGIS-derived, sub-meter resolution, forest-canopied, discontinuous LULCs along an expanding agro-irrigated, African, riceland agro-ecosystems. The apparent simplicity may conceal the fact that important assumptions are being made when undertaking the bootstrap analysis (e.g. independence of African, *Ae. aegypti,* riceland habitat transitioned form forest canopied inhomogeneous samples) for optimally forecasting seasonal, hyperproductive, grid-stratified, oviposition sites where these foci would be more formally stated in other approaches.

A robust regression analyses employing geo-spectrotemporally uncoalesced, immature, *Ae.aegypti,* sub-meter resolution, eco-epidemiological, LULC productivity data as quantitated on geoclassifiable, sub-meter resolution, geo-spatiotemporal LULC, sylvatic, endmember estimators using nonparametric bootstrap percentile confidence intervals may infer the observed significance level of the discontinuous, forest-canopied, regression effects when correlating YF with expanding, agro-irrigated, agro-ecosystem, riceland complex estimators. The multiple linear regression may be performed with 1000 bootstrap replications, for example, by fixing the design matrix and resampling from the possible responses conditional on each treatment combination. The bootstrap orthogonal distribution of each endmember, specified, African, riceland, dscontinuous, forest-canopied, sub-meter resolution, geoclassifiable, LULC, regression coefficient may be compiled, and the 5th and 95th percentiles of the empirical distribution may be employable to form the limits for the 95% bootstrap percentile confidence interval for seasonally targeting hyperproductive, oviposition, *Ae aegypti,* LULC sites for quantitating YFV transmission based on geolocations of capture points in a expanding irrigated, African, riceland environment. Plots for the effects may be described in an ArcGIS cyberenvironment. If the confidence interval fails to include 0, then the p-value may be deemed to be less than or equal to 0.05, and the effect of the closer interface between the forest-canopied, inhomogeneous LULCs and the irirgated, agro-riceland, ecosystem complex.

Bootstrapping allows for estimation of statistics through the repeated resampling of LULC data (e.g., eco-georeferenceable,eco-epidemiological, seasonal, hyperproductive, geo-spectrotemporal or geo-spatiotemporal, autoregressive, sub-resolution, forest-canopy and or riceland corridor, sylvatic, *Ae aegypti*, forecast, vulnerability, sub-meter resolution, endmember estimators). An arbobiologist, a medical entomologist or YFV researcher may demonstrate several methods of bootstrapping a confidence interval about an R-squared statistic in SAS. The hsb2 dataset may be usable. The sub-meter resolution, YFV, eco-epidemiological, forecast, vulnerability model construction may begin with running an OLS regression for optimally regressively predicting LULC, weather variables and levels of forest-canopy geo-spatiotemporally or geo-spectrotemporally associated with immature capture point, *Ae. aegypti*, oviposition, eco-georeferenecable, eco-epidemiological, seasonal, hyperproductive, immature habitat and saving the R-squared value in a dataset called t0. The R-squared value in this regression would be 0.5189.

```
ods output FitStatistics = t0;
proc reg data = hsb2;
  YFV model = Ae egypti LULC;
run;
quit;
```

The REG Procedure
Model: MODEL1
Dependent Variable: read reading score

Number of LULC habitat Observations   200

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	10855	2713.73294	52.58	<.0001
Error	195	10064	51.61276		
Corrected Total	199	20919			

Root MSE	7.18420	R-Square	0.5189	
Dependent Mean	52.23000	Adj R-Sq	0.5090	
Coeff Var	13.75493			

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	6.83342	3.27937	2.08	0.0385
Ae.egypti		1	-2.45017	1.10152	-2.22	0.0273

LULC	score	1	0.45656	0.07211	6.33	<.0001
weather	w score	1	0.37936	0.07327	5.18	<.0001
levels of canopy		1	1.30198	0.74007	1.76	0.0801

```
*store the estimated r-square;
data _null_;
 set t0;
 if label2 =  "R-Square" then
 call symput('r2bar', cvalue2);
run;
```

To bootstrap a confidence interval about this R-squared value, an arbovirologist, a medical entomologist or YFV reseachist may first need to resample the empirical, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiologically geosampled data. This step would involve sampling with replacement from the original empirical YF dataset to generate a new optimizable, eco-epidemiological, eco-georeferenceable, forecast dataset (e.g., seasonal, hyperproductive, geo-spectrotemporal or geo-spatiotemporal, autoregressive, sub-resolution, forest-canopy and or riceland corridor, sylvatic, *Ae aegypti*, forecast, vulnerability, endmember, LULC, risk model estimators) which optimally should be the same size as the original dataset. For each of these samples, the same regression as above must be run for quantitating the R-squared value. PROC SURVEYSELECT allows resampling in one step (www.sas.edu).

Before carrying out the risk model tabulations, it may be important for an arbobiologist, medical entomologist or YFV researchist to outline the assumptions about the geosampled, endmember, LULC data (e.g., eco-georenceable,eco-epidemiological, seasonal, hyperproductive, geo-spectrotemporal or geo-spatiotemporal, autoregressive, sub-resolution, forest-canopy and or riceland corridor, *Ae. aegypti*, forecast, vulnerability, endmember, model estimators). For example, it may be assumed that the LULC observations in the YF dataset are independent. Further, it may be assumed that the statistic are being estimated is asymptotically normally distributed in the YF model.

An output sylvatic, YFV, eco-epidemiological sub-meter resolution, LULC geoclassifed, empirical, grid-stratified, sub-meter resolution, orthogonal, wavelength frequency dataset may reveal a seed, a sampling method, and the number of replicates in SAS/GIS. The sampling method may indicate, unrestricted random sampling, or sampling with replacement. The samprate may indicate how large each YF regression sample should be relative to the input eco-epidemiological, grid-stratified, orthogonal dataset. A samprate of 1 means that a geosampled entomological, sub-meter resolution, LULC-oriented, endmember, parameterizable, covariate estimator dataset should be of the same size as the input dataset (Jacob et al. 2007). So in this eco-georeferenceable, YFV –related, capture point, eco-epidemiological, LULC forecast, vulnerability, eco-georeferenecable, uncoalesced, spectral, wavelength, frequency model example below 500 covariates were generated for   multiple endmember, sub-meter resolution, estimator datasets, so our output dataset bootsample had over 100,000 observations.

```
%let rep = 500;
proc surveyselect data= hsb2 out=bootsample
    seed = 1347 method = urs
            samprate = 1 outhits rep = &rep;
run;
ods listing close;
```

The SURVEYSELECT Procedure

Selection Method   Unrestricted Random Sampling

Input Data Set   HSB2
Random Number Seed   1347
Sampling Rate   1
Sample Size   200
Expected Number of Hits   1
Sampling Weight   1
Number of Replicates   500
Total Sample Size   100000
Output Data Set   BOOTSAMPLE

We used these African, riceland, *Ae aegypti,* ovisposition, sub-meter resolution, grid-stratified, orthogonal, endmember LULCs Specifying by replicate allowed the model to be run separately for each of the 500 sample datasets. After that, a data step may be employable to convert the R-squared values to numeric values (e.g., finite discrete integers of immature, sylvatic, *Ae aegypti* in forecasted, seasonal, African, riceland, oviposition sites along a forest-corridor).

```
ods output  FitStatistics = t (where = (label2 =  "R-Square"));
proc reg data = bootsample;
  by replicate;
  model read =  yellow fever write ses;
run;
quit;
* converting character type to numeric type;
data t1;
  set t;
  r2 = cvalue2 + 0;
run;
```

An arbovirologist, medical entomologist or YFV researchist may create a confidence interval in an endmember , sub-mter resolution, African, riceland, *Ae aegypti,* oviposition model using the normal distribution theory. This assumes that the R-squared values follow a t distribution, so a 95% confidence interval can be optimally rendered about the mean of the R-squared values based on quantiles from a t-distribution with 499 degrees of freedom, for

example. It may be found, for example, the critical t values for the confidence interval may be robustly multiplied by the standard deviation of the R-squared values that arose from the 500 replications in an endmember, YF, remotely sensed, grid-stratified, eco-epidemiological, forecast, vulnerability model. A confidence interval employing this method may be symmetric about the R-squared value in a forecast, vulnerability, LULC, sub-meter resolution, regression model. The 95% confidence interval (0.432787, 0.605013) may be parsimoniously detremined. Calculating the bias in the original value of R-squared rendered from the YF, sub-meter resolution, LULC, African, riceland environment, the difference between that value and the mean of the 500 R-squareds in a bootstrap sample may reveal an elucidative, explanatorial, covariate estimator for targeting hyperproductive, sylvatic, *Ae. aegypti*, oviposition geolocations along an expanding African, ricleland, agro-village complex.

```
* creating confidence interval, normal distribution theory method;
* using the t-distribution;
%let alphalev = .05;
ods listing;
proc sql;
  select  &r2bar as r2,
       mean(r2) - &r2bar as bias,
                 std(r2) as std_err,
       &r2bar - tinv(1-&alphalev/2, &rep-1)*std(r2) as lb,
       &r2bar + tinv(1-&alphalev/2, &rep-1)*std(r2) as hb
   from t1;
quit;
```

r2	bias	std_err	lb	hb
0.5189	0.006616	0.043829	0.432787	0.605013

Another way to generate a bootstrap 95% confidence interval from the sample of 500 R-squared values in an eco-epidemiological, YF forecast, eco-georferenceable, LULC, vulnerability model for optimally targeting seasonal, hyperproductive, *Ae. aegypti,* oviposition, sub-meter resolution, LULC sites in an expanding agro-village complex is to look at the 2.5th and 97.5th percentiles in this distribution. (i.e., the percentile confidence interval). This approach to the confidence interval has some advantages over the normal approximation used above. This interval may not be symmetric about the original estimate of the R-squared in a YF, forecast, vulnerability model and this method may be hence unaffected by monotonic transformations on the estimated statistic (e.g., forecasted, hypeproductive, sylvatic, *Ae. aegypti,* immature, riceland habitat sub-meter resolution, density count). The first advantage is relevant because of the original, eco-epidemiological, diagnostic, YFV, orthogonally regressable, LULC regression estimate may be subject to bias. Bias in an entomological, regression, orthogonally forecastable, vulnerability, uncoalesced, signature model can render misspecications (Hosmer and Lemeshew 2000). The second advantage is less relevant in the example provided ascompatd with an instance where the estimate might be subject to a transformation in the eco-epidemiological risk model. The bootstrap estimates that form the bounds of the interval in a geo-spectrotemporal, endemic, YFV, geo-spatiotemporal, eco-epidemiological, forecast,

vulnerability, sub-meter resolution, geoclassified, endmember, LULC signature, African, riceland, agro-ecosystem, grid-stratified, risk model may be transformed in the same way to create the bootstrap interval of the transformed estimate.

An arbobiologist, medical entomologist or YFV reseachist can easily generate a percentile confidence interval in SAS using PROC UNIVARIATE after creating some clicnal, field or remote, geo-spectrotemporal or geo-spatiotemporal, endmember, YFV –related, LULC macro variables for optimally quantitating the percentiles of interest and using them in the output statement. In so doing, the confidence interval from this method may be quantitated (0.436, 0.6017). This information may be placed into a new dataset, method. The standard output may then be omitted in PROC UNIVARIATE as:

```
%let alphalev = .05;
%let a1 = %sysevalf(&alphalev/2*100);
%let a2 = %sysevalf((1 - &alphalev/2)*100);
* creating confidence interval, percentile method;
proc univariate YFV data = t1 alpha = .05;
  var r2;
  output out=pmethod mean = r2hat pctlpts=&a1 &a2 pctlpre = p pctlname = _lb _ub ;
run;
```

<... output omitted ... >

```
data t2;
  set pmethod;
  bias = r2hat - &r2bar;
  r2 = &r2bar;
run;
ods listing;
proc print data  = t2;
  var r2 bias p_lb p_ub;
run;
```

Obs	r2	bias	p_lb	p_ub
1	0.5189	.0066164	0.436	0.6017

An arbovirologit, medical entomologist or YF experimenter can also correct for bias when calculating confidence interval in an eco-epidemiological, sub-meter resolution, sylvatic, *Ae. aegypti,* oviposition, geoclassified LULC, African, riceland, YF eco-epidemiological, model. We have calculated bias in the previous method as the difference between the R-squared in the example provided. We observed in our initial regression and the mean of the 500 R-squared values from the bootstrap samples. The R-squared estimate from the initial regression may thus be assumed to be an unbiased, explanative, orthogonal estimate of the true R-squared. If an arbovirologist, medical entomologist or YF experimenter desires to correct for the bias in

calculating a confidence interval, he or she could go through the steps below. These have been described by Cameron and Trivedi in Microeconomics Using Stata.

Initially there would be a requirement to calculate the proportion of the bootstrap R-squareds that are less than the original value in an endemic, forecast vulnerability probabilistic, YF, sub-mter resolution, African riceland, forest-canopied, sub-meter resolution, LULC model. Thereafter the percentiles would have to readjusted and then employed to define the confidence interval trasnmission-oriented, sub- LULC model based on how this proportion differs from 0.5. Henceforth, calculating the probit of this proportion (z0) and the proportion associated with a alpha level (zalpha) would be necessitated in order to optimally define seasonal, hyperproductive, sylvatic, *Ae. aegypti*, oviposition sites on transitional, inhomogenous, partially forest-canopied, sub-meter resolution, grid-stratifiable LULCs Next, an abovirologist, medical entomologist or YF experimenter may calculate the percentiles that may be applicable to quantiate a confidence interval, p1 and p2, from these regressed values. Thereafter, PROC UNIVARIATE would calculate   an interval. From this method, for the example provided, the interval rendered would be (0.40575, 0.5936).

```
%let alphalev = .05;
%let alpha1 = %sysevalf(1 - &alphalev/2);
%put &alpha1;
proc sql;
  select sum(r2<=&r2bar)/count(r2) into :z0bar
  from t1;
quit;

   0.44

data _null_;
  z0 = probit(&z0bar);
  zalpha = probit(&alpha1);
  p1 = put(probnorm(2*z0 - zalpha)*100, 3.0);
  p2 = put(probnorm(2*z0 + zalpha)*100, 3.0);
  output;
  call symput('a1', p1);
  call symput('a2', p2);
run;

* creating confidence interval, bias-corrected method;
proc univariate data = t1 alpha = .05;
  var r2;
  output out=pmethod mean = r2hat pctlpts=&a1 &a2 pctlpre = p pctlname = _lb _ub ;
run;
```

<... output omitted ...>

```
data t2;
  set pmethod;
  bias = r2hat - &r2bar;
  r2 = &r2bar;
run;

ods listing;

proc print data  = t2;
  var r2 bias p_lb p_ub;
run;


Obs    r2       bias      p_lb    p_ub
 1    0.5189   .0066164   0.40575  0.5936
```

If the oviposition, African, riceland, ecosystem, sylvatic, forecast, vulnerability, eco-epidemiological, geo-spectrotemporal or geo-spatiotemporal, sub-meter resolution, *Ae aegypti* immature productivity series has a strong seasonal, eco-epidemiological, ecogeoreferenceable, LULC pattern this may be constant from year to year; thus, seasonal adjustment may be an appropriate way to estimate and extrapolate the pattern in an eco-epidemiological, YF, forecast, vulnerability model. The advantage of seasonal adjustment is that it models the seasonal pattern explicitly while providing the option of studying the seasonal indices and the seasonally adjusted grid-stratified, orthogonal, endmember, LULC data. The disadvantage is that this decomposition process would require the estimation of a large number of additional parameters (particularly for monthly meteorological, sub-meter resolution data), and it provides no theoretical rationale for the calculation of "correct" confidence intervals. In any statistcal geodatabase Out-of-sample validation is especially important to reduce the risk of over-fitting the past data through seasonal adjustment (Neter, 1990).

If the the regressed, grid-stratfied, orthogonal, oviposition, eco-epidemiological, endmember, uncoalesced, signature LULC, data representing the ecogeorefernceable, immature, seasonal hypeproductive, *Ae.aegypti,* riceland, agro-village, ecosystem model estimators geosampled on geoclassifiable, sub-meter resolution, forest-canopied, discontinuous LULCs along an expanding agro-irrigated, African, riceland cosystem is strongly seasonal but the arbrovirologist, medical entomologist or YFV experimenter does not choose an adjustment in a SAS or R ethe alternatives may then be to either (i) employ a seasonal autoregressive integrated moving average (ARIMA) model, which implicitly may forecast the seasonal pattern using lags and differences, or (ii) use a model, which estimates time-varying seasonal indices.

In statistics and econometrics, and in particular in time series analysis, an autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). They are applied in some cases where data show evidence of non-stationarity (e.g., seasonal transitioning ,oviposition, hyerproductive, *Ae. aegypti* habitats in geoclassified, forest canopy LULCs in peripheral

AGRO-irrigated, African, riceland, agro-village ecosystems), where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to reduce the non-stationarity in the emprical datasets.

Non-seasonal, time series, YF, explanatorial, ARIMA models may be denoted ARIMA(*p,d,q*) where endmember,orthogonal, grid-stratifiable LULC, sylvatic, *Ae. aegypti*, geo-spectrotemporal or geo-spatiotemporal parameters *p*, *d*, and *q* may be non-negative integers, *p* the order of the autoregressive model, *d* the degree of differencing, and *q* the order of the moving-average model. Seasonal ARIMA forecast models are usually denoted ARIMA(*p,d,q*)(*P,D,Q*)$_m$, where *m* could refer to the number of sample periods in each agro-irrigated, riceland, seasonal paradigm and the uppercase *P,D,Q* could refer to the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA, vulnerability model representing the immature oviposition, productivity, *Ae. aegypti*, geosampled count data.When two out of the three terms are zeros, the model may be referred to based on the non-zero,explanatory parameters, dropping "AR", "I" or "MA" from the acronym describing the model (Cressie 1993) For example, ARIMA (1,0,0) is AR(1), ARIMA(0,1,0) is I(1), and ARIMA(0,0,1) is MA(1). ARIMA models can be estimated following the Box-Jenkins approach. In time series analysis, the Box–Jenkins method, applies autoregressive moving average ARMA or ARIMA models to find the best fit of a time-series model to past values of a time series (Neter et al. 1990).

In Jacob et al. (2013b) results from both a Poissonian and a negative binomial (i.e., an explanatorial, Poisson random variable with a gamma distrusted mean) revealed that malaria related, district-level, time series, sub-meter resolution, regressively parameterizable covariates rendered from a predictive vulnerability endemic model for Uganda were significant, but furnished virtually no predictive power. Inclusion of indicator explanatory variables denoting the time sequence and the district geolocational spatial structure was then articulated with Thiessen polygons in ArcGIS which also failed to reveal meaningful iterative, quantitatively, interpolatable, LULC diagnostic covariates. Thereafter, an ARIMA model was constructed in PROC ARIMA which revealed a conspicuous but not very prominent first-order, temporal, autoregressive structure in the individual, district-level, eco-epidemiological, time-series, endmember, LULC, dependent data. A random effects term was then specified employing monthly, Box-Jenkins, time-series, modeling methodology to explicatively regressively quantitate, geo-spectrotemporal and geo-spatiotemporal, endmember, dependent data. This specification included a district-specific intercept term that was a random deviation from the overall intercept term which was based on a draw from a normal frequency distribution. The notation $AR(p)$ indicated an autoregressive model of order *p* which was defined as

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

where $\varphi_1, \cdots, \varphi_p$ were the unbiased parameter estimators of the model, c was a constant, and ε_t was white noise. The latter variable was equivalently written

$$X_t = c \quad \sum_{i=1}^{p} \varphi_i B^i X_t \mid \varepsilon_t$$

optimally employing the backshift operator *B* as By moving the summation term to the left side of the equation a polynomial notation was rendered as $\phi(B)X_t = c + \varepsilon_t$. In so doing, the autoregressive model was viewed as the output of an all-pole

infinite impulse response filter whose input was white noise. Infinite impulse response (IIR) is a property applying to many linear time-invariant systems (Cressie 1993).

The random effects specification revealed a non-constant mean across the Ugandan geosampled districts. This random intercept represented the combined effect of all omitted descriptive covariates that caused districts to be more prone to the malaria prevalence than other districts. Additionally, inclusion of a random intercept assumed random heterogeneity in the districts' propensity or, underlying risk of malaria prevalence which persisted throughout the entire duration of the time sequence under study. This random effects term displayed no spatial autocorrelation, and failed to closely conform to a bell-shaped curve. Autocorrelation is the correlation among values of a single variable strictly attributable to their relatively close geolocational positions on a two-dimensional surface, introducing a deviation from the independent observations assumption of classical statistics (Griffith 2003). The model's variance, however, implied a substantial variability in the prevalence of malaria across districts in the Ugandan, eco-epidemiological, study site. The estimated model contained considerable overdispersion (i.e., excess Poisson variability): quasi-likelihood scale = 76.565. The following equation was then employed to forecast the expected value of the prevalence of malaria at the district-level: prevalence = $\exp[-3.1876 + (\text{random effect})_i]$.

Exponential smoothing is a rule of thumb technique for smoothing time series LULC data, particularly for recursively applying as many as three low-pass filters with exponential window functions. Such techniques have broad application that is not intended to be strictly accurate or reliable for every geospatial modelling, sub-mtyer resolution, LULC situation. It is an applied procedure in most statistical packages for approximately calculating or recalling some value, or for making some determination based on prior assumptions by the user (seasonality of hyerproduction of *Ae. aegypti,* oviposition, capture points density counts, overlaid on sub-meter resolution geoclassifiable, discontinuous, forest-canopied, sparse shaded LULCs along an expanding irrigated, African, riceland agroecosystem corridor). Like any application of repeated low-pass filtering, the observed phenomenon may be an essentially random process, or it may be an orderly, but noisy, process. Whereas in the simple moving average the past observations are weighted equally, exponential window functions assign exponentially decreasing weights over time (Griffith 2003). The use of three filters is based on empirical evidence and broad application (Cressie 1993).

Exponential smoothing is commonly applied to smoothed vector entomological data (Jacob et al. 200, Griffith 2005) LULC data, as many window functions are in signal processing, acting as low-pass filters to remove high frequency noise. This method parrots Poisson's use of recursive exponential window functions in convolutions from the 19th century, as well as Kolmogorov and Zurbenko's use of recursive moving averages from their studies of turbulence in the 1940s.

The Kolmogorov–Zurbenko (KZ) filter is a series of iterations of a moving average filter of length *m*, where *m* is a positive, odd integer. The KZ filter belongs to the class of low-pass filters. The KZ filter has two parameters, the length *m* of the moving average window and the number of iterations *k* of the moving average itself. It also can be considered as a special window function designed to eliminate spectral leakage (spatially overflow of geo-spectrotemporal.

uncoalesced, sub-meter resolution, *Ae. egypti* oviposition sites along a forest canopied LULCs from an agro- irrigated, riceland expanding ecosystem). The impulse response function of the product of filters would be the convolution of impulse responses. The coefficients of the KZ filter can be interpreted as a distribution obtained by the convolution of k uniform discrete distributions on the interval [$-(m - 1)/2$, $(m - 1)/2$] where m is a regressable LULC discrete integer value. Therefore, the coefficient forms a tapering window which has finite support [$(m - 1)k + 1$]. The KZ filter a has main weight concentrated on a length of $m\sqrt{k}$ with weights vanishing to zero outside. The impulse response function of the KZ filter has $k - 2$ continuous derivatives and is asymptotically Gaussian distributed. Zero derivatives at the edges for the impulse response function make from it a sharply declining function, what is resolving in high frequency resolution.

The KZ filter is robust and nearly optimal. Because its operation is a simple moving average, the KZ filter performs well in a missing data environment, especially in multidimensional time and space where missing data can cause problems arising from spatial sparseness. Another nice feature of the KZ filter is that the two, eco-epidemiological, YF geo-spectrotemporal or geo-spatiotemporal, LULC parameters each have clear interpretations so that it can be easily adopted by an arbovirologist, medical entomologist or YFV experimenter in different LULC area interreptations in a single sampled riceland seasaon. Software implementations for time series, longitudinal and spatial data have been developed in some popular statistical package R which can then facilitate the use of the KZ filter and its extensions for robustly regressively quantitating uncoalesced, sub-meter resolution, oviposition, *Ae egypti*, LULC riceland agro-irrigated, sites along a partially forest-canopied LULCs from a irrigated, African, expanding agroecosystem

KZ filter can be used to smooth the periodogram. In statistical signal processing, the goal of spectral density estimation (SDE) is to estimate the spectral density (also known as the power spectral density) of a random signal from a sequence of time samples of the signal. Intuitively speaking, the spectral density characterizes the frequency content of the signal (Jensen 2005). Spectrum analysis, also referred to as frequency domain analysis or spectral density estimation, is the technical process of decomposing a complex signal into simpler parts. Many bio-geophysical, explanative, geospatial processes of geo-spectrotemporally uncoalesced, sub-meter resolution, *Ae aegypti* oviposition sites along a forest-canopied, discontinuous, LULCs from a irrigated, African, expanding agroecosystem may be optimally described as a sum of many individual frequency components. Any process that quantifies the various amounts (e.g. amplitudes, powers, intensities, or phases), versus frequency can be called spectrum analysis.

Spectrum, endmember, sub-meter resolution, orthogonal, LULC sgnature analysis can be performed on the entire sub-meter resolution, *Ae aegypti* oviposition ,capture point, along a forest-canopied, discontinuous, LULCs from an agro-irrigated, African, expanding, agroecosystem, signature signal. Alternatively, the riceland signal can be broken into short segments (sometimes called *frames*), and spectrum analysis may be applied to these individual endmember LULC segments. Periodic functions are particularly well-suited for sub-division analyses (Cressie 1993). A function $f(x)$ is said to be periodic (or, when emphasizing the presence of a single period instead of multiple periods, singly periodic) with period P if $f(x) = f(x + np)$ for $n = 1$, (Equation 1.3) (Hazewinkle 2001) For example, the sine function $\sin x$,

for the equation 1.3 would be periodic with least period (often simply called "the" period) 2π(as well as with period -2π, 4π, 6π, etc.). The constant function $f(x)=0$ would then also be periodic with any period $R$for all nonzero geosampled,geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, frequency, wavelength, YF, forecast, vulnerability, sub-meter resolution, discontinuous, forest-canopied or riceland, agrovillage, capture point, sylvatic, *Ae .aegypti* , seasonal, hyperproductive foci R, so there is no concept analogous to the least period for constant functions.

General mathematical techniques for analyzing non-periodic functions fall into the category of Fourier analysis. The Fourier transform of a function produces a frequency spectrum which contains all of the information about the original signal, but in a different form(Griffith 2003). This means that the original function in an oviposition, forecast vulnerability, *Ae egypti*, epidemiological, sub-meter resolition LULC, signature, YF model for targeting hyerproductive, seasonal habitats in an agro-irirgated, expanding, African, riceland agroecosystem can be completely reconstructed (synthesized) by an inverse Fourier transform.

The Fourier transform is a generalization of the complex Fourier series in the limit as $L \to \infty$.Fourier series is an expansion of a periodic function $f(x)$in terms of an infinite sum of sines and cosines. (Hosmer and Lemeshew 2002). Using the method for a generalized Fourier endmeber LULC series, the usual Fourier series involving sines and cosines may be obtainable by taking $f_1(x)=\cos x$and $f_2(x)=\sin x$. Since these functions form a complete orthogonal system over$[-\pi, \pi]$, the Fourier series of any eco-epidemiological, YF sub-meter resolution,LULC, oviposition, seasonal, hypeproductive, foci,, function $f(x)$ may be regressively optimally geo-spectrotemporally or geo-spatiotermporally rendered by

$$f(x)=\frac{1}{2}a_0+\sum_{n=1}^{\infty}a_n\cos(nx)+\sum_{n=1}^{\infty}b_n\sin(nx),$$ where$a_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)dx$,$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos(nx)dx$,and $b_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin(nx)dx$ and $n=1$, 2, 3, .

By replacing the discrete $A_n$with the continuous $F(k)dk$while letting $n/L \to k$ in a sub-meter resolution, remotely sensed, forecasting vulnerability, probabilistic,oviposition, *Ae egypti*, LULC paradigm, the sum to an integral, and the equation would become $f(x)=\int_{-\infty}^{\infty}F(k)e^{2\pi ikx}dk$ and $F(k)=\int_{-\infty}^{\infty}f(x)e^{-2\pi ikx}dx$. Here, $F(k)=\mathcal{F}_x[f(x)](k)=\int_{-\infty}^{\infty}f(x)e^{-2\pi ikx}dx$ would be called the forward $(-i)$ Fourier transform, and $f(x)=\mathcal{F}_k^{-1}[F(k)](x)=\int_{-\infty}^{\infty}F(k)e^{2\pi ikx}dk$ which would be the *inverse* $(+i)$ Fourier transform. The notation $\mathcal{F}_x[f(x)](k)$ was introduced in Trott (2004), and $\hat{f}(k)$and $\check{f}(x)$are sometimes also used to denote the Fourier transform and inverse Fourier transform, respectively.

For perfect construction of a sylvatic, sub-meter resolution, LULC, *Ae. aegypti,* forecast, vulnerability, eco-epidemiological, signature, endmember model to optimally target hyperproductive*, georefernceable, oviposition sites on sparsely canopied, LULC along an expanding, agro-irrigated, African, riceland agroecosyem, a spectrum analyzer must preserve both th wamplitude and phase or each frequency component reflected from the geoclassified

LULCs. These pieces of sub-meter resolution LULC information may represent as a 2-D vector as a complex number of a magnitude for optimally phasing in ground corrdinates of discontinuous forest canopy which may be gelocations of hypeproductive seasonal *Ae. aegypti* ovispoition sites along expanding irrigated, African, riceland agroecosystems.

A common technique in signal processing is to consider the squared amplitude, or power; in this case the resulting plot is referred to as a power spectrum. For a given signal, the power spectrum gives a plot of the portion of a signal's power (energy per unit time) falling within given frequency bins. The most common way of generating a power spectrum is by using a discrete Fourier transform, but other techniques such as the maximum entropy method can also be used.

In practice, nearly all software and electronic devices that generate frequency spectra apply a fast Fourier transform (FFT), which is a specific mathematical approximation to the full integral solution. Formally stated, the FFT is a method for computing the discrete Fourier transform of a sampled signal. The continuous Fourier transform is definable as $f(v) = \mathcal{F}_t[f(t)](v) = \int_{-\infty}^{\infty} f(t)\,e^{-2\pi i v t}\,dt.$ Now consider generalization to the case of a discrete function, $f(t) \to f(t_k)$ in a YF eco-epidemiological, signature, sub-meter resolution, LULC model by letting $f_k \equiv f(t_k)$, where $t_k \equiv k\Delta$, with $k = 0, ..., N-1$. Writing this out would render the discrete Fourier transform $F_n = \mathcal{F}_k\left[\{f_k\}_{k=0}^{N-1}\right](n)$ as $F_n = \sum_{k=0}^{N-1} f_k\, e^{-2\pi i n k/N}.$ The inverse transform $f_k = \mathcal{F}_n^{-1}\left[\{F_n\}_{n=0}^{N-1}\right](k)$ would then be $f_k = \frac{1}{N}\sum_{n=0}^{N-1} F_n\, e^{2\pi i k n/N}.$ Discrete Fourier transforms (DFTs) are extremely useful because they reveal periodicities in input data as well as the relative strengths of any periodic components(Griffth 2003). There are however a few subtleties in the interpretation of discrete Fourier transforms for a YF endmeber African riceland agro-irrigated, forecast, vulnerability, sub-meter resolution, LULC model. In general, the discrete Fourier transform of a real endmember sequence of discrete integer values will be a sequence of complex numbers (gfeo-spectrotemporally or geo-spatiotemporally geosampled, sylvatic, seasonal hypeproductive, eco-epidemiological, eco-georefenceable, *Ae. aegypti,* riceland, predicted, larval sites) of the same length. In particular, if f_k are real, then F_{N-n} and F_n in a YF model then the geoclassified LULCs (e.g., agro-irrgated tillering habitats or forest, inhomogeenously canopied, capture points) may be related by $F_{N-n} = \overline{F}_n$ for $n = 0, 1, ..., N-1$, when $\overline{z}$ denotes the complex conjugate in the remotely sensed, sub-meter resolution optimizable model.

Because of reversibility, the Fourier transform is called a representation of the function, in terms of frequency instead of time; thus, it is a frequency domain representation (Cressei 1993) . Linear operations that could be performed in the time domain od a YF eco-epidemiological, eco-georefenceable, sub-meter resolution, geo-spectrotemporal or geo-spatiotemporal , endmember, LULC signature paradigm may have counterparts that can often be performed more easily in the frequency domain. Frequency analysis simplifies the understanding and interpretation of the effects of various time-domain operations, both linear and non-linear (Griffith 2003). For instance, only non-linear or time-variant operations can create new

frequencies in the frequency spectrum of a sub-meter resolution, YF, forecast, vulnerabilitym signature, eco-epidemiological, LULC model.

The Fourier transform of a stochastic (random) waveform (noise) is also random. Some kind of averaging is required in order to create a clear picture of the underlying frequency content (frequency distribution)(Cressie 1993). Typically,a forecast. vulnerability, eco-epidemiological, uncoalesced, sub-meter resolution, signature, LULC frequency, wavelength, endmember, sylvatic, *Ae. aegypti,* data would be divided into time-segments (e.g., hyperproductive, capture points) of a chosen seasonal duration, and henceforth transformed onto each one. Then the magnitude or (usually) squared-magnitude components of the transforms are summed into an average transform. This is a very common operation performed on digitally sampled time-domain data, using the discrete Fourier transform.

In mathematics, the discrete Fourier transform (DFT) converts a finite sequence of equally spaced samples of a function into the list of coefficients of a finite combination of complex sinusoids, ordered by their endmember wavelength frequencies, that has those same sample values. It can be said to convert the sampled function from its original domain (often time or position along a line) to the frequency domain.

The input samples are complex numbers (dicrete integer value rendered from a forecast, vulnerability, eco-epidemiological, YF model), and the output coefficients would be complex as well. The frequencies of the output sinusoids are integer multiples of a fundamental frequency, whose corresponding period is the length of the sampling interval (Cressie 1993). The combination of sinusoids obtained through the DFT is therefore periodic with that same period. The DFT differs from the discrete-time Fourier transform (DTFT) in that its input and output sequences are both finite; it is therefore said to be the Fourier analysis of finite-domain (or periodic) discrete-time functions.

The DFT is the most important discrete transform, used to perform Fourier analysis in many practical applications. In digital signal processing, the function is any quantity or signal that varies over time, such as the pressure of a sound wave, a radio signal, or daily temperature readings, sampled over a finite time interval (often defined by a window function). In image processing, the samples can be the values of pixels along a row or column of a raster image (www.esri.com). The DFT may be used to efficiently solve partial differential equations, and to perform other operations such as convolutions or multiplying large integers when constructing a robust, YF, forecast, vulnerability, eco-epidemiological, riceland, *Ae egypti,* sub-meter reolution, LULC, ecoepidemiological model for remotely targeting, seasonal, hypeproductive habitats on transitioning inhomogeneous, forest-canopied, peripheral corridors in these African agro-irrigated ecosystems.

Since it deals with a finite amount of data, it can be implemented in computers by numerical algorithms or even dedicated hardware. These implementations usually employ efficient fast Fourier transform (FFT) algorithms so much so that the terms "FFT" and "DFT" are often used interchangeably. Prior to its current usage, the "FFT" initialism may have also been used for the ambiguous term "finite Fourier transform".

For a class of diagnostic, geo-spatiotemporal, geospectrotemporal, explanatorial, diagnostic, stochastic, YFV-related, grid-stratified, LULC processes, the only information available may be African, riceland, *Ae. aegypti* –related, variable, spectral density and smoothness which may be quantified by Hölder condition. A function $\phi(t)$ may satisfy the Hölder condition on two orthogonal, *Ae.aegypti* capture points t_1 and t_2 on an arc L when $|\phi(t_2) - \phi(t_1)| \leq A |t_2 - t_1|^{\mu}$, with A and μ positive real constants. In some literature, functions ϕ satisfying the Hölder condition are sometimes said to be (locally) μ-Hölder continuous; moreover, μ and A are sometimes called the Hölder exponent and Hölder constant of ϕ, respectively. The Hölder condition comes up frequently in several branches of mathematics, notable among which is the study of Brownian motion in probability.

A real-valued stochastic process $\{B(t) : t \geq 0\}$ is a Brownian motion which starts at $x \in \mathbb{R}$ if the following properties are satisfied: 1. $B(0) = x$. 2. For all times $0 = t_0 \leq t_1 \leq t_2 \leq \cdots \leq t_n$, the increments $B(t_k) - B(t_{k-1})$, $k = 1$, ..., n, are independent random variables. 3. For all $t \geq 0$, $h > 0$, the increments $B(t + h) - B(t)$ are normally distributed with expectation value zero and variance h. 4. The function $t \mapsto B(t)$ is continuous almost everywhere. The Brownian motion $B(t)$ is said to be standard if $B(0) = 0$. It is easily shown from the above criteria that a Brownian motion has a number of unique natural invariance properties including scaling invariance and invariance under time inversion. Moreover, any Brownian motion $B(t)$ satisfies a law of large numbers so that $\lim_{t \to \infty} \frac{B(t)}{t} = 0$ almost everywhere. Moreover, despite looking ill-behaved at first glance, Brownian motions are Hölder continuous almost everywhere for all values $\alpha < 1/2$. Contrarily, any Brownian motion is nowhere differentiable almost surely. The above definition is extended naturally to get higher-dimensional Brownian motions. More precisely, given independent Brownian motions $B_1, \cdots, B_d$ which start at $x_1, \cdots, x_d$, one can define a stochastic process $\{\beta(t) : t \geq 0\}$ by

$$\beta(t) = \begin{bmatrix} B_1(t) \\ \vdots \\ B_d(t) \end{bmatrix}.$$

Such a β is called a d-dimensional Brownian motion which starts at $(x_1, \ldots, x_d)^{\mathrm{T}} \in \mathbb{R}^d$.

If the series shows compound growth and/or a multiplicative, oviposition, eco-epidemiological, YF, African riceland, agro-village environment, syvaltic, *Ae. aegypti*, discontinuous, capture point, eco-epidemiological, seasonal pattern, a logarithm transformation may be helpful in addition to or lieu of deflation. Logging the grid-stratifiable, orthogonal, LULC data will not flatten an inflationary growth pattern, but it will straighten it out it so that it can be fitted by a linear model (e.g., a random walk or ARIMA model with constant growth, or a linear exponential smoothing model). Also, logging will convert multiplicative, transitional, African, agro-irrigated, riceland, seasonal patterns (flooding, post-tillering) to additive patterns, so that if an arbovirologist, medical entomologist or YF researcher performs seasonal adjustment after logging, he or she should use the additive type for regressively quantiating vital diagnostic forest-canopied, sub-meter resolution, transitioned. *Ae aegypti*, oviposition geolocations. Logging deals with inflation in an implicit manner; if a inflation is to be modeled explicitly (i.e., if if an arbovirologist, medical entomologist or YF researcher wants the inflation rate to be a visible LULC, parameter in the time series, *Ae aegypti* , forecast, vulnerability, LULC model), or if plots of deflated data need to be viewed. In th elater case deflating rather

than logging would be more vital for optimally regressively ge-spectrotemporal, or geo-spatiotemporal quantitate sub-meter resolution, newly transitioned, African, riceland, *Ae. aegypti,* immature habitats to forest-canopied, hyperproductive foci.

Another important use for the log transformation in an African agro-village, YF eco-epidemiological, endemic, transmission-oriented, vulnerability-oriented, predictive model, is linearizing relationships amongst the uncoalesced, forest canopy, inhomogeneous and riceland, *Ae. aegypti,* predictor variables in the model. For example, if the explanatory, dependent variable is a multiplicative rather than additive function of the independent variables, or if the relationship between dependent and independent variables is linear in terms of percentage changes rather than absolute changes in an oviposition, YF, eco-epidemiological, forecast, orthogonal, vulnerability-related, entomological, endmember, sub-meter resolution, riceland, *Ae. aegypti,* LULC model then applying a log transformation to one or more variables may be appropriate for geolocation, seasonal, hyperproductive foci.

By logging rather than deflating, the need to incorporate an explicit entomological, residual, explanative, iterative, interpolative, epidemiological, forecast (e.g., targeted), *Ae. aegypti*, hypeproductive, forest-canopied, discontinuous, eco-georeferenceable, capture point along an agro- irrigated, expanding, African, riceland ecosystem is unnecessary : merely lumping the geosampled data together with any other sources in the original data would be suffice. Logging the data before fitting a random walk model yields a so-called geometric random walk—(i.e., a random walk with geometric rather than linear growth) (Cressie 1993). A geometric random walk is the default forecasting model that may be applicable for optimizing an eco-epidemiological, eco-georeferenceable, YF-related, diagnostic, riceland model targeting, forest-canopied, hyperproductive, grid-stratified, *Ae. aegypti* seasonal, oviposition, sub-meter resolution, LULC sites.

Regressively probabilistically quantitated targeted, hyperproductive, ecogeoreferenceable, explanative, oviposition, *Ae. aegypti*, forest canopied, discontinuous, capture point data may show periods of increased or decreased volatility over time. This is normal and is often modeled with so-called ARCH (auto-regressive conditional heteroscedasticity) models in which the error variance is fitted by an autoregressive model. Such models are beyond the scope of this discussion, but a simple fix would be to work with shorter intervals of YF geosampled data in which volatility is more nearly constant. Non-quantitated, remotely sensed, endmember, heteroskedastic, immature, productive, count data can also be a byproduct of a significant violation of the linearity and/or independence assumptions in an eco-epidmiological, YF, geo-spectrotemporal, geospatial, forecast, vulnerability, entomological, probabilistic. Regression, time series, orthogonal paraidgm, in which case it may also be fixed as a byproduct of fixing those problem. During immature stages, vector species differ in their priorities for habitat (Gu and Novak 2005).

Changes in land-use such as those experienced in growing agro-irrigated, riceland, African agro-villages due to anthropogenic pressue may also permit the colonization of new riceland, oviposition, *Ae. aegypti* habitats in forest-canopied LULCs thereby causing expansion or reduction of the range of the vector. The LULC transitions may also modify the composition of the mosquito vector community in these transitional eco-zones. Agriculture is the largest driver

for land-use change across the earth (Patz et al. 2008). Croplands and pastures together have become one of the largest terrestrial biomes on the planet, rivaling forest, and cover in extent of 40% of the land surface area (Ramankutty and, Foley 1999). Land-use/land-cover patterns and socio-cultural practices have a strong relation with vector, arthropod-related, seasonal, immature habitats (Wood and Washino 1994). Seasonal, explanatorial, geospectrotemporally uncoalesced, LULC characteristics can influence the risk of contracting YF as they modify mosquito populations along agro-irrigated, expanding, African, riceland   agro-ecosystem, peripheral paddies and canals (Jacob et al. 2007). Land-use change is a major constituent of global environmental change that can potentially affect human health in relation to mosquito borne diseases by influencing the mosquito's habitat (Patz et al. 2008). The distribution and abundance of vectors concomitantly mediates human-mosquito interactions including biting rate (Gu et al. 2006).

Ideal *Aedes* larval habitats include artificial or natural water containers (water storage containers, flower pots, discarded tires, plates under potted plants, cemetery vases, flower pots, buckets, tin cans, clogged rain gutters, ornamental fountains, drums, water bowls for pets, birdbaths, etc.) that are all (www.cdc.gov). This species has also been found in underground collections of water such as open or unsealed septic tanks, storm drains, wells, and water meters.

Yellow fever virus pathogens may respond to changing seasonal, sub-meter resolution, geoclassifiable, grid-stratified, LULC dynamics on multiple transmission levels and appear to increase in disturbed systems, yet current knowledge of mosquito diversity and the relative abundance of vectors as a function of immature habitat, capture point, is limited for implementing seasonal control strategies in agro-irrigated, expanding African riceland ecosystems. Describing seasonal, *Ae. aegypti*, time series, mosquito-community diversity employing geoclassifiable, sub-meter resolution, LULC change, site data with other covariates (meterological, census etc) may aid in implementation of an eco-epidemiological, seasonal, YF signature, surveillance system in these ecosytems. Additionally SAS or R may be employable for identification of prolific, eco-georeferenceable, seasonal, eco-epidemiological, LULC, capture point, geoclassified, habitat types and immature productivity of the oviposition site, especially in landscapes experiencing different levels of anthropogenic and economic pressures (e.g., partially canopied, forested, LULC eco-zones along peripherial, African agro-irrigated, riceland agro-villages). The modification of the urban landscape influences the local (microscale), mesoscale, and even the macroscale climate (Brazel et al., 2000; Quattrochi et al., 2000; Voogt and Oke, 2003).

Assumptions of a general larval population or identical habitats are untenable for a realistic evaluation of larval control interventions or the development of models to be empirically tested under field conditions (Gu and Novak 2005). Diverse spatial LULC sub-meter data exist, varying in extent and resolution. Remotely-sensed, sub-meter, resolution, mosaicked data provides the most robust extensive coverage for unveiling heterogeneity and endmember signatures (Jensen 2005) of disturbed landscapes where seasonally explanatorial, hyperproductive, vector, arthropod-related, immature habitat, data points are commonly georeferenced (Jacob et al. 2011, Jacob et al. 2007). An endmember LULC can reflect the spectrum of a pure ground component (e.g., vegetation, soil, water); thus these sub-mixed pixel (mixel) fractions are typically interpreted as ground components (Jensen 2005).

It is well documented that urbanization results in increased amount of impervious surfaces (Brabec et. al., 2002) that augment the intensity, volume, temperature, and duration of irrigated, storm water runoff (Booth and Reinfelt, 1993; Schueler, 1994). Urban rice-agro-village complex, seasonal, storm water runoff may cause or contribute to water quality degradation by changing natural eco-geohydrologic, explanative, seasonal, LULC patterns (Driver and Troutman, 1989), accelerating natural stream flows (Booth and Jackson, 1997), decreasing stream bank erosion (May et al., 1997), increasing aquatic habitat (Booth and Reinelt, 1993; Horner et. al., 1997), degrading stream water quality (Schueler, 1994; Booth and Jackson, 1997; May et al., 1997), increasing temperature (Galli, 1990), and elevating pollutants. Hence, different land uses can modify contact of susceptible humans with infectious vectors (e.g., YFV) which can modify human cases distribution. Biodiversity loss affects infectious disease risk by disrupting normal relationships between hosts and pathogens on specific landscapes (Patz 2000).

Combining publicly available data on sub-resolution, explanatively, iterative interpolative, seasonal, geoclassfiable, LULC, geo-spectrotemporal, sylvatic, YF, parameterizable covariates and anthropogenic, geo-spatiotemporal, grid-stratified, population data with *Ae. aegypti* surveillance data may produce a spatially disaggregated, YF-related, time series, endemic, transmission, forecast, vulnerability, eco-georeferenceable map representing the vector-to-human host ratio, a common quantity originating from process-based, eco-epidemiological, time series, forecasting, vulnerability models. The vector-to-host ratio allows an evaluation of the level of risk of contact even in the absence of the pathogen since it does not rely on sparse epidemiological data (www.who.gov). Thus, understanding geospatial, seasonal, LULC patterns may enable quantitating human risk of exposure to YFV in SAS or R which may be critical for targeting limited prevention, surveillance, and control resources (e.g., spatial targeting of vaccination, drug administration, or education campaigns; use of sentinel sites to monitor vector abundance; and identifying areas for most effective use of pesticides). Treatments of habitat perturbations should be based on surveillance of larvae in the most productive areas of an ecosystem (Gu and Novak 2005). These evaluations may be particularly important in areas that experience occasional, YF ,epidemic transmission such as in an expanding irrigated, African riceland agro-village ecosystem due to more contact with the sylvatic vector.

One essential element leading to eradication of sylvatic,YF in an expanding, agro-irrigated African, riceland ecosystem may be an real-time, statistical, geodatabase cyberenvironment, surveillance system capable of providing uncoalesced, geoclassfiable, endmember, LULC, sub-meter, wavelength frequencies [visible and near infra red (NIR) measurements] of uncoalesced, time series, eco-epidemiological, vulnerability eco-georeferenecable forecasts detailed, from sub-meter resolution, cartographic and logistic regression models. These parmeterized renderings may regard the human host, the parasite, and the *Aedes,* mosquito, oviposition site on a specific, newly transitioned, sub-meter resolution, geoclassified LULC (e.g., ploughing habitats along side a periperial, agro-irrigated, discontinuous, forest canopied, partially canopied, seasonal eco-zone border. Probabilistically regressed distributions of vector-borne diseases tend to be highly spatially heterogeneous, varying according to the often-heterogeneous distribution of transmission systems components: vectors, pathogens, and hosts (Gu and Novak 2005). Regression-related, real-time seasonal, logistic, LULC patterns, eco-cartographically quantized in a SAS/GIS geodatase, statistical cyberenvironment can reval human risk of exposure to arthropod vectors and their associated

pathogens which may be critical for targeting limited prevention, surveillance, and control resources (e.g, Jacob and Novak 2014, Gu et al. 2006).

Yellow Fever disease risk may be modeled with high predictive accuracy by employing regression models constructed in SAS, R or other geostatistcal packages employing abundances of *Ae. aegypti,* immature, syvalitic habitat, optimally regressively parameterizable, time series, covariate coefficients and their eco-epidmiological, oviposition, LULC, wavelength frequency distribution data. In these probabilistic paradigms the explanative, dependent variable may either be a clinical, field or remote, unbiased, YFV specified, logistic/Poission, regression, response, remotely sensed LULC estimator (e.g., prevalance or incidence, seasonal, larval count, capture point, sub-meter resolution, geoclassified LULC, waveband, iterative, interpolative, transmittance signature). Hence, pathogen, reservoir, quantitative, LULC data may be mathematically exploitable as these estimators may be employeable as exogenous or endegenous, explanatory, independent, eco-epidemiological, time series, explicative, diagnostic regressors in a linear framework. Linear and non-linear, forecasting, vulnerability, eco-epidemiological, geo-spectrotemporal, probabilistic, unmixed models constructed in a real-time, geostatistical cyberenvironments can model human exposure to vector-borne pathogens, which can have high accuracy for robustly forecasting LULC areas of elevated risk without overestimating risk coverage, based on regressable, time series, unbiased, eco-epidmiological, grid-stratified, orthogonal, endmember descriptors of abundance and distribution of vectors (Jacob and Novak 2014).Such models may be employable to geolocate eco-georeferenceable, seasonally eco-epidemiological, capture point, hyperproductive, sylvatic, *Ae. aegypti*, sub-meter resolution, LULC, time series, oviposition sites within a residual, explicative, optimally diagnostic, geostatistical dataset of endmember, sub-meter resolution, iterative, stochastic or deterministic, fractionalized, endmember, signature variables in an iterative interpolator (e.g., weighted, inerse distance matrix) for remotely optimally distinguishing unknown, ungeosampled, hyperproductive, forest-canopy,immature habitats along an agro-irrigated, African, riceland, peripheral, geoclassifiable, geosampled eco-zone.

Unfortunately, currently in literature two major geostatistical algorithmic approaches have only been employed to investigate seasonal eco-environmental, eco-epidemiological factors related to endmember, YF, endemic, transmission: empirical statistical models and mechanistic models. Empirical geostatistical models are based on inductive approaches and quantitative statistical analyses, including in a spatially explicit context (Vanwambeke et al., 2007a, Vanwambeke et al., 2010. Eisen and Lozano-Fuentes, 2009). The drawback of these models is that their inference is limited due to propagational, unquantiated probabilistic, geo-spectrotemporal LULC uncertainties associated with violations of regression assumtpions (e.g., multicolinear or heteroskedastic residuals) and the inefficient domains employed to construct the paradigms.

SPSS is employed often by applied researchers for constructing eco-epidmiological, vulnerability, forecasting sylvatic, grid-stratified, forecast, vulnerability, YF models from remotely sensed, geoclassifiable LULCs and other parameterizable,clinical, remote and field, covariate uncoalesced, data points. SPSS has many built-in functionalities for managing, analysing and visualizing, YFV-related, eco-georeferenceable, geo-spectrotemporal geoclassified, sub-meter resolution, African riceland or discontinuous, forest-canopied,

oviposition, geoclassified covariates. It is also possible to write syntax code and make some personal scripts for remotely targeting hyperproductive, seasonally eco-georeferenceable, riceland, *Ae. aegypti,* oviposition, sub-meter resolution, LULC, discontinuous, forest -canopied sites. However, it definitely has limitations when it comes to advanced modelling and development of statistical approaches for targeting unknown, ungeosampled, eco-georeferenceable, geoclassifiable, hyperproductive, uncoalesced, agro-irrigated, African, sub-meter resolution, LULC sites along discontinuous, forest canopied, sparsely canopied LULCs. Running tests in SPSS 11.5 on a very large geospatialized, time series, eco-epidemioloigical, sub-meter resolution, elucidative dataset of YF , clinical cases in an agro-irrigated, expanding, riceland, agro-village African ecosystems may be efficient in mid-high level computers (e.g., 1 GB RAM,Files set to 2 GB min and max, dual core processors each 3 GHz).

The point-and-click stage of SPSS may be less tedious for the arbovirologist, medical entomologist or YF experimenter than iteratively quantitatively interpolating an uncoalesced, sub-meter resolution, discontinuous, forest-canopied, sparsely shaded, sub-meter resolution, seasonal spectral, sub-meter resolution, LULC, oviposition, endmember, signature geosampled along the periphery of an African, agro-irrigated, eco-georeferenced, riceland, expanding complex due to anthropogenic pressures. However, without scripting there would be covariance matrix weight limtations in the forecast, vulnerability, unmixing algorithmic, geo-spectrotemporal, probabilistic paradigm. A random vector is a variable with multiple dimensions where each element of the vector is a scalar random variable (Cressie 1993). Each element may have either a finite number of explicative, observed, geometric, endmember uncoalesced, sub-meter resolution, geoclassifiable, orthogonal, grid-stratified, LULC, empirical values or a finite or infinite number of potential values which may be subsequently specified by a joint probability distribution. In a geo-spectrotemporal, forecast, vulnerability, sylvatic, *Ae aegypti,* oviposition, signature, seasonal, sub-meter resolution, eco-epidemiological, capture point, immature, hyperproductive, sub-meter resolution, LULC site on an ecogeoreferenecable, geoclassifiable, discontinuous, forest canopy, sparsely canopied, LULC along an agro-irrigated, riceland, peripherally, grid-stratified polygon. In a joint probability distribution table, numbers in the cells of the table eco-cartographically and/or regressively specifying the probability that particular LULC values ((e.g., targeted, forest canopied, sparsely shaded, ecogeoreferenceable, hyperproductive, seasonal, *Ae. aegypti* oviposition, geoclassified, LULC, eco-epidemiological, capture point along the peripheral of an agro-irrigated, African, riceland agro-ecosystem) i of X and Y may occur together in geospace.

An arbovirologist, medical entomologist or YF experimenter could construct a robust covariance matrix for an eco-epidemiological, forecasting, vulnerability, probabilistic, sub-meter resolution, uncoalesced LULC, wavelength, frequency paradigm to remotely optimally, identify, seasonal, hypeproductive, *Ae. aegypti,* oviposition sites, on geoclassifiable, sub-meter resolution, grid-stratified, forest-canopied, inhomogeneous LULCs along an agro-irrigated, African, riceland, peripheral corridor. He or she could do so by employing the Reliability procedure, the Correlation procedure, or the Regression procedure in SAS which could help quantitate uncoalesced, iteratively interpolative, geo-spectrotemporal, fractionizalied,sub-meter resolution, geoclassified, endmember, wavelength, frequency LULC habitat signatures geosampled along the periphery of the agrovillage complex. The Reliability procedure is

somewhat simpler in that the covariance matrix is automatically printed as a separate table (www.ibm.com/SPSS_Statistics). The Correlation procedure would combine the correlation, significance, cross-product deviations, covariance, and the quantitated expanatorial, LULC, delineated, orthogonal, sub-meter resolution, grid-stratified polygons with an eco-georeferenecable, uncoalesced dataset of endmember empirical, geosampled, clinical, field or remote-specified, geospecrotemporal or geo-spatiotemporal, YFV predictior explanatory variable pair into a single table cell. There is a way to re-organize this table so that the model covariance matrix is printed as a separate table or sub-table. It should be noted that the Reliability procedure employs listwise deletion of cases with missing values, (e.g., the covariances will be computed with only those YF cases that have no missing values on the variables in the analysis). The correlation procedure would employ pairwise deletion by default, whereby each covariance would be optimally calculated with all cases that have valid values on that explicative, eigen-decomposable, predictor variable pair and each variance would be therafter calculated with all cases that have valid values on that variable. Listwise deletion is available as an option in the Correlation procedure. (See SPSS Technote 1475199, which addresses this distinction between the options). The Regression procedure must be run from syntax for the covariance matrix option to be included in the forecasting, vulnerability, probabilistic, endmember, YFV, signature, LULC paradigm. If an arbovirologist, medical entomologist or YF experimenter wants listwise deletion in the model output and deires the covariance matrix to be printed in a separate table, then the Reliability procedure may be the simplest solution. The Covariance matrix will print as a subtable in the Correlations table. Note that regression employs listwise deletion by default but pairwise deletion in a seasonal, African, riceland, eco-epidemiological, YF, forecast, oviposition, vulnerability model can be requested from the "Regression: Options" dialog or by replacing "LISTWISE" with "PAIRWISE" in the REGRESSION command.

Unfortunately with large cyberenvionment geodatabases of continuous, geo-spectrotemporal or geospatiotemporal, geosampled, eco-georeferenceable, uncoalesced, agro-irrigated, riceland, African ecosystem, iterative, interpolative, seasonal, LULC, sub-meter resolution, parameterizable, explanative, forecasting vulnerability, forest canopied, discontinuous, grid-stratified   diagnostic model, endmember estimators may not have unique values. SPSS limitations would be hard to avoid when regression modeling sub-meter resolution, frequency wavelength, endmember, explicative, time series, unmixed, geo-spectrotemporal, wavelength frequencies and other time series, endmember,geo-spatiotemporal, YFV-related,explanatorial, clinical, field or remote LULC data for optimally identifying hyperproductive, sylvatic, *Ae egypti,* oviposition sites along forest canopied peripheral borders in agro-irrigated, African, riceland, agroecosystem environments employing Analysis of variance (ANOVA).

Analysis of variance is a collection of statistical models used to analyze the differences among group means and their associated procedures (such as "variation" among and between groups). The ANOVA procedure automatically produces graphics as part of its ODS output (www.sas.edu). In the ANOVA setting, the observed variance in a particular observational, signature, endmember, predictor, sub-meter resolution, geoclassified, sub-meter resolution, LULC variable (e.g., seasonal, geosampled, *Ae. aegypti*, oviposition, hyperproductive foci

along a discontinuous, forest-canopied, LULC adjacent to to an expanding, agro-irrigated, African, riceland village complex) may be partitioned into frequency wavelength components. In so doing, different sources of eco-geographical, seasonal variation attributable to hypeproductive, *Ae aegypti*, capture point, oviposition sites may be remotely identifiable employing sub-meter resolution data. In its simplest form, ANOVA could provide a statistical test of whether or not the discontinuous, forest-canopied, hyperproductive immature habitats along the expanding African, riceland, agro-village corridor tabulated means of several regression groups are equal, and thereafter generalize the t-test to more than two groups. ANOVAs are useful for comparing (testing) three or more means (groups or variables) for statistical significance in a vector entomological, forecast, vulnerability, endemic, transmission-oriented, model (Jacob et al. 2007). It would be conceptually similar to multiple two-sample t-tests, but would render less conservative (e.g., less Type I error) and thus would be suitable to a wide range of practical problems in a prognosticative, hyperproductive, sub-meter resolution, LULC, *Ae. aegypti*, ovipoistion, vulnerability model.

The ANOVA procedure is one of several procedures available in SAS/STAT software for quantitating the analysis of variance for a forecasting, eco-epidemiological, probabilistic endmember, vulnerability, sub-meter resolution, signture LULC paradigm for seasonally targeting ecogeoreferenceable, hypeproductive, *Ae.aegypti*, geoclassifiable, oviposition, ecological sites on geoclassifiable, forest-canopied LULCs along an agro-irrigated, African, riceland, peripheral corridor. The ANOVA procedure is designed to handle balanced data (that is, data with equal numbers of observations for every combination of the classification factors), whereas the generialized linear model (GLM) procedure can analyze both balanced and unbalanced data. Because PROC ANOVA would take into account the special structure of a balanced design in the oviposition, sylvatic, *Ae aegypti*, remotely sensed, forecast, vulnerability, LULC model, it would be faster and uses less storage than PROC GLM for balanced data.

PROC ANOVA may be employable for the analysis of balanced geosampled, eco-georeferenceable, sylvatic, YFV-related clinical, field or remote-specified, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, LULC , sub-meter resolution data only, with the following exceptions: one-way analysis of variance, Latin square designs, certain partially balanced incomplete block designs, completely nested (hierarchical) designs, and designs with cell frequencies that are proportional to each other and are also proportional to the background population. These exceptions have designs in which the factors are all orthogonal to each other. For further discussion, see Searle (1971, p. 138).

PROC ANOVA works for designs with block diagonal $\mathbf{X'X}$ matrices where the elements of each block all have the same value. The procedure would partially test the eco-epidemiological, probabilistic endmember, sub-meter, resolution, signture paradigm for equal cell means. However, this test is imperfect: some LULC designs for targeting *Ae aegypti* oviposition sites along inhomogeneous, forest-canopy sites in expanding, African, agro-irrigated ecoystems that cannot be analyzed correctly might pass the test, and designs that can be analyzed correctly might not pass. If the model design does not pass the test, PROC ANOVA will produce a warning message to tell an arbovirologist, medical entomologist or YF experimenter that the design is unbalanced and that the ANOVA analyses might not be valid. Complete validation of

designs for a robustifiable, sylvatic, *Ae. aegypti*, sub-meter resolution, eco-epidemiological, forecasting, vulnerability, endmember model may not be performed, in PROC ANOVA since this would require the whole $\mathbf{X'X}$matrix; Alternatively, if PROC ANOVA is employed for the YF analysis of unbalanced, time series, clinical, field or remote, LULC data, an arbovirologist, medical entomologist or YF experimenter must assume responsibility for the validity of the results (e.g., remotely targeted, *Ae. aegypti*, seasonal, eco-georeferenecable, eco-epidemiological, oviposition, capture point on a geoclassified, partially canopied, forested LULC along an expanding African riceland agro-village complex).

Further, if an arbovirologist, medical entomologist or YF experimenter constructs an ANOVA employing eco-epidemiological, uncoalesced, sub-meter resolution, *Ae egypti,* oviposition, sub-meter resolution, wavelength, trasmittance frequencies for optimally targeting hyperproductive,immature habitats along geoclassifiable, forest-canopied LULCs along an African, ricefield, peripheral eco-zone in SPSS, the resulting F value and significance level would only detail one group in the analysis as different from at least one other. Analysis of Variance, is a statistical method employed to compare the means of more than two sets of data, to see if they are statistically different from each other (www.spss.com).

SPSS would allow the use of a one-way ANOVA for quantitating the, inhomogeneous, forest-canopied, *Ae. aegypti,* immature, seasonal productivity, non-homogenous, LULC counts along a peripheral, expanding, African, riceland corridor for regrerssively predicting, high risk eco-zones in theses schemes. Unfortunately, SPSS will not detail how many geoclassifiable LULCs or other geoclassfied group, or groups, of parameterizable, time series, regression covariates differ statistically. In order to determine this, follow-up comparisons must be performed. This is rarely a problem in small analyses, but the higher the number of groups included in the follow-up test, the greater the chance of making a Type I error in the eco-epidemiological, YFV model, which is like assuming an effect where there is not one in the paradigm. In statistical hypothesis testing, a Type I error is the incorrect rejection of a true null hypothesis (a "false positive", i.e., accepting a false hypothesis as correct), while a Type II error is the failure to reject a false null hypothesis (a "false negative", i.e., rejecting a true hypothesis as incorrect) (Sheskin, 2004).

Another limitation of ANOVA is that it would assume that the testable, YFV, entomological and land cover, geo-spectrotemporally or geo-spatiotemporally, geoclassified groups may have the same, or very similar, standard deviations. The greater the difference in standard deviations between groups, the greater chance that the conclusion of the test is inaccurate (Hosmer and Lemshew 2002). Like the normal distribution assumption, this would not be a problem in an prognosticative, vulnerability, geometric, endmember, LULC model as long as the standard deviations of the uncoalesced, capture point (i.e, seasonal, hyperproductive, eco-georeferenceable, African, riceland, *Ae. aegypti*, capture point, immature habitat) are not hugely different, and the sample sizes of each group are roughly equal in a sub-meter resolution, eco-epidemiological, forecasting, vulnerability, spectral, wavelength, frequency transmittance model. If this is not the case, a Welch test is a better option.

In statistics, Welch's t-test, or unequal variances t-test, is a two-sample location test which is employable to test the hypothesis that two populations have equal means. For conducting a YF –related Welch's t-test an arbovirologist, medical entomologist or other

experimenter may employ an uncoalesced, iteratively qualitatively interpolative, quantitative,e co-epidemiological, orthogonal dataset of eco-georeferenceable, sub-meter resolution, endmember LULC, signature variables of forest, discontinuously canopied, capture point, *Ae. aegypti,* oviposition sites in an expanding, agro-irrigated riceland environment, if 1) the first sample of size n_1 in the LULC, forecast, vulnerability, orthogonal, eco-epidemiological, probabilistic paradigm is drawn from a normal population with mean μ_1 and variance σ^2, 2)the second sample of size n_2 is also drawn from a normal population with a mean μ_2 : and, 3) the two samples are independent. (see Welch 1947). Welch's t-test is an adaptation of Student's t-test, that is, it has been derived with the help of Student's t-test and is more reliable when the two samples have unequal variances and unequal sample sizes (Hosmer and Lemeshew 2002).

A t-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis. In probability and statistics, Student's t-distribution (or simply the t-distribution) is any member of a family of continuous probability distributions that arises when estimating the mean of a normally distributed population in situations where the sample size is small and population standard deviation is unknown (Fotheringham 2002). The distribution may be employable to determine if two sets of geosampled LULC datasets (e.g., uncoalesced, sub-meter resolution, geometric, endmember, coefficients extracted from an eigen-decomposition of a discontinuous, forest-canopied, geospectral wavelength frequency, hyperproductive, seasonal, *Ae egypti,* oviposition, capture point, and its neighboring agro-irrigated, African riceland ecosystem, foci) are significantly different from each other in sampled immatures. Whereas a normal distribution describes a full population, t-distributions describe samples drawn from a full population; accordingly, the t-distribution for each sample size is different, and the larger the sample, the more the distribution resembles a normal distribution The t-distribution plays a role in a number of widely used statistical analyses, including Student's t-test for assessing the statistical significance of the difference between two sample means, the construction of confidence intervals for the difference between two population means, and in linear regression analysis (Hosmer and Lemeshew 2002). The Student's t-distribution also arises in the Bayesian analysis of data from a normal family (Gelman et al 1995), A t-test is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an orthogonalized grid-stratified, *Ae. aegypti,* oviposition, geoclassified, sub-meter resolution, expanding African, riceland, regression estimate along a forest corridor based on the LULC data, the test statistics (under certain conditions) may follow a Student's t distribution.

NORPAR CORR in SAS can store the entire array of uncoalesced, endmember, geo-spectrotemporal, sub-meter resolution, fractionalized, geosampled, eco-georeferenceable, intermittently canopied, LULC, orthogonal, wavelength frequencies rendered from an eco-georefernced, capture point, *Ae. aegypti,* oviposition, site geosampled in an irrigated, African, riceland expanding ecosystem. NONPAR CORR stores the entire array of variables requested in memory to get their rank (www. sas.edu). If explanatory, discrete, YFV, geosampled, numerical, finite integer values exist in a geosampled, LULC dataset an arbovirologist, medical entomologist or YF experimenter may optimally regressively quantitate, nonparametric correlations using CROSSTABS which may or may not store all the time series, explanative, YFV data in memory. Note that standard Perason correlation tests take very little memory for

constructing a forecast, vulnerability, probabilistic, endmember paradigm (see Jacob et. al. 2011).

In statistics, the Pearson product-moment correlation coefficient (sometimes referred to as the PPMCC or PCC or Pearson's r) is a measure of the linear correlation between two variables X and Y, giving a value between +1 and −1 inclusive, where 1 is total positive correlation, 0 is no correlation, and −1 is total negative correlation [Kendall and Stuart, (1973)]. It is widely employed in entomological sciences as a measure of the degree of linear dependence between two geoampled, field-operational, LULC variables. Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations (Kenney,and Keeping 1951). The form of the definition involves a "product moment", that is, the mean (the first moment about the origin) of the product of the mean-adjusted, time series, random variables; hence, the modifier product-moment in the name.The absolute values of both the sample and population Pearson correlation coefficients are less than or equal to 1. Correlations equal to 1 or −1 may correspond to hyperproductive, *Ae. aegypti*, oviposition, seasonal, capture points in expanding, African, riceland, agro-irrigated ecosystems lying exactly on a line (in the case of the sample correlation), or to a bivariate distribution entirely supported on a line (in the case of the population correlation). The Pearson correlation coefficient is symmetric: corr(X,Y) = corr(Y,X) (Hosmer and Lemeshew 2002).

A key mathematical property of the Pearson correlation coefficient is that it is invariant to separate changes in location (e.g., seasonal hypeproductive, *Ae. aegypti,* capture point, sub-meter resolution and LULC scale in the two variables). That is, a arbovirologist, medical entomologist or YF experimenter may transform X to a + bX and transform Y to c + dY, where a, b, c, and d are constants with b, d ≠ 0, without changing the correlation coefficient in a vulnerability, orthogonal, forecast model for targeting, prolific, seasonal, oviposition sites. This fact would holds for both the geosampled, geo-spatiotemrpoal or geo-spectrotemporal, mosquito, vector population and sample Pearson correlation coefficients. Note that more general linear transformations would change the correlation coefficints of the regressors in the eco-epidemiological, time series, YF model. The correlation coefficients would range from −1 to 1. A value of 1 would imply that a linear equation can optimally describe the relationship between X and Y perfectly, with all geosampled oviposition, LULC, data points lying on a line for which Y increases as X increases. A value of −1, on the other hand would imply that all geosampled, African, riceland, forest-canopied capture points converted to agro-irrigated habitat lie on a line for which Y decreases as X increases. A value of 0 would imply that there is no linear correlation between the geo-spectrotemporally or geo-spatiotemporally geosampled, YF, predictor variables.

More generally, note that $(X_i − X)(Y_i − Y)$employed for regressing an empirical dataset of eco-georeferenceable sub-meter resolution, geoclassifable LULCs employed for optimally, remotely, targeting hyperproductive, seasonal, *Ae. aegypti* in an expanding, African, riceland environment is positive if and only if X_i and Y_i lie on the same side of their respective means. Thus, the correlation coefficient is positive if X_i and Y_i tend to be simultaneously greater than, or simultaneously less than, their respective means in the eco-epidemiological, YF, forecast vulnerability, eco-georferenceable model. Conversely, the correlation coefficient would be negative if X_i and Y_i tend to lie on opposite sides of their respective means in the YF model.

Moreover, the stronger the tendency in either direction in the model, the larger is the absolute value of the correlation coefficient rendered.

The Pearson correlation is +1 in the case of a perfect direct (increasing) linear relationship (correlation), −1 in the case of a perfect decreasing (inverse) linear relationship (anticorrelation), and some value between −1 and 1 in all other cases, indicating the degree of linear dependence between the variables (Fotheringham 2002). As it approaches zero there is less of a relationship (closer to uncorrelated). Hence, the closer the coefficient is to either −1 or 1 in a YF model for optimally targeting, prolific, oviposition, *Ae aegypti* sites in an expanding, African, agro-village into inhomogeneous, forest-canopied, geoclassified LULCs, the stronger the correlation between the variables.

If the geo-spectrotemporal and or geo-spatiotemporal LULC variables are independent in a YF, forecast, vulnerability,time series, eco-epidemiological, sub-meter resolution model, the Pearson's correlation coefficient would be 0, but the converse is not true because the correlation coefficient would detect only linear dependencies between any two geosampled *Ae. aegypti* variables. For example, suppose a agro-irrigated geosampled immature habitat explanatorial, time series, random variable X is symmetrically distributed about zero, and $Y = X^2$. Then Y would be completely responsive to X, since X and Y are perfectly dependent, but their correlation in the YF model would be zero; (i.e., uncorrelated). However, in the special case when X and Y are jointly normal, uncorrelatedness would be equivalent to independence in the YF model.

Unfortunately, the Pearson correlation in a forecast, YF, vulnerability model for targeting seasonal hyperproductive, sylvatic, *Ae. aegypti* oviposition, sub-meter resolution, LULC sites would be optimally definable only if both of the standard deviations are finite and nonzero. It would be a corollary of the Cauchy–Schwarz inequality that the correlation realized by the eco-epidemiological, remotely sensed YF, model cannot exceed 1 in absolute value. The correlation coefficient is symmetric: $corr(X,Y) = corr(Y,X)$ (Hosmer ande Lemeshew 2002). In mathematics, the Cauchy–Schwarz inequality, also known as the Cauchy–Bunyakovsky–Schwartz inequality, is a useful inequality encountered in many different settings, such as linear algebra, analysis, probability theory, vector algebra and other areas. It is considered to be one of the most important inequalities in all of mathematics

The Cauchy-Schwarz inequality is an elementary inequality and at the same time a powerful inequality which can be stated as follows: Theorem. Let (a1, a2, . . . , an) and (b1, b2, . . . , bn) be two sequences of real numbers (e.g., geosampled, sylvatic, *Ae aegypti* oviposition finite, riceland agro-irrigated integer values then Xni=1a2i! Xni=1b2i!_Xni=1aibi!2, (1)with equality if and only if the sequences (a1, a2, . . . , an) and (b1, b2, . . . , bn) are proportional, [i.e., there is a constant _ such that ak = _bk for each k 2 {1, 2, . . . , n}].

There is a way to construct an eco-epidemiological, YF forecast, vulnerability model in SPSS although it is a little tricky and, must use the syntax window. The partial correlation command can take a matrix as its input, so an arbovirologist, medical entomologist or YF exprimneter must calculate the rank-order correlation matrix without partialling any geo-spectrotemporal or geospatiotemporal, time series, explanatory, LULC variables out. In such

circumstances an input rank-order matrix into the partial correlation procedure would optimally remotely target, hyperproductive, *Ae. aegypti,* oviposition, seasonal sites on sub-meter reolution, inhomogeneous, regressively, forst-canopy data.

Further, for plotting normal or non-Gaussian probability distributions rendered from a YF, forecast, vulnerability, oviposition, time series model for optimally targeting prolific, sylvatic, *Ae. aegypti* habitats on transitional remotely demarcated, forest-corridor, LULC sites with viable confidence intervals a individual score(e.g, 95% pseudo R^2 value) in SAS could render diagnostic, explanative, uncoalesced, geo-spectrotemporal wavelength, frequencies X which may be optimally adjusted in regression space such that the tabulated scores may be qualitatively quantized, geo-spectrotemporally or geo-spatiotemporally. The only way to adjust the scores however in the YF model would be to run a regression in which the covariates predict the X and thereafter save and the diagnotstic plot with residual variance. In that case, software packages such as R and Matlab may be much more helpful than SPSS. These packages, however, would require additional effort from an arbovirologist, medical entomologist or YF investigator since there would be no longer a "point-and-click" orientation.

In applications of classic, linear regression, explanative, metamodels and their concomitant seasonal, experimental designs for explicatively, probabilistically regressing, time series, forecastable, ecogeoreferenceable, heuristically optimizable, YF, endemic transmission, uncoalesced, LULC, iteratively interpolative, signature covariates, the assumumptions of a univariate (not multivariate) simulation response may not include white noise. A medical entomological, explanatively regressable, random vector (i.e., a partially indeterminate process that produces vectors of optimally parameterizable, sub-meter resolution, diagnostic, time series, elucidative, YFV clincial, field or remote specfied, geo-spectrotemporal or geo-spatiotemporal autoregressive explanator) is said to be a white noise vector or white random vector if its fractionalizable, reference, target, sub-mixel signature components each have a probability distribution with zero mean and finite variance. Additionally, statistically independence would be mandatory that is, their joint probability distribution would have to be the product of the distributions of the individual geo-spectotemporal or geo-spatiotemporal, sub-meter reolution, endmember, LULC components (see Jacob 2007).

In the study of probability in a forecasting, eco-epidemiological, geo-spectrotemporal and or geo-spatiotemporal, explanative, diagnostic, seasonal, linear, probabilistic, uncoalesced, YF, vulnerability-oriented, eco-epidemiological, regressively optimizable, forecast model, given at least two explictively, randomized, exogenous, oviposition, African riceland, expanding geosampled regressands constructed from sub-meter resolution, geoclassifiable, ecogeoreferenceable, *Ae. aegypti,* iteratively, quantatively, explanatively, diagnostically interpolative, uncoalesced, LULC, endmember signature variables X, Y, ..., that are defineable on a probability space, the joint probability distribution for X, Y, ... would be a probability distribution that effectively renders the probability that each of X, Y, ... falls in any particular range or explicit, discrete, parameterizable, coefficient values. A probability space in an eco-epidemiological, seasonal, regression-related, predictive, medical entomological, probabilistic, eco-epidemiological, endmember, sub-meter resolution, LULC risk model can be a measure space such that the measure of the whole space is equal to 1 (Gu and Novak 2005, Griffith 2005, Jacob et al. 2009).

If a regression-related, probabilistic-oriented, time series, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, YF, forecasting, vulnerability, sub-meter resolution, LULC, endmember model is multivariate and some sub-meter resolution, descriptive, iterative, quantiative, interpolative, explicatively uncoalesced, LULC time series, explanatorial, African, expanding agro-irrigated agro-ecoystem, predictor covariates are measured with some error, the effects of the measured parameterizable data may be be biased, (see Carroll et al., 1985; Gleser et al., 1987). The direction of the bias would then depend on the correlation amongst the time series, eco-epidemiological, YFV, African riceland geo-spatiotemporal or geo-spatiotemporal, geosampled, geoclassified LULC covariates. Moreover, heteroskedastic, undiagnosed varaiables may cause a loss of power for detecting signals amongst the uncoalesced, quantitatable, endmembers and may mask important features of the regressed, ecogeoreferenceable, sylvatic, *Ae. aegypti*, oviposition, LULC,capture point, immature site productivity counts   on ArcGIS geoclassifiable, forest-canopied, sub-meter resolution, peripheral, grid-stratfiied LULCs along a, sparsely shaded, agro-irrigated, African, expanding riceland corridor due to anthropogenic pressure. Regression error is often completely ignored or not treated properly. One reason might be that abovirologists, medical entomologists, or YF experimenters, often pay very little attention to violation of regression assumptions in eco-epidemiological, vector, arthropod-related, geo-spatiotemporal or geo-spectrotemporal, forecasting, vulnerability, eco-epidemiological endmember , sub-meter resolution, models although the problems have been recognized for a long time

Virus pathogens of YF may respond to changing seasonal, sub-meter resolution, endmember, grid-stratifiable,sub-meter resolution, LULC dynamics on multiple endmeic, transmission levels which may increase in disturbed agro-irrigated, ecosystems, yet current knowledge of mosquito diversity and the relative abundance of vectors as a function of habitat capture point is limited for implementing seasonal control strategies in expanding, African, riceland ecosystems.   Describing seasonal, *Ae. aegypti,* mosquito-community diversity employing geoclassifiable, sub-meter resolution, endmember, oviposition LULC, change site data with other covariates (e.g., meteorological, census etc) in ArcGIS may aid in implementation of a sylvatic, YF, oviposition, endmember signature, LULC, surveillance system in   these expanding agro-irrigated ecosytems. Additionally SAS or R may be optimally employable   for identification of prolific, eco-georeferenceable, seasonal capture point, immature habitat types and immature productivity of the foci, especially in landscapes experiencing different levels of anthropogenic and economic pressures (e.g., partially canopied, forest-canopied LULC eco-zones along peripherial, agro-irrigated, riceland agro-villages). The modification of the urban landscape influences the local (microscale), mesoscale, and even the macroscale climate (Brazel et al., 2000; Quattrochi et al., 2000; Voogt and Oke, 2003).

Assumptions of a general, aquatic, larval population or identical habitats are untenable for a realistic evaluation of larval control interventions or the development of models to be empirically tested under field conditions (Gu and Novak 2005). Diverse spatial data exist, varying in extent and resolution, but sub-meter, remotely-sensed, mosaicked sub-meter resolution, geocalssified, LULC data in ArcGIS provides the most robust extensive coverage for unveiling heterogeneity and quantitated endmember signatures(Jensen 2005) of disturbed landscapes where seasonally explanatorial, hyperproductive, vector, arthropod-related, immature, habitat data are commonly eco-georeferenced (Jacob et al. 2011, Jacob et al. 2007). An

endmember is the spectrum of a pure ground component (e.g., vegetation, soil, water); thus, these sub-mixed pixel (mixel) fractions are typically interpreted as ground LULC components.

It is well documented that urbanization results in increased amount of impervious surfaces (Brabec et. al., 2002) which can augment the intensity, volume, temperature, and duration of irrigated, storm water runoff (Booth and Reinfelt, 1993; Schueler, 1994). Urban, African, riceland, agro-village complex, seasonal, storm water runoff may cause or contribute to water quality degradation by changing natural, geohydrologic, seasonal, LULC patterns (Driver and Troutman, 1989), accelerating natural stream flows (Booth and Jackson, 1997), decreasing stream bank erosion (May et al., 1997),increasing aquatic habitat (Booth and Reinelt, 1993; Horner et. al., 1997), degrading stream water quality (Schueler, 1994; Booth and Jackson, 1997; May et al., 1997), increasing temperature (Galli, 1990), and elevating pollutants. Hence, different land uses can modify contact of susceptible humans with infectious vectors (e.g., YFV) which can modify human cases distribution. Biodiversity loss affects infectious disease risk by disrupting normal relationships between hosts and pathogens on specific landscapes (Patz 2000).

Landscape patterns in seasonal, vector, arthropod-related, uncoalesced, remotely sensed, explanative dataset of observational endmember predictors are a common source of heteroscedasticity in the errors (Jacob et al. 2006, Jacob et al. 2005) due to unexplained variations in dependent variables throughout the course of an African rice-cycle season. The response variable may be consistent in percentage rather than absolute terms, in which case larger uncertainty errors in the regressed *Ae aegypti,* sub-meter resolution, endmember dataset may occur in eco-epidemiological, residual, explicative, diagnostic forecasts in ArcGIS (e.g., targeted, oviposition, hyperproductive, LULC sites in a partially shaded, forest-canopied, sub-meter resolution, geoclassifiable LULC along the periphery of an expanding, agro-irrigated, African, riceland environment where activity is greater, which could show up as a LULC pattern of changing variance on the residual-vs-time plots in PROC LOGISTIC. A log transformation may be employable to address this problem. For example, if the seasonal, YF, endemic, topological pattern is being modeled through the use of dummy variables for months or quarters of the year, a log transformation applied to the dependent variable (e.g., incidence of of YF cases)will convert the LULC coefficients of the dummy variables to multiplicative adjustment factors rather than additive adjustment factors. In so doing, the errors in prognosticated logged LULC variables will be roughly interpretable as percentage errors based on the original geosampled, geo-spectrotemporal or geo-spatiotemporal, geoclassifiable, hypeproductive, *Ae aegypti,* oviposition, sub-meter resolution, geoclassifiable, endmember, grid-stratifiable, orthogonal, sub-meter resolution, LULC, diagnostic variable. Seasonal adjustment of all geosampled riceland and or inhomogenous, forest-canopied, endmember, grid-stratfiable,orthogonal LULC data prior to fitting the regression model might be another option when targeting hyperproductive, seasonal, *Ae aegypti* oviposition, LULC sites in an expanding African rice-agro-village complex. If a log transformation has already been applied to an eco-georeferenceable, explanatorial, eco-epidemiological,geo-spectrotemporal, or geo-spatiotemporal, sub-meter resolution, diagnostic, clinical, field or remote-specified, exogenous, endemic, YFV transmission-oriented, explanatory regressor, then additive rather than multiplicative seasonal adjustments may be employable in PROC LOGISTIC for targeting, seasonal, hyperproductive, capture points along the forest canopy periphery of the expanding ecosystem.

Homoscedasticity of the errors would have to exist in the endemic, forecast, vulnerability, YF , eco-georeferenceable, eco-epidemiological, real-time, ArcGIS or SAS model which would be included in the estimates of the sylvatic, *Ae. aegypti*, hyperproductive, seasonal, explanative, capture point, forest canopies, LULC sub-meter imaged, ovipositions sites nearest to the agro-irrigated, expanding borders of the African, riceland agro-village ecosystem, if: (a) common variance versus time in the time series, explanatory, geo-spectrotemporal or geo-spatiotemporal, YFV, eco-epidemiological, diagnostic data is equal (b) common variance in the predictions (e.g., seasonally targeted, eco-georeferenceable, hyperproductive, grid-stratified, capture point, agro-irrigated, immature habitat) (c) geomorphological, terrain-related, explanatory independent variable,   three (3) –dimensional (D) catchment watershed, parameterizable, covariate estimator. Finally, the eco-epidemiological, forecast, vulnerability YF model would have to display normality of the error distribution.

Violations of normality in an explanatory, residual, YF explicative, time series, diagnostic, forecast, paradigm output remotely targeting, sub-meter resolution, forest canopied, sparsely shaded, *Ae. aegypti* seasonally hyperprodutive, capture point, seasonal, eco-georeferenceable, immature habitats can create problems for optimally determining whether vulnerability, probabilistic, iteratively, interpolative, uncoalesced, sub-meter resolution, model coefficients are significantly different from zero when calculating confidence intervals. Sometimes the error distribution may be "skewed" by the presence of a few, large, geospatial outliers (i.e, “extereme” *Ae. aegypti* capture point, immature habitat observations)   in a forecasting vulnerability, seasonal, vector, arthropod-related, eco-epidemiological, LULC, regressively predictive, probabilistic, YF model. Since parameter estimation would be based on the minimization of squared error in the regressed,   geo-spectrotemporally geosampled, fractionalized, geometric, endmember datset of time series, clinical, field or remote, diagnostic, explanative, YFV geosampled, sub-meter resolution, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, endmember, LULC variables, a few extreme observations can exert a disproportionate influence on optimal parameter estimation. Calculation of confidence intervals and various significance tests for LULC coefficients in most entomological, immature, capture point, habitat-based, reference, signature, targeted, intervention models are all based on the assumptions of normally distributed errors (Gu and Novak 2005). If the error distribution is significantly non-normal, confidence intervals may be too wide or too narrow, the targeted, sylvatic, forest-canopied, hyperproductive, seasonal, *Ae. aegypti* oviposition, endmember, LULC sites in geo-specifically geoclassified, sub-meter resolution, agro-irrigated, African, expanding riceland LULCs, due to anthropogenic or economic pressure, may be misspecified.

Technically, the normal distribution assumption is not necessary if an arbovirologist, medical entomologist or other YF experimenter are willing to assume that the eco-epidmiological, regression, grid-stratified, orthogonal, eigen-decomposable, vulnerability model geo-spectrotemporal or go-spatiotemporal iterative, frequency, wavelength, sub-meter resolution, endmember, signature predictive risk equation is correct when field-verifying (“ground truthing’) remotely sensed LULC targets where the primary goal is to quantitate oviposition, paramterizable covariate coefficients to generate robust, stochastic or deterministic, signature explanators   (e.g., hyperproductive, *Ae. aegypti* sites) in such a way as to minimize mean squared error. The formulas for optimally estimating heuristically parameterizable, YF, time series, uncoalesced, sub-meter resolution, iteratively, qualitatively interpolative,   LULC

signature coefficients would require no more than that, since normally distributed errors is amongst the key regression assumption. Generally an arbovirologist, medical entomologist or other YF experimenter may be interested in making inferences about an empirical dataset of diagnostic, forecasting, vulnerability, endmember, latent model descriptors for optimally approximating the probability that a given residual, explanative, geo-spatiotemporal or geo-spectrotempoeal propogagtional, diagnostic, probabilistic, uncertainty estimate error will exceed some threshold in a particular direction, in which case distributional assumptions would be pertinent. Also, a significant violation of the normal distribution assumption is often a "red flag" indicating that there is some other problem with the model assumptions and/or that there are a few unusual data points that should be studied closely such that a better model may be constructed(Fotheringham 2002) .

The optimal test for determining normally distributed error in a eco-epidemiological, eco-georefereneceable, YFV, forecast, vulnerability, regression, eco-epidemiological, LULC, geometric, endmember, quantiative model for targeting hyperproductive, eco-georeferenceable, sylvatic, *Ae aegypti*, seasonal, capture point, immature habitats and senteniel sites in a forest-canopied, geoclassifiable, sub-meter resolution, ArcGIS map of an expanding agro-irrigated, African riceland, agro-village complex, is a normal probability plot or normal quantile plot of the residuals. These are plots of the fractiles of error distribution versus the fractiles of a normal distribution having the same mean and variance. If the distribution is normal in the forecasting YF, endmember, time series paradigm, the points on the residual, diagnostic, explanative, diagnostic, regression plots should fall close to the diagonal reference line. A bow-shaped pattern of deviations from the diagonal would indicate that the YF-related, clinical, field or remote specified, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, eco-georferenceable, sub-meter resolution, explanative, endmember LULC regressed residuals have excessive skewness (i.e., they are not symmetrically distributed, with too many large errors in one direction). An S-shaped pattern of deviations would indicate that the riceland, oviposition, iterative residuals have excessive kurtosis—(i.e., there are either too many or two few large errors in both directions). Sometimes the problem may be that there are a few data points on one or both ends that deviate significantly from the reference line ("outliers"), in which case they should get close attention (Griffth 2003).

There are a variety of statistical tests for normality, in SAS for testing, predictive, YF, sub-meter resolution, LULC, seasonal models for optimally geostatistically delineating hyperproductive, *Ae aegypti*, oviposition, sub-meter resolution, geocalssified LULC sites along an agro-irrigated, African, riceland ecosystem expanding periphery including the Kolmogorov-Smirnov test, the Shapiro-Wilk test, the Jarque-Bera test, and the Anderson-Darling test. The Anderson-Darling test (which is the one used by RegressIt) is generally considered to be optimal for seasonal, vector, arthropod-related, endmember, iterative stochastic or deterministic, time series, iterative quantitative interpolators, as it is specific to the normal distribution (unlike the K-S test) and it would study whole YF, *Ae aegypti* covariate regressed distribution rather than just the skewness and kurtosis (like the J-B test).

The unfortunate assumptions by previous arbovirologists, medical entomologists and other YF experimenters is that they have assumed, rather unwisely, that diagnostic, time series, clinical, field or remote *Ae. aegypti,* oviposition, geo-spectrotemporal or geo-spatiotemporal,

eco-epidemiological, sub-meter resolution, LULC explanators rarely have probabilistic endogenous errors. Thus, the variables selected for regression for optimally remotely targeting, seasonal, hyperproductive, *Ae. aegypti,* oviposition, hypeproductive sites in an expanding African, riceland environment are assumed to be perfectly normally distributed. This justification of the authors in these contributions have in most circumstances fit the eco-georeferenceable, YF, forecast, vulnerability, endmember, sub-meter resolution, LULC, eco-epidemiological, model whose errors did not violate the normality assumption at the 0.05 level of significance. However a backward stepwise multicolinarity diagnostic check in PROC REG may reveal non-homoskedacistic, normalized, *Ae. aegypti*, ovispoition African riceland geosasampled parameters. According to Griffth (2005), frequency and non-frequency tests conducted on, time series endmember, vector arthropod-related, iteratively quantitatively interpolative, eigen-decomposable, descriptive, LULC regressors contain inconspicuous probabilistic propagational non-normalities (e..g, latent negative autocorrelation). These hidden violations of non-linear regression assumptions can influence the generation of few outliers. Thus, normality assumed in the unquantitated eco-epidemiological, eco-georeferenceable, forecast datatsets of endmember, seasonal, hyperproductive, *Ae aegypti*, oviposition, sites on sub-meter resolution, parameterizable, orthogonal, geoclassified, forest-canopied, grid-stratified LULCs along an agro-irrigated, expanding, African, riceland community would be mathematically erroneous.

A PROC REG normal probability plot or normal quantile plot could draw precise conclusions about whether a regression parameter estimation problem is serious in an eco-epidemiological, endmember, YF forecast, vulnerability geo-spectrotemporal or geo-spatiotemporal, signature model employing a backward analyses If any of regression assumptions is violated in a YF model (i.e., if there are nonlinear explanative, relationships between dependent and independent explanatory variables, or the diagnostic errors exhibit correlation, heteroscedasticity, or non-normality) then the residually forecasted, confidence intervals, yielded by the regression model may be (at best) inefficient or (at worst) seriously biased or misleading. Ideally, mathematical software automatically provide charts and statistics that test whether regression assumptions are satisfied for any given LULC model. Unfortunately, many software packages do not provide such outputs by default (i.e., additional menu commands must be executed or code must be written) and, some such as Excel's built-in regression add-in offer only limited options. Unfortunately, many seasonal, vector, arthropod-related, real-time, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, iteratively quantitatively interpolative, sub-meter resolution, forecast, vulnerability LULC models violates most regression assumptions yet they exist in literature as they are accepted by a naïve user on the basis of a large value pseudo R^2.

The normal quantile plots from seasaonal, forecasting, YF, vulnerability probabilistic, endmember, uncoalesced, orthogonal, grid-stratfied paradigms constructed from *Ae. aegypti* riceland, geosampled, geo-spectrotemporal or geo-spatiotemporal, sub-meter resolution, LULC covariate estimators may be shown employing quantile-quantile (Q-Q) plots for determing geolocatons of *Ae aegypti* immature, capture point, hyperproductive habitats along an expanding African agro-irrigated environment. A Q–Q is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. First, the set of intervals for the quantiles is chosen. A point (x, y) on the plot may be correspond to one of

the quantiles of the second distribution (y-coordinate of a geo-spectrotemporal or geo-spatiotemporal, endmember, YF, geometric model) plotted against the same quantile of the first distribution (x-coordinate). Thus, the line is a parametric curve in the forecast, vulnerability, eco-epidemiological YF, model for optimally targeting partially discontinuous, diagnostic, explanatory probabilistic, sub-meter, resolution, African, riceland, sylvatic, *Ae. aegypti*, oviposition, inhomogeneous, forest-canopied, LULC, parameterized estimators would be the (number of the) interval for the quantile. If the two YF distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line y = x. If the distributions are linearly quantiatively related, the geosampled oviposition, eco-epidemiological, LULC, signature, sub-meter resolution, capture points in the Q–Q plot will approximately lie on a line, but not necessarily on the line y = x. Q–Q plots can also be used as a graphical means of estimating geosampled geo-spatiotemporal or geo-spectrotemporal, oviposition, endmember parameters in a capture point, seasonal, hyperproductive, geolocation-scale family of distributions (Anselin 1995).

The QQPLOT statement can create and compare ordered variable medical entomological oviposition, optimally regressable, sub-meter resolution, African, riceland, grid-stratfiable, endmember, LULC values in a sub-meter resolution, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, YF-related, iterative, quantiatively, interpolative, ecogeoreferenceable, risk model with quantiles of a specified theoretical distribution. If the geosampled orthogonal LULC data distribution matches the theoretical distribution, the forecasted prolific, eco-epidemiological, seasonal, hyperproductive, immature habitat, capture points on the plot would form a linear pattern. Thus, an arbovirologist, medical entomologist or other YF experimenter may employ a Q-Q plot to quantitate how well a eco-epidemiological prognosticative, LULC distribution, model could optimally render a set of oviposition, immature habitat measurements for targeting hyperproductive, *Ae aegypti,* sites in geoclassified empirical dataset of eco-georeferenceable, geolocations along a forest canopy corridor on the periphery of a agro-irrigated, African, riceland, agro-village ecosystem.

Q-Q plots are similar to probability plots, which can be created with the PROBPLOT statement. Q-Q plots may be preferable for graphical estimation of endmember LULC distribution parameters related to geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, eco-georeferenceable, sub-meter resolution, *Ae. aegypti,* prolific, seasonal sites on sub-meter resolution, geoclassified data; whereas, probability plots are preferable for graphical estimation of percentiles. Any number of QQPLOT statements in the UNIVARIATE procedure can generate an explanatorial, residual, diagnostic, graphical, regression test for optimally deducing if an empirical endmember LULC dataset of time series dependent, clinical, field or remote geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, geosampled, eco-georeferenceable, uncoalesced, iteratively quantitatively interpolative variables can optimally delineate a seasonal, hyerproductive, *Ae. aegypti*, oviposition LULC site, on an forest-canopied, sub-meter resolution, geoclassifiable LULC along a peripherial geolocation of an expanding, agro-irrigated, African, riceland, ago-village community due to anthropogenic pressure.

The UNIVARIATE procedure may render the following: descriptive statistics based on moments for parimoniously, robustly, regressively, prognosticating, experimental, time series, dependent, orthogonal, residual forecasts of targeted hyperproductive, sylvatic, *Ae. aegypti,* oviposition, seasonal, grid-stratieid, LULC sites in transitional expanding, African, agro-village ecosystem into inhomogeneously, forest-canopied, sub-meter resolution geoclassified LULCs with high statistical significance (e.g.,confidence interval of 95%). The model diagnostic eco-epidemiological, output would   include skewness and kurtosis, quantiles or percentiles (such as the median), frequency tables, and extreme valued histograms that optionally can be fitted with probability density curves for quantitating visually various distributions (e.g., regressed ecogeoreferenceable, oviposition, capture point, hyperproductive, foci, signature geolocations with kernel density estimates and cumulative distribution function plots of diagnostic, geosampled, clincial, field or remote geospectrotemporal or geo-spatiotemporal, YFV, geosampled, sub-meter resolution, geoclassified LULCs. Optionally, these can be superimposed with probability distribution curves for various distributions. Q-Q plots, probability plots, and probability-probability plots (P-P plots) may aid in unveiling inconspicious, propagational, probabilistic, erroneous, time series, capture point, *Ae aegypti*, immature habitat, seasonal variables geosampled along an expanding, African, riceland eco-environment, peripheral, discontinuous, forest-canopy, LULC corridors

A Q–Q plot is used to compare the shapes of endmember LULC distributions, providing a graphical view of how properties such as location, scale, and skewness are similar or different in the two distributions (Anselin 1995). Q–Q plots can be employed to compare collections of endmember geo-spectrotemporal or geo-spatiotemporal eco-epidemiological, geosampled, YF data, or theoretical distributions geosampled in expanding African, riceland, discontinuous corridors as they geospatially spillover into inhomogeneous partially canopied, peripheral forestlands. The use of Q–Q plots would then compare two samples of the data which may be viewed with a non-parametric algorithm to compare their underlying distributions. A Q–Q plot is generally a more powerful approach than the common technique of comparing histograms of the two samples, but requires more skill to interpret (Anselin 1995). Q–Q plots are commonly used to compare a data set to a theoretical model (Griffith 2003). The indictaive signature , sub-meter reoslution, eco-epidemiological, signature, oviposition paradigms can provide an assessment of "goodness of fit" that is graphical, rather than reducing to a numerical summary which would not necesarily remoptely target, prolific, sylvatic, seasonal, hyperproductive, *Ae. aegypti,* immature habitats along an eco-georeferenceable partially, canopied forest, peripheral corridor. Q–Q plots are also usable to compare two theoretical distributions to each other (Anselin 1995) . Since Q–Q plots would compare regressed YF distributions in an expanding, African, agro-rice-village complex there is no need for the geo-spatiotemrpoal or geo-spectrotemporal oviposition signature LULC sub-meter resolution values to be observed as pairs, as in a scatter plot, or even for the numbers of parameterizable, endmember, sub-meter resolution, LULC, endmember, covariate values in the two groups being compared to be equal.

The term "probability plot" sometimes refers specifically to a Q–Q plot, sometimes to a more general class of plots, and sometimes to the less commonly used P–P plot. The probability plot correlation coefficient is a quantity derived from the idea of Q–Q plots, which measures the agreement of a fitted distribution (Cressie 1993). These plots may facilitate the comparison of a regression-related, capture point, seasonal, hyperproductive, *Ae. aegypti,* oviposition,eco-

epidemiological, eco-georerenceable, eigen-decomposable, LULC distribution with various theoretical distributions. Goodness-of-fit tests for a variety of regressed YF, signature, endmember, LULC distributions including the normal has the ability to yield summary statistics on plots which could include the ability to analyze empirical geosampled, geo-spectrotemporal or geo-spatiotemporal, uncoalesced wavelength, sub-meter resolution frequencies. This diagnostic fit could enable arbovirologist, medical entomologist or YF experimenter employing a geoclassifiable, sub-meter resolution, African, riceland, forest-canopy, discontinuous, LULC, explicative, predictor variable the ability to create output endmember datasets containing summary statistics, histogram intervals, and parameters of fitted curves on with observed expanding agro-village data which may be usable as a means of fitting a YF case distribution to geosampled sylvatic, *Ae. aegypti,* oviposition dataset.

For optimally tabulating precise quantiles for obtaining comparative distributions typically the formula $k / (n + 1)$ is used. Several different formulas have been proposed as affine symmetrical plotting positions for LULc model. Such formulas have the form $(k - a) / (n + 1 - 2a)$ for some iteratively quantiatively interpolative, sub-meter resolution signature-related YF value of a in the range from 0 to 1/2, which would render a range between $k / (n + 1)$ and $(k - 1/2) / n$. Other expressions include: $(k - 0.3) / (n + 0.4)$, $k - 0.3175) / (n + 0.365)$, $(k - 0.326) / (n + 0.348)$, $(k - ⅓) / (n + ⅓)$, $(k - 0.375) / (n + 0.25)$, $(k - 0.4) / (n + 0.2)$, $(k - 0.44) / (n + 0.12)$, $(k - 0.5) / (n)$, $(k - 0.567) / (n - 0.134)$, $(k - 1) / (n - 1)$. For large eco-epidemiological, time sreies, YF case sample size, n, there may be little difference between these various LULC expressions especially for targeting, seasonal, hyperproductive, *Ae. aegypti*, oviposition, African, riceland, expanding, agro-village, geoclassifiable, LULC, capture points on geoclassified sub-meter resolution.

Non-noisy, Poissonian and autorgressive, iterative, Bayesian, estimation paradigms have been successfully employed in eco-epidemiological, vector-borne, disease, grid-stratified, sub-meter resolution, orthogonal regression, forecast, vulnerabiltiy models particularly for mapping vector or disease case distribution. (see Griffith 2005, Jacob et al. 2005), In probability theory and statistics, the Poisson distribution theory is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event (Haight 1967). Inferencial Bayesian orthogonal paradigms belongs to the category of evidential probabilities which may evaluate the a hypothesis (e.g., expanding anthropogenic, African, agro-village, agro-irrigated, ecosystems geospatial LULC spillovers into bordering, forest canopy peripheral corridors can generate hyperproductive, seasonal, *Ae. aegypti,* oviposition, sites seasonally). Bayesian probabilist specifies some prior probability, which is then updated to a posterior probability in the light of new, relevant evidence (Cox 2001). These models unbiasing techniques (e.g., stepwise backward analyses, latent autocovariate weighted, orthogonal grid matrices) may be optimal for treating propogational regression spatial error (e.g., autoccorelation) in an YF, forecast, vulnerability mathematical, grid-stratified, orthogonal model for optimally targeting, prolific, seasonal, *Ae. aegypti,* oviposition, eco-geographic, LULC sub-meter resolution, geoclassified sites. In statistical classification, the Bayes error rate is the lowest possible uncertainty rate for any classifier of a random outcome (one of two, diagnostic, eco-georefernceable, time series, eco-epidemiological, sub-meter resolution, uncoalesced, riceland, African, expanding, time series, grid-stratified, LULC categories such as post-harvesting and

tillering) which would be analogous to the irreducible error. The stochastic nature of most computational, vector, arthrpod-related, eco-epidemiological, forecast vulnerability models would gives rise to an instance of the Probabilistic Model Checking (PMC) problem. Suppose M is a stochastic *Ae aegypti*, sub-meter resolution, forecasting vulnerability, geo-spectrotemporal or geo-spatiotemporal, geosampled, eco-epidemiological, sub-meter resolution geoclassifiable, eco-georeferenceable, LULC model over a set of states S, s0 would be then a starting state, whenst φ is a dynamic property expressed as a formula in temporal logic, and $\theta \in [0, 1]$ is the probability threshold. The PMC problem would then render the 4-tuple(M, s0, φ, θ), to decide algorithmically whether M, s0 $\models P_{\geq\theta}(\varphi)$ which could be usable to regressively target seasonal, hyperproductive, immature, seasonal, *Ae. aegypti*, oviposition, capture points in African, riceland agro-village ecosystems along inhomogeneous, forest canopied geocalssified corridors.

Existing algorithms for solving the PMC problem may fall in a predictive, grid-stratified, algorithmic, orthogonal YF, endmember, orthogonal, eigen-decomposable, sub-meter resolution LULC paradigm into one of two categories. The first category would comprise of numerical methods which ma compute the probability with which the property holds with high precision in an eco-epidemiological, YF signature, iterative, interpolative,endmember model. In a deforested, sylvatic, capture point, *Ae. aegypti*, sub-meter resolution, uncoalesced, forecast, vulnerability, African. riceland, inhomogeneous, LULC, probabilistic, signature model for optimally targeting, hyperproductive, immature, eco-epidemiological,YF, capture point, seasonal, oviposition sites, the number of states can be excessive. As such an arbovirologist, medical entomologist or other YF experimenter may seek alternative, geo-statistical, endmember algorithms for solving the PMC problem in an approximate fashion in a grid-stratifiable, sub-meter resolution, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, LULC model. Approximation probabilistic, residual, algorithmic methods in SAS may work by sampling a set of traces from the model. In so doing, ecogeoreferenceable endmember, itearatively interpolative, LULC targets, (e.g., hyperproductive, sylvatic, *Ae. aegypti* oviposition sites) may be optimally rendered from the paradigm summary statement.

For example, an arbovirologist, medical entomologist or YF experimenter may employ PROC CALIS to fit some measurement error, regression eco-epidemiological, forecast, vulnerability signature models. Latent, explicative, diagnostic, sub-meter resolution, grid-stratifiable, orthogonal, time series, LULC variables rendered from these paradigms may optimally define "true" scored, uncoalesced, geo-spectrotemporal or geo-spatiotemporal, iteratively quantitatively, krigable variables that are measureable without errors. In statistics, originally in geostatistics, Kriging or Gaussian process regression is a method of interpolation for which the interpolated values are modeled by a Gaussian process governed by prior covariances, as opposed to a piecewise-polynomial spline chosen to optimize smoothness of the fitted values (Cressie 1993).

A prior distribution may be optimally assigned to the varying sylvatic, YFV, clinical, field or remote-specified, geo-spectrotemporal, regression LULC coefficients, and the oviposition, grid-stratifiable, endmember parameters of that prior distribution themselves may be given as a hyperprior. In Bayesian statistics, a hyperprior is a prior distribution on a hyperparameter, that is, on a parameter of a prior distribution. As with the term hyperparameter, the use of hyper is to distinguish it from a prior distribution of a parameter of the model for the

underlying system (Gelman 1995). These may arise particularly in the use of conjugate priors in a time seies, eco-epidemiological, YF forecast, vulnerability, signature, iteratively interpolative, sub-mter resolution LULV model.

Orthogonal, Bayesian, geometric, endmember, YF, sylvatic, *Ae aegypti,* African, riceland, agro-village, sub-meter resolution, signature, eco-epidemiological, LULC models can be constructed in PROC MCMC in general terms as follows: data ingroups j = 1, . . . , J. For each group j, there would be response vector yj (e.g., targeted, forecast, of an African, riceland, *Ae aegypti,* oviposition, sub-meter, seasonal hyperproductive, eco-georeferenceable LULC along with two data matrices, Xj and Zj), that may have fixed and varying oviposition, endmember, signature, parameterizable, covarite coefficients, respectively. The data model would be p(yj |Xj_ + Zjbj), where the vector of fixed coefficients and bj is the vector of regression coefficients that varies by group. The vectors bj would then be optimally modeled as independent draws from a prior distribution, p(bj), given some hyperparameters when   targeting seasonal, hypeproductive, discontinuous, forest-canopied, sylvatic, *Ae aegypti*, oviposition sub-meter resolution, grid-stratifiable LULCs, in expanding, African, riceland environments.   By assuming a Gaussian model for the varying geosampled, geo-spectrotemporal or geo-spatiotemporal eco-epidemiological, eco-georeferenceable, endmember, LULC sub-meter resolution, riceland coefficients, so that bj ~ N(0,_), the model could also include a nonzero mean vector or a group-level regression structure for the tabulated, hyperprior, sylvatic, endemic, YF distribution, but these can be folded into the fixed coefficients in the entomological, forecast, vulnerability model without loss of generality.

There is a rich literature on full Bayesian inference for hierarchical regressions.There is also an empirical Bayes version in which the hyperparameters (e.g, seasonal, hyperproductive, eco-georferenced, capture point, sylvatic, *Ae. aegypti* , sub-meter resolution LULC oviposition sites in an expanding African riceland environment due to anthropogenic influences)are estimated via ML . In such circumstances the inference for the endmember LULC coefficients may be regressively conditionally performed on the estimated prognosticators. In so doing, seasonal, prolific, *Ae. aegypti,* capture point, habitats may be remotely distinguished on sub-meter resolution geoclassified LULCs along inhomogeneous, partially forest canopied peripheral corridors in these ecosystems. From this perspective, the empirical Bayes approach would be sub-optimal (e.g., eco-georeferenceable, vulnerability, eco-epidemiological, forecasts of hyperprpoductive, *Ae. aegypti,* capture point, immature, hyperproduictive , seasonal habitats would reveal only < than 50% in a sensitivity and or specificity test) , because the Bayesian perspective avoids the use of any prior information and because the model would understate the posterior uncertainty.

When fitting hierarchical, grid-stratified, orthogonal, time series, regression endmember models, ML estimation has computational (and, for some users, philosophical) advantages compared with full Bayesian inference, but whenst the number of groups is small, estimates of the covariance matrix of group-level varying coefficients are often degenerate.   An arbovirologist, medical entomologist or YF experimenter can get a better performanace in a Bayesian estimation model employing a purely point-estimation perspective, and a prior distribution or penalty function. In this paper, we use a Bayes model estimation to obtain positive definite covariance matrix estimates. A class of Wishart (not inverse-Wishart) priors with a

default choice of hyperparameters was employed to regressively distinguish, seasonal, hyperproductive, sylvatic, *Ae. agypti* oviposition, capture point, LULC sites along an expanding forest corridor. The degrees of freedom in the model was set equal to the number of varying geosampled LULC coefficients plus 2, and the scale matrix was the identity matrix multiplied by a value that was large relative to the scale of theimaged oviposition, seasonal hypeproductive *Ae. aegypti*. This prior was equivalent to independent gamma priors for the eigenvalues with a shape parameter 1.5 and rate parameter close to 0. It was also equivalent to independent gamma priors for the variances with the same hyperparameters multiplied by a function of the signature correlation coefficients. With this default prior, the posterior mode was positive definite in the sylvatic,YF model. We noted that the resulting uncertainty for the fixed coefficients in the model was less underestimated than under classical ML or restricted ML. An extension of was used with prior information which was available for some of the variances or correlations in the eco-epidemiological YF, forecast, vulnerability, signature, LULC model.

Therefater the sylvatic, *Ae egypti* hypeproductive, immature, capture point, explanative, erroneous residual ovipositionparameterized,uncoalesced, sub-meter resolution, geoclassified LULC covariates (e.g., size of densely shaded tillers) ,on sub-meter resolution, seasonally geoclassifiable, deforested, peripheral, agro-irrigated, African, rice- village LULCs may be identified, in an ArcGIS geodatabase cyberenvironment employing, endmember, parameter estimators fixed values in the PATH model specification.

Consider a simple linear regression, *Ae. aegypti,* eco-georeferenceable, geo-spectrotemporally, geo-spatiotemporally geosampled, sub-meter resolution, forecasting, vulnerability, eco-epidemiological, unmixed, sub-meter resolution, LULC model with dependent variable *y* and predictor variable *x*. The path diagram for this simple linear regression model would be depicted as follows:

Suppose the following SAS –based, time series, geo-spectrotemproally or geo-spatiotemporally, geosampled, sub-meter resolution, endmember, *Ae. aegypti,* grid-stratified, LULC dataset for the endmember, regression analysis is presented for *y* for quantitating statistical significance values (pvalues at a 95% confidence interval) for *x*:

```
data measures;
  input x y @@;
  datalines;
7.91736   13.8673   6.10807   11.7966   6.94139   12.2174
7.61290   12.9761   6.77190   11.6356   6.33328   11.7732
7.60608   12.8040   6.65642   12.8866   6.26643   11.9382
7.32266   13.2590   5.76977   10.7654   5.62881   11.5041
7.57418   13.2502   7.17305   13.3416   8.23123   13.9876
7.17199   13.1750   8.04604   14.5968   5.77692   11.5077
5.72741   11.3299   6.66033   12.5159   7.14944   12.4988
7.51832   12.3588   5.48877   11.2211   7.50323   13.3735
7.15814   13.1556   7.35485   13.8457   8.91648   14.4929
```

   5.37445   9.6366   6.00419   11.7654   6.89546   13.1493
;

This YF data set contains 30 agro-irrigated, African, riceland, deforested, discontinuous observations for the *x* and *y* variables. The experimenter may fit the simple linear regression model to the measured data varaibles by the PATH model specification of PROC CALIS, as shown in the following statements:

```
proc calis data=measures;
  path
    x ---> y;
  run;
```

In so doing, each explciative, independent variable would be evaluated to determine whether it satisfied the properties of regression. The forecasts may also be employable for satisfying traces used to(approximately) to decide whether M, s0 |= P$\geq\theta(\varphi)$ could be employable to remotely target seasonal, hyperproductive, *Ae. aegypti*, oviposition geolocations on sub-meter resolution, geoclassifiable, grid-stratifiable, LULC, African, riceland, agro-village, ecosystem.

Approximate PMC methods can be further divided into two sub-categories:(i) those that seek to estimate the probability that the property holds and then compare that estimate to θ (e.g., [26,39]), and (ii) those that reduce the PMC problem to a hypothesis testing problem (e.g., [46,47]). In terms of an *Ae egypti*, sub-meter resoluton, georeferenceable, oviposition LULC vulnerability, forecasting, eco-epidemiological, linear model this would be ascertained by deciding between two hypotheses — H0 : P$\geq\theta(\varphi)$ versus H1 : P$<\theta(\varphi)$. Hypothesis-testing eco-epidemiological based YF regression methods may be more efficient than those based on estimation when θ (which may be specified by a sub-meter resolution LULC experiment) which may be significantly different than the true probability that the property holds which may be determined by M and s0 for targeting seasonal, hyperproductive, *Ae. aegypti*, African, riceland, inhomogeneous, sub-meter resolution, forest-canopy, geoclassified LULCs.

We assumed such data could provide a higher level of detail than census-like data, which would geo-spatiotemporally or geo-spectrotemporally aggregate heterogeneous LULC variables (e.g., hyperproductive, *Ae. aegypti* capture points on agro-irrigated, African, deforested, sub-meter resolution, geoclassified LULCs) at the level of administrative units. We also assumed that census-like optimally regressed, time series, uncoalesced, iteratively quantitative, interpolative, grid-stratifiable data may be focused on residence (e.g., eco-georeferenceable, agro-irrigated, African, rice–village complexes) where seasonal, hyperproductive, immature, capture point, sub-meter resolution, LULC habitat existed. Quantizing temporal variables of humans who spend time data in other neighboring LULCs (e.g., discontinous forested-canopy) than their residence may have potentially profound implications for YF epidemiological, forecast, vulnerability modelling in expanding African agro-irrigated, riceland environments.

Regression and PMC checking may be performed in PROC REG to determine the best scale for optimal, capture point, *Ae aegypti*, rice-cycle, seasonal, LULC observations to determine influential sub-meter resolution, uncoalesced, optimizable, endmember, LULC variables associated with immature hyperproductivity. A radius buffer was determined in

Geospatial Analyst [TM] for cartographically quantitating the sub-meter resolution, imaged, seasonal, iterative, interpolative, geo-spectrotemporal, geospatialized LULCs and mosquito abundance relationship, *Ae. aegypti* presence may be positively associated with structure and medium height trees and negatively associated with bare earth, for example in an agro-irrigated, African, riceland, immature, capture point, eco-georeferenceable forecast, vulnerability analyses. Such findings may emphasize quantitating vegetation, impervious surfaces, and soil influences on *Aedes* mosquito presence in an urban expanding agro-irrigated, African ricleand agro-ecosystem due to anthropogenic pressure. The land cover mosquito abundance relationships was quantititated in ArcGIS to produce precise, eco-epidemiological, sub-meter resolution, geoclassified,orthogonal, grid-stratfiable, LULC maps of seasonal presence that highlight high risk areas (e.g., eco-georeferenecable geolocations of hyperproductive, oviposition, eco-epidemiological, LULC sites) which may can be useful for focusing mosquito control program actions.

Inventory of georeferenced, sparsely canopied, forestland corridor, agro-irrigated, inhomogeneous, riceland, seasonal, aquatic, habitats of sylvatic, *Ae aegypti*, regarding their larval productivity may be also regressed within autoregressive forecasting paradigm covariance matrices employing elucidative, eco-georeferenceable, time series, parameterizable, geoclassifiable, uncoalesced, iteratively interpolative, uncoalesced, oviposition, endmember, LULC signature, uncoalesced, covariate coefficient values in ArcGIS. These probabilistically regressed explanative time series, eco-epidemiological, endmember residuals can provide critical information for optimally characterizing species-specific oviposition, LULCs sites on geoclassified sub-meter resolution, satellite imagery. In so doing, selection and planning of integrated mosquito managements may be optimally targeted. Larval control will probably have little impacts on incidence, if interventions are untargeted and levels of coverage limited Gu and Novak 2005).

The strategy of sub-meter resolution, LULC-oriented, capture point, habitat-based, ArcGIS based oviposition, geostatistically targeted, eco-epidemiolgical interventions for an expanding agro-irrigated, African, riceland, agro-village complex due to anthropogenic and economic pressure may be to recognize the importance of variation employing time series, regression models for quantitating immature, mosquito production amongst peripheral, transitioned, geoclassifiable, eco-geographically discontinuous, sparsely inhomgeneous, forest-canopy to rice agriculture, LULCs. The intimate relationship between the schedule of rice husbandry and mosquito breeding has long been recognized in several countries(Surtees1970, Grainger 1947, Russell and Rao 1940). Sen (1948) attributed the immature habitat patterns of *Anopheles* spp. to rice cultivation practices which Jacob et al. 2005 remotely quantitated at the habitat level employing QuickBird visible and NIR 061 meter (m) data. In Japan, Makiya (1967) related breeding densities of *Culex tritaeniorhynchtts* Giles to increase in rice height, and Reuben(1971)observed a similar pattern in the breeding of culicines in rice fields of Madras, India. In Californian rice fields, Womeldorf and Whitesell (1922) demonstrated the relationship between *Culex tarsalis* Coq. and *Anopheles freeborni* Aitken and water depth and height of the rice

In more recent studies on mosquitoes of the the Mwea rice scheme (Kenya), Muturi et al. (2007) indicated that in the rice field environments, numbers of the main malaria vector in the

region *Anopheline arabiensis s.s.* were higher than in agro-irrigated geoclassified LULC zones specifically in post-tillering habitats. Mwangangi et al. (2007) assessed the impact of rice husbandry on mosquito, immature productivity and identified indigenous biocontrol agents with potential for controlling malaria and filaraisis mosquito habitats and in the scheme. The study established a close relationship between the schedule of the farming practices (particularly the the post-tillering phase) and seasonal, anopheline, immature hyerproductivity.

Exploring the relationship between mosquito, vector, relative abundance and uncoalesced, geo-spectrotemporal, seasonal, LULC characteristics of eco-georeferenceable, geo-spatiotemporal, *Ae. egypti* immature habitats in agro-irrigated, riceland and discontinuous, partially shaded, inhomogeneous, forest-canopied, sub-meter resolution, geoclassified, LULC changes sites employing specific cyberenvironments (e.g., ENVI C++) may allow optimizing predictive YF maps for remote targeting sub-meter resolution imaged, hyperendemic unknown, un-geosampled, hyperproductive, targeted LULC geolocations employing stochastic or deterministic, iteratively, quantitative interpolators (e.g, co-kriging, invese distance matrix) which may in turn facilitate implementing control intervention. Ultimately, mosquito and host distribution and diversity can affect vector behaviour and vector-borne disease endemic transmission risk on, seasonally agro-irrigated, satellite detected, LULC change sites(Jacob et al. 2007, Griffith 2005, Hay 2000). Understanding vector community dynamics in the face of anthropogenic shifts in irrigated African, riceland environments could form the basis for understanding the emergence and persistence of YF in various ecosystems.

Unfortunately, one of the major barriers to developing an accurate YFV surveillance program in an agro-irrigated African, expanding, riceland environment into forest-shaded, canopy LULC may be the design and implementation of a orthogonal or non-orthogonal grid-stratified, non-noisy, time series, ArcGIS, weighted spatial matrix for optimally regressively monitoring, *Aedes aegypti* eco-geographically geosampled quantitatvely eigen-decomposable, optimally parameterizable, explanatorial, geo-spectrotemporally uncoalesced, sub-meter resolution, geoclassifiable, endmember, LULC, covariates. For regressively forecasting seasonal transitions in discontinuous, deforested, canopied, sub-meter resolution, imaged landscapes caused by expanding agro-irrigated, African, agro-village riceland anthropogenic pratices, grid cell information may be pertinent. Studies on mosquito communities have mainly focused on either medically important genera such as *Aedes* spp. or specific habitats such as rice fields swamp forests and rural villages.

Overlaying an ArcGIS grid on remotely sensed, sub-meter, resolution data can help organize and characterize mosquito larval habitats (Jacob et. al. 2006). A grid is constructed by applying a mathematical algorithm in order to fit a continuous and bounded surface consisting of equidistant estimates of a quantity from a field geosampled LULC attribute. ArcGIS grid-based data files consist of columns and rows ofuniform cells coded according to data values(www.esri.com). Each grid cell within a weighted matrix contains an attribute value as well aslocation coordinates (Griffith 2003). The spatial location of each cell is hence implicitly contained within the ordering of the matrix in a YF regression-related, *Ae.aegypti*, oviposition, LULC model. As such, riceland, African, aquatic habitat, capture points containing the same spatial attribute value are easily recognized. Regular grids or lattices are frequently used to study ecosystems, for observations, experiments and simulation predictive, vulnerability modeling

(www.esri.com).   A sub-meter resolution (e.g., panchromatic QuickBird at 0.61m spatial resolution) digitized, orthogonal, grid cell data was employed to determine abundance riceland *An. arabiensis s.s.* larvae in the paddy and canal habitats in the Mwea riceland agro-ecosyetem study site in Jacob et al. 2006.

Unfortunately, a thorough literature search does not reveal any studies that have investigated the diversity of *Aedes* mosquito communities remotely optimally employing the relative abundances of vectors, and their vertebrate communities across productive or non-productive, ecogeoreferenceable oviposition, sub-meter resolution, geoclassifiable LULC sites for cartographically or regressively quantitating seasonal clustering tendencies in newly, deforested, sparsely discontinuously canopied, immature, *Ae egypti*, capture points in expanding, irrigated, agro-cosystem rice –village into discontinuous forest canopied LULC in Africa. Thus, forecasting, georeferenceable, geo-spectrotemporal, eco-epidemiological, capture point, seasonally prolific, oviposition sites of aquatic, immature habitats associated with *Ae aegypti* in literature is currently non-existent

Four factors that make the implementation of a YF surveillance programs a logistical challenge in expanding, African, irrigated, riceland agro-ecosystems due to anthropogenic pressures may be 1) the magnitude of the *Aedes* habitats [i.e., variable sizes of habitat, eco-epidemiological capture point, and sentinetinel sites within a single ecosystem (e.g., pre-flooded,rice agriculture, sparsely shaded forest canopied); 2) the low density of mosquito larvae and pupae in specific geolocations (e.g., posted-harvest. peripheral riceland deforested, immature habitats)) 3) the short duration of the mosquito vectors' aquatic life cycle; and, 4) limited georeferenced data on seasonal, immature, habitat productivity. This underscores the need to develop new strategies for estimating both *Aedes* larval and pupal density distributions and adult vector productivity from their immature, aquatic habitats as well as cooperative geoclassified, sub-meter resolution, LULC imaged areas and georeferenceable, communal, prolific, oviposition, habitat sites.

Developing the basic elements for a surveillance system for detection of endemic and exotic zoonotic pathogens carried and transported by *Ae. egypti* mosquitoes on newly transitioned, sub-meter resolution discontinuously, deforested, sparsely or dense canopied, employing sub-meter, satellite, remote sensing information, eco-epidemiological, georeferenceable, field-geosampled, empirical data and linear/ non-linear, geo-statistical algorithms in ArcGIS, ENVI and C++. may identify and regressively quantitate, probabilistic regressors that regulate the abundance and distribution of *Aedes* mosquito habitats in expanding, irrigated, African, riceland areas. In so doing, the geo-spectrotemporal, geo-spatiotemporal, explanative, seasonal, ecogeoreferenceable distribution of immature and adult, *Aedes* mosquitoes and their association with sub-meter resolution, geoclassifiable, ovipositions, aggregation areas within the LULC may be employed to elucidate patterns and specific seasonlly hyperoductive foci. Targeted environmental management interventions in these anthropogenically influenced, ArcGIS geoclassifiable newer, irrigated, deforested, riceland aquatic, immature habitats may be thus based on a sound understanding of the heterogeeity in mosquito productivity.

However, deficiencies in field methodology for predictively measuring productivity hamper our progress in understanding of immature Aedes mosquito productivity on newly transitioned , sub-meter resolution, eco-epidemiological, georeferenceable, geo-spectrotemporal, geospatial, capture, point imature habitats. To address these issues, an experimenter may develop a framework of habitat-based interventions initially in ArcGIS and the data may then be exported into other statistical geodatabase statistical cyberenvironments (e.g., PROC REG) employing a landscape logistical cartographic approach to elucidate mechanisms underlying immature, seasonal, mosquito productivity on neowly trasnistioned deforested, irrigated African riceland agro-ecosystem, hypeproductive , seasonal, proilic, habitats of sylvatic, immature *Ae egypti*. The importance of vigorously quantitative estimation of the immature *Ae egypti* productivity geolocations may be seasonally highlighted . Spatial models may be proposed to examine the interrelationship between mosquito productivity and oviposition of gravid mosquitoes in newly transitioned, forest, discontinuously canopied, sub-meter resolution, LULC to rice agriculture LULC. Environmental management approaches must take into account variability in productivity, in efforts to improve feasibility, cost-effectiveness, and sustainability of such approaches, particularly when implemented along with other malaria control measures (Gu et al. 2006).

Species sanitation is defined as environmental management of the main vector species by targeting the preferred habitats based on an understanding of the characteristic breeding habitats (Takken et al. 1991). For example, selective elimination of *Anopheles umbrosus*, the main vector in Malaysia, was achieved by targeting the preferred shaded habitats in wooded areas. Malaria control was obtained without having to eliminate all larval habitats. This successful strategy was abandoned once dichlorodiphenyltrichloroethane (DDT) and other powerful insecticides were discovered and became the backbone of the global malaria eradication era during the 1950s and 1960s(*Bradley 1994.)* Substantial variabilities in productivity of *An. gambiae* should be explored to form the basis of habitat-based intervention programs (Gu and Noval 2006).

For implementing newer remote strategies for catographically regressively , elucidatively quantitating hyperproductive seasonal, sylvatic,*Ae egypti* , ovipoistion sites on a newly transitioned, sub-meter resolution, georeferenceable, rice-agriculture, geoclassifiable LULC site and , several critical issues need to be addressed. First of all, immature deforestes riceland oviposition sites should be evaluated on the basis of quantitative measures of mosquito productivity. Nevertheless, this task is complicated because the notion of *Ae .aegypti* productivity has been conceived differently among experimeneters. Evidently, "productivity" is nothing but the rate of adults emerging from individual habitats. Indices of the productivity currently employed include presence/absence or density/abundance of larvae or pupae. However, the accuracy of these indices is largely unknown for taregting newlky deforeasted, irrigated riceland agro-ecosyustem, seasaonl immature habitats. Second, understanding of mosquito productivity cannot be formulated without properly accounting for mosquitoes' foraging for oviposition. Currently habitat surveys in literature of sylvatic, *Ae. aegypti* focus on inherent physicochemical, explanative, time series variables of habitat sites, such as size, turbidity, vegetation coverage, etc. Because egg-laying is a spatial process depending upon the geolocation of the focal habitat relative to sources of gravid mosquitoes, habitats closer to human inhabitations may tend to receive more eggs of *Ae. aegypt* , and thus are more productive if

conspecific competition is negligible in newly ntrasnitioned irrigated riceland habitats. Therefore, elucidation of variability in mosquito productivity requires spatial accounts of seasoanl, oviposition processes which may be optimally deterimined in explicative, probabilistic, vulnerability geo-spectrotemporal, geospatial, diagnostic, clinical, field or remote YFV-related, regression-oriented, paradigms Thereafter, forecasted areas of seasonally hyperproductive seasaonl capture points based on iteratively interpolated, sub-meter resolution, georeferenceable, fractionalized, endmember biosignature may be extensively ground-truthed to determine the validity of the model outputs. On a long-term basis, a pilot study may seek to expand yellow fever surveillance systems throughout other urban ecosystems using interpolated, sub-meter resolution, biosignatures attained from irrigated, hyperproductive, ovipoistion habitats in iterative interpolative stochastic or deterministic geo-cyberenvironments. The proposed modeling methods and resulting information (e.g., targeted, hyperpoductive, georeferenceable, *Ae egypti,* eco-geographic, geolocations on a transitioned, discontinuous, sparsely or dense, shaded, forest-canopied, pre-flooded LULC to rice plot which may provide public health officials in Africa with the tools to accurately identify factors regulating outbreaks of yellow fever in expanding irrgated ecosystems due to anthropogenic pressures.

A web-based interface may be developed in SAS/GIS for public health officials in Africa which may include:1)Real-time, syndrome-based reporting tool, 2)Automated, immediate 'Alerts' to public health officials, 3)Health 'Alerts' to doctors, hospitals and schools, 4) A web-based tool for data entry and communication, 5) Geographic mapping of yellow fever outbreaks, 6) Environmental predictive maps determining the areas at risk based on locations of *Aedes* mosquito and their resvoirs (e.g., wild monkey populations). Oviposition, sub-meter resolution, geo-spectrotemporally imaged, geoclassified, ArcGIS-derived, LULC sites may be quantified based on land cover types, (e.g., dense forest canopied, elevated areas), specific areas (e.g.,waste tire dump sites) and meterological uncoalesced, iteratively interpolative, endmember, proxy covariates such as evaporation demand for determining length of time for standing water present to allow mosquitoes to complete life cycle which may be optimally deduced within a regression matrix in SAS.

By regressively qualitatively quantitating precise probability space of discretely, geo-spectrotemporal geosampled, *Ae egypti* hypeproductive, georeferenceable, immature habitats along peripheral irrigated, African, agro-ecosystems, the oviposition, habitat, seasonal immature productivity may be ideally measured from paramterizable, sub-meter resolution, covariate estimates of emerging adults from individual habitats in SAS/GIS delineated, immature, prolific, eco-epidemiological, capture point maps. In most situations, a practical approach may be to employ larval density as a dependent (response) variable in a regression equation (e.g., PROC LOGISTIC statement) as an alternative for obtaining robustly, parsimoniously, regressed probabilistic estimates of emerging sylvatic,*Ae. aegypti* adults from newly transitioned, discontinuous, forest-canopy to irrigated, riceland, agro-village discontinuous, geo-classifiable LULCs. However, experimental studies show that *An. gambiae* Giles display density-dependent regulations with delayed developmental rates of larvae and smaller body sizes of emerging adults when they reared in crowding conditions in artificial habitats (Novak et al. 2012) although this phenomenon may be uncommon in natural habitats of *Ae egypti*-related African irirgated ricefield, seasonal ovipoistion sites. It may be emphasized for

a particular irrigated African, georeferenceable agro-ecosystem, seasonal LULC that large habitats with low density of emerging adults may be more productive than small water bodies with high density (e.g., a post-havested capture point of low larval density probably ranking higher than a pre-flooded habitat with high density). In large habitats like rice paddies, elucidative distribution patterns of larvae are useful to accurately estimate productivity (Gu and Novak 2005, Mwangangi et al. 2008, Muturi et al. 2007, Mururi et al. 2007). For example, *Ae egypti* larvae may tend to aggregate along edges of deforested newly transitioned immature ricefield thus estimation of seasoanl productivity could be optimally quanatiated by regressivelly probabilsitically stratifying gridded, georeferenceable immature capture points in terms of larval density and obtaining corresponding estimates from each strata.

It should be noted that geo-spectrotemporally geosampled, ArcGIS and SAS-based, explanatively quantitative, georeferenceable, optimal predictions rendered from sub-meter resolution, LULC models have a combination of adopted assumptions and values of parameters. For example, assuming contacts between hosts and blood feeding mosquitoes in a robust, cartographic, regression sub-meter resolution, eco-epidemioloigical, diagnostic, clinical, field or remote, YFV model constructed for revealing seasaonlly hyperproductive immature *Ae egypti* habitats in newly transitioned, deforested, geoclassifiable LULCs along peripheral edges of irrigated African, riceland expanding communities may reveal uniformly distributed focal area ; whereas, AUTOREG probabilistic paradigm outputs (e.g., decomposed eigenfunction values) of blood feedings of sylvatic, *Ae egypti* mosquitoes in these landscapes may reveal prolific, aggregations of sparse or densely shaded, forest-canopied, iteratively interpolative, LULC seasonal geolocations in geospace Various assumptions of cartographic and regressively probabilistically spatially weighted values of some parameterizable, immature, decomposeable YFV-related covariate estimators (e.g., the daily mortality rates and the recovery rates), may substantially alter numerical predictions rendered from vulnerability, forecasting, eco-epidemiological, sub-meter resolution, explicative models for precisely, remotely targeting georferenceable, hyerproductive, *Ae egypti,* oviposition sites. Using noiseless, linear and non-linear, eco-epidemiological, non-Gaussian/ Gaussian distribution, probabilistic, regression models in ArcGIS and SAS may provide qualitative understandings of larval interventions from the habitat perspective, which should hold when productivity parameterized, endmember LULC values remotely quantizable in African rice fields.

Although targeted, remotely sensed, sub-meter resolution, larval interventions have a great potential in helping to reduce transmission intensity and incidence of yellow fever in georeferenceable, irrigated, African, riceland agro-agrosystems, it should not be proposed that a geodatabase statistical or logistic, cartographic cyberenvironment forecasted, specied larval control is a panacea for combating the disease in all settings of iririgated expanding African agro-ecosystems. Combined with other interventions in an integrated manner, larval interventions for yellow fever prevention may be successful in situations where major habitats are limited and manageabls in these ecosystems especially in peripheral arease where agro-village anthropogenic pressure may reval more contact with sylvatic *Ae egypti* by farmers and their family members. Emphasizing informed larval interventions guided by remotely sensed, habitat-based, cartographic and regression, eco-epidemioloigical, ArcGIS and SAS models may play an important role in managing entomological features of local yellow fever transmission in newly transitioned discontinuous, deforested, sparsely or dense canopied, geo-classifiable,

LULCs to rice agriculture, hyperproductive, sylvatic,*Ae egypti*, oviposition sites. Given the constraint of resources throughout Africa, targeted remotely sensed larval interventions have a great potential for combating yellow fever trasnmission especially in areas of low to intermediate transmission in irrigated expanding agroecosystems, due to anthropogenic pressures.

Logistically specified, predictive, immature, entomological population models constructed in other relational datbased (Python, C++) employing georeferenced, *Ae. egypti,* sub-meter resolution, eco-epidemiological, transitioned LULC, capture point, decomposeable data collections may also be a useful tool to provide qualitative and quantitative understandings of influences of larval interventions on yellow fever transmission. For these purposes, a predictive modeling framework may be developed for conceiving a quantity of the total productivity of *Ae. aegypti* ovipoistion sites on in a new peripheral African irrigated, deforeseted, rice landscapes, which, in turn, may be partitioned into its constituent parts in ArcGIS based on data geo-spectrotemrpoally geosampled from georeferenceable, individual, oviposition sites.

Three field-operational simulated scenarios of larval interventions were evaluated in relation to impacts on parasitological indicators of malaria transmission in Gu and Novak (2005). Their results revealed that it is unnecessary to manage all aquatic habitats to obtain significant reductions in incidence and prevalence of malaria in situations of low and intermediate levels of transmission. The authors highlight that informed larval interventions by identifying and targeting prolific habitats can play a critical role in combating malaria in Africa.

Jacob et al. (2015) employed the modeling tactics of Gu and Noval (2005) to construct a robust, onchocerciasis-related, geo-spectrotemporal, probability-oriented, explanatory, endemic transmission, sub-meter resolution, heterogenous spatial filter,orthogonalizable, eigenvector, entomological, oviposition, foci, hyperproductive,sub-meter resolution, LULC predictive, risk map in AUTOREG employing a prevalence responsible variable [i.e., aggregate counts over a riverine tributary, eco-geographical region subdivided by administrative boundaries (e.g., districts)] in northern Uganda. Onchocerciasis, or river blindness, is historically one of the most important causes of blindness worldwide. (Thylefors 1978). The parasite is transmitted by black flies (*Simulium damnosum s.l.*) that breed in fast running rivers and streams. In linear algebra, an eigenvector or characteristic vector of a linear transformation T from a vector space V over a field F into itself is a non-zero vector that does not change its direction when that linear transformation is applied to it. In other words, if v was a vector that is not the zero vector in a vulnerability, vector arthropod-related, explicative, probabilistic, eco-epidemiological, forecasting stochastic or deterministic interpolator model then it would be an eigenvector of a parameterizable, linearizable, seasonal, covariate explataive estimator, log transformation T if $T(v)$ is a scalar multiple of v in the residual forecasts.

In Jacob et al. (2015) this condition was written in AUTOREG as the mapping $T : \mathbf{v} \mapsto \lambda\mathbf{v}$,where λ was a scalar in the field F,(i.e., known as the eigenvalue or characteristic value associated with the, riverine larval habitat synthetic eigenvector v. A generalized autoregressive conditional heteroscedasticity (GARCH) model was employed to model the time

series with heteroscedastic errors in *the S. damnosum* s.l. forecasting model. The GARCH regression model with autoregressive errors was written in AUTOREG as. $y_t = \mathbf{x}_t'\boldsymbol{\beta} + v_t$

$$v_t = \varepsilon_t - \varphi_1 v_{t-1} - \ldots - \varphi_m v_{t-m} \quad \varepsilon_t = \sqrt{h_t}\, e_t \quad h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \gamma_j h_{t-j} \quad e_t \sim \mathrm{IN}(0,1)$$

This model combined the **m**th-order autoregressive error model with the GARCH(p,q) variance model in AUTOREG. It was denoted as the AR(m)-GARCH(p,q) regression model.

The tests for the presence of ARCH effects (namely, Q and LM tests, tests from Lee and King (1993) and tests from Wong and Li (1995)) helped determine the order of the ARCH model appropriate for the data. For example, the Lagrange multiplier (LM) tests shown to be significant $(p < 0.0001)$ through order 12, which indicates that a very high-order ARCH model wasneeded to model the heteroscedasticity in the resiudl forecasts of the geo-spectrotemporal, endmember, forecasting *S. damnosum* s.l. immature habitat, vulnerability , forecasting model

The basic ARCH(q) model $(p=0)$ was a short memory process in that only the most recent **q** squared, geosampled *S. damnosum* s.l. habitat, endmember LULC residuals awere used to estimate the changing variance. The GARCH model $(p>0)$ allowed long memory processes, which usd all the past squared residuals to estimate the variance in the geo-spectrotemrpoal, vector arthropod-related probahbistic paradidgm. . The LM tests suggested the use of the GARCH model $(p>0)$ instead of the ARCH model for remotely taregteing hypeproductive riverine tributary oviposition sites along a coordor of the Achwa r basin> .

The GARCH(p,q) model was specified with the GARCH=(P=p, Q=q) option in the MODEL statement. The basic ARCH(q) model was the same as the GARCH$(0,q)$ model and was specified with the GARCH=(Q=q) option. The following statements fit an AR(2)-GARCH$(1,1)$ model for the Y series that is regressed on TIME. The GARCH=(P=1,Q=1) option specifies the GARCH$(1,1)$ conditional variance, *S. damnosum* s.l., endmember, sub-mter resolution, model. The NLAG=2 option specifies the AR(2) error process. Only the maximum likelihood method is supported for GARCH models; therefore, the METHOD= option is not needed (www.sas.edu). The CEV= option in the OUTPUT statement stores the estimated conditional error variance at each geo-spectrotemporal, time period in the variable VHAT in an output eco-epidemioloigical, data set named OUT. The data set is the same as in the section Testing for Heteroscedasticity (www.sas.edu).

Initially, the GENMOD procedure fit a generalized linear model(GLM) to the geo-spectrotemporally geosampled, time series, ovipoition, sub-mter resolution, *S. damnosum* s.l. geo-spectrotemporally geosampled, riverine tributary, endmember, entomological, sub-meter resolution, geoclassified LULC data by maximum likilhood estimation of the parameter vector β using miultiple bio-geophysical, eco-epidemioliogical, time series, clinical, field and remote diganostic data geosampled in an agro-village, riverine tributary in northern Togo. The GENMOD procedure fit multiple *S. damnsoum* s.l.- generalized linear, immature habitat capture point, eco-epidemiolgical, forecasting , vulnerability models, similar in covariance-matrix of coefficients similar as those designed by Nelder and Wedderburn (1972).) The linear models

allowed the mean of the geo-spectrotemporally , geospatially geosampled, immature *Similium* population to depend on an explanative, diagnosizable, linearizable, optimal, probabilistic, LULC predictor(e.g., percentage of partailly shaded trailing vegtation), through a nonlinear link function which allowed the response probability distribution to be any member of an exponential family of distributions. To use the method of maximum likelihood, in GEN MOD the authors of Jacob et al. (2015) specified the joint density function for all the georeferenceable, *S. damnsoum* s.l. immature ,capture point, trailinmg vegatation immature habitat Precambrian rock observations.

In the study of probability, given at least two random variables vector arthropod-related *X, Y, ...*, that are defined on a probability space, the joint probability distribution for *X, Y, ...* is a probability distribution that would render the probability that each of *X, Y, ...* falls in any particular range or discrete set of explanative values specified for that variable. In the case of only two random variables, this is called a bivariate distribution, but the concept generalizes to any number of entomological, randomizable explanative variables for rendering a multivariate distribution. The joint probability distribution can be expressed either in terms of a joint cumulative distribution function or in terms of a joint probability density function (in the case of continuous variables) or joint probability mass function in the case of discrete variables (Hosmer and Lemeshew 2002). These in turn can be employed to find two other types of distributions: the marginal distribution giving the probabilities for any one of the variables with no reference to any specific ranges of values for the other variables, and the conditional probability distribution giving the probabilities for any subset of the variables conditional on particular values of the remaining variables

For an independent and identically distributed YF, *Ae .aegypti*, ovispoition, geo-spectrotermporal or geo-spectrotemrporal , eco-epidemiological, LULC sample, this joint density function may be optimally quanatitated for rendering geolocations of seasonal, hypeproductive, *Ae. aegypti* , discontinuous , forest-canopied, or rice-agro-village, grid-stratified, capture points in an African expanding, agro-irrigation, eco-epidemiological study site. This function came from considering the observed values $x_1, x_2, \ldots, x_n$ to be fixed "parameters" of this function, whereas θ was the function's variable which the authors allowed to vary freely using the

$$\mathcal{L}(\theta\,;\,x_1,\ldots,x_n) = f(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta).$$

likelihood Note that " ;" denoted a separation between the two input arguments: θ and the geosampled,immature observations $x_1, \ldots, x_n$. In practice it is often more convenient to work with the logarithm of the likelihood function in optimizable, regression forecast models, called the log-likelihood:

$$\ln \mathcal{L}(\theta\,;\,x_1,\ldots,x_n) = \sum_{i=1}^{n} \ln f(x_i \mid \theta),$$

or the average log-likelihood: $\hat{\ell} = \frac{1}{n} \ln \mathcal{L}.$ (Feller 1968).

In Jacob et al. (2015) the GENMOD procedure estimated the seasonal-geosampled, explicative, eigen-decomposeable, spatial filter, orthogonal, parameterizable, *Similium* immature, habitat covariates numerically through an iterative fitting process. The dispersion parameter was robustly estimated by the residual deviance and by Pearson's chi-square divided by the degress of freedom (df). Pearson's chi-squared test (χ^2) is a statistical test applied to sets of categorical

data to evaluate how likely it is that any observed difference between the sets arise by chance (Hosmer and Lemeshew 2002) which may be regressively applicable for unpaired eco-epidemiological, yellow fever, time series, data anlyses from large samples. Covariances, standard errors, and p-values were then computed for the geosampled $S.$ $damnosum$ s.l., capture point, immature, habitat uncoalesced, endmember, remotely sensed covariate coefficients based on the asymptotic normality derived from the maximum likelihood estimation. A sequence of parametric statistical models $\{P_{n,\theta}: \theta \in \Theta\}$ is said to be locally asymptotically normal at θ if there exist matrices r_n and I_θ and a random vector $\Delta_{n,\theta} \sim N(0, I_\theta)$ such that, for every converging sequence $h_n \rightarrow h$, (Ibragimov, and Has'minskiĭ, 1981).

The standard asymptotics told that the maximum-likelihood estimator in GEN MOD that $\sqrt{n}$-was consistent in the eco-epidemiological dataset of uncoalesced, iteratively interpolative, ,sub-meter resolution, $S.$ $damnosum$ s.l.,immature, eco-epidemiological,, eub-mter resolution, habitat model and asymptotically efficient, meaning that it reached the Cramér–Rao bound: $\sqrt{n}(\hat{\theta}_{mle} - \theta_0) \xrightarrow{d} \mathcal{N}(0, I^{-1})$ where I was the Fisher information matrix: $I_{jk} = \mathrm{E}_X\left[-\frac{\partial^2 \ln f_{\theta_0}(X_t)}{\partial\theta_j\,\partial\theta_k}\right]$. The Fisher information matrix is a $d \times d$ matrix with element $I_{m,k}$ defined $I_{m,k} = \mathrm{E}\left[\frac{\partial}{\partial\theta_m}\log f(x;\theta)\frac{\partial}{\partial\theta_k}\log f(x;\theta)\right] = -\mathrm{E}\left[\frac{\partial^2}{\partial\theta_m\partial\theta_k}\log f(x;\theta)\right]$ (Rao, and Radakrishna 1994).The authors then let $T(X)$ be an estimator of any vector function of the geosampled, $S.$ $damnsoum$ s.l. habitat regressors $T(X) = (T_1(X),\ldots,T_d(X))^T$, and denoted its expectation vector $\mathrm{E}\,T(X)$ by $\psi(\theta)$. The Cramér–Rao bound in GEN MOD then stated that the covariance matrix of $T(X)$ satisfied $\mathrm{cov}_\theta(T(X)) \geq \frac{\partial\psi(\theta)}{\partial\theta}[I(\theta)]^{-1}\left(\frac{\partial\psi(\theta)}{\partial\theta}\right)^T$ where t he matrix inequality $A \geq B$ was intrepreted to mean that the matrix $A - B$ was positive semidefinite in the paradigm. In linear algebra, a symmetric $n \times n$ real matrix M is said to be positive definite if the scalar $z^T M z$ is positive for every non-zero column vector z of n real numbers(Hosmer and Lemshew 2002). In the eco-epidemiological $S.$ $damnsoum$ s.l. habitat forecast, vulnerability, model, z^T denoted the transpose of z.[Horn 1990]. Further, $\partial\psi(\theta)/\partial\theta$ was the Jacobian matrix whose ij element I was given by $\partial\psi_i(\theta)/\partial\theta_j$.In vector calculus, the Jacobian matrix is the matrix of all first-order partial derivatives of a vector-valued function (Cressie 1993). If $T(X)$ was an unbiased, geo-spectrotemporal, geosampled, georeferenceable, $S.$ $damnosum$ s.l. immature habitat, eco-epidemiological, explanative, time series, capture point, estimator of θ(i.e., $\psi(\theta) = \theta$), then the Cramér–Rao bound (CRB) reduced any time series, parameterizable predictor (e.g., percentage of shaded trailing vegetion) associated to any hypeproductive, riverine tributary, agro-village, oviposition site. In estimation theory and statistics, the CRB expresses a lower bound on the variance of estimators of a deterministic parameter.In, the targeted, seasonal, $S.$ $damnsoum$ s.l., eco-epidemiological, residual forecasts the bias of the maximum-likelihood estimator was equal to zero up to the order $n^{-1/2}$.

Interestingly, $T = t(X)$ may be an explanative unbiased, georeferenceable, geo-spectrotemporal, immature. habitat, paramterizable covariate estimator for the value $\psi(\theta)$ based on the geosampled, .immature, capture point, eco-epidemiological ,geo-spectrotemporal, explicative time series, habitat, observations X, and so $\mathrm{E}(T) = \psi(\theta)$. The goal was to prove that,

for all θ, $\text{var}(t(X)) \geq \frac{[\psi'(\theta)]^2}{I(\theta)}$. Letting X be a geo-spectrotemporally geosampled, randomized *Ae egypti,* irrigated, riceland oviposition, explanatorial, predictor variable with probability density function $f(x;\theta)$, then $T = t(X)$ may be a statistic which is employable as an estimator in a yellow fever LULC, sub-meter resolution, forecasting, vulnerability model for $\psi(\theta)$. In so doing, V may be defined as $V = \frac{\partial}{\partial\theta}\ln f(X;\theta) = \frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)$ where the chain rule may be used in the final equality tabulations. In calculus, the chain rule is a formula for computing the derivative of the composition of two or more functions. That is, if *f* and *g* are functions, in a forecasting, yellow fever, vulnerability, LULC agro-ecosystem model; then the chain rule expressed by the derivative of their composition *f* ∘ *g* (the function which maps *x* to *f*(*g*(*x*)) in terms of the derivatives of *f* and *g* and the product of functions maybe be optimally defined as follows: $(f \circ g)' = (f' \circ g) \cdot g' \cdot$ (see Hazewinkle 2001). In so doing, the expectation of V, written $\mathrm{E}(V)$, in GEN MOD was zero in the explanative, model georeferenceable, time series, lineraized, explicative, diagnostic, eco-epidemiological,residual forecasts. This is because: $\mathrm{E}(V) = \int_x f(x;\theta)\left[\frac{1}{f(x;\theta)}\frac{\partial}{\partial\theta}f(x;\theta)\right]dx = \frac{\partial}{\partial\theta}\int_x f(x;\theta)dx = 0$ where the integral and partial derivative are interchanged in GEN MOD which is justifiable by the second regularity condition .In order for a minimum point x^* to satisfy the above regularity conditions, the problem should satisfy some regularity conditions linearity constraint qualification (Krantz, 2002),

Interestingly, g_i and h_j are affine functions which may be useable in a forecasting, vulnerability, eco-epidemiological, YFsub-meter resolution, probabilistic LULC, paradigm, then no other condition may be needed in order to obtain robust explanative time series estimator of seasaonlly hyeproductive, immature capture points on newly deforested, agro-ecosystem, sub-meter resolution, geoclassifiable LULCs In so doing, linear independence constraint qualification (LICQ) may quanatites the gradients of the active inequality constraints and the gradients of the equality constraints in the entomological, predictive, risk model which may be linearly independent at x^*. A Mangasarian–Fromovitz constraint qualification (MFCQ) where the gradients of the active inequality constraints and the gradients of the equality constraints occur may be positive-linearly independent at x^* in the paradigma. The MFCQ is satisfied at $\bar{x} \in C$ if(i) the system of vectors { $\nabla fi(x)$, $i \in I1$} has constant rank in a neighbourhoodof $\bar{x}$;(ii) there exists a $z \in \mathbb{R}n$ such that $\langle\nabla fi(\bar{x}), z\rangle = 0$, $i \in I1$, $\langle\nabla fi(\bar{x}), z\rangle < 0$, $i \in I2(\bar{x})$ van der Vaart, Aad W. (1998).

A Constant rank constraint qualification (CRCQ) may be cartographically illustratable in an ArcGIS geodatabase cyberenvironment for each subset of optimizable LULC gradients in an eco-epidemioloigical forecast, vulnerability, geo-spectrotemporal, yellow fever -related, eco-epidemiological, time series, uncoalesced, LULC, sub-meter resolution paradigm of the active inequality constraints where the gradients of the equality constraints the rank at a vicinity of x^* is constant. A constant, positive, linear dependence, constraint qualification for each subset of ArcGIS delineated gradients of the active inequality constraints and the gradients of the equality constraints, would be is positive-linear dependent at a vicinity of x^* in the targeted oviposition sites on LULC, time series, change sites . Quasi-normality constraint qualification (QNCQ) may

be applicable for regressively quanatiting   the gradients of the active inequality constraints and the gradients of the equality constraints for determination of positive-linearly quantitable latent dependeny at x^* with associated multipliers λ_i for equalities and μ_j for inequalities in the resdiually forecasted estimates. In so doing any non-sequenceable features (e.g., $x_k \to x^*$) such that $\lambda_i \neq 0 \Rightarrow \lambda_i h_i(x_k) > 0$ and $\mu_j \neq 0 \Rightarrow \mu_j g_j(x_k) > 0$ may be herusitically optimized. Slater conditions may be evauted in the model for identifying any   convex problems, for example, where non-normalities exist at a   point x such that $h(x) = 0$ and $g_i(x) < 0$. $((v_1, \cdots, v_n)$ is positive-linear dependent. This may be pertinent if there exists $a_1 \geq 0, \ldots, a_n \geq 0$ such that $a_1 v_1 + \cdots + a_n v_n = 0$. (see Nocedal and Wright, (2006). It may be shown that LICQ⇒MFCQ⇒CPLD⇒QNCQ, LICQ⇒CRCQ⇒CPLD⇒QNCQ (and the converses are not true) in a geo-spectrotemrpoally , geosampled, vector arthrpod-related , sub-meter resolution , eco-epidemiological, forecast, vulnerability , probabilistic paradigm although MFCQ may not be equivalent to CRCQ (see Boyd, and Vandenberghe. 2004). In practice weaker constraint qualifications are preferred in seasonal, robustifiable, vector arthropod-related, iterative endmember interpolators since they provide stronger optimality conditions (see Jacob et al. 2005 and Griffth 2005).

The authors in Jacob et al.2013 then considered the covariance $\text{cov}(V,T)$ of V and T. in a forecasting vulnerability model very similar to Jacob et al. 2015 for determing district level malarial explanatorial georferenceable, time series, paramterizable, explicative covariates associated to sub-meter resolution geoclassified LULCs and other covariates associated to prevalance by districts in Uganda. Traditionally, univariate statistics and regression models have been generated from the satellite data to determine covariates (e.g., rainfall) related to monthly malarial prevalence rates (Hay 1997, Washino and Wood 1994). Specific district-level prevalence measures however, can be forecasted using autoregressive specifications and spatiotemporal data collections for targeting districts that have higher prevalence rates(Jacob et al. 2007). In Jacob et al. (2013)initially, case, as counts, were used as a response variable in a Poisson probability model framework for quantifying datasets of district-level covariates (i.e., meteorological data, densities and distribution of health centers, etc.) sampled from 2006 to 2010 in Uganda. Results from both a Poisson and a negative binomial (i.e., a Poisson random variable with a gamma distrusted mean) revealed that the covariates rendered from the model were significant, but furnished virtually no predictive power. Inclusion of indicator variables denoting the time sequence and the district location spatial structure was then articulated with Thiessen polygons which also failed to reveal meaningful covariates. Thereafter, an Autoregressive Integrated Moving Average (ARIMA) model was constructed which revealed a conspicuous but not very prominent first-order temporal autoregressive structure in the individual district-level time-series dependent data. A random effects term was then specified using monthly time-series dependent data. This specification included a district-specific intercept term that was a random deviation from the overall intercept term which was based on a draw from a normal frequency distribution. The random effects specification revealed a non-constant mean across the districts. This random intercept represented the combined effect of all omitted paramterizable time series, district-level, sub-meter resolution (i.e. panchromatic QuickBird) uncoalesced, iteratively quantitaively, interpolable, LULC covariates that caused districts to be more prone to the malaria prevalence than other districts. Additionally, inclusion of a random intercept assumed random heterogeneity in the districts' propensity or, underlying risk of malaria prevalence which

persisted throughout the entire duration of the time sequence under study. This random effects term displayed no spatial autocorrelation, and failed to closely conform to a bell-shaped curve. The model's variance, however, implied a substantial variability in the prevalence of malaria across districts. The estimated model contained considerable overdispersion (i.e., excess Poisson variability): quasi-likelihood scale = 76.565. The following equation was then employed to forecast the expected value of the prevalence of malaria at the district-level: prevalence = exp[-3.1876 + (random effect)i] .

In Jacob et al. (2013) the district-level, predictive, malarial, risk model rendered $\mathrm{cov}(V,T) = \mathrm{E}(VT)$, because $\mathrm{E}(V) = 0$. Expanding this expression in GEN MOD led to

$$\mathrm{cov}(V,T) = \mathrm{E}\left(T \cdot \left[\frac{1}{f(X;\theta)}\frac{\partial}{\partial\theta}f(X;\theta)\right]\right) = \int_x t(x)\left[\frac{1}{f(x;\theta)}\frac{\partial}{\partial\theta}f(x;\theta)\right]dx = \frac{\partial}{\partial\theta}\left[\int_x t(x)\,dx\right] = \psi'(\theta)$$

again because the integration and differentiation operations commute (i.e., second condition). The Cauchy–Schwarz inequality was written in vector form as $\|\mathbf{a}\cdot\mathbf{b}\| \leq \|\mathbf{a}\|\,\|\mathbf{b}\|$. In two-dimensions, which became

$(a^2 + b^2)(c^2 + d^2) \geq (ac + bd)^2$ which was proven by solving $\sum_{i=1}^{n}(a_i x + b_i)^2 = \sum_{i=1}^{n}a_i^2\left(x + \frac{b_i}{a_i}\right)^2 = 0.$ If b_i/a_i is a constant c, then $x = -c$. If it is not a constant, then all terms cannot simultaneously vanish for real x, so the solution is complex and can be found using the quadratic equation

$x = \frac{-2\sum a_i b_i \pm \sqrt{4(\sum a_i b_i)^2 - 4\sum a_i^2 \sum b_i^2}}{2\sum a_i^2}$. (homer and Lmeshew 2002) The model showed

that $\sqrt{\mathrm{var}(T)\mathrm{var}(V)} \geq |\mathrm{cov}(V,T)| = |\psi'(\theta)|$ therefore $\mathrm{var}(T) \geq \frac{[\psi'(\theta)]^2}{\mathrm{var}(V)} = \frac{[\psi'(\theta)]^2}{I(\theta)}$ The Cauchy–Schwarz inequality states that for all vectors u and v of an inner product space it is true that $|\langle u, v\rangle|^2 \leq \langle u, u\rangle \cdot \langle v, v\rangle$ where $\langle\cdot,\cdot\rangle$ is the inner product. (Gerard 1972). In linear algebra, an inner product space is a vector space with an additional structure called an inner product *Axler*, (Sheldon 1997).

Recall that the dimension of an inner product space in a regression forecasting model is the cardinality of a maximal orthonormal system.which may be based on Zorn's lemma since it contains at least one, one maximal element.and any two will have the same cardinality. In mathematics, the cardinality of a set is a measure of the "number of elements of the set". For example, the set A = {2, 4, 6} contains 3 elements, and therefore A has a cardinality of 3. There are two approaches to cardinality – one which compares sets directly using bijections and injections, and another which uses cardinal numbers. The cardinality of a set is also called its size, when no confusion with other notions of size is possible.The cardinality of a set A is usually denoted $|A|$, with a vertical bar on each side; this is the same notation as absolute value and the meaning depends on context. Alternatively, the cardinality of a set A may be denoted by n(A), $\overline{\overline{A}}$, card(A), or #A.

Zorn's lemma was used to show that every nontrivial ring R with unity contains a maximal ideal in a malarial district-level, geo-spectrotemporal, geospatial, eco-epidemiological, forecasting, vulnerability model for targeting paramterizable covariates associated to hyper/hypo prevalance, geoclassified district level LULC areas. In the terminology the set P consisted of all (two-sided) ideals in R except R itself, which was not empty since it contained at least the trivial

ideal {0}. This decomposition in an spatial filter, eigevector dataset of sub-meter resolution, eco-georeferenceable, cartographic, dataset of heuristically optimizable variables for a decomposed probabilistic sub-meter resolution, discontinuous forest canopy to rice agriculture LULC may be partially ordered by set inclusion. Finding a maximal ideal in a geo-spectrotemporal, vector, arthropod-related, forecast vulnerabilty model may be the same as finding a maximal element in P. The ideal R may be excluded in the LULC model because maximal ideals by definition are not equal to R.

To apply Zorn's lemma to a a non-empty totally ordered subset T of P catalytically in a geo-spectrotemrpoal, geosmapled, vector arthropos-related, eco-epidmeioloigical, vulnerability, forecasting, probabilistic paradigm It may be necessary to show that T has an upper bound in the model covariance that is, there exists an ideal $I \subseteq R$ which is bigger than all members of T in the LULC model but still smaller than R otherwise it would not be in P. (see *Krantz, Steven G.(2002)*, If an expeimnter takes I to be the union of all the ideals in T in the vector artrhrpos-related model T contains at least one element, and that element would contain at leasta 0, thus the union I would contain at least 0 and would not be empty (*Zorn, 1935)* To prove that I is an ideal, note in a georeferenceable, sub-meter resolution, discontinuous, forest canopied, to an rice irrigated agriculture dataset of uncoalesced, geo-spectrotemporal, LULCs a and b maut be elements of I, in the eco-epidemiological, model then there would exist two ideals $J, K \in T$ such that a is an element of J and b is an element of K. Since T may be totally ordered, $J \subseteq K$ or $K \subseteq J$. In the first case, both a and b are members of the ideal K in the paradigm therefore their sum $a + b$ would be a member of K, which may reveal that $a + b$ is member of I in the LULC entomological, time series model. In the second case, both a and b would be members of the ideal J, and thus $a + b \in I$. Further, if $r \in R$, then ar and ra would be elements of J and hence elements of I. Thus, I would an ideal in R in a robust forecasting, vulnerability, endmeic, transmission-oriented, explicative, georeferenceable, Se egypti trageting, sub-meter resolution model on newly trasnitioned, deforested canopied, LULCs to rice irrigated, agro-acosystem LULCs

Now, an ideal is equal to R if and only if it contains 1 (Hazewinkle 2001) . Thus it is clear that if an explanatorial, georeferenceable, residual,time series forecast rendered from a a sub-meter resolution LULC, explicative, bio-geophysical, model for regressively quantitating hypeproductive seasonl, *Ae egypti* immature, eco-epidemiological, capture points is equal to R, then it must contain 1; on the other hand, if it contains 1 and r is an arbitary element of R, then $r1 = r$ is an element of the ideal output and this ideal predictor would be equal to R. So, if I were equal to R, then it would contain 1, and that means one of the members of T (i.e., an orthogonally explicatively decomposed, sub-meter resolution, uncoalesced, iteratively intrepolative, geoclassified, ArcGIS, deforested forest-canopied, seasonal LULC to rice agriculture LULC would contain 1 which then would thus be equal to R – but R is explicitly excluded from P. Based on Zorn's lemma thus there would be a maximal element in P, in other words a maximal ideal in R.

Note, that the quantizable noise would depend on the fact that the ring R has a multiplicative unit 1 in the residiual eco-epidemiological forecasts (e.g., hypeproductive *Ae .egypti* oviposition sites on a newly deforested discontinuously canopied, georefernceable LULC to riceland plots along the periphery of an expanding irrigated African ecosystem due to

anthropogenic pressures . Without this, the model would be misspecified and wouldnot be able to remopte traget hypeproductive seasaonl habitats. For example, the ring with $\mathbb{Q}$ as additive with only proper usage of trivial multiplication (i. e. $ab = 0$ for all a, b) for remote deduction of hypeproductive, explicative, seasaonlly hyperprodouctive, sylvatic, *Ae egypti* , ovipoistion, seasonal LULC sites in an irrigated African, riceland agro-ecosystem has no maximal ideal (and of course no 1). Thus, ideal forecasting model outputs would be precisely based on the additive subgroups in the regression framework.

Suppose that *R* is a prime ring with the center *Z* and the extended centroid *C* in a georeferenceable, geo-spectrotemrporal , sub-meter resolution deforested canopied LULC to rice agriculture LULC for regressively quanatiting *Ae. egypti* oviposition sites in an expanding African irrigated ecosystem In abstract algebra, a nonzero ring *R* is a prime ring if for any two elements *a* and *b* of *R*, *arb = 0* for all *r* in *R* which may imply that either *a = 0* or *b = 0*. This definition may be remotely regarded as a simultaneous generalization of both integral domains and simple rings in diagnostic, newly transitioned, sub-meter resolution, sylvatic, *Ae egypti,* oviposition, geoclassifiable, LULC site. An integral domain in ArcGIS is a nonzero commutative ring in which the product of any two nonzero elements is nonzero (www.esri.com) Integral domains are generalizations of the ring of integers and provide a natural setting for studying divisibility (Hazewinkle 2001). In an integral domain the cancellation property may hold for multiplication by a nonzero element *a* in a vector, arthrpod-related, forecasting ,vulnerability, eco-speidemiological, probabisltic paradigm that is, if a ≠ 0, an equality *ab = ac* implies *b = c*. A simple ring is a non-zero ring that has no two-sided ideal besides the zero ideal and itself (Hazewinkle 2004).

The idempotents of *R* may have an important connection to decomposition of *R* modules in a remotely sesnitive deforested, ago-ecosystem, geoclassi3ed, LULC , sub-mter resolution paradigm that is constructed in ArcGIS for fporecasting hypeporductive, seasaonl Ae egeypti, oviposition sites in an expanding, African, agro- irrigation ecosyetm due to anthropogenic pressures. . If *M* is an *R* module and $E = \text{End}_R(M)$ is its ring of endomorphisms, then $A \oplus B = M$ if and only if there is a unique idempotent *e* in *E* such that $A = e(M)$ and $B = (1 − e)(M)$ in the trasnitioned deforested, non-continuous , geoclassifed sub-meter resol ution LULCs. Clearly then, *M* is directly indecomposable if and only if 0 and 1 are the only idempotents in *E* in the risk model for optimally targeting prolific immature, georeferenceable, geo-spectrotemporal, geospatial, eco-epidemioloigical, diagnostic, sylvatic, YFV-related, clinical, field or remote georefverenceabl, seasaonl model.

In the case when *M = R* the endomorphism ring $\text{End}_R(R) = R$, where each endomorphism arises as left multiplication by a fixed ring element. In mathematics, an endomorphism is a morphism (or homomorphism) from a mathematical object to itself. For example, an endomorphism of a vector space V is a linear map f: V → V, and an endomorphism of a group G is a group homomorphism f: G → G. In mathematics, a linear map (also called a linear mapping, linear transformation or, in some contexts, linear function) is a mapping $V \rightarrow W$ between two modules (including vector spaces) that preserves (in the sense defined below) the operations of addition and scalar multiplication.

Linear maps can often be represented as matrices, and simple examples include rotation and reflection linear transformations.An important special case is when $V = W$, in which case the map is called a linear operator, or an endomorphism of V. Sometimes the term *linear function* has the same meaning as *linear map*, while in analytic geometry it does not.A linear map always maps linear subspaces onto linear subspaces (possibly of a lower dimension); for instance it maps a plane through the origin to a plane, straight line or point.In the language of abstract algebra, a linear map is a module homomorphism. In the language of category theory it is a morphism in the category of modules over a given ring.

Let V and W be vector spaces over the same field K. in a sub-meter resolution, hypeproductive, *Ae. aegypti,* geo-spectrotemrpoal or geo-spatiotemrpoal, African, riceland agro-village complex, larval habitat, capture point, oviposiition, LULC model. Then a function $f : V \to W$ would be be a linear map if for any two vectors x and y in V and any scalar α in K, the following two conditions are satisfied: additivity and homogeneity of degree. This is equivalent to requiring the same for any linear combination of vectors, (i.e. that for any vectors $x_1, ..., x_m \in V$ and scalars $a_1, ..., a_m \in K$, the following equality holds:Denoting the zero elements of the vector spaces V and W by 0_V and 0_W respectively, in the sylvatic, YF eco-epidemiological, LULC model, it follows that $f(0_V) = 0_W$ as letting $\alpha = 0$ in the equation for homogeneity of degree 1,Further, V and W may be considered to be vector spaces over different fields in the entomological model output It may be necessary to specify which ground fields is being used in the definition of "linear" in the YF model. If V and W are considered as spaces over the field K in the model , the forecasted targets of seasonal hypeproductive foci would be rendered in as K-linear maps. For example, the conjugation of complex numbers is an R-linear YF forecast, vulnerability map C $\to$ C, would not be C-linear. A linear map from V to K (with K viewed as a vector space over itself) is called a linear functional (Cressie 1993).

These statements generalize to any left-module $_RM$ over a ring R without modification, and to any right-module upon reversing of the scalar multiplication With this modification of notation, $A \oplus B = R$ as right modules if and only if there exists a unique idempotent e such that $eR = A$ and $(1 - e)R = B$. Thus every module direct summand of R is generated by an idempotent.

If a is a central idempotent,in sub-meter resolution, hypeproductive, sylvatic, *Ae. aegypti,* geo-spectrotemrpoal or geo-spatiotemrpoal, African, riceland agro-village complex, larval habitat, capture point, oviposiition, LULC model. then the corner ring $aRa = Ra$ is a ring with multiplicative identity a. Just as idempotents determine the direct decompositions of R as a module, the central idempotents of R determine the decompositions of R as a direct sum of rings. If R is the direct sum of the rings $R_1,...,R_n$, then the identity elements of the rings R_i in a YF eco-3epidemiological, forecast, vulnerability, model central idempotents would be pairwise orthogonal, and their sum would be 1. Conversely, given central idempotents $a_1,...,a_n$ in R that are pairwise orthogonal and have sum 1, then R would be the direct sum of the rings $Ra_1,…,Ra_n$. So in particular, every central idempotent a in a YF endmember model with R would rise to a decomposition of R as a direct sum of the corner rings aRa and $(1 - a)R(1 - a)$. As a result, a ring R is directly indecomposable in a African ricland, discontinuous, forest-canopy, YF eco-epidemiological model as a ring if and only if the identity 1 is centrally primitive.

An additive subgroup A of R is said to be invariant under special automorphisms if $(1+t)A(1+t)^{-1} \subseteq A$ for all $t \in R$ such that $t^2=0$. Assume that R possesses nontrivial idempotents in a sub-meter resolution, hypeproductive, sylvatic, *Ae. aegypti,* geo-spectrotemrpoal or geo-spatiotemrpoal, African, riceland agro-village complex, larval habitat, capture point, oviposition, LULC model. . Then an arbovirologist, medical entomologist of other experimenter may employ abstract algebra for quantiating an LULCV element x of a set with a binary operation [i..e, an idempotent element (or just an idempotent) if $x * x = x$]. In so doing the YF model ouptut would reflect the idempotence of the binary operation on that particular LULC element.Idempotents are especially prominent in ring theory(Cressie 1993). For general rings, elements idempotent under multiplication are tied with decompositions of modules, as well as to homological properties of the ring. In Boolean algebra, the main objects of study are rings in which all elements are idempotent under both addition and multiplication (Hazewinkle 2001)..

It may be proven the ch$R \neq 2$ or if$RC \neq C_2$, then any noncentral additive subgroup of$R$ invariant under special automorphisms in a eco-epidemiological, sub-meter reoslution *Ae. aegypti*, ovispoition, African , ricland, agro-village, inhomogeneous forest-canopied, grid-stratfiied, LULC model that contains a noncentral Lie ideal. If ch$R$=2,RC=C_2 and$C \neq \{0, 1\}$, then the following two conditions may be equivalent in the forecast estimator dataaset of seasonal, hypeproductive foci : (i) any noncentral additive subgroup invariant under special automorphisms would contains a noncentral Lie ideal; (ii) there is$\alpha \in Z / \{0\}$ such that$\alpha^2 Z \subseteq \{\beta^2 : \beta \in Z\}$ in the resdiuals.

The factor group $\mathbb{Q}/A$ by a proper subgroup A is a divisible group, hence certainly not finitely generated, hence has a proper non-trivial subgroup, which gives rise to a subgroup and ideal *containing A.(* Kelley, 1950). The Zorn lemma is a proposition of set theory that states that a partially ordered set containing upper bounds for every chain (that is, every totally ordered subset) necessarily contains at least one maximal element (see Hazewinkle 2001).

An orthonormal basis is certainly a maximal orthonormal system, but the converse need not hold in a geo-spectrotemporal, eco-epidemiological, vector, arthropod forecast, vulnerability, iteratively interpolative, probabistic paradigm.. Thus, if G is a dense subspace of an inner product space H, in a explanative, geo-spectrotemporal, YF, eco-epidemiological, risk model for determining prolific, seasonal, immature *Ae egypti* oviposition geolocations on newly transitioned, discontinuous,s forest-canopied to rice agriculture, sub-meter resolution LULCs, for example, then any orthonormal basis for G would be automatically an orthonormal basis for H. Thus, it suffices to construct an inner product space H with a dense subspace G in a d diagnostic, clinical, field or remote geo-spectrotemrpoal, forecasting vulnerability, endemic transmission-oriented, eco-epidemiological, YFV model whose dimension is strictly smaller than that of H. Let K be a Hilbert space of dimension $\aleph_0$(for example $K = \ell^2(N)$) in a geo-spectrotemporal, yellow fever, LULC, probabilistic, risk model analyses in ArcGIS. Then let E be an orthonormal basis of K, so $|E| = \aleph_0$. By extending E to a Hamel basis in the model then $E \cup F$ for K, where $E \cap F = \emptyset$. The Hamel dimension is a natural generalization of the dimension of Euclidean space, since E^n is a vector space of dimension n over R (the reals). the Hamel dimension depends on the base field, so while R has dimension 1 when considered as a vector space over itself, it has dimension c (the cardinality of the continuum) when considered as a

vector space over Q *(*Itzkov, 2009).Since it is known that the Hamel dimension of *K* is *c*, the cardinality of the continuum, it must be that $|F| = c$ in a vector artrhopos-related iterative stochastic or deterministic, endmember, sub-relation interpolator Let *L* be a Hilbert space of dimension *c* (for example $L = \ell^2(\text{R})$) in a yellow fever model. Then *B* would be an orthonormal basis for *L*, and $\phi : F \to B$ would be a bijection. In so doing, there would be a linear transformation $T : K \to L$ such that $Tf = \phi(f)$ for $f \in F$, and *Te* = 0 for $e \in E$ (Ciesielski, Krzysztof 1997)..Further, let $H = K \oplus L$. by letting $G = \{(k, Tk) : k \in K\}$ be the graph of *T* then $\bar{G}$ would be the closure of *G* in *H* in the resdiual forecasted, eco-epidemiological, time series dataset for robustly iteraively quanatiatively interpolatively and targeting prolific *Aedes* habitat, immature, sylvatic, seasonal, capture points, on deforested sub-meter resolution, geoclassifiable , Riceland irrigated, LULCs. Since for any $e \in E$ there would be $(e, 0) \in G$, (see Hazewinkle 2001) it follows that $K \oplus 0 \subset \bar{G}$ in any *Ae egypti* oviposition, hyperproductive, probabilistic , forecasting paridigm.Next, if $b \in B$, then $b = Tf$ for some $f \in F \subset K$, so $(f, b) \in G \subset \bar{G}$; since $(f, 0) \in \bar{G}$ as well, we also have $(0, b) \in \bar{G}$. It follows that $0 \oplus L \subset \bar{G}$, so $\bar{G} = H$, and *G* would be dense in *H*.Finally, $\{(e, 0) : e \in E\}$ would be a maximal orthonormal set in *G*; in the output if $0 - \langle (e, 0), (k, Tk) \rangle - \langle e, k \rangle + \langle 0, Tk \rangle - \langle e, k \rangle$ for all $e \in E$ then certainly $k = 0$, so $(k, Tk) = (0, 0)$ (i.e., the zero vector in *G*0. Hence, the dimension of *G* is $|E| = \aleph_0$, in the LULC model whereas it would be clear that the dimension of *H* is *c* in the sub-meter resolution, hypeproductive, *Ae. aegypti,* geo-spectrotemporal or geo-spatiotemporal, African, riceland agro-village complex, larval habitat, capture point, oviposition, LULC model.

However, when an experimenters consider the higher-order terms in the expansion of the distribution of a hypeproductiv , georeferenceable , sylvatic, *Ae egypti* oviposition , sub-meter resolution LULC, optimally parameterizable geo-spectrotemporally uncoalesced, iteratively intrepolative, signature estimator,derived from an ArcGIS geoclassifed discontinuous sparsely or dense shaded, LULC to rice agriculture LULC in an irrigated African ricleand agro-ecosystem , the residual forecasts may reveal θ_{mle} in the bias of order n^{-1}. This bias may be quanatized to by

$$b_s \equiv \text{E}[(\hat{\theta}_{\text{mle}} - \theta_0)_s] = \frac{1}{n} \cdot I^{si} I^{jk} \left(\frac{1}{2} K_{ijk} + J_{j,ik} \right)$$

employing Einstein's summation convention over the repeating indices. In so doing, I^{jk} would denote the *j,k*-th component of the *inverse* Fisher information matrix I^{-1}, and

$$\frac{1}{2} K_{ijk} \mid J_{j,ik} = \text{E}_X \left[\frac{1}{2} \frac{\partial^2 \ln f_{\theta_0}(X_t)}{\partial \theta_i \, \partial \theta_j \, \partial \theta_k} \mid \frac{\partial \ln f_{\theta_0}(X_t)}{\partial \theta_j} \frac{\partial^2 \ln f_{\theta_0}(X_t)}{\partial \theta_i \, \partial \theta_k} \right].$$

Using these formulas in GEN MOD would then make it possible to estimate the second-order bias of the maximum likelihood estimator, and correct for that bias by subtracting it: $\hat{\theta}^*_{\text{mle}} = \hat{\theta}_{\text{mle}} - \hat{b}$. in the sub-meter resolution, LULC model output. This estimator would be unbiased up to the terms of order n^{-1}, would be the bias-corrected maximum likelihood estimator.This bias-corrected estimator is second-order efficient (at least within the curved exponential family), meaning that it has minimal mean squared error among all second-order bias-corrected georeferenceable *Ae egypti* oviposition, sub-meter resolution LULC, parameterizable geo-spectrotemporally uncoalesced, iteratively intrepolative, signature, error estimator in a probabistic ArcGIS paradigm cartographically illustrating seasonally trasnitioned, discontinuous canopied, deforeseted LULC to rice agriculture LULC.

Each cell (except the final one, whose value is completely determined by the others) in the vulnerability, forecasting, *S. damnosum* s.l., eco-epidemiological, geo-spectrotemporal model was treated as an independent binomial variable, and their contributions were summed and each contributes revealed one df. The distribution indeed approaches asymptotically when the χ^2 distribution of the number of observations approaches infinity((Jaynes 2003) . The authors of Jacob et al. (2015) let *n* be the number of immature, capture point, seasonal, georeferenceable, elucidative, time series observations, *m* the number of cells and p_i the probability of an observation to fall in the i-th cell, for $1 \leq i \leq m$. The authors denoted $\{k_i\}$ the configuration where for each *i* in the habitat, forcasting, vulnerability model, there were k_i observations in the *i*-th cell. Note that $\sum_{i=1}^{m} k_i = n$ and $\sum_{i=1}^{m} p_i = 1$. The authors then let $\chi^2_P(\{k_i\}, \{p_i\})$ be Pearson's cumulative test statistic for such a configuration, and let $\chi^2_P(\{p_i\})$ be the distribution of this statistic. In so doing the authors were able to reveal how the latter probability approached the χ^2 distribution with *m*-1df , as n approachedinfinity.For any arbitrary value T in the *S. damnsoum* s.l., immature habitat, capture point the values rendered was

$$P(\chi^2_P(\{p_i\}) > T) = \sum_{\{k_i\} | \chi^2_P(\{k_i\}, \{p_i\}) > T} \frac{n!}{k_1! \cdot k_2! \cdot \ldots k_m!} \cdot \prod_{i=1}^{m} p_i^{k_i}$$

The authors of Jacob et al. (2015) employed a procedure similar to the approximation in de Moivre–Laplace theorem.to efficiently regressively quantitate the parameterizable, geosampled, *S. damansoum* s.l., seasonal covariates. In probability theory, the de Moivre–Laplace theorem, which is a special case of the central limit theorem, states that the normal distribution may be used as an approximation to the binomial distribution under certain conditions . As *n* grew large, for *k* in the neighborhood of *np* the authors of Jacob et al. (2015) were able to regressively approximate $\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}, \qquad p + q = 1, \ p,q > 0$ in the sense that the ratio of the left-hand side to the right-hand side converges to 1 as $n \to \infty$. Contributions from small k_i-s were of subleading order in the black-fl,y vector, probabilistic paradigm *n* ;thus, Stirling's formula for both n! and k_i! were tabulated as :

$$P(\chi^2_P(\{p_i\}) > T) \sim \sum_{\{k_i\} | \chi^2_P(\{k_i\}, \{p_i\}) > T} \prod_{i=1}^{m} \left(\frac{n \cdot p_i}{k_i}\right)^{k_i} \cdot \sqrt{\frac{2\pi n}{\prod_{i=1}^{m} 2\pi k_i}}$$

In mathematics, Stirling's approximation (or Stirling's formula) is an approximation for factorials. Abramowitz and Stegun 2002) The formula as typically used in applications is $\ln n! = n \ln n - n + O(\ln n)$ (in big O notation). The next term in the O(ln n) is1/2 ln(2πn); a more precise variant of the formula is therefore $n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$.

By substituting for $x_i = (k_i - n \cdot p_i)/\sqrt{n}$ for i = 1...m-1, the authors of Jacob et al. (2015) were able to approximate for *n* in the sum over the k_i-s by an integral over the x_i-s. Noting that $k_m = n \cdot p_n - \sqrt{n} \cdot \sum_{i=1}^{m-1} x_i$ the geo-spectrotemporal, residual forecasts renderings were

$$P(\chi^2_P(\{p_i\}) > T) \sim \sqrt{\frac{2\pi n}{\prod_{i=1}^{m} 2\pi k_i}} \int_{\chi^2_P(\{\sqrt{n} x_i + n p_i\}, \{p_i\}) > T} \left(\prod_{i=1}^{m-1} \sqrt{n} dx_i\right) \prod_{i=1}^{m-1} \left(1 + \frac{x_i}{\sqrt{n} \cdot p_i}\right)^{-\left(n p_i + \sqrt{n} x_i\right)} \cdot \left(1 - \frac{\sum_{i=1}^{m-1} x_i}{\sqrt{n} \cdot p_n}\right)^{-\left(n p_n - \sqrt{n} \sum_{i=1}^{m-1} x_i\right)}$$

and then

$$= \sqrt{\frac{2\pi n}{\prod_{i}^{m}\left(2\pi n \cdot p_i + 2\pi\sqrt{n}\cdot x_i\right)}} \int_{\chi_P^2(\{\sqrt{n}x_i + n\cdot p_i\},\{p_i\})>T} \left(\prod_{i=1}^{m-1}\sqrt{n}dx_i\right)\prod_{i=1}^{m-1}\exp\left[-\left(n\cdot p_i + \sqrt{n}\cdot x_i\right)\cdot\ln\left(1+\frac{x_i}{\sqrt{n}\cdot p_i}\right)\right]\cdot\exp\left[-\left(n\cdot p_n - \sqrt{n}\sum_{i=1}^{m-1}x_i\right)\cdot\ln\left(1-\frac{\sum_{i=1}^{m-1}x_i}{\sqrt{n}\cdot p_n}\right)\right]$$

By expanding the logarithm in the *S. damnoum* s.l. eco-epidemiological, forecast model and taking the leading terms in *n* gave

$$P(\chi_P^2(\{p_i\})>T) \sim \frac{1}{\prod_{i=1}^{m-1}\left(\sqrt{p_i}\right)}\cdot\frac{1}{(2\pi)^{m-1}\sqrt{p_n}}\int_{\chi_P^2(\{\sqrt{n}x_i + n\cdot p_i\},\{p_i\})>T}\left(\prod_{i=1}^{m-1}dx_i\right)\prod_{i=1}^{m-1}\exp\left[-\frac{1}{2}\left(\sum_{i=1}^{m-1}x_i^2/p_i + (\sum_{i=1}^{m-1}x_i)^2/p_n\right)\right]$$

Now, it should be noted that Pearson's chi, $\chi_P^2(\{k_i\},\{p_i\}) = \chi_P^2(\{\sqrt{n}\cdot x_i + n\cdot p_i\},\{p_i\})$, was precisely the argument of the exponent except for the -1/2 in the black fly model in Jacob et al. (2015).. Further, the final term in the exponent's argument was equal to $(k_m - n\cdot p_n)^2/(n\cdot p_n)$ This argument was written as:

$$-\frac{1}{2}\sum_{i,j=1}^{m-1}x_i A_{ij} x_j$$

, in GEN MOD where i,j = 1... m-1 and $A_{ij} = \delta_{ij}/p_i + 1/p_n$. A is a regular symmetric (m-1)x(m-1) matrix, and hence diagonalizable(Griffth 2003). It was therefore possible to make a linear change of the geospectrotemporally geosampled, *S. damnsoum* .s.l. capture point, immature habitat, predictors {x_i} so as to get m-1 new variables {x'_i} so that:

$$\sum_{i,j=1}^{m-1}x_i A_{ij} x_j = \sum_{i=1}^{m-1}x_i'^2$$

. This linear change in the onchocerciasis, forecasting vulnerability, model merely multiplied the integral by a constant Jacobian, which subsequently rendered $P(\chi_P^2(\{p_i\})>T) \sim C\cdot\int_{\sum_{i=1}^{m-1}x_i'^2>T}\left(\prod_{i=1}^{m-1}dx_i'\right)\prod_{i=1}^{m-1}\exp\left[-\frac{1}{2}\left(\sum_{i=1}^{m-1}x_i'^2\right)\right]$ where C was a constant. In vector calculus, the Jacobian matrix is the matrix of all first-order partial derivatives of a vector-valued function (Hazewinkle 2001).

Note that the sample size *N* completely dropped out of the probability function, which in this research had the same functional form for all the geosampled parameterized covariate estimator indicator values (i.e., v). As expected, the Poisson distribution was normalized so that the sum of probabilities equaled 1. The ratio of probabilities was then determined by $\sum_{n=0}^{\infty}P_v(n) = e^{-v}\sum_{n=0}^{\infty}\frac{v^n}{n!} = e^{-v}e^v =$ which was then subsequently expressed as

$$\frac{P_v(n=i+1)}{P(n=i)} = \frac{\frac{v^{i+1}e^{-v}}{(i+1)!}}{\frac{e^{-v}v^i}{i!}} = \frac{v}{i+1}.$$

The Poisson distribution revealed that the explanatory covariate coefficients reached a maximum when $\frac{dP_v(n)}{dn} = \frac{e^{-v}n(\gamma - H_n + \ln v)}{n!} = 0$, where γ was the Euler-Mascheroni constant and H_n was a harmonic number, leading to the transcendental equation $\gamma - H_n + \ln v = 0$. The regression model also revealed that the Euler-Mascheroni constant arose in the integrals as

$$\gamma = -\int_0^{\infty}e^{-x}\ln x\,dx = -\int_0^1\ln\ln\left(\frac{1}{x}\right)dx = \int_0^{\infty}\left(\frac{1}{1-e^{-x}} - \frac{1}{x}\right)e^{-x}\,dx = \int_0^{\infty}\frac{1}{x}\left(\frac{1}{1+x} - e^{-x}\right)dx$$

. The Euler-

Mascheroni constant γ, sometimes also called 'Euler's constant' or 'the Euler constant' (but not to be confused with the constant $e = 2.718281\ldots$) is defined as the limit of the sequence $\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \ln n\right) = \lim_{n\to\infty}(H_n - \ln n),$ where H_n is a harmonic number (Haight 1967). It was first defined by Euler (1735), who used the letter C, and stated that it was "worthy of serious consideration" [133]. The symbol γ was first used by Mascheroni (1790) where γ has the numerical value $0.5772156649015328606065120900824024310421\ldots$

If the vector space V is finite-dimensional, then the linear transformation T in the vulnerability, yellow fever, sub-meter resolution, LULV, forecasting paradigm can be represented as a square matrix A, and the vector v by a column vector, rendering the above mapping as a matrix multiplication on the left hand side and a scaling of the column vector on the right hand side in the equation $A\mathbf{v} = \lambda\mathbf{v}.$ There is a correspondence between n by n square matrices and linear transformations from an n-dimensional vector space to itself. For this reason, it is equivalent to define eigenvalues and eigenvectors using either the language of matrices or the language of linear transformations (Griffith 2003) .Geometrically, an eigenvector corresponding to a real, nonzero eigenvalue points in a direction that is stretched by the transformation and the eigenvalue is the factor by which it is stretched. If the eigenvalue is negative, the direction is reversed(Anselin 1995) If two-dimensional space is visualized as a piece of entomological material (e.g., geo-spectrotemporally geosampled, ,Ae geypti, seasonally hyperproductive, georeferenceable, immature, capture point, sentinenel site on a geoclassfiiable, georferenceable, newly transitioned, sub-meter resolution, discontinuous, forest-canopied, dense or sparasely shaded LULC to rice agriculture LULC being stretched by the ArcGIS, geospatially, weighted matrix, the eigenvectors would make up the line along the direction the materail etched in and the line of cloth at the center of the stretching, whose direction isn't changed by the stretching either(Jacob et al. 2008). The eigenvalues for the first line would give the scale to which the material is stretched, and for the second line the scale to which it is tightened. As such, a habitat LULC, sub-meter resolution polygon, reflection of deforeseted, riceland agroecosystems may be viewed as stretching a line to scale −1 while shrinking the axis of reflection to scale 1. For 3D rotations, the eigenvectors form the axis of rotation, and since the scale of the axis is unchanged by the rotation, their eigenvalues are all 1 (www .esri.com).

In Jacob et al. (2013) a Poisson regression model in GEN MOD was constructed employing the geo-spectrotemporal, seasonal-geosampled district-level, explanatorial, covariate coefficient measurement values dory every district in Uganda. The malaria, eco-epidemiological, forecasting, vulnerability model was generalized by introducing an unobserved heterogeneity term for each geosampled district-level observation i at the study site. The weights were then assumed to differ randomly in a manner that was not fully accounted for by the other seasonal-geosampled, parameterizable, geo-spectrotemporal covariates. This district-level process was formulated as $E(y_i \mid x_i, \tau_i) = \mu_i \tau_i = e^{x_i'\beta + \varepsilon_i}$ where the unobserved heterogeneity term $\tau_i = e^{\varepsilon_i}$ was independent of the vector of regressors x_i. Then the distribution of y_i was conditional on x_i and had a Poisson specification with conditional mean and conditional variance

$\mu_i \tau_i : f(y_i \mid x_i, \tau_i) = \dfrac{\exp(-\mu_i \tau_i)(\mu_i \tau_i)^{y_i}}{y_i!}$. The authors then let $g(\tau_i)$ be the probability density function of τ_i. Then, the distribution $f(y_i \mid x_i)$ was no longer conditional on τ_i Instead it was obtained by integrating $f(y_i \mid x_i, \tau_i)$ with respect to $g(\tau_i)$: $f(y_i \mid x_i) = \int_0^\infty f(y_i \mid x_i, \tau_i) g(\tau_i) d\tau_i$.

Jacob et al. (2013) found that an analytical solution to this integral existed in the, district-level malarial Poission model when τ_i was assumed to follow a gamma distribution. The model also revealed that y_i, was the vector of the geosampled predictor covariate coefficients while x_i, was independently Poisson distributed with $P(Y_i = y_i \mid x_i) = \dfrac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, y_i = 0, 1, 2, \ldots$ and the mean parameter — that is, the mean number of paramterizable, time series, explicative, district-level covariates per geosampled period was given by $\mu_i = \exp(x_i' \beta)$ where β was a $(k+1) \times 1$ parameter vector. The intercept in the model was β_0 and the coefficients for the k regressors were $\beta_1, \ldots, \beta_k$. The quantitated exponential of $x_i' \beta$ ensured that the mean parameter μ_i was nonnegative. Thereafter, the conditional mean was provided by $E(y_i \mid x_i) = \mu_i = \exp(x_i' \beta)$.

In Jacob et al. (2013),the district-level explanative, geo-spectrotemporal or geo-spatiotemporal, eco-epidemiological, parameterizable estimators were then evaluated using $\ln[E(y_i \mid x_i)] = \ln(\mu_i) = x_i' \beta$. Note, that the conditional variance of the count random variable was equal to the conditional mean (i.e., equidispersion) in the malarial model [i.e., , $V(y_i \mid x_i) = E(y_i \mid x_i) = \mu_i$]. In a log-linear model the logarithm of the conditional mean is linear[Hosmer and Lemeshew 2002]. The marginal effect of any district-level regressor in the geo-spectrotemporal, malarial model was then provided by $\dfrac{\delta E(y_i \mid x_i)}{\delta x_{ji}} = \exp(x_i' \beta) \beta_j = E(y_i \mid x_i) \beta_j$ in GEN MOD. Thus, a one-unit change in the jth regressor in the model led to a proportional change in the conditional mean $E(y_i \mid x_i)$ of β_j.

In Jacob et al. (2013), the standard estimator for the malarial, Poisson, forecasting , vulnerability model was the maximum likelihood estimator. Since the district-level explanatorial, observations were independent, the log-likelihood function in the model was $= \sum_{i=1}^{N}(-\mu_i + y_i \ln \mu_i - \ln y_i!) = \sum_{i=1}^{N}(-e^{x_i' \beta} + y_i x_i' \beta - \ln y_i!)$. Given the geosampled dataset of district-level, explicative, time series, caovariate,unbiased, parameter estimators (i.e., θ) and an input vector x, the mean of the predicted Poisson distribution was then provided by $E(Y \mid x) = e^{\theta x}$. In so doing, the Poisson distribution's probability mass function was then optimally rendered

$$p(y \mid x; \theta) = \frac{e^{y(\theta'x)} e^{-e^{\theta'x}}}{y!}$$

by $\quad$ The probability mass function in a targeted spatiotemporal, predictive, seasonal, malaria risk model can be the primary means for defining a discrete probability distribution, and, as such, functions could exist for either scalar or multivariate field-sampled random variables, given that the distribution is discrete [Gu and Novak 2005].

Since in this research, the geospectrotemrpoally geosampled data consisted of m vectors $x_i \in \mathbb{R}^{n+1}, i = 1, \ldots, m,$, along with a set of m values $y_1, \ldots, y_2 \in \mathbb{R}$ then, for the district-level, parameter estimators θ, the probability of attaining this particular set of the geosampled observations was provided by the equation $p(y_1, \ldots, y_m \mid x_1, \ldots, x_m; \theta) = \prod_{i=1}^{m} \frac{e^{y_i(\theta'x_i)} e^{-e^{\theta'x_i}}}{y_i!}$ .Consequently, the authors found the set of θ that made this probability as large as possible in the model estimates. To do this, the equation was first rewritten as a likelihood function in terms of θ: $p(y_1, \ldots, y_m \mid x_1, \ldots, x_m; \theta) = \prod_{i=1}^{m} \frac{e^{y_i(\theta'x_i)} e^{-e^{\theta'x_i}}}{y_i!}$ in PROC REG.Note the expression on the right hand side in our model had not actually changed. Next, the authors optimally employed a geo-spectrotemporal,

$$\ell(\theta \mid X, Y) = \log L(\theta \mid X, Y) = \sum_{i=1}^{m} \left(y_i (\theta'x_i) - e^{\theta'x_i} - \log(y_i!) \right)$$

log-likelihood[i.e., $\qquad$. Because the logarithm is a monotonically increasing function, the logarithm of a function achieves its maximum value at the same points as the function itself, and, hence, the log-likelihood can be used in place of the likelihood in maximum likelihood estimation and related techniques[Fox 1997]. Finding the maximum of a function in a malarial-related model often involves taking the derivative of a function and solving for the parameter estimator being maximized, and this is often easier when the function being maximized is a log-likelihood rather than the original likelihood function [Draper and Smith,1998].

Notice that the parameters θ only appeared in the first two terms of each term in the summation in PROC REG. Therefore, given that the authors in Jacob et al. (2013) were only interested in finding the best value for θ in the forecasting, vulnerability, time series, eco-epidemiological, geo-spectrotemporal, ecogeorefernceable, district-level, predictive, malarial-related, regression model, the authors dropped the $y_i!$ and simply wrote

$$\ell(\theta \mid X, Y) = \sum_{i=1}^{m} \left(y_i (\theta'x_i) - e^{\theta'x_i} \right)$$

. Thereafter, to quantitate a maximum, the authors solved an equation $\frac{\partial \ell(\theta \mid X, Y)}{\delta \theta} = 0$ which had no closed-form solution. However, the negative log-likelihood (LL)[i.e., $-\ell(\theta \mid X, Y)$] was a convex function, and so standard convex optimization was applied to find the optimal value of θ .

In Jacob et al. (2012) the Poisson process in a   regression model the limit of a binomial

distribution was $P_p(n \mid N) = \dfrac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$.   Viewing the distribution as a function of the

expected number of successes[i.e., $\nu \equiv Np$] in the model, instead of the sample size N for fixed

P, then rendered the equation (2.1) which then became $p_{\nu/N}(n \mid N) = \dfrac{N!}{n!(N-n)!}\left(\dfrac{\nu}{N}\right)^n \left(1 - \dfrac{\nu}{N}\right)^{N-n}$ Our
model revealed that as the sample size N become larger, the distribution approached P when the
following   equations

aligned $\lim\limits_{N\to\infty} p_p(n \mid N)$ $\lim\limits_{N\to\infty} \dfrac{N(N-1)\ldots(N-n+1)}{n!} \dfrac{\nu^n}{N^n}\left(1-\dfrac{\nu}{N}\right)^N \left(1-\dfrac{\nu}{N}\right)^{-n}$ $\lim\limits_{N\to\infty} \dfrac{N(N-1)\ldots(N-n+1)}{N^n} \dfrac{\nu^n}{n!}\left(1-\dfrac{\nu}{N}\right)^N \left(1-\dfrac{\nu}{N}\right)^{-n}$

$1 \cdot \dfrac{\nu^n}{n!} \cdot e^{-\nu} \cdot 1 \,\text{and}\quad \dfrac{\nu^n e^{-\nu}}{n!}$. Note, in this research, that the sample size N had completely dropped out of
the probability function, which had the same functional form for all values of ν in the model.

Thereafter, as expected, the Poisson regression distribution was normalized so that the sum

of probabilities was equal to 1, since $\sum\limits_{n=0}^{\infty} P_\nu(n) = e^{-\nu} \sum\limits_{n=0}^{\infty} \dfrac{\nu^n}{n!} = e^{-\nu} e^{\nu} = 1$ The ratio of probabilities

was then provided by the equation $\dfrac{P_\nu(n=i+1)}{P(n=i)} = \dfrac{\dfrac{\nu^{i+1} e^{-\nu}}{(i+1)!}}{\dfrac{e^{-\nu} \nu^i}{i!}} = \dfrac{\nu}{i+1}$. Our model revealed that the

Poisson distribution reached a maximum when $\dfrac{dP_\nu(n)}{dn} = \dfrac{e^{-\nu} n (\gamma - H_n + \ln \nu)}{n!} = 0$ where g was the

Euler-Mascheroni constant and H_n was a harmonic number, leading to the equation
$\gamma - H_n + \ln \nu = 0$ which could not be solved exactly for n.
Next, the moment-generating function of the Poisson distribution was given by

$\mathrm{M} = M = e^{-\nu} e^{\nu e^t} = e^{\nu(e^t - 1)}$; $M = \nu e^t e^{\nu(e^t - 1)}$ and $\mathrm{M} = M = (\nu e^t)^2 e^{\nu(e^t-1)} + \nu e^t e^{\nu(e^t-1)}$, when

$R = \nu(e^t - 1)$, $R' = \nu e^t$ so $R = R'(0) = \nu$. The raw moments were also computed directly by
summation, which yielded an unexpected connection with the exponential polynomial

$\varphi_n(x)$ and Stirling numbers of the second kind[i.e. $\phi_n(x) = \sum\limits_{k=0}^{\infty} \dfrac{e^{-x} x^k}{k!} k^n = \sum\limits_{k=1}^{n} x^k S(n,k)$] which in
this research was the Dobiński's formula.

In combinatorial mathematics, Dobinski's formula states that the number of partitions of

a set of n members is $\dfrac{1}{e}\sum\limits_{k=0}^{\infty}\dfrac{k^n}{k!}$ This number has come to be called the nth Bell number B_n, where the proof is rendered as an adaptation to probabilistic language as given by Rota[11]. In our

malarial-based regression model the formula $\phi_n(x)=\sum\limits_{k=0}^{\infty}\dfrac{e^{-x}x^k}{k!}k^n=\sum\limits_{k=1}^{n}x^k S(n,k)$ was then

viewed as a particular case, for $x=0$, employing the relation $\dfrac{1}{e}\sum\limits_{k=x}^{\infty}\dfrac{k^n}{(k-x)!}=\sum\limits_{k=0}^{n}\binom{n}{k}B_k x^{n-k}$. The expression given by the model's Dobinski's formula was then revealed as the n th moment of the Poisson distribution with expected value 1. In this research, Dobinski's formula was the number of partitions of a set of the sampled malarial parameter estimator size (i.e.,n) which equalled the nth moment of that distribution. We used the Pochhammer symbol $(x)_n$ to denote the falling factorial $(x)_n=x(x-1)(x-2)...(x-n+1)$. If x and n are nonnegative integers, $0 \le n \le x$, then $(x)_n$ is the number of one-to-one functions that map a size-n set into a size-x set[1]. At this junction we let f be any function from a size-n set A into a size-x set B. Thus, in the model. $u \in B$ .We then let $f^{-1}(u) = \{v \in A : f(v) = u\}$. Then $\{f^{-1}(u) : u \in B\}$ was a partition of A. This equivalence relation was the "kernel" of the function f. Any function from A into B factors in to one function that maps a member of A to that part of the kernel to which it belongs, and another function, which is necessarily one-to-one, that maps the kernel into B[2]. In this research the first of these two factors was completely determined by the partition π, that is the kernel. The number of one-to-one functions from π into B was then $(x)_{|\pi|}$, in the district-level malarial regression model when $|\pi|$ was the number of parts in the partition π. Therefore, the total number of functions from a

size-n set A into a size-x set B was $\sum\limits_{\pi}(x)_{|\pi|}$ in the model when the index π ran through the set of all partitions of A. On the other hand, the number of functions from A into B was clearly x^n.

$$x^n = \sum_{\pi}(x)_{|\pi|}$$

Thus, the authors of Jacob et al. (2015) had Since X was a Poisson-distributed spatiotemporal-seasonal, malarial-related, district-level, random variable with

expected value 1, then the nth moment of this probability distribution was $E(X^n)=\sum\limits_{\pi}E\left((X)_{|\pi|}\right)$ but all of the factorial moments $E((X)_k)$ of this probability distribution was equal to 1 in the model

also. Thereafter, we had, $E(X^n)=\sum\limits_{\pi}1$,which was the number of partitions of the set A in the

model. Therefore, in the model, $v(1+v)$, $v(1+3v+v^2)$, and $v(1+7v+6v^2+v^3)$.

Thereafter, the central moments in the malarial model was computed as $v(1+3v)$ so the mean, variance, skewness, and kurtosis were

$\dfrac{\mu_3}{\sigma^3}=\dfrac{v}{v^{3/2}}=v^{-1/2}$, $\dfrac{\mu_4}{\sigma^4}-3=\dfrac{v(1+3v)}{v^2}-3$ and $\dfrac{v+3v^2-3v^2}{v^2}=v^{-1}$, respectively. The characteristic

function for the Poisson distribution in the district -level Poisson predictive autoregressive model was then revealed as $\phi(t) = e^{\nu\left(e^{it}-1\right)}$ and the cumulative distribution function was $K(h) = \nu\left(e^h - 1\right) = \nu\left(h + \frac{1}{2!}h^2 + \frac{1}{3!}h^3 + ...\right)$ so $\kappa_r = \gamma$. The mean deviation of the Poisson distribution mode was then rendered by $MD = \frac{2e^{-\nu}\nu^{[\nu]+1}}{[\nu]!}$. The cumulative distribution functions of the Poisson and chi-squared distributions were then related in the district-level model as $F_{Poisson}(k; \lambda) = 1 - F_{\chi^2}(2\lambda; 2(k+1))$ integer k and $P_r(X=k) = F_{\chi^2}(2\lambda; 2k) - F_{\chi^2}(2\lambda; 2(k+1))$. The Poisson distribution was then expressed in terms of $\lambda \equiv \frac{\nu}{x}$ whereby, the rate of changes were equal to the equation $P_\nu(n) = \frac{(\lambda x)^n e^{-\lambda x}}{n!}$. The moment-generating function of the Poisson distribution generated from the sampled, district-level, explanatory, predictor variables was also rendered by $M(t) = e^{(\nu_1 + \nu_2)(e^t - 1)}$ Given a random variable x and a probability distribution function $P(x)$, if there exists an $h > 0$ such that $M(t) \equiv \langle e^{tx} \rangle$ for $|t| < h$, where $\langle y \rangle$ denotes the expectation value of y, then $M(t)$ is called the moment-generating function[2]. Commonly, for a continuous distribution in a seasonal linear regression-based time-series dependent regression model $\int_{-\infty}^{\infty} e^{tx} P(x)\,dx \int_{-\infty}^{\infty}\left(1 + tx + \frac{1}{2!}t^2 x^2 + ...\right) P(x)\,dx$ the equation $1 + t m_1' + \frac{1}{2!}t^2 m_2' + ...$ is is used where m_r' the r the raw moment.[5]. For quantifying independent X and Y, the moment-generating function in a robust model must satisfy the equation $M_{x+y}(t) = \langle e^{t(x+y)} \rangle, \langle e^{tx} e^{ty} \rangle, \langle e^{tx} \rangle \langle e^{ty} \rangle$ and $M_x(t) M_y(t)$ if, the independent variables $x_1, x_2, ..., x_N$ have Poisson distributions with parameters $\mu_1, \mu_2, ..., \mu_N$ and $X = \sum_{j=1}^{N} x_j$ [3].In this research this was evident since the cumulant-generating function was $K \equiv \sum_j K_j(h) = \left(e^h - 1\right)\sum_j \mu_j = \mu\left(\left(e^h - 1\right)\right)$.

In Jacob et al. (2013), the eco-epidemiological, endemic, district-level, malaria model had a directed Kullback-Leibler (K-L) divergence between Pois(λ) and Pois(λ_0) which was provided by $D_{KL}(\lambda \| \lambda_0) = \lambda_0 - \lambda + \lambda \log\frac{\lambda}{\lambda_0}$. In probability theory and information theory, the K-L divergence along with information divergence, information gain, relative entropy are a non-symmetric measures of the difference between two probability distributions P and Q in a model. For quantifying the malaria, forecasted, probability distributions P and Q of a sampled discrete random variable the K–L divergence was defined by $D_{KL}(P \| Q) = \sum_i P(i)\ln\frac{P(i)}{Q(i)}$ .in Jacob et al. (2013) The model revealed that the average of the logarithmic difference between the

probabilities P and Q was the average quantified using the probabilities P whisly quantiating all significant paramterizable covariates. The K-L divergence is only defined if P and Q both sum to 1 and if $Q(i) > 0$ for any i such that $P(i) > 0$ [Griffth 2003]. In the district-level, spatiotemporal, malaria-based, forecast, vulnerability, regression-based model, if the quantity 0 ln 0 appeared it was interpreted as zero. For distributions P and Q of the continuous random variable in the sampled datasets K-L divergence was defined to be the integral[i.e.,

$$D_{KL}(P \| Q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx]$$

where p and q denoted the geosampled district level immature densities of P and Q. More generally, since P and Q were probability measures over the geosampled dataset X, and Q which was absolutely continuous with respect to P, then the K-L

divergence from P to Q was defined as $D_{KL}(P \| Q) = -\int_x \ln \frac{dP}{dQ} dP$ in the model output where $\frac{dQ}{dP}$ was the Radon–Nikodym derivative of Q with respect to P, provided the expression on the right-hand side existed. In mathematics, the Radon–Nikodym theorem is a result in measure theory that states that given a measurable space (i.e., X,Σ), if a σ-finite is measured on (i..e, X,Σ) then the expression is absolutely continuous with respect to a σ-finite measure μ on (X,Σ). In so doing, in this research a measurable function f was rendered on X $(0,\infty)$, such that $v(A) = \int_A f d\mu$ for any other measured value which then revealed the statistical significance of the sampled district-level malaria-related, vulnerability, covariate coefficients.

Likewise, since P was absolutely continuous with respect to Q in the district-level malaria, regression model. The explanatory, predictor, covariate coefficients were then defined employing: $D_{KL}(P \| Q) = \int_x \ln \frac{dP}{dQ} dP = \int_x \frac{dP}{dQ} \ln \frac{dP}{dQ} dQ$ which in Jacob et. al. (2015) was recognized as the entropy of P relative to Q. The authors found that if μ was any measure on X in the model then

$p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ existed, and the K-L divergence from P to Q was given as

$D_{KL}(P \| Q) = \int_x p \ln \frac{p}{q} d\mu$. The bounds for the tail probabilities of the Poisson random variable were then derived in the district-level malarial regression model employing a Chernoff bound

argument as $X \sim Pois(\lambda)$ $P(X \geq x) \leq \frac{e^{-\lambda}(e\lambda)^x}{x^x}$, for $x < \lambda$ and as $P(X \geq x) \leq \frac{e^{-\lambda}(e\lambda)^x}{x^x}$ for $x < \lambda$.

In probability theory, the Chernoff bound, provides exponentially decreasing bounds on tail distributions of sums of independent random variables. It is a sharper bound than the known first or second moment based tail bounds such as Markov's inequality or Chebyshev inequality, which only yield power-law bounds on tail decay. However, the Chernoff bound required that the variates be independent - a condition that neither the Markov nor the Chebyshev inequalities require. In probability theory, Markov's inequality gives an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant[5].

In this research, we let X_1, ..., X_n be independent Bernoulli random variables, each having probability $p > 1/2$. Then the probability of simultaneous occurrence of more than $n/2$ of the district-level sampling events had an exact value S in the model

$$S = \sum_{i=\lfloor \frac{n}{2} \rfloor + 1}^{n} \binom{n}{i} p^i (1-p)^{n-i}.$$

when   The Chernoff bound revealed that S had the following lower bound: $S \geq 1 - e^{-2n\left(p - \frac{1}{2}\right)^2}$. We noticed that if X was any sampled district-level random variable and $a > 0$, then $\Pr(|X| \geq a) \leq \frac{E(|X|)}{a}$. In the language of measure theory, Markov's inequality states that if (X, Σ, μ) is a measure space, f is a measurable extended real-valued function, and $\epsilon \geq 0$, then $\mu(\{x \in X : |f(x)| \geq \epsilon\}) \leq \frac{1}{\epsilon} \int_X |f| \, d\mu$ [2] We then used the Chebyshev's inequality to determine the variance bound to the probability that the spatiotemporal-seasonal sampled random variable deviated far from the mean in the model. Specifically we used $\Pr(|X - E(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}$ for any $a>0$. In this research, Var(X) was the variance of X, defined as: $\text{Var}(X) = E[(X - E(X))^2]$. Chebyshev's inequality follows from Markov's inequality by considering the random variable $(X - E(X))^2$ for which Markov's inequality also reads $\Pr((X - E(X))^2 \geq a^2) \leq \frac{\text{Var}(X)}{a^2}$ (Gelman 1995). Further, in Markov's inequality if x takes only nonnegative field-sampled malarial values, then $P(x \geq a) \leq \frac{(x)}{a}$. can be re-written

$$(x) = \int_0^{\infty} x P(x) \, dx = \int_0^a x P(x) \, dx + \int_a^{\infty} x P(x) \, dx.$$

However, since $P(x)$ is a prevalence rate value in a spatiotemporal malarial regression-based model, it must be ≥ 0. Thus, it must be stipulated that $x \geq 0$ so

$$(x) = \int_0^a x P(x) \, dx + \int_a^{\infty} x P(x) \, dx \geq \int_a^{\infty} x P(x) \, dx \geq \int_a^{\infty} a P(x) \, dx = a \int_a^{\infty} P(x) \, dx = a P(x \geq a).$$

in order to determine district–level covariate coefficients of statistical significance

The authors in Jacob et al. (2013) then considered the Euler product $\zeta(s) = \prod_{k=1}^{\infty} \frac{1}{1 - \frac{1}{p_k^s}}$ where $\zeta(s)$ was the Riemann zeta function and p_k was the k the prime. $\zeta(1) = \infty$. Thereafter, by taking the finite product up to k=n in the district-level malarial regression model and pre-multiplying by a factor $1/\ln p_n$, we were able to employ $n \to \infty$ to render $\lim_{n \to \infty} \frac{1}{\ln p_n} \prod_{k=1}^{n} \frac{1}{1 - \frac{1}{p_{k=e^\gamma}}} = e^\gamma$ which was equivalent to 1.781072…..By doing so, g became the Euler-Mascheroni constant which in this research also represented the limit of the sequence g= $\gamma = \lim_{n \to \infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln n \right) = \lim_{n \to \infty} (H_n - \ln n) =$ in the residuals where H_n was the harmonic number which in this research had the form $H_n = \sum_{k=1}^{n} \frac{1}{k}$ in the district-level malarial regression model. A harmonic number can be expressed analytically as $H_n = \gamma + \psi_0(n+1)$ where γ is the Euler-Mascheroni constant and $\Psi(x) = \psi_0(x)$ is the digamma function[2]. Our model revealed that the Euler product attached to the Riemann zeta function

$\zeta(s)$ represented the sum of the geometric series rendered from the spatiotemporal-sampled empirical dataset of explanatory predictor covariate coefficients as $\prod_{p}(1-p^{-s})^{-1} = \prod_{P}\left(\sum_{n=0}^{\infty} p^{-ns}\right) = \sum_{n=1}^{\infty}\frac{1}{n^s} = \zeta(s)$. A closely related result was also obtained by noting that

$$1 + \frac{1}{\mu_k} = \frac{1 - \frac{1}{p_k^2}}{1 - \frac{1}{p_k}}.$$

Jacob et al. (2012) constructed a Poissonian, *S. damnosum* s.l. habitat, vulnerability model in SAS GEN MOD The authors employed a regression matrix, for qualitatively quantitating an operational dataset of seasonally geosampled, productive, *S. damnosum* s.l., trailing vegetation, sparsely canopied LULC, riverine, larval habitat, unmixed remotely sensed forecasting, eco-epidemiological, explanatorily, signature, risk model variables with autocorrelated disturbances as follows: Yt = Xtiβ +Vt In these equations, Ytwere log-transformed, dependent, decomposed, satellite, wavelength values, Xt was a column vector of the decomposed, probabilistic, regressor variables, β was a column vector of structural parameters, where ϵt was normally and independently distributed with a mean of 0 and a variance of σ2. Note that in the parameterization, the signs of the autoregressive *S. damnosum* s.l. immature uncoalesced, iteratively interpolative, explanatorial estimators were reversed from the parameterization documented in literature. ArcGIS probabilistic, uncertainty-oriented, photosynthetic, endmember estimation methods for the optimizable, geo-spectrotemporal geosampled, canopied, *S. damnosum* s.l., georeferenced, riverine, larval habitat, time series, ecogeographical, explanatorial, error model initially employed a default method. Yule-Walker (YW) estimation which was performed in ArcGIS. The Yule–Walker equations are the following

$$\gamma_m = \sum_{k=1}^{p} \varphi_k \gamma_{m-k} + \sigma_\varepsilon^2 \delta_{m,0},$$

set of equations. where $m = 0, ..., p$, yielding $p + 1$ equations. Here γ_m is the autocovariance function of X_t, σ_ε is the standard deviation of the input noise process, and $\delta_{m,0}$ is the Kronecker delta function. In mathematics, the Kronecker is a function of two variables, usually just positive integers. The function is 1 if the variables are equal, and 0 otherwise:

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

where the Kronecker delta δ_{ij} is a piecewise function of variables i and j. For example, $\delta_{1\,2} = 0$, whereas $\delta_{3\,3} = 1$

In linear algebra, the $n \times n$ identity matrix I has entries equal to the Kronecker delta: $(I)_{ij} = \delta_{ij}$ where i and j take the values 1, 2, ..., n, and the inner product of vectors can be written as $\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{ij} a_i \delta_{ij} b_j$. (Agarwal 2007)The Kronecker delta has the so-called *sifting* property that for $j \in \mathbb{Z}$: $\sum_{i=-\infty}^{\infty} a_i \delta_{ij} - a_j$. and if the integers are viewed as a measure space, endowed with the counting measure, then this property coincides with the defining property of the Dirac delta function $\int_{-\infty}^{\infty} \delta(x-y)f(x)\,dx - f(y)$, and in fact Dirac's delta was named after the Kronecker delta because of this analogous property. Theodore (2012)In signal processing it is

usually the context (discrete or continuous time) that distinguishes the Kronecker and Dirac "functions". And by convention, $\delta(t)$ generally indicates continuous time (Dirac), whereas arguments like *i, j, k, l, m*, and *n* are usually reserved for discrete time (Kronecker). Another common practice is to represent discrete sequences with square brackets; thus: $\delta[n]$ (see Lovelock, and Rund (1989) It is important to note that the Kronecker delta is not the result of directly sampling the Dirac delta function (Griffith 2003) .The Kronecker delta forms the multiplicative identity element of an incidence algebra (Spiegel, and O'Donnell 1997),

In terms of the, geospectrotemporally, geosampled, sparsely canopied, trailing vegation, geoclassified uncoalesced *S. damnsoum* s.l. remote, regressive, endmemberm LULC ,indices, the authors in Jacob et al. (2015) employed a classfication regression scheme such that:

$$\delta^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_p} = \begin{cases} +1 & \text{if } \nu_1\ldots\nu_p \text{ are distinct integers and are an even permutation of } \mu_1\ldots\mu_p \\ -1 & \text{if } \nu_1\ldots\nu_p \text{ are distinct integers and are an odd permutation of } \mu_1\ldots\mu_p \\ 0 & \text{in all other cases.} \end{cases}$$

The authors then let $\mathfrak{S}_p$ be the symmetric group of degree *p* in the riverine forecasting paradigm where

$$\delta^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_p} = \sum_{\sigma\in\mathfrak{S}_p} \text{sgn}(\sigma)\, \delta^{\mu_1}_{\nu_{\sigma(1)}}\cdots\delta^{\mu_p}_{\nu_{\sigma(p)}} = \sum_{\sigma\in\mathfrak{S}_p} \text{sgn}(\sigma)\, \delta^{\mu_{\sigma(1)}}_{\nu_1}\cdots\delta^{\mu_{\sigma(p)}}_{\nu_p}.$$

Using anti-paramter symmetrization rendered: $\delta^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_p} = p!\,\delta^{\mu_1}_{[\nu_1}\ldots\delta^{\mu_p}_{\nu_p]} = p!\,\delta^{[\mu_1}_{\nu_1}\ldots\delta^{\mu_p]}_{\nu_p}$ In terms of a *p* × *p*

$$\delta^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_p} = \begin{vmatrix} \delta^{\mu_1}_{\nu_1} & \cdots & \delta^{\mu_1}_{\nu_p} \\ \vdots & \ddots & \vdots \\ \delta^{\mu_p}_{\nu_1} & \cdots & \delta^{\mu_p}_{\nu_p} \end{vmatrix}.$$

determinant the model employed Using the Laplace expansion (Laplace's formula) of determinant, the residual forecasts targeting the prolific *S. damnosum* s.l. immature, capture point, seasaonal habitats was recursively, optimally, geospectrotemporally defined as:

$$\delta^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_p} = \sum_{k=1}^{p}(-1)^{p+k}\, \delta^{\mu_p}_{\nu_k}\delta^{\mu_1\cdots\check{\mu}_k\cdots\check{\mu}_p}_{\nu_1\cdots\check{\nu}_k\cdots\nu_p}$$

which was coincidentally equivalent to

$$\delta^{\mu_p}_{\nu_p}\delta^{\mu_1\cdots\mu_{p-1}}_{\nu_1\cdots\nu_{p-1}} - \sum_{k=1}^{p-1}\delta^{\mu_p}_{\nu_k}\delta^{\mu_1\cdots\mu_{k-1}\,\mu_k\,\mu_{k+1}\cdots\mu_{p-1}}_{\nu_1\cdots\nu_{k-1}\,\nu_p\,\nu_{k+1}\cdots\nu_{p-1}},$$

where ˇindicates an index that is omitted from the sequence.When *p* = *n* (the dimension of the vector space), in terms of the Levi-Civita symbol $\delta^{\mu_1\cdots\mu_n}_{\nu_1\cdots\nu_n} = \varepsilon^{\mu_1\cdots\mu_n}\varepsilon_{\nu_1\cdots\nu_n}$ The two-dimensional Levi-Civita symbol is defined by:

$$\varepsilon_{ij} = \begin{cases} +1 & \text{if } (i,j) \text{ is } (1,2) \\ -1 & \text{if } (i,j) \text{ is } (2,1) \\ 0 & \text{if } i = j \end{cases}$$

The values can be arranged into a 2 × 2 antisymmetric matrix: $\begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} \\ \varepsilon_{21} & \varepsilon_{22} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$

The Kronecker delta is also called degree of mapping of one surface into another.[9] Suppose a mapping takes place from surface S_{uvw} to S_{xyz} that are boundaries of regions, R_{uvw} and R_{xyz} which is simply connected with one-to-one correspondence. In this framework,

if s and t are parameters for S_{uvw}, and S_{uvw} to S_{xyz} are each oriented by the outer normal n: $u = u(s,t), v = v(s,t), w = w(s,t)$ while the normal has the direction of: $(u_s i + v_s j + w_s k) \times (u_t i + v_t j + w_t k)$. Let x=x(u,v,w),y=y(u,v,w),z=z(u,v,w) be defined and smooth in a domain containing S_{uvw}, and let these equations define the mapping of S_{uvw} into S_{xyz}. Then the degree δ of mapping is $1/4\pi$ times the solid angle of the image S of S_{uvw} with respect to the interior point of S_{xyz}, O. If O is the origin of the region, R_{xyz}, then the degree, δ is given by the integral:

$$\delta = \frac{1}{4\pi} \iint_{R_{st}} \frac{\begin{vmatrix} x & y & z \\ \frac{\partial x}{\partial s} & \frac{\partial y}{\partial s} & \frac{\partial z}{\partial s} \\ \frac{\partial x}{\partial t} & \frac{\partial y}{\partial t} & \frac{\partial z}{\partial t} \end{vmatrix}}{(x^2 + y^2 + z^2)\sqrt{x^2 + y^2 + z^2}} ds dt.$$

According to Jacob et al. (2012) ,in an empirical probabilistic, regressed datset of geosampled, immature, *S. damnosum* s.l., riverine, larval habitat-related, explanatorial, ecoepidemiological, time series dependent, uncertainty-oriented, nonlinear, risk model forecasts, the vector of autoregressive parameters ϕ is known the matrix V which can be computed from decomposed, probabilistic, non-optimizable spatially pseudo-replicated uncertainty probabilities. Σ which may be then delineated byσ2V . Given Σ , the efficient emissivity transmisstance estimates of the autoregressive, explantorial, endmember, canopied, *S. damnosum* s.l.,, riverine tributary, larval habitat, time series dependent, regression parameters β were computed using generalized least squares (GLS). The GLS estimates yielded the unbiased elucidative,time series estimate of the variance σ2 in the riverine, larval habitat, forecast model spatially structured random intercept which subsequently accounted for the effect of the missing predictors in the model derivatives.

Thereafter, the YW alternated estimation of β employing the GLS with iterative equations which rendered the sample autocorrelation function. The YW method formed the OLS estimate of β. Next, φ was estimated from the sample autocorrelation function of the empirically autoregressed. explanatorial, field and remote, unmixed, immature S. damnosum s.l. riverine, larval habitat, photosynthetic and NPV, time series dependent, geospatial probabilistic, uncertainty –oriented, optical properties and OLS residuals. Then V was tabulated from the estimate of φ and Σ was estimated from Vand the OLS estimate ofσ2. The autocorrelation eigenfunction decomposition algorithm corrected estimates of the canopied, *S. damnosum* s.l., shaded, larval habitat, time series dependent, regression parameters β which were then computed by GLS, employing the estimated Σ weighted matrix. These were the Yule-Walker estimates. Other methods were the niterated YW, unconditional least squares (ULS), and maximum likelihood (ML). The ULS method is also referred to as nonlinear least squares (NLS) or exact least squares (ELS)Neter et al. 1990).

The authors in Jacob et al. (2012) then defined the transformed error, e, as e = L−1n where n = y −Xβ .in ArcGIS The unconditional sum of squares for the model, S, was S =

n'V−1n = e'e The ULS estimates were computed by minimizing S with respect to the geosampled, decomposed, canopied, S. damnosum s.l., shaded, larval habitat regression parameters β and ' φ . The full log likelihood function for the autoregressive error model waswhere V denoted the determinant of V for the ML method, the likelihood function was maximized by minimizing an equivalent sum-of-squares function. Maximizing l with respect to σ2 (and concentrating σ2 out of the likelihood) and dropping the constant term then rendered the concentrated log likelihood function Rewriting the S. damnosum s.l.-related decomposed, endmember, regression variable term within the logarithm then rendered the objective function which in Jacob et al. (2012) was expressed as a explanatory diagnostic geo-spatiotemrpoal variable. The sample autocorrelation function was computed from the structural excessive, decomposed, partially canopied, *S. damnosum* s.l., shaded, larval habitat noise which was then subsequently qualitatively regressively quantitated employing where b was an estimate of β . The sample autocorrelation function was the sum of all available lagged products of ntof order j divided by l+j, where l was the number of such products. The calculation of V from φ for the generalized, canopy decomposition, endmember, AR(m) risk model was complicated, and the size of V was dependent on the number of photosynthetic and NPV canopied observations. Instead of actually calculating Vand performing GLS in the usual way, a Kalman filter algorithm was instead used to transform the geosampled empirical data which was then employed to compute the GLS results through a recursive process.

The Kalman filters was based on linear dynamic systems discretized in the time domain. The filters were modeled on a Markov chain built on linear operators perturbed by the unmimxed, biosignature-related, geosampled, decomposed, canopied, *S. damnosum* s.l., shaded, larval habitat, regression errors including the Gaussian noise. Gaussian noise is statistical noise having a probability density function (PDF) equal to that of the normal distribution, which is also known as the Gaussian distribution (Cressie 1993). In other words, the values that the noise can take on are Gaussian-distributed. The state of the system was then eco-geographically represented as a vector of the endmember, time series, photosynthetic and NPV, covariate parameter estimator, time series, reflectance emissivity coefficinet values. At each discrete time increment, a linear operator was applied to the state to generate the new state, with some noise mixed in, and optionally some information from the controls on the system. Then, another linear operator mixed with more noise rendered the observed outputs from the true ("hidden") state. The Kalman filter may be regarded as analogous to the hidden Markov model, with the key difference that the hidden state variables take values in a continuous space (as opposed to a discrete state space as in the hidden Markov model) (Gelman 1995). There was a strong duality between the equations of the Kalman Filter and those of the hidden canopied, S. damnosum s.l., shaded, larval habitat, Markov model .

The algorithm estimated the internal state of the decomposed endmember riverine, canopy, larval habitat explanatorial regressors employing a sequence of noisy, decomposed, canopy endmember, autocorrelation observations in accordance with the framework of the Kalman filter in ArcGIS. This meant specifying the following matrices: Fk, the state-transition model; Hk, the observation model; Qk, the covariance of the process noise; Rk, the covariance of the observation noise; and Bk, the control-input model, for each timestep, k,

Ellipses representing the seasonally eco productive, canopied, *S. damnosum* s.l., shaded, larval habitat, multivariate, normal distributions with the mean and covariance matrix were enclosed in ArcGIS. Unenclosed values were then qualitatively regressively quantitated as vectors. In the simple case, the various matrices were constant with time, and thus the subscripts were dropped, but the Kalman filter allowed any of the decomposed, riverine, larval habitat, sub-meter resolution, canopied, endmember biosignature, optical properties to change at each time step. The Kalman filter model assumed the true state at time k which was evolved from the state at (k − 1) according to where Fkwas the state transition model which was subsequently applied to the previous state xk−1, where Bk was the control-input model which was applied to the control vector uk. In the geo-spatiotemporally, geosampled, S. damnosum s.l., shaded, larval habitat, , spatially,probabilistically, regressed model wk was the process noise which the authors assumed to be drawn from a zero mean, multivariate, normal distribution with covariance At time k then a decomposed, georefernced canopied, *S. damnosum* s.l., riverine larval habitat canopied endmember observation zk of the true state xk was parameterized according to where Hk was the decomposed observation model which mapped the true state space into the observed space when vk was the observation noise which Jacob et al. (2015) assumed to be zero mean Gaussian white noise with covariance in a balsk fly ovispoition, eco-epidemiological, forecast, vulnerability model. The initial state, and the noise vectors at each step {x0, w1, …, wk, v1 … vk} were all then deemed to be mutually independent.

In all of the algorithmic estimation methods, the original decomposed, georefernced, canopied, *S. damnosum* s.l., shaded, larval habitat endmember data were transformed by the inverse of the Cholesky root of V in ArcGIS. Let L denote the Cholesky root of V then, V = LL' with L lower triangular [Hazewinkle 2001. For the AR(m) eco-epidemiological, decomposed, forecasting, operationizable, emissivity transmisstance, reflectance, wavelenght, risk model, L-1 was a band diagonal matrix with m anomalous rows at the beginning and the autoregressive unmixed parameters along the remaining rows. Therefore, if there were no missing values, after the first m-1 canopy endmember observations the regressed data were transformed as . The transformation was carried out employing the Kalman filter, and the lower triangular matrix L which was never directly computed in ArcGIS. Although L was not computed explicitly, for ease of residual presentation the uncertainty probabilisticforecasts were spatially defined in termsof L. If there are missing values, then the submatrix of L consisting of the rows and columns with nonmissing values are used to generate the transformations (www.esri.com).

The ULS and ML estimates employed a Gauss-Newton algorithm to minimize the sum of squares and maximize the log-likelihood, respectively. The relevant optimization wasperformed in ArcGIS simultaneously for both the regression and AR parameters. The OLS estimates of β and the Yule-Walker estimates of φ were employed as starting values for the uncertainty, first-order auto-evaluation. The Gauss-Newton algorithm requires the derivatives of e or (L)1/N e with respect to the decomposed, explanatorial, canopied parameters [2]. The derivatives with respect to the parameter vector β were then The derivatives with respect to φ were then computed by differentiating the Kalman filter recurrences and the equations for the initial conditions. For the Yule-Walker method, the estimate of the error variance, S2, was the

error sum of squares from the application of GLS, divided by the error degrees of freedom (i.e, number of decomposed, canopied, S. damnosum s.l., shaded, larval habitat endmember observations N minus the number of free parameters). The variance-covariance matrix for the components of b was taken as for the Yule-Walker method. For the ULS and ML methods, the variance-covariance matrix of the canopied, decomposed, larval habitat, parameter estimates was then computed as For the ULS method, J was the matrix of the geospectral derivatives of e with respect to the unmixed, riverine, canopied, larval habitatdecomposable parameters. For the ML method, Jwasthe matrix of derivatives of e divided by . The estimate of the variancecovariance matrix of b assuming that ϕ was then

Spatially autocorrelated discontionus canopy species abundance or distribution datasets may generate spatially autocorrelated residuals in generalized linear models (GLMs) thus, a broader modelling framework may be required to remotely, qualitatively, regressively, quantitate geospectrally decomposed, *S. damnosum*.s.l., larval habitat, canopy bisignatures. Auto-logistic and related auto-models, implemented approximately as autocovariate regression, provide simple and direct modeling of endmember, spatialized, probabilistic *S. damnosum* s.l.- related population processes. However, Dormann questioned the validity of auto-logistic regression for fully observed decomposed endmember data, giving examples of apparent underestimation of covariate parameter estimators in residual analysis of simulated data. Dormann et al. extended this critique to auto-Poisson and certain autonormal models, finding again that autocovariate-regressed endmember estimates for time series covariate parameter estimators bore little resemblance to values employed to generate 'snouter' data. Jacob et al. [2015] acclaimed that compound probabilistic regression uncertainties associated may be associated other factors (e.g., sampling error) such as algorithm selection, presence data, and variable collinearity. We note that all the above studies employed neighborhood weighting schemes inconsistent with auto-model definitions; in the auto-Poisson case, a further inconsistency was the failure to exclude cooperative interactions. Investigating the impact of implementation errors on auto-model probabilistic estimation employing both empirical, and simulated datasets of geospectrally decomposed, resampled, productive, shade canopied, S. damnosumsol s.l., riverine, larval habitats may show that when spatially "re-adjusted" endmember canopy data are re-analyzed employing valid weightings, very different residually forecasted, emissivity, transmisstance, reflectance wavelengtht, estimates are obtained for photosynthetic and NPV predictors. For auto-logistic and auto-normal S. damnosum s.l. riverine larval habitate co-epidemiological, forecasting canopy, risk models, the new estimates may agree closely with values used to generate the 'snouter' simulations. A substantial fraction of papers employing auto-logistic regression use these invalid neighborhood weightings, which have been embedded as default options in ArcGIS.

The Poisson process in the vulnerability diagnostic analyses was provided by the limit of a binomial distribution of the sampled district-level explanatory predictor covariate coefficient

$$P_p\left(n \mid N\right) = \frac{N!}{n!\left(N-n\right)!}\, p^n \left(1 - p\right)^{N-n}.$$

estimates using   The authors viewed the distribution as a function of the expected number of count variables using the sample size *N* for quantifying the fixed *p* in equation (2.1), which was then transformed into the linear equation:

$$P_{\frac{v}{N}}(n|N) = \frac{N!}{n!(N-n)!}\left(\frac{v}{N}\right)^n \left(1-\frac{v}{N}\right)^{N-n}.$$

Based on the N, the distribution approached $P_v(n)$ was

$$\lim_{n\to\infty} P_p(n|N) = \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{n!}\frac{v^n}{N^n}\left(1-\frac{v}{N}\right)^N\left(1-\frac{v}{N}\right)^{-n} = \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{N^n}\frac{v^n}{n!}\left(1-\frac{v}{N}\right)^N\left(1-\frac{v}{N}\right)^{-n} = 1\cdot\frac{v^n}{n!}.$$

$$e^{-v}\cdot 1 = \frac{v^n e^{-v}}{n!}.$$

The GENMOD procedure then fit a generalized linear model (GLM) to the geosampled geo-spectrotemporal, entomological datasets by maximum likelihood estimation of the parameter vector β in the onchocerciasis, predictive model. The GENMOD procedure optimally estimated the seasonal-geosampled parameters of each georferenceable, district-level time series, onchocerciasis risk model numerically through an iterative fitting process. The dispersion parameter was then estimated by the residual deviance and by Pearson's chi-square divided by the degrees of freedom (d.f.). Covariances, standard errors, and *p*-values were then computed for the uncoalesced LULC covariate coefficients based on the asymptotic normality derived from the maximum likelihood estimation.

Note, that the sample size N completely dropped out of the probability function, which in Jacob et al. (2015) had the same functional form for all the geosampled, explanative, district-level, parameter estimator, time series, indicator values (i.e., v). As expected, the Poisson distribution was normalized so that the sum of probabilities equaled 1. The ratio of probabilities was then optimally determined by $\sum_{n=0}^{\infty} P_v(n) = e^{-v}\sum_{n=0}^{\infty}\frac{v^n}{n!} = e^{-v}e^v =$ which

$$\frac{P_v(n=i+1)}{P(n=i)} = \frac{\frac{v^{i+1}e^{-v}}{(i+1)!}}{\frac{e^{-v}v^i}{i!}} = \frac{v}{i+1}.$$

was then subsequently expressed as

The Poisson distribution revealed that the district-level, geo-spectrotemporally geosampled, explanatory covariate coefficients reached a maximum when $\frac{dP_v(n)}{dn} = \frac{e^{-v}n(\gamma - H_x + \ln v)}{n!} = 0,$ where γ was the Euler-Mascheroni constant and H_n was a harmonic number, leading to the transcendental equation $\gamma - H_n + \ln v = 0,$ for each district in the riverine tributray eco-epidemiological, study site. The regression model revealed that the Euler-Mascheroni constant arose in the integrals as

$$\gamma = -\int_0^\infty e^{-x}\ln x\,dx = -\int_0^1 \ln\ln\left(\frac{1}{x}\right)dx = \int_0^\infty\left(\frac{1}{1-e^{-x}} - \frac{1}{x}\right)e^{-x}\,dx = \int_0^\infty\frac{1}{x}\left(\frac{1}{1+x} - e^{-x}\right)dx$$

.Commonly, integrals that render γ in combination with geo-spectrotemporal geosampled constants include

$$\int_0^\infty e^{-x^2}\ln x\,dx = -\frac{1}{4}\sqrt{\pi}(\gamma - 2\ln 2)$$

which is equal to

$$\int_0^\infty e^{-x}(\ln x)^2\,dx = \gamma^2 + \frac{1}{6}\pi^2\ [2].$$

(Haight 1967). Thereafter, the double integrals in the district-level, seasonal, eco-epidemiological, *S. damnosum* s.l. regression model included

$$\gamma = \int_0^1 \int_0^1 \frac{x-1}{(1-xy)\ln(xy)}\,dx\,dy.$$

[1.2]. An interesting analog of equation (1.2) in the regression-based paradigm was then calculated as

$$\ln\left(\frac{4}{\pi}\right) = \sum_{n=1}^{\infty}(-1)^{n-1}\left[\frac{1}{n}-\ln\frac{n+1}{n}\right] = \int_0^1 \int_0^1 \frac{x-1}{(1+xy)\ln(xy)}\,dx\,dy = 0.241564...\gamma$$

. This solution was also

$$e^{\gamma} = \lim_{n\to\infty}\frac{1}{\ln p_n}\prod_{i=1}^{n}\frac{1}{1-\frac{1}{p_i}}$$

provided by incorporating Mertens theorem[i.e., $\qquad$ where the product was aggregated over the district-level, geosampled trailing vegation, *S. damnosum* s.l. immature, habitat, capture point, regressable values found in the empirical, eco-epidmiological, datasets.

Mertens' 3rd theorem: $\displaystyle\lim_{n\to\infty}\ln n\prod_{p\le n}\left(1-\frac{1}{p}\right)=e^{-\gamma}$ is related to the density of prime numbers where γ is the Euler–Mascheroni constant[Haight 1967].By taking the logarithm of both sides in the vulnerability, eco-epidemiological, forecast, sub-meter resolution model, an explicit formula

for γ was then optimally derived employing $\displaystyle\gamma = \lim_{x\to\infty}\left[\sum_{p\le x}\ln\left(\frac{1}{1-\frac{1}{p}}\right)-\ln\ln x\right]$. This expression was also rendered coincidently by quantifying the data series employing Euler, and equation 1.2 by first

replacing $\ln n$, $\ln(n+1)$, in the equation $\displaystyle\gamma = \sum_{k=1}^{\infty}\left[\frac{1}{k}-\ln\left(1+\frac{1}{k}\right)\right]$ and then generating

$$\lim_{n\to\infty}\left[\ln(n+1)-\ln n\right]=\lim_{n\to\infty}\ln\left(1+\frac{1}{n}\right)=0$$

. The authors in Jacob et al. [2015] then substituted the

telescoping sum $\displaystyle\sum_{k=1}^{n}\ln\left(1+\frac{1}{k}\right)$ for $\ln(n+1)$ which then generated $\ln\left(1+\frac{1}{k}\right)=\ln(k+1)-\ln k$. In mathematics, a telescoping series is a series whose partial sums eventually only have a fixed number of terms after cancellation }Thomson and Andrew M. Bruckner,2008}. Thereafter, the

product was $\displaystyle\lim_{n\to\infty}\left[\sum_{k=1}^{n}\frac{1}{k}-\sum_{k=1}^{n}\ln\left(1+\frac{1}{k}\right)\right]_{\gamma} = \lim_{n\to\infty}\sum_{k=1}^{n}\left[\frac{1}{k}-\ln\left(1+\frac{1}{k}\right)\right]$.

The objective of Jacob et al. (2013) was to optimally determine seasonal geo-spectrotemporal, regressive, immature habitat explanators (e.g., rainfall) related to monthly, district-level, prevalence rates. Specific prevalence measures were optimally forecasted using autoregressive, spatial filter, eigenvector specifications and data collections in ArcGIS and SAS for targeting districts that had higher prevalence rates. Initially, case, as counts, were employed as a response variable in a Poisson probability model framework in PROC REG for quantifying eco-epidemiological, optimizable, datasets of district-level covariates (e.g., distribution of health centers) geosampled from 2006 to 2010 in Uganda.

A discrete stochastic variable X is said to have a Poisson distribution with parameter $\lambda > 0$, if $k = 0, 1, 2$, while the probability mass function of X is rendered by: $f(k; \lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$ where e is the base of the natural logarithm ($e = 2.71828...$) and $k!$ is the factorial of k[Hosmer and Lemeshew 2002]. The mode of the malaria-related, Poisson-distributed, district-level, geo-spectrotemporally geosampled, sub-meter resolution, predictive variable in Jacob et al (2013) with a non-integer λ was equal to $\lfloor \lambda \rfloor$, which represented the largest integer less than or equal to λ in the model. This was written as floor (λ) in the model. Floor(x) = is the largest integer less than or equal to x and ceiling(x) = is the smallest integer greater than or equal to x [Fotheringham 2002]. In so doing, the floor function $\lfloor x \rfloor$ was the greatest integer function or integer value generating the largest integer less than or equal to x in the explanative, georeferenceable, time series, heuristically optimizable, explicative, residual forecasts remotely targeting statistically significant, district-level, eco-epidemiological sub-meter resolution, LULC covariates. The floor and ceiling functions then mapped a field-geosampled malarial-related, parameterizable, elucidative, covariate coefficient value to the largest previous where floor(x) = $\lfloor x \rfloor$ and was the largest integer not greater than x and ceiling(x) = $\lceil x \rceil$ was the smallest integer not less than x[Haight 1967]. Since λ was a positive integer in the geo-spectrotemporal, geosampled, district-level, regression-based, malarial, eco-epidemiological, vulnerability, forecasting model, the modes were λ and $\lambda - 1$. Hence, all of the cumulants of the Poisson distribution in the malarial model was equal to the expected value λ calculated at each geosampled district-level geolocation at the Ugandan study site.

Additionally, the explanatorial, time series, heuristically parameterizable, sub-meter resolution, predictor, covariate coefficient of variation in the Poisson-specified, geospatiotemporal, malaria-related, regression model was $\lambda^{-1/2}$ while the index of dispersion was 1. Thereafter, the mean deviation was explicatively geo-spectrotemporally quantitated about the mean in the malarial model which then expressed $E|X - l| = 2\exp(-1)\frac{l^l}{\lfloor l \rfloor!}$ for optimally determining statistical significance of the geosampled estimators.

Further, other series in for the Mwea irrigated, ricefield agro-ecosystem in Central Kenya in Jacob and Novak (2014) included the equation ($\diamondsuit$) where $\gamma = \sum_{n=2}^{\infty} (-1)^n \frac{\zeta(n)}{n} = \ln\left(\frac{4}{\pi}\right) + \sum_{n=1}^{\infty} \frac{(-1)^{n-1}\zeta(n+1)}{2^n(n+1)}$ where $\zeta(z)$ was $\gamma = \sum_{n=1}^{\infty} (-1)^n \frac{[\lg n]}{n}$ plus the Riemann zeta function for constructing a geo-spatiotemporal, explanative, eco-epidmiological, malaria-related, *Anopheline arabiensis* district-level, LULC , sub-meter resolution, (i.e., panchromatic Quickbird), forecasting vulnerability, regression model for forecasting hypeproductive habitats in a ricevillage ecosysem in central Kenya (Karima agro-village complex) The Riemann zeta function ζ(s) is a function of a complex variable$s$ that analytically continues the sum of the infinite series $\sum_{n=1}^{\infty} \frac{1}{n^s}$ which converged in the model when the real part of s was greater than 1 where lg was the logarithm to base 2 and the $\lfloor x \rfloor$ is the floor

function(Hosmer and Lemeshew 2002). Nielsen(1897) earlier provided a series equivalent to

$$\gamma = 1 - \sum_{n=1}^{\infty} \sum_{k=2^{n-1}}^{2^n-1} \frac{n}{(2k+1)(2k+2)}$$ and, thereafter $\frac{1}{(2k+1)(2k+2)} = \frac{1}{2k+1} - \frac{1}{2k+2}$ which was then added to

$$0 = -\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$$ to render Vacca's formula. The sums in the *An. arabiensis* model then was

$$\sum_{n=1}^{\infty} \sum_{k=2^n}^{\infty} \frac{(-1)^k}{k} \sum_{k=1}^{\infty} \frac{1}{2^{k+1}} \sum_{j=0}^{k-1} \left(2^{k-j} + j \atop j \right)^{-1}$$ Jacob and Novak(2104) employed the

sums $$\gamma = \sum_{n=1}^{\infty} \sum_{k=2^n}^{\infty} \frac{(-1)^k}{k} = \sum_{k=1}^{\infty} \frac{1}{2^{k+1}} \sum_{j=0}^{k-1} \left(2^{k-j} + j \atop j \right)^{-1}$$ with $k - j$ by replacing the undefined *I* and then rewrote the equation as a double series for applying the Euler's series transformation to each of the sampled time-series dependent, explanatorial, geo-spectrotemporal, sub-meter resolution, LULC, rice field, geo-spectrotemporal, geospatial, explanatorial,*An. arabienis*, grid-stratified, covariate coefficient, regression estimates. Given a convergent alternating series with sum

$$S = \sum_{k=0}^{\infty} (-1)^k a_k,$$ Abramowitz and Stegun (1972, p. 16) define Euler's transformation as

$$S = \sum_{k=0}^{\infty} \frac{(-1)^k \Delta^k a_0}{2^{k+1}},$$ where Δ is the forward difference operator $\Delta^k a_0 = \sum_{m=0}^{k} (-1)^m \left(k \atop m \right) a_{k-m}$ and $\left(k \atop m \right)$ is a binomial coefficient.

An alternate formulation was provided in Jacob and Novak (2014) which defined the vulnerability, geo-spectrotemporal, geosampled, eco-epidemiological, eco-georeferenceable, *An. arabiensis*, forecasting, sub-meter resolution, LULC regression, vulnerability model transformation as $$S = \sum_{k=0}^{\infty} \frac{\nabla^k a_0}{2^{k+1}},$$ where ∇ wa the backward difference operator $\nabla^k a_0 \equiv \sum_{m=0}^{k} (-1)^m \left(k \atop m \right) a_m.$ Jacob and Novak (2014) gives examples of different types of convergence behavior upon application of the transformation: $\sum_{n=0}^{\infty} \frac{(-1)^n}{2^n} = \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{4^n}$ in PROC REG for rendering faster convergence, of geo-spectrotemporally geosampled, riceland agroecosystem, sub-meter resolution lineraizable ,time series,prognosticative LULC covariates , employing$\sum_{n=0}^{\infty} \frac{(-1)^n}{3^n} = \frac{1}{2} \sum_{n=0}^{\infty} \frac{1}{3^n}$ which can provide the same rate of convergence, and $\sum_{n=0}^{\infty} \frac{(-1)^n}{4^n} = \frac{1}{2} \sum_{n=0}^{\infty} \left(\frac{3}{8} \right)^n$ Euler's formula, is a mathematical formula in complex analysis that establishes the fundamental relationship between the trigonometric functions and the complex exponential function (Haight 1967).

Euler's formula, is a mathematical formula in complex analysis that establishes the fundamental relationship between the trigonometric functions and the complex exponential

function. Euler's formula states that, for any real number x:where e is the base of the natural logarithm, i is the imaginary unit, and cos and sin are the explanativem trigonometric functions cosine and sine respectively, with the argument x given in radians (Cressie 1993) . This complex exponential functionmay be edentoable in a YF, *Ae.aegypti,* ovispoition, sub-meter resolutiuon, eco-epidemiological, LULC, sub-mtere resolution, African, riceland, capture point, discontinuous,forest- canopied model employing x ("cosine plus i sine"). The formula would be still even if x is a complex geo-spectrotemrpoal or geo-spatiotemrpoal, eco-epidemiological, log-transformed, discrete, finite, integer values (e.g., montly YF case distribution)

The Euler formula, sometimes also called the Euler identity (e.g., Trott 2004, p. 174), states $e^{ix} = \cos x + i \sin x$, where i is the imaginary unit. Note that Euler's polyhedral formula is sometimes also called the Euler formula, as is the Euler curvature formula. The equivalent expression $ix = \ln(\cos x + i \sin x)$ had previously been published by Cotes (1714). The special case of the formula with $x = \pi$ gives the beautiful identity $e^{i\pi} + 1 = 0$, an equation connecting the fundamental numbers i, pi, e, 1, and 0 (zero), the fundamental operations $+$, $\times$, and exponentiation, the most important relation $=$, and nothing else. Gauss is reported to have commented that if this formula was not immediately obvious, the reader would never be a first-class mathematician (Derbyshire 2004, p. 202).

$$e^{if} = x + iy$$

Interestingly, if may tabulate an empirical orthogonal, sub-meter resolution dataset of uncoalesced, *Ae egypti* riceland habitats along the periphery of these ecosystems for quantiate geoclassified LULCs that have protuded geospatially onto discontinuous, forest-canopied, LULCs in a forecast vulbnerability, remote sesning model by replacing i with $-i$. This may be because $-i$ is as good a square root of -1 as i.,(Neter 1990). So an arbovirologist, medical entomologist or YF experimenter may generate a riceland, African, ecosystem, ovisposition *Ae.*

aegypti risk model as $1 = e^{if}e^{-if} = (x - iy)(x + iy) = x^2 + y^2 = |e^{if}|^2$ so that $|e^{if}| = 1$

Figure 1.1. Euler's formula, for any, parameterizable, Poissonian variable using $x: e^{ix} = \cos x + i \sin x$ **where *e* is the base of the natural logarithm, *i* is the imaginary unit, and cos and sin are the trigonometric functions cosine and sine respectively, with the argument *x* given in radians**
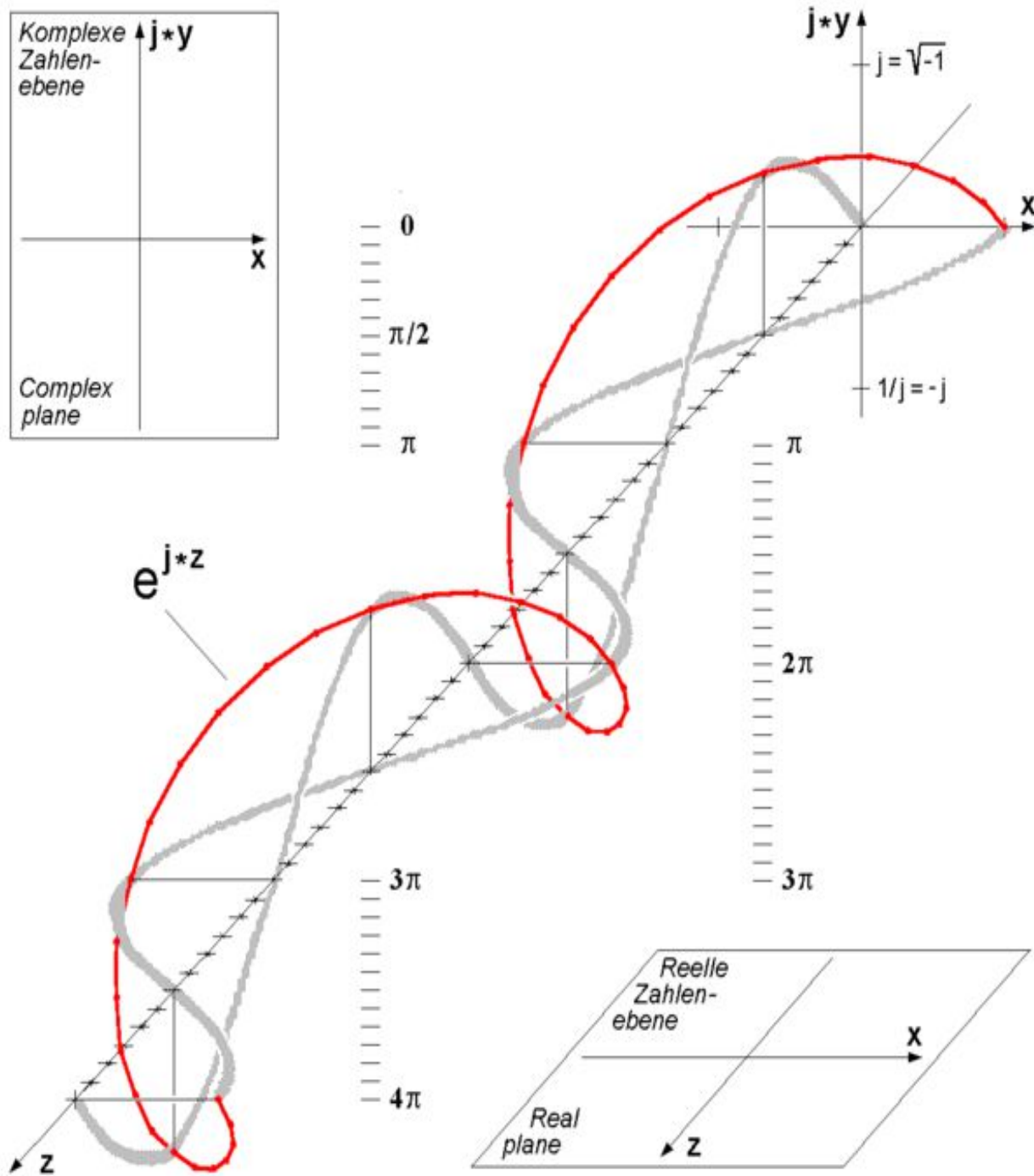
So if $e^{i\psi} = \cos\psi + i\sin\psi$ for some angle ψ in an uncoalesced, frequency, wavelength, geo-spectrotemporal or geo-spatiotemrpoal, eco-epidemiological, sub-meter resolution, grid-startfield, sub-meter resolution, LULC model , $\psi = \theta$ could occur in the vulnerability foreasts(e.g., , newly transitioned discontinuous, forest-canopied,, eco-epidemiological capture point into a seasonal, , hyperproductive African ricland *Ae .aegypti* oviposition site).Recall that the exponential function e^x is well approximated by the linear function $1+x$ when x is very small in any entomological, geo-spatioemporal, eco-epidemiological, eco-georferenceable, sub-mter resolution entomological, vector arthropod, prognosticative ovispoition probabilisc paradigm (see Jacab et al. 2009). A arbovirologist, medical entomologist or YF experimenter may hence assume that the function e^x retains this property for complex z with small modulus. $e^z \approx 1 + z$ for small $|z|$ so in particular $e^{i\theta} \approx 1 + i\theta$ for small θ .Thus, an arbovirologist, medical entomologist or YF experimenter could employ two more approximations which may be smallgeocampled paramterizable , ovispoition LULC sampled values of θ. ((e.g., $\cos\theta \approx 1$ or $\sin\theta \approx \theta$ These approximations may become increasingly more precise as $|\theta|$ decreases which may allow the Euler formula to holds when $\theta = 0$. Putting all these processes together, a sub-meter resolution, *Ae egypti*, ovipsoition LULC, ,model may be expressed as $e^{i\theta} = \cos\theta + i\sin\theta$ if θ is sufficiently small .,Furthermore by de Moivre's formula $e^{in\theta} = (\cos\theta + i\sin\theta)^n = \cos(n\theta) + i\sin(n\theta)$ may also exist in a dataset of eco-georferenceable, sub-meter resolution, *Ae egypti,* ovispoition discontinuous, forest-canopied, De Moivre's Theorem states that for every real number θ and every positive integer n, we have$(\cos θ + i \sin θ)^n = \cos nθ + i \sin n$

To see why the Euler transformation work in a forecasting, time series, eco-epidemiolgical, eco-georferenecable, YF, sub-meter resolution, LULC, regression, probabilistic,vulnerability, endmember, optimizable, geo-spectrotemporal, paradigm an arbovirologist, medical entomologist or other experimenter may, consider convention for difference operator and write $S = u_0 - u_1 + u_2 - \ldots = \frac{1}{2}u_0 + \frac{1}{2}\left[(u_0 - u_1) - (u_1 - u_2) + (u_2 - u_3) - \ldots\right]$. in order to optimally quantitate seasonal endmember, hyperporductive discontinuous, eco-georeferenceable, *Ae egypti* ovispition sites on sub-meter resolution geoclassified LULCs and their wavelength, transmittance gid-stratied frequencies in an African, riceleand, expandng, agro-irrigated, village ecosystem. Optimally the process may be repeated in SAS on the series in brackets to quantitatively, qualitatively, geo-spatiotemporallyy or geo-spectrotemporally obtain $S = \frac{1}{2}u_0 + \frac{1}{4}(u_0 - u_1) + \frac{1}{4}\left[(u_0 - 2u_1 + u_2) - (u_1 - 2u_2 + u_3) + (u_2 - 2u_3 + u_4) - \ldots\right]$, in a robust eco-epidemiological, YF , forecast, vulnerability, model In so doing ,sub-meter resolution discontinuous, sub-mter resolution, geoclassifiable, seasonal, forest-canopy to rice agriculture *Ae egypti* , hyperproductive, LULC, eco-georferenceable, oviposition sites may be iteratively, interpolatively identified. Each finite step in the derivation, in the regressed diagnostic,oviposition, capture point, hyperproductive, YFV-related time series, geosampled, geoclassified, clinical, field or remote geo-spectrotemporally geosampled, ecogeorefereceable

residual, sub-meter resolution, parameterizable LULC forecast space would "continue to infinity" without use of a limit.

In Jacob and Novak. (2014) used a binomial coefficient, rearranged to achieve the conditionally convergent series in a geo-spatiotemporal, forecast, vulnerability, uncoalesced, sub-meter resolution, endmember, wavelength district-level, malaria-related linear, *An. arabiensis* regression specified, frequency   model for precisely predicting post-tillering, agro-irrigated, hyperproductive, seasonal habitats had the highest immature productivity seasaonl counts. Tillering: extends from the appearance of the first tiller until the maximum tiller (5 - 9) number is reached in Mwea ricelands (Mwangangi et al. 2007). Jacob and Novak (2014) constructed multiple PROC REG regression models for comparing transmission wavelength frequency transmittance intensity at the gris-stratified, sub-meter resolution LULC levels for the Karima eco-epidemiological study site which extended 2 kilometers from the poltical boundary of the intervention study village. The regression analyses assumed independent counts (i.e., n_i), taken at the sampled anopheline larval habitat locations $i = 1, 2, L, n$, where each of the explanatory predictor endmeber wavelength covariates represented in the linear endmember geometric framework was from a Poisson distribution. The Poissonian regression assumed the response explanatory entomological sub-meter resoluton, grid-startified, LULC variable Y had a Poisson distribution and assumed the logarithm of its expected value was modeled by a linear combination of the geosampled, eco-epidemiological, capture point larval habitat sampled parameterized covariate   estimates. This expression was written more compactly as $\log\left(E\left(Y|x\right)\right)$ where x was an n + 1-dimensional vector consisting of n independent variables concatenated to 1 and θ, which Jacob and Novak (2014) simply expressed as a linearly linked to b. Hence, in the Poissonian model θ was an input vector x and, the predicted mean of the associated oviposition seasonal, hypeproductive foci distribution was given by $E\left(Y|x\right) = a^{\theta'x}$, if x Î R^n was a vector of the sampled independent explanatory, sub-meter resolution, geoclassified, LULC variables. Thereafter, the Poissonian model took the from $\log\left(E\left(Y|x\right)\right) = a'x + b$ where a Î R^n and b Î R. Positing salient estimators using the maximization of an auto-Gaussian log-likelihood function and a set of eigenvectors where lambda was a subspace of R^n may identify important explanatory predictor covariates associated to productive habitats based on geo-spectrotemporal, geo-spatiotemporal field-geosampled entomological count *Ae. aegypti* data in a a seasonal, hyperproductive, YF eco-epidemiological, forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model.

After the regression, Jacob and Novak (2014) attempted to ascertain whether the proportions of the geosampled *An. arabiensis,* grid-stratified, immature data differed by QuickBird,digitized grid cells in the riceland agro-village, eco-georeferenceable, study site. The model assumed a random sample between Y_i, (i.e., geosampled, capture point, seasonal, hyperproductive, immature habitat, total larval count, density data), and the regressors X_{i1}, L, X_{ip}. A disturbance term, ε_i, which was a random variable, in the forecast vulnerability endmeber analyses was added to this assumed relationship to optimally capture the

136

influence of all the parameterizable endmember, sub-mter resolution, covariate estimators geosampled on Y_i and on variables other than X_{i1}, L, X_{ip}. The random error term, ε, in the regression models was assumed to be normally distributed with mean zero and variance σ^2. Statistical characteristics of the geosampled immature anopheline LULC data were examined in PROC UNIVARIATE. The PLOT option in the PROC UNIVARAITE statement generated histograms and boxplots. The NORMAL option was employed to test whether the clinical, field and remote-geosampled explanatorial, endemic transmission-oriented, uncoalesced, LULC covariate coefficients had a normal distribution. The regression equation for optimally quantitating the riceland, African, eco-epidemiological, eco-epidemiological, study site was: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_1 X_{2i}, i = 1, 2, L, n$. It was important to distinguish the geo-spatiotemporal-sampled *An. arabiensis* aquatic larval habitat, remotey sensed LULC models in terms of random variables and the observed values of the random variables. Thus, Jacob and Novak (2014) geo-spectrotemporally determined p + 1 parameters β_0, L, β_p in an eco-cartographic, geostatistical cyberenvironment. In order to estimate the sampled parameters, it was also useful to employ the matrix notation $Y = X\beta + \varepsilon$, where Y was a column vector that included the geo-spatiotemporal-sampled endmember LULC, esub-mter resolution, *An. arabiensis* aquatic larval habitat count values of Y_1, L, Y_n, geosampled in the study site, which in Jacob and Novak (2014) included the unobserved stochastic endmember, geoclassified, LULC components $\varepsilon_1, L, \varepsilon_n$ and the matrix X. This matrix was the observed larval habitat parame-

$$X = \begin{pmatrix} 1 & x_{1i} & L & x_{1p} \\ 1 & x_{2i} & L & x_{2p} \\ M & M & O & M \\ 1 & x_{ni} & L & x_{np} \end{pmatrix}$$

ter values of the explanatorial, uncoalesced, regressors may be expressed as in a a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model.

Thereafter, the sampled explanatory predictor covariate coefficient values were log-transformed to normalize the distribution and minimize standard error in the residual, *An. arabiensis,* aquatic, larval habitat, eco-georeferenecable, capture point forecasts. Multicollinearity diagnostics from the COLLIN option in SAS were estimated. Residual-based diagnostics for univariate and multivariate conditional heteroscedastic endmember, wavelength frequency, LULC models previously constructed from clustering clinical, field and remote geosampled vector insect habitat parameter estimates have revealed that errors in variance, probability, endmember, uncertainty estimation can substantially alter numerical predictions models due to multicollinearity (Griffith 2005, Jacob et al. 2009) . In Jacob and Novak (2014), the SAS COLLIN option rendered eigenvalues and condition index, as well as proportions of variances with respect to individual-sampled , sub-meter resolution, *An. arabiensis,* aquatic, larval habitat, explanatory, predictor covariates in the LULV geo-spctrotemporal, forecast, vulnerability models. The conditional index scores indicated no significant multicollinearity in either model output. The same quantiation of non-independent explanators may be identified in a

in a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model.

Extra-Poisson variation was detected in the sub-meter resolution, diagnostic, residual variance estimates of both *An. arabiensis* aquatic larval habitat, eco-epidemiological, ecogeoreferenecable, LULC, sub-meter resolution, endmember models. Extra-binomial (i.e., extra-Poisson) variation occurs when discrete data comes in the form of counts or proportion that display greater variability than would be predicted when fitting a model (Haight 1967). When eco-epidemiological, geosampled vector, mosquito, aquatic, larval habitat, uncoalesced, remotely sensed regression, geoclassified LULC data are overdispersed, the square root and logarithmic transformations may be less effective at making the mean and variance independent (see Jacob et al. 2005). As such, Jacob and Novak (2014) employed a negative binomial regression with a gamma distributed non-homogenous mean constructed in PROC GENMOD to account for the overdispersion in both the geo-spatiotemporal, eco-epidemiological, prognostication *An. arabiensis* larval habitat models. The negative binomial distribution arises as a continuous mixture of Poisson distributions in a vector mosquito larval habitat distribution model where the mixing distribution of the Poisson rate is a gamma distribution (Haight 1967). The negative binomial model is a quadratic function of the mean and the variance which commonly affect the weights in an iteratively weighted least-squares algorithm for fitting geospectiotemporal-sampled, vector, mosquito, aquatic larval, habitat data (Hosmer and Lemeshew 2000). Jacob and Novak (2014) employed the negative binomial regression model with a non-homogenous gamma distributed mean to linearly adjust the geosampled sub-mter resolution LULC data in the Karima, rice agro-village, complex, eco-epidemiological, study site. The negative binomial distribution [i.e., NB(r, p)] was represented as a compound Poisson distribution by letting Y_n, $n \in N_0$ denote a sequence of the independent and identically distributed, aquatic, larval, habitat, *An. arabiensis,* randomized, optimizable, geoclassifiable, grid-stratified, oviposition, LULC variables, each one having the logarithmic distribution Log(p). The negative binomial for a linear vector entomological prognosticative endmember YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC , oviposition , discontinuous, forest-canopied, African riceland model. LULC model may be a Poisson (λ) distribution, where λ is itself a random variable, distributed according to Gamma(r, p/(1 − p) .

In Jacob et al. (2013), the Bessel function $J_n(z)$ was defined by the contour integral

$$J_n(z) = \frac{1}{2\pi i} \oint e^{(z/2)(t-1/t)} t^{-n-1} dt$$

where the contour enclosed the origin and was traversed in a counter-clockwise direction in a. S. damnosum s.l. habitat , sub-mter resolution, oviposition, forecast, vulnerability, endmember model. This function generated:

$$J_0\left(2i\sqrt{z}\right) = \frac{1}{\pi} \int_0^\pi e^{(1+z)\cos\theta} \cos\left[(1-z)\sin\theta\right] d\theta \quad z \equiv 1 - z' \text{ and } \quad y \equiv 1 + z'.$$

In mathematics, Bessel functions are canonical solutions y(x) of Bessel's differential equation:

$$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + (x^2 - \alpha^2)y = 0$$

for an arbitrary real or complex number α (i.e., the order of the Bessel function); the most common and important cases are for α an integer or half-integer (Hosmer and Lemeshew 2002).   The Bessel functions of the first kind were defined as the

solutions to the Bessel differential equation: $$x^2\frac{d^2y}{dx^2} + x\frac{dy}{dx} + \left(x^2 - n^2\right)y = 0$$ in the immature habitat, LULC model which were non-singular at the origin. The n-th order Bessel function of x

then was: $$J_n(x) = \sum_{k=0}^{\infty}\frac{(-1)^k}{k!\Gamma(n+k+1)}\left(\frac{x}{2}\right)^{n+2k}$$ where: $$\Gamma(n+k+1) = \int_0^{\infty}e^{-k}x^{n+k}dx$$ was the Gamma function. In mathematics, the gamma function (represented by the capital Greek alphabet letter Γ) is an extension of the factorial function, with its argument shifted down by 1, to real and complex numbers

Thereafter, to optimally regressively quantitate the equivalence in the geo-spectrotemporal, explicative, regression-based, orthogonalized, onchocerciasis parameterizable, covariate, sub-meter resolution, LULC estimators, the authors of Jacob et al. (2013) expanded $1/(1+x)$ in a geometric explanatory time series and multiplied the district-level geosampled uncoalesced, oviposition, endmember, data feature attributes by x^{2^n-1}, and integrated the term wise as in Sondow and Zudilin (2006). Other series for γ then

included $$\gamma = \frac{3}{2} - \ln 2 - \sum_{m=2}^{\infty}(-1)^m\frac{m-1}{m}\left[\zeta(m) - 1\right]$$ and $$\gamma = \frac{2^n}{e^{2^n}}\sum_{m=0}^{\infty}\frac{2^{mn}}{(m+1)!}\sum_{t=0}^{m}\frac{1}{t+1} - n\ln 2 + o\left(\frac{1}{2^ne^{2^n}}\right)$$ A rapidly converging limit for γ was then optimally provided by

$$\gamma = \lim_{n\to\infty}\left[\frac{2n-1}{2n} - \ln n + \sum_{k=2}^{n}\left(\frac{1}{k} - \frac{\zeta(1-k)}{n^k}\right)\right] = \lim_{n\to\infty}\left[\frac{2n-1}{2n} - \ln n + \sum_{k=2}^{n}\frac{1}{k}\left(1 + \frac{B_k}{n^k}\right)\right]$$ and

$$\gamma = \lim_{n\to\infty}\left[\frac{2n-1}{2n} - \ln n + \sum_{k=2}^{n}\left(\frac{1}{k} - \frac{\zeta(1-k)}{n^k}\right)\right] = \lim_{n\to\infty}\left[\frac{2n-1}{2n} - \ln n + \sum_{k=2}^{n}\frac{1}{k}\left(1 + \frac{B_k}{n^k}\right)\right]$$ where B_k was a Bernoulli number.   Another limit formula was then provided by the equation

$$\gamma = -\lim_{n\to\infty}\left[\frac{\Gamma\left(\frac{1}{n}\right)\Gamma(n+1)n^{1+1/n}}{\Gamma\left(2+n+\frac{1}{n}\right)} - \frac{n^2}{n+1}\right]$$ ...

In mathematics, the Bernoulli numbers Bn are a sequence of rational numbers with deep connections to number theory, whereby, values of the first few Bernoulli numbers are B0 = 1, B1 = ±1/2, B2 = 1/6, B3 = 0, B4 = −1/30, B5 = 0, B6 = 1/42, B7 = 0, B8 = −1/30 (Hosmer and Lemeshew 2002). The Bernoulli numbers B_n are a sequence of signed rational numbers that can

be defined by the exponential generating function $$\frac{x}{e^x - 1} \equiv \sum_{n=0}^{\infty}\frac{B_n x^n}{n!}.$$ These numbers arise in the

series expansions of trigonometric functions, and are extremely important in number theory and analysis. There are actually two definitions for the Bernoulli numbers. To distinguish them, the Bernoulli numbers as defined in modern usage (National Institute of Standards and Technology convention) are written B_n, while the Bernoulli numbers encountered in older literature are written B_n^* (Gradshteyn and Ryzhik 2000). In each case, the Bernoulli numbers are a special case of the Bernoulli polynomials $B_n(x)$ or $B_n^*(x)$ with $B_n = B_n(0)$ and $B_n^* = B_n^*(0)$. The Bernoulli number and polynomial should not be confused with the Bell numbers and Bell polynomial, which are also commonly denoted B_n and $B_n(x)$, respectively. Bernoulli numbers defined by the modern definition are denoted B_n and sometimes called "even-index" Bernoulli numbers. These are the Bernoulli numbers returned, by example, by the Wolfram Language function BernoulliB[n]. The Bernoulli number B_n may be defined by the contour integral

$$B_n = \frac{n!}{2\pi i} \oint \frac{z}{e^z - 1} \frac{dz}{z^{n+1}},$$

where the contour encloses the origin, has radius less than 2π (to avoid the poles at $\pm 2\pi i$), and is traversed in a counterclockwise direction in a YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-mter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model.

Jacob and Novak (2013) found if m and n are geosampled district-level, eco-epidemiological, georeferenceable, African irrigated rice-agro-village complex, *S. damnosum* s.l. immature, hyperproductive, oviposition, capture point, time series, eigen-decomposable, regressable values and f(x) is a smooth sufficiently robust differentiable explanatorial, diagnostic function in a seasonal, LULC, forecast, vulnerability endmember, sub-meter resolution model framework in an SAS/GIS geodatabase cyberenvironment all the geosampled, geospectrotemporal, iterative interpolative values of x in the interval $[m, n]$, can be based on the integral $I = \int_m^n f(x)\, dx$. Furthermore, this realization can be approximated by the sum (or vice versa) $S = \frac{1}{2}f(m) + f(m+1) + \cdots + f(n-1) + \frac{1}{2}f(n)$. The Euler–Maclaurin formula will subsequently provide optimal optimal LULC, sub-meter resolution, immature, capture point, eco-georeferenceable, eco-epidemiological, habitat expressions for the quantizable tabulated difference between the sum and the integral in terms of the higher derivatives f(k) at the capture points and of the interval m and n in any SAS/GIS, predictive, geospatial, geostatistical paradigm.

In Jacob et al. (2016), the Euler–Maclaurin formula provides a powerful connection between integrals and sums in a empirical dataset of uncoalesced, geosampled, *S. damnosum* s.l. endmember, geo-spectrotemporal, geometric, forecasting, vulnerability model unbiased estimators which was subsequently employed to approximate integrals by finite sums, or conversely to evaluate finite sums and infinite series using integrals and the machinery of calculus. Thereafter, for the explanatorial, district-level time series, geosampled, immature habitat, sub-meter resolution, uncoalesced, iteratively interpolative, geosampled immature, capture point, oviposition habitat values, *p*, had

$$S - I = \sum_{k=2}^{P} \frac{B_k}{k!} \left(f^{(k-1)'}(n) - f^{(k-1)'}(m) \right) + R$$

where B1 = −1/2, B2 = 1/6, B3 = 0, B4 = −1/30, B5 = 0, B6 = 1/42, B7 = 0, B8 = −1/30, and R which was an error term.

Note in Jacob et al. (2016) $-B_1 (f(n) + f(m)) = \frac{1}{2} (f(n) + f(m))$ Hence, the authors re-wrote the regression-based formula as follows:

$$\sum_{i=m}^{n} f(i) = \int_{m}^{n} f(x)\, dx - B_1 (f(n) + f(m)) + \sum_{k=1}^{P} \frac{B_{2k}}{(2k)!} \left(f^{(2k-1)'}(n) - f^{(2k-1)'}(m) \right) + R$$

They then rewrote the equation more elegantly as

$$\sum_{i=m}^{n} f(i) = \sum_{k=0}^{P} \frac{1}{k!} \left(B_k f^{(k-1)'}(n) - B_k^* f^{(k-1)'}(m) \right) + R$$

with the convention of $f^{(-1)'}(x) = \int f(x)\, dx$ (i.e. the -1th derivation of f is the integral of the function). Limits to the explicative, district-level, eco-epidemiological, sub-meter resolution, time series, *S. damnosum* s.l. oviposition LULC, geoclassified, eco-epidemiolgical, forecast, vulnerability, endmember, regression model was then rendered by $\gamma = \lim_{x \to \infty} \zeta(\zeta(z)) - 2^x + \left(\frac{4}{3} \right)^x + 1$ where $\zeta(z)$ was the Riemann zeta function. The Bernoulli numbers appear in the Taylor series expansions of the tangent and hyperbolic tangent functions, in formulas for the sum of powers of the first positive integers, in the Euler–Maclaurin formula and in expressions for certain values of the Riemann zeta function (Haight 1967).

Another connection with the primes in the malaria district-level, optimizable data in Jacob et al (2016) was provided by $d(n) = \sigma_0(n)$ for the district-level, geo-spectrotemporal, explanative, diagnostic, numerical values from 1 to n in the empricial geosampled, ovipsoition LULC, *S. damnosum* s.l. dataset which the authors was found to be asymptotic.

De laValléePoussin[1898] proved that if a large number n is divided by all primes <n , then the average amount by which the quotient is less than the next whole number is g. An identity for g in the district-level, regression-based, eco-epidemiological, YFV,geospectrotemporal, model may then be optimally provided by $\gamma = \frac{S_0(z) - K_0(z)}{I_0(z)} - \ln\left(\frac{1}{2} z \right)$ where $I_0(z)$ is a modified Bessel function of the first kind, $K_0(z)$ is a modified Bessel function of the second kind, and $S_0(z) \equiv \sum_{k=0}^{\infty} \frac{\left(\frac{1}{2} z \right)^{2k} H_k}{(k!)^2}$ where H_n is a harmonic number. For non-integer α, Yα(x) is related to Jα(x) by: $Y_\alpha(x) = \frac{J_\alpha(x) \cos(\alpha \pi) - J_{-\alpha}(x)}{\sin(\alpha \pi)}$. (Griffith 2003). In the case of integer order n, the function is defined by taking the limit as a non-integer α tends to n: $Y_n(x) = \lim_{\alpha \to n} Y_\alpha(x)$ (Haight 1967).

Since the Bessel functions of the second kind, may be denoted by $Y\alpha(x)$ as in Jacob et al. (2016) and by $N\alpha(x)$, which are actually solutions of the Bessel differential equation employing a singularity at the origin (x = 0) , an entomological, YF, geoclassifiable, *Ae egypti,* sub-meter resolution, iteratively interpolative, uncoalesced, LULC biosignature, (i.e, panchromatic QuickBird), , riceland agro-village ovipositon, geometric endmember, forest canopy, discontinuous, immature, eco-georferenecable habitat model may be able to identify forest-canopy foci along African ricalend expanding corridors. An iterative algorithm for g by computing $B_k = \dfrac{B_{k-1}n^2}{k^2}$ $=$ $A_k = \dfrac{1}{k}\left(\dfrac{A_{k-1}n^2}{k}+B_k\right)$ $=$ $U_kV_k = U_{k-1}+A_k$ and $V_k = V_{k-1}+B_k$ $A_0 = -\ln n B_0 = 1 U_0 = A_0$ and $V_0 = 1$ Reformulating   this   identity   optimally   rendered   the   limit

$$\lim_{n\to\infty}\left|\sum_{k=0}^{\infty}\frac{\left(\frac{n^k}{k!}\right)^2 H_k}{\sum_{k=0}^{\infty}\left(\frac{n^k}{k!}\right)^2} - \ln n\right| = \gamma$$

Infinite products involving g also arose from the Barnes G-function using the positive integer n. In mathematics, the Barnes G-function G(z) is a function that is an extension of superfactorials to the complex numbers which is related to the Gamma function[Haight 1967]. In Jacob et al. 2016, this function provided $\prod_{n=1}^{\infty}e^{-1+1/(2n)}\left(1+\dfrac{1}{n}\right)^n = \dfrac{e^{1+\gamma 2}}{\sqrt{2\pi}}$ and also the equation $\prod_{n=1}^{\infty}e^{-2+2/n}\left(1+\dfrac{2}{n}\right)^n = \dfrac{e^{3+2\gamma}}{2\pi}$ for optimally remotely quantitaing the regressively paramterizable *S. damnosum* s.l. , capture point, immature , seasonal eco-georferenecable, hypeproductive habitats The Barnes G-function was then linearly defined in the time-series dependent, explantive, district-level, immature habitat, forecast, vulnerability endmembr LULC, ovip[osiition, geo-spectrotemporal, uncoalesced   model which then generated

$$G(z+1) = (2\pi)^{z/2}\exp\left(-\left(z(z+1)+\gamma z^2\right)/2\right)\times\prod_{n=1}^{\infty}\left|\left(1+\frac{z}{n}\right)^n\exp\left(-z+z^2/(2n)\right)\right|$$

where γ was the Euler–Mascheroni constant, exp(x) = ex, and $\prod$ was capital pi notation. The Euler-Mascheroni constant was then rendered by the expressions $\gamma = -\Gamma'(1) = -\psi_0(1)$ where $\psi_0(x)$ was the digamma   function $\gamma = \lim_{s\to 1}\left[\zeta(s) - \dfrac{1}{s-1}\right]$ and   the   asymmetric   limit   form of $\gamma = \lim_{s\to 1^+}\sum_{n=1}^{\infty}\left(\dfrac{1}{n^s}-\dfrac{1}{s^n}\right)$ and $\gamma = \lim_{x\to\infty}\left[x - \Gamma\left(\dfrac{1}{x}\right)\right]$. In mathematics, the digamma function is defined as the logarithmic derivative of the gamma function: $\psi(x) = \dfrac{d}{dx}\ln\Gamma(x) = \dfrac{\Gamma'(x)}{\Gamma(x)}$ where it is the first of the polygamma functions (Hosmer and Lemeshew 2002). In Jacob et al. (2016) the vulnerability, *S. damnosum* s.l., vulnerability, eco-epidemiolgical, forecasting, sub-meter resolution, immature, eco-georeferenceable capture point, habitat model the digamma function, $\psi 0(x)$ was related to the harmonic numbers in that $\psi(n) = H_{n-1} - \gamma$ where Hn was the

nth harmonic number, and γ was the Euler-Mascheroni constant. In mathematics, the n-th harmonic number is the sum of the reciprocals of the first n natural numbers (Hosmer and Lemeshew 2002).The difference between the nth convergent in equation ($\diamondsuit$) and γ the eco-epidemiological, geo-spectrotemrpoal, entomological, district-level regression-based model was then otpimally calculated by $\sum_{k=1}^{n}\frac{1}{k}-\ln n-\gamma=\int_{n}^{\infty}\frac{x-|x|}{x^2}dx$ where $\lfloor x\rfloor$ was the floor function which satisfied the inequality $\sum_{k=1}^{n}\frac{1}{k}-\ln n-\gamma=\int_{n}^{\infty}\frac{x-|x|}{x^2}dx$. The symbol g was then $\gamma'\equiv e^{\gamma}\approx 1.781072$. This led to the radical representation of the geosampled, district-level, eco-epidemiological, iteratively interpolative, unmixed, wavelength, frequency LULC eigendecomposed, covariate coefficients as $e^{\gamma}=\left(\frac{2}{1}\right)^{1/2}\left(\frac{2^2}{1\cdot 3}\right)^{1/3}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/4}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/5}$ which coincidentally related to the double series $\gamma=\sum_{n=1}^{\infty}\frac{1}{n}\sum_{k=0}^{n-1}(-1)^{k+1}\binom{n-1}{k}\ln(k+1)$ and $\binom{n}{k}$,a binomial coefficient.

Thereafter, another proof of product in the geo-spectrotemporal, district-level, onchcerciasis regression-related, predictive risk model model was provided by the equation $\frac{\pi}{2}=\left(\frac{2}{1}\right)^{1/2}\left(\frac{2^2}{1\cdot 3}\right)^{1/4}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/8}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/16}$. The solution was then made even clearer by changing $n\rightarrow n+1$. In Jacob et al. (2016) both these regression-based formulas were analogous to the finalized product for e which was then rendered by calculating $e=\left(\frac{2}{1}\right)^{1/1}\left(\frac{2^2}{1\cdot 3}\right)^{1/2}\left(\frac{2^3\cdot 4}{1\cdot 3^3}\right)^{1/3}\left(\frac{2^4\cdot 4^4}{1\cdot 3^6\cdot 5}\right)^{1/4}$.

A negative binomial regression model in GEN MODwith non-homogenous means and a gamma distribution was employed in Jacob et al. (2012) for treating multiple overdispersed. *Culex quinquefasciatus* ans *Aedes alpopictus* forecasting, regression regression models . The authors incorporated $\alpha=\frac{1}{\theta}(\alpha>0)$ to the regression equation statement in PROC REG.

Initially, Poisson regression models were generalized by introducing an unobserved heterogeneity term for observation i. Thus, the individuals in the Cx. quinquefascistus and Ae alo assumed to differ randomly in a manner that is not fully accounted for by the observed covariates. This is formulated as $E(y_i|\mathbf{x}_i,\tau_i)=\mu_i\tau_i=e^{\mathbf{x}_i'\beta+\varepsilon_i}$ where the unobserved heterogeneity term $\tau_i=e^{\varepsilon_i}$ was independent of the vector of regressors $\mathbf{x}_i$. Then the distribution of y_i conditional on $\mathbf{x}_i$ and τ_i was Poisson with conditional mean and conditional variance $\mu_i\tau_i$: $f(y_i|\mathbf{x}_i,\tau_i)=\frac{\exp(-\mu_i\tau_i)(\mu_i\tau_i)^{y_i}}{y_i!}$ Tha authors in Jacob et al. (2012) Let $g(\tau_i)$ be the probability density function of τ_i. In probability theory, a probability density function (PDF), or density of a

continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value.

A probability density function is most commonly associated with absolutely continuous univariate distributions. A random variable X has density f_X, where f_X is a non-negative Lebesgue-integrable function, if: $\Pr[a \leq X \leq b] = \int_a^b f_X(x)\, dx.$ Hence, if F_X is the cumulative distribution function of X, then: $F_X(x) = \int_{-\infty}^x f_X(u)\, du,$ and (if f_X is continuous at x) $f_X(x) = \frac{d}{dx} F_X(x).$ Intuitively, one can think of $f_X(x)\, dx$ as being the probability of X falling within the infinitesimal interval $[x, x + dx]$.

For continuous *Ae egypti* immature habitats on newly trasnitioned , sub-meter resolution, iteratively interpolative, deforesed LULC to rice agriculture LULC uncoalesced, randomized, time series immature capture point, geo-spectrotemporal, geospatial, georrferenceable variables $X_1, \ldots, X_n$, it may be also possible to define a probability density function associated to the set as a whole, (i.e., joint probability density function). This density function may be optimally defined in ArcGIS or SAS as a function of the n geosampled entomological explanative LULC variables, such that, for any domain D in the n-dimensional space of the geosampled immature pordictivity values of the variables $X_1, \ldots, X_n$, may be robustly optimized based on the probability that a realization of the set variables falls inside the domain D which may be subsequently quantized employing $\Pr(X_1, \cdots, X_N \in D) = \int_D f_{X_1,\cdots,X_N}(x_1, \cdots, x_N)\, dx_1 \cdots dx_N.$ If $F(x_1, \ldots, x_n) = \Pr(X_1 \leq x_1, \ldots, X_n \leq x_n)$ is the cumulative distribution function of the vector $(X_1, \ldots, X_n)$, then the joint probability density function can be computed as a partial derivative $f(x) = \left.\frac{\partial^n F}{\partial x_1 \cdots \partial x_n}\right|_x$ in the eco-epidemiological,, time series, yellow fever ,forecasting model. The cumulative distribution function of a real-valued random variable X is the function given by $F_X(x) = P(X \leq x),$ where the right-hand side represents the probability that the random variable X takes on a value less than or equal to x.[Hosmer and Lemeshew 2002] The probability that X lies in the semi-closed interval $(a, b]$, where $a < b,$ in a YFV risk model then may be $P(a < X \leq b) = F_X(b) - F_X(a).$

In Jacob et al. (2012) the probability of a randomized *Culex quinquefasciatusor* or *Ae albopictus* geo-spectrotemporally, geosampled, sub-meter resoltion, gridded, stratifed variable falling within a particular range of georeferenceable, predictive variable values was given by the integral of this variable's density over that range given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the range.The probability density function is nonnegative everywhere, and its integral over the entire space is equal to one[Hosmer and Lemeshew 2002]. In so doing the distribution $f(y_i|\mathbf{x}_i)$ (no longer conditional on τ_i) was obtainable by integrating $f(y_i|\mathbf{x}_i, \tau_i)$ with respect to τ_i: $f(y_i|\mathbf{x}_i) = \int_0^\infty f(y_i|\mathbf{x}_i, \tau_i) g(\tau_i)\, d\tau_i$

However, an analytical solution in COUNTREG was employed to this integral which only existed when τ_i was assumed to follow a gamma distribution. This solution was the negative binomial distribution. When the *Cx. quinquefascitus* and *Ae albopictus*,ovipsoition, eco-epidemiological, immature habitat, QuickBird viisble and NIR, forecasting, vulnerability, eco-epidemiological models contained a constant term, it was necessary to assume that $E(e^{\varepsilon_i}) = E(\tau_i) = 1$, in order to identify the mean of the distribution. Thus, the authors in Jacob et al. (2012) assumed that τ_i followed a gamma(θ, θ) distribution with $E(\tau_i) = 1$ and $V(\tau_i) = 1/\theta$:

$$g(\tau_i) = \frac{\theta^\theta}{\Gamma(\theta)} \tau_i^{\theta-1} \exp(-\theta \tau_i)$$ where $\Gamma(x) = \int_0^\infty z^{x-1} \exp(-z) dz$ was the gamma function and θ was a positive geosampled, georeferenceable, explanative parameter in both paradigms. Then, the

density of y_i given $\mathbf{x}_i$ was derived as $\int_0^\infty f(y_i|\mathbf{x}_i, \tau_i) g(\tau_i) d\tau_i \frac{\theta^\theta \mu_i^{y_i}}{y_i! \Gamma(\theta)} \int_0^\infty e^{-(\mu_i+\theta)\tau_i} \tau_i^{\theta+y_i-1} d\tau_i$

$$\frac{\theta^\theta \mu_i^{y_i} \Gamma(y_i+\theta)}{y_i! \Gamma(\theta)(\theta+\mu_i)^{\theta+y_i}} \frac{\Gamma(y_i+\theta)}{y_i! \Gamma(\theta)} \left(\frac{\theta}{\theta+\mu_i}\right)^\theta \left(\frac{\mu_i}{\theta+\mu_i}\right)^{y_i}.$$

The authors in Jacob et al. (2012) then let $g(\tau_i)$ be the probability density function of τ_i in the model. Then, the distribution $f(y_i|\mathbf{x}_i)$ was no longer conditional on τ_i. Instead it was obtained by integrating $f(y_i|\mathbf{x}_i, \tau_i)$ with respect to τ_i: $f(y_i|\mathbf{x}_i) = \int_0^\infty f(y_i|\mathbf{x}_i, \tau_i) g(\tau_i) d\tau_i$. The distribution in the linearized, eco-epidemiological, district-level, malarial, regression model was

$$f(y_i|x_i) = \frac{\Gamma(y_i+\alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}} \left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i}, y_i = 0,1,2.$$

then The negative binomial non-homogenous distribution for both models was thus derived as a gamma mixture of Poisson random variables. The conditional mean in the model was then $E(y_i|x_i) = \mu_i = e^{x_i'\beta}$ and the variance in the residual probabilistic, regressed estimates were

$$V(y_i|x_i) = \mu_i\left[1 + \frac{1}{\theta}\mu_i\right] = \mu_i[1+\alpha\mu_i] > E(y_i|x_i)$$ To further estimate the district-level, eco-epidemiological models, the authors specified DIST=NEGBIN (p=1) in the MODEL statement in PROC REG. The negative binomial model NEGBIN1 was set p=1 , which revealed the variance function $V(y_i|x_i) = \mu_i + \alpha\mu_i$ was linear in the mean of the models. The log-likelihood function of the NEGBIN1 model was then provided by $L = \left\{\sum_{j=0}^{y_i-1} \ln(j + \alpha^{-1}\exp(x_i'\beta))\right\}$ Additionally, In Jacob et al. (2013) the equation $-\ln(y_i!) - (y_i + \alpha^{-1}\exp(x_i'\beta))\ln(1+\alpha) + y_i\ln(\alpha)$ was employed by both models for robust linear quantiattion of the geosampled covariates. The gradient for the geo-spatiotemporal, malarial-based, forecasting regression model was then quantified employing $\frac{\delta L}{\delta\beta} = \sum_{i=1}^N \left\{\left(\sum_{j=0}^{y_i-1} \frac{\mu_i}{(j\alpha+\mu_i)}\right)x_i - \alpha^{-1}\ln(1+\alpha)\mu_i x_i\right\}$ and $\frac{\delta L}{\delta\alpha} = \sum_{i=1}^N \left\{-\left(\sum_{j=0}^{y_i-1} \frac{\alpha^{-1}\mu_i}{(j\alpha+\mu_i)}\right) - \alpha^{-2}\mu_i\ln(1+\alpha) - \frac{(y_i+\alpha^{-1}\mu_i)}{1+\alpha} + \frac{y_i}{\alpha}\right\}$

Thereafter, the negative binomial regression models with variance function $V(y_i|x_i) = \mu_i + \alpha\mu_i^2$, wawere referred to as the NEGBIN2 model.s To estimate this regression-based models, wthe

145

authors specified DIST=NEGBIN (p=2) in the MODEL statements. A test of the Poisson distribution was then performed by examining the hypothesis that $\alpha = \dfrac{1}{\theta_i} = 0$.

A Wald test was also provided using the reported t statistics for the *Cx. quinquefascitus* and *Ae albopictus*, forecats, vulnerability model estimates. Under the Wald statistical test, the maximum likelihood estimate $\hat{\theta}$ of the parameter(s) of interest θI mwas compared with the proposed value θ_0, with the assumption that the difference between the two will be approximately normally distributed[2]. The log-likelihood function of the regression models (i.e., NEGBIN2) was then generated by the equation:

$$L = \sum_{i=1}^{N}\left\{\sum_{j=0}^{y_i-1}\ln\left(j+\alpha^{-1}\right)-\ln\left(y_i!\right)=-\left(y_i+\alpha^{-1}\right)\ln\left(1+\alpha\exp\left(x_i'\beta\right)\right)+y_i\ln(\alpha)+y_ix_i'\beta\right\}$$

whose gradient

was $\dfrac{\delta L}{\delta\beta}=\sum_{i=1}^{N}\dfrac{y_i-\mu_i}{1+\alpha\mu_i}x_i$ .The variance in our models was then assessed by

$$\frac{\delta L}{\delta\alpha}=\sum_{i=1}^{N}\left\{-\alpha^{-2}\sum_{j=0}^{y_i-1}\frac{1}{\left(j+\alpha^{-1}\right)}\alpha^{-2}\ln\left(1+\alpha\mu_i\right)+\frac{y_i-\mu_i}{\alpha\left(1+\alpha\mu_i\right)}\right\}$$

. The final mean in the models was calculated as: $\dfrac{pr}{1-p}$, the mode as; $\begin{cases}\left\lfloor\dfrac{p(r-1)}{1-p}\right\rfloor & \text{if } r>1 \\ 0 & \text{if } r\leq 1\end{cases}$, the variance as $\dfrac{pr}{(1-p)^2}$, the skewess as $\dfrac{1+p}{\sqrt{pr}}$, the kurtosis as $\dfrac{6}{r}+\dfrac{(1-p)^2}{pr}$, the moment generating function as $\left(\dfrac{1-p}{1-pe^t}\right)^r$ for $t<-\log p$, the characteristic function as $\left(\dfrac{1-p}{1-pe^{it}}\right)^r$ with $t\in\mathbb{R}$; and, the probability generating function as $\left(\dfrac{1-p}{1-pz}\right)^r$ for $|z|<\dfrac{1}{p}$.

In Jacob et al. (2013b) results from both a Poisson and a negative binomial (i.e., a Poisson random variable with a gamma distrusted mean) revealed that malarial related district-level, time series, sub-metere resolution, paramterizable covariates rendered from a predictive model for Uganda were significant, but furnished virtually no predictive power. Inclusion of indicator variables denoting the time sequence and the district geolocational spatial structure was then articulated with Thiessen polygons in ArcGIS which also failed to reveal meaningful iteratively quantitative, interpolatable covariates. Thereafter, an Autoregressive Integrated Moving Average (ARIMA) model was constructed in PROC ARIMA which revealed a conspicuous but not very prominent first-order, temporal, autoregressive structure in the individual, district-level, eco-epidemiological, time-series, endmember, dependent data. A random effects term was then specified using monthly, Box-Jenkins, time-series, modeling methodology to explicativey regressively quantitate, dependent data. This specification included a district-specific intercept term that was a random deviation from the overall intercept term which was based on a draw from a normal frequency distribution. The notation $AR(p)$ indicated an autoregressive model of

order p which was defined as $X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$ where $\varphi_1, \cdots, \varphi_p$ were the unbiased parameters of the model, c was a constant, and ε_t was white noise. Was equivalently written using the backshift operator B as $X_t = c + \sum_{i=1}^{p} \varphi_i B^i X_t + \varepsilon_t$ so that, moving the summation term to the left side was evaluated using polynomial notation, which then rendered $\phi(B)X_t = c + \varepsilon_t$. In so doing, the autoregressive model was viewed as the output of an all-pole infinite impulse response filter whose input was white noise. Infinite impulse response (IIR) is a property applying to many linear time-invariant systems (Cressie 1993)

The random effects specification revealed a non-constant mean across the Ugandan districts. This random intercept represented the combined effect of all omitted covariates that caused districts to be more prone to the malaria prevalence than other districts. Additionally, inclusion of a random intercept assumed random heterogeneity in the districts' propensity or, underlying risk of malaria prevalence which persisted throughout the entire duration of the time sequence under study. This random effects term displayed no spatial autocorrelation, and failed to closely conform to a bell-shaped curve. Autocorrelation is the correlation among values of a single variable strictly attributable to their relatively close geolocational positions on a two-dimensional surface, introducing a deviation from the independent observations assumption of classical statistics (Griffith 2003). The model's variance, however, implied a substantial variability in the prevalence of malaria across districts in Ugandan study site. The estimated model contained considerable overdispersion (i.e., excess Poisson variability): quasi-likelihood scale = 76.565. The following equation was then employed to forecast the expected value of the prevalence of malaria at the district-level: prevalence = exp[-3.1876 + (random effect)$_i$] .

Compilation of additional and accurate, time series, sub-meter resolution, vector arthropod-related, seasonal abundance and iteratively quantitatively interpolative, georefernceable geo-spectrotemporal, distribution, data can allow continual updating of the random effects term estimates allowing research intervention teams to bolster the quality of the immature, capture point, forecasts for future district-level malarial, probabilistic, optimizable , risk modelling efforts employing principles of Gu and Novak (2005). In practice, seasonal, larval control measures are applied to vector arthropod, immature, seasonal habitats individually . Therefore, the effect of any intervention is reflected by treatment-induced changes in the adult productivity. If the assumption that treatments applied to individual habitats are 100% effective in elimination of emerging adults, that is, treated immature habitats produce zero contribution to the total productivity, defined as the total emerging female mosquitoes for example,in an ArcGIS predictive map, it may be deduced that the probability of endemic disease transmission also reduces. Larval control measures can range from resource reduction to environmental manipulation of habitats to the application of microbial larvicides (www.who.gov).

Signature interpolation may geolocate unknown, un-geo-sampled, prolific seasonal geo-spectrotemporally geosampled, vector arthropod data in ArcGIS maps using principles of Gu and Novak (2005). Spectral unmixing algorithms have proliferated in a variety of ecological disciplines by exploiting, sub-meter resolution, remotely sensed eco-epidemiological,

wavelength, sub-meter resolution, immature, oviposition, uncoalesced, habitat data. Medically important entomological capture points (e.g., east African ricefield agro-ecosystems, aquatic habitats of *Anopheles arabiensis*, a major vector of malaria in sub-saharan Africa) utilize semi-permanent to temporary habitats (e.g., floodwater areas, vernal pools, hoof prints) (Mwangangi et al. 2008, Muturi 2007, mururi 2007). Jacob et al. (2011) geo-spectrotemporally othogonally, quantitatively decomposed a sub-meter, spatial resolution, eco-epidemiological dataset of QuickBird wavebands representing multiple, georefereced, riceland, *An. arabiensis,* habitat mixels for predicting hyperproductive, georeferenceable habitats in a riceland environment in central Kenya. Initially, the authors constructed a regression model which revealed that paddy preparation, *An. arabiensis* habitats were the most productive based on geo-spectrotemporal, field-geosampled, count data. Individual, uncoalesced, mixel, spectral reflectance, regresseable estimates from a QuickBird visible and NIR dataset of a paddy preparation, *An. arabiensis,* eco-epidemiological, georeferenceable, capture point, oviposition, immature habitat were then extracted by using a Li-Strahler geometric-optical model in ArcGS. The model used three scene components: sunlit canopy (C), sunlit background (G) and shadow (T) generated from the riceland image. The G, C, T components' classes were remotely estimated using ENVI, an object-based classification algorithm. In ENVI$^{®}$, the Digital Number (DN) of the mixel in every QuickBird band was viewed using the z-profile from a spectral library. After making an atmospheric correction from the image for the eco-epidemiological, study site, the DN was converted into ground reflectance. A convex geometrical model was also employed for endmember validation of the geo-spectrally decomposed, paddy preparation, capture point, immature habitat. An ordinary, kriged-based, iterative interpolation was performed in ArcGIS Geostatistical Analyst ^{TM} employing the reference signature generated from the unmixing models. Linear unbiased, eco-epidemiological, georeferenceable, elucidative predictors and variance estimates may be derived in a similar fashion fora seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model.

Spatially predictive techniques such as kriging was used to develop a three-dimensional model and maps of chemical distributions in soil and to produce reliable definitions of remedial extent. Spatial models or more commonly referred to as 3-D geostatistical models have practical applications in remedial engineering as well as regulatory development of cleanup goals(www.esri.gov). Geostatistical models and various other computer simulations are becoming indispensable as industry and regulators rely more heavily on cost-effective, risk-based, and privatized approaches to site regulation and remediation. Sampling plans and analyses designed specifically to meet the needs of defensible geostatistical models are expected to play a large part in environmental studies and remedial engineering.

Spectral unmixing tools in ArcGIS and ENVI have been   also used to decompose sub-meter resolution, visible and NIR, mixel reflectance of other unknown, un-geosampled, seasonally hyperproductive, immature, vector arthropod, georeferenceable, larval habitats to implement control strategies. The identification of a spectral signature characteristic of black–fly vector of onchocerciasis, *Similium damnosum* s.l., positive aquatic sites has been previously described Jacob et al. (2011). Onchocerciasis, or river blindness, has historically been one of the most important causes of blindness worldwide which is caused by the filarial parasite *Onchocerca volvulus* (Thylefors 1978). This spectral signature is characteristic of the habitat

features found at known positive sites, which include fast flowing water passing over a substrate of Precambrian rock. The incidence of onchocerciasis is not distributed uniformly across the country, but there is a clear correlation with basic geology (Crosskey1981) because the riverine conditions which create suitable larval sites for the vector, *Simulium damnosum* s.l., are most common where the Precambrian basement rock is exposed to break the flow of the water and create rapids(Mafuya et al 2006).

In order to geospatially characterize the sparsely shaded sub-meter resolution wavelength, QuickBird visble and NIR data and obtain stable solutions of within-canopy multiple scattering, Jacob et al. (2013b) decomposed the spectrally extracted Red Edge NDVI biosignature into three parts; unscattered radiance $I^0(\tau,\Omega)$, single scattering radiance $I^1(\tau,\Omega)$, and multiple scattering radiance $I^M(\tau,\Omega)$ $I(\tau,\Omega) = I^0(\tau,\Omega) + I^1(\tau,\Omega) + I^M(\tau,\Omega)$ in ArcGIS.

A simple scheme was then represented by $I^0(\tau,\Omega)$ which was denoted by 1, and was not scattered by the atmosphere, but reflected directly by the within canopy surface features. $I^1(\tau,\Omega)$ represented the various Red Edge, NDVI, canopy, 5m, biosignature radiance values either scattered once by the atmosphere, denoted by 2, or once by the within canopy, structural, spectral variables which was denoted by 3. The variable $I^M(\tau,\Omega)$ was the most complicated component, which included all of other imaged, riverine, larval habitat, canopied, operationizable, georeferenced components in the radiation field of the coupled medium.

Unscattered sunlight radiances $I^0(\tau,\Omega)$ were then characterized by the following radiative transfer equation and corresponding boundary conditions. When T < Ta, the radiative transfer model rendered:

$$
\begin{cases}
-\mu \dfrac{\partial I^0(\tau,\Omega)}{\partial \tau} + I^0(\tau,\Omega) = 0 \\
I^0(0,\Omega) = \delta(\Omega - \Omega_0) i_0 \qquad \mu < 0 \\
I^0(\tau_a^{bot},\Omega) = I^0(\tau_c^{top},\Omega) \qquad \mu > 0
\end{cases}
$$

, where τ_a^{bot} and τ_c^{top} were the optical depths at the bottom of the atmosphere and the ToA of the habitat canopy, respectively. Here different notations from the 5m imager were used to indicate the physical meaning of the canopy boundary conditions. The model provided the upper boundary condition, which meant only parallel sunlight illuminated the atmosphere at the top of the riverine larval habitat canopy in the direction Ω_0.

When $\tau > \tau_\alpha$, the residuals were:

$$
\begin{cases}
-\mu \dfrac{\partial I^0(\tau,\Omega)}{\partial \tau} + h(\tau,\Omega) G(\Omega) I^0(\tau,\Omega) = 0 \\
I^0(\tau_c^{top},\Omega) = I^0(\tau_\alpha^{bot},\Omega) \qquad \mu < 0 \\
I^0(\tau_t,\Omega) = fs(\Omega_0,\Omega)|\mu_0| I^0(\tau_t,\Omega_0) \qquad \mu > 0
\end{cases}
$$

Jointly solving the above equations with these boundary conditions rendered:

$$I^0(\tau,\Omega)=$$

$$
\begin{cases}
I^0_{d1}(\tau,\Omega)= i_0 \exp(-\tau/|\mu|)\delta(\tau,\Omega_0) \quad \mu<0, \tau \le \tau_\alpha \\
\quad\quad I^0_{d2}(\tau,\Omega)= I^0_{d1}(\tau,\Omega) \\
\quad \cdot \exp[-C(\Omega)(\tau-\tau_\alpha)/|\mu|] \quad \mu<0\, \tau_\alpha<\tau\le\tau_t \\
\quad\quad\quad I^0_{u2}(\tau,\Omega)= i_0 \\
\quad\quad \cdot \exp\left[-\dfrac{\tau_\alpha+(\tau_t-\tau_\alpha)G(\Omega_0)}{|\mu_0|}\right] \\
\cdot fs(\Omega_0,\Omega)|\mu_0|\exp[-\varepsilon(\tau,\Omega)] \quad \mu>0, \tau_\alpha<\tau\le\tau_t \\
I^0_{u1}(\tau,\Omega)= I^0_{u2}(\tau_\alpha,\Omega) \\
\cdot \exp[-(\tau_\alpha-\tau)/\mu] \quad \mu>0, \tau\le\tau_\alpha
\end{cases}
$$

The biosignature information was expressed as $I^0_{u2}(\tau,\Omega)$, which represented the upwelling sunlight radiance within the georeferenced habitat canopy, and the function $\varepsilon(\tau,\Omega)$. The extinction coefficients of the canopy endmembers were modified, and we then incorporated the extracted within-canopy radiance values including the floating, hanging, and surrounding dead vegetation canopy geospectral components employing :

$$\varepsilon(\tau,\Omega)=\frac{1}{\pi}\int_\tau^{\tau_t} h(t,\Omega)G(\Omega)dt \quad = G(\Omega)\frac{\tau_t-\tau}{\mu}-\left[\sqrt{\frac{G(\Omega_0)G(\Omega)}{\mu|\mu_0|}}\frac{kH}{\Delta(\Omega_0,\Omega)}\right]t_0$$

,where t_0 was defined

$$t_0=\exp\left[-\frac{\Delta(\Omega_0,\Omega)\tau}{kH}\right]-\exp\left[-\frac{\Delta(\Omega_0,\Omega)\tau_t}{kH}\right]$$

as .

The QuickBird endmember biosignature model in ArcGIS revealed that for single scattering radiances, unscattered sunlight became the scattering source, and the boundary conditions. These conditions were then determined based on the fact that no incident single scattering radiances originated from above ToA or below the bottom of the canopy. When T < Ta occurred in the model, the residuals rendered:

$$
\begin{cases}
-\mu\dfrac{\partial I^1(\tau,\Omega)}{\partial\tau}+I^1(\tau,\Omega)= \\
\dfrac{\omega i_0}{4\pi}p(\Omega_0\to\Omega)\exp\left(-\dfrac{\tau}{\mu_0}\right) \\
\quad r^1(0,\Omega)=0 \quad \mu<0 \\
\quad I^1(\tau_\alpha^{bot},\Omega) \quad \mu>0
\end{cases}
$$

Additionally, when T > Ta the decomposition trailing vegetation, *S. damnosum* s.l. larval hábitat model rendered:

$$\begin{cases} -\mu \dfrac{\partial \tau^1(\tau,\Omega)}{\partial \tau} + h(\tau,\Omega)G(\Omega)\mathrm{I}^1(\tau,\Omega) \\[2mm] \dfrac{i_0'}{\pi}\Gamma(\Omega_0,\rightarrow \Omega)\exp\left[-(\tau-\tau_\alpha)\dfrac{G(\Omega_0)}{|\mu_0|}\right] \\[2mm] \mathrm{I}^1(\tau_c^{top},\Omega)=\mathrm{I}^1(\tau_\alpha^{bot},\Omega) \qquad \mu<0 \\[2mm] \mathrm{I}^1(\tau_t,\Omega)=0 \qquad \mu>0 \end{cases}$$

, where i_0' was the incident solar net flux arriving at the top of the

$$i_0' = i_0 \exp\left(-\tau_\alpha/|\mu_0|\right)$$

habitat canopy . In the downward direction $\mu<0$, the solution was easily derived. When T < Ta, the riverine larval habitat QuickBird, endmember biosignature decomposition model was solved using:

$$\mathrm{I}^1(\tau,\Omega)=$$

$$\begin{cases} \dfrac{\omega F o p(\Omega_0 \rightarrow \Omega)|\mu_0|}{4(|\mu_0|\leftarrow|\mu|)}\left[\exp\left(-\dfrac{\tau}{|\mu_0|}\right)-\exp\left(-\dfrac{r}{|\mu|}\right)\right] & \Omega \neq \Omega_0 \\[4mm] \dfrac{\omega F_0 \tau}{4|\mu_0|}P(\Omega_0 \rightarrow \Omega)\exp\left(-\dfrac{\tau}{\mu_0}\right) & \Omega = \Omega_0 \end{cases}$$

. Then $\tau_\alpha < \tau < \tau_t$ the model was solved using the equation:

$$\mathrm{I}^1(\tau,\Omega)=$$

$$\begin{cases} \dfrac{i_0'|\mu_0|\Gamma(\Omega_0 \rightarrow \Omega)}{\pi[G(\Omega)|\mu_0|-G(\Omega_0)|\mu|]}t_1 + \Delta \mathrm{I}^1(\tau,\Omega) & \Omega \neq \Omega_0 \\[4mm] \dfrac{(\tau-\tau_\alpha)i_0'\Gamma(\Omega_0 \rightarrow \Omega)}{\pi|\mu_0|}\exp\left[-G(\Omega_0)\dfrac{\tau-\tau_\alpha}{|\mu_0|}\right] \\[2mm] \qquad\qquad + \Delta \mathrm{I}^1(\tau,\Omega) \end{cases}$$

where t_1 was defined by the

$$t_1 = \exp\left[-G(\Omega_0)\dfrac{\tau-\tau_\alpha}{|\mu_0|}\right]-\exp\left[-G(\Omega)\dfrac{\tau-\tau_\alpha}{|\mu|}\right]$$

equations and $\Delta \mathrm{I}^1(\tau,\Omega)=\mathrm{I}^1(\tau_\alpha,\Omega)\exp\left[-G(\Omega)(\tau-\tau_\alpha)/|\mu|\right]$,

which represented the single scattering riverine, habitat canopy radiances emerging from the atmosphere without scattering in the riverine, trailing vegetation habitat canopy.

In the upward direction (p > 0), the solutions were a little more complicated because due

$$\begin{cases} \dfrac{1}{\mu}\int_r^{\tau_t} F(\tau',\Omega) \\[2mm] \cdot \exp\left[-\dfrac{1}{\mu}\int_\tau^{\tau'} h(\varepsilon,\Omega)G(\Omega)d\varepsilon\right]d\tau' & \tau_\alpha \leq \tau \leq \tau_t \\[2mm] \dfrac{\omega F_0 p(\Omega_0 \rightarrow \Omega)|\mu_0|}{4(\mu+|\mu_0|)}t_2 \\[2mm] + \mathrm{I}^1(\tau_\alpha,\Omega)\exp\left(\dfrac{\tau-\tau_\alpha}{\mu}\right) & \tau < \tau_\alpha \end{cases}$$

to the hotspot effect, which was determined by $\mathrm{I}^1(\tau,\Omega)=$,

where t_2 was $t_2 = \exp\left[-\dfrac{\tau}{|\mu_0|}\right] - \exp\left[\dfrac{\tau}{\mu} - \left(\dfrac{1}{\mu_0} + \dfrac{1}{\mu}\right)\tau_\alpha\right]$, and the second integration at Ta < T < Tt in the above equation (see Figure 9). This equation was then explicitly obtained by means of an alternative integer and range, which was solved using $F(\tau', \Omega) = \dfrac{i_0'}{\pi}\Gamma(\Omega_0 \to \Omega)\exp\left[-G(\Omega_0)(\tau - \tau_\alpha)/|\mu_0|\right]$. The radiance $I^1(\tau, \Omega)$ at Ta < T <Tt derived for the QuickBird endmember, biosignature material was then numerically evaluated without further assumptions. An explicit approximation to $I^1(\tau, \Omega)$ was then derived and used for inversion in the canopy, biosignature decomposition model.

The first term in equation $K_c C = \dfrac{1}{2}\left(1 + \langle i, v \rangle\right)\left(1 - e^{-\lambda \pi R^{*2} \sec \theta_v}\right)C$ ignored the problem of mutual shading of the geosampled trailing vegetation, sparsely shaded, *S. damnosum s.l.* riverine habitat canopy and the ripple water components. Strahler and Jupp [1990] handled this problem through multiple integration, in which the mutual shadowing of canopies and other associated objects were treated in the same way as the mutual shading of leaves. However, our objective in this research was to derive a simple approximation to describe the effect of the *S. damnosum s.l.* larval habitat based on collections of individual discrete reflectance surface values (i.e., Precambrian rock and ripple water components). To carry this out, we developed an approach that applied one-stage geometric optics to deal with the spatial relationship between the sub-pixel endmember reflectance spectra of the decomposed *S. damnosum s.l.* habitat surface, the ripple water components that were mutually shaded in the illumination direction, and the parts mutually shaded in the view direction. We then quantified mutual shadowing proportions generated from the *S. damnosum s.l.* habitat and its associated ripple water components. In Li and Strahler [1985] and Li [1985], simulation and mathematics simplified to the one-dimensional case was proved so that for the nadir-viewed cone model, mutual shadowing of illumination would not change the ratio $K_c / (1 - K_g)$. In this research, this ratio was itself denoted K_c, which we used to generate u A_c / A for determining consistency with K_g, where the mutual shadowing in illumination and viewing directions was independent A_c / A for consistency with K_g.

We then considered the proportion of the *S. damnosum s.l.* habitat which was mutually shadowed by the Precambrian rock and ripple water components. In the direction of illumination, the immature, *S. damnosum s.l.* riverine habitat had an area of $\pi R^2 \sec \theta_i'$, and the total projected LULC area of the habitat was then calculated to be $\lambda \pi R^2 \sec \theta_i'$, if there was no mutual shadowing. Because of mutual shadowing, however, the net projected area was $1 - e^{\lambda \pi R^2 \sec \theta_i'}$. The difference indicated the total mutual shadowing of the *S. damnosum s.l.* riverine habitat attribute. Thus, the authors in Jacob et al. [2013b] defined the quantity M_i, the mutual shadowing

$$M_i = 1 - \frac{1 - e^{-\lambda \pi R^2 \sec \theta_i'}}{\lambda \pi R^2 \sec \theta_i'}$$

proportion in the illumination direction, as . M_i, which revealed the degree of mutual shadowing in the illumination direction. In other words, each spheroid, on average, had a proportion M_i of the imaged *S. damnosum s.l.* habitat surface area that was not

sunlit. This part of the habitat was concentrated at the lower part of the spheroid. We then generated a boundary drawn on the habitat surface of the spheroid with the area comprising M_i located below it. Similarly, we defined M_v as the mutual shadowing proportion of the ripple water components in the view direction as

$$M_v = 1 - \frac{1 - e^{-\lambda \pi R^2 \sec\theta_v'}}{\lambda \pi R^2 \sec\theta_v'}$$

. The viewing shadows were concentrated at the lower part of the spheroid so we were able to define the M_v boundary. The proportion of sunlight the QuickBird sensor captured corresponded to the area above both M_I and M_v boundaries, which were then dependent on both zenith and azimuth differences between the illumination and view directions. At the hotspot, M_i and M_v boundaries overlapped and the QuickBird data revealed no mutual shadowing of the *S. damnosum s.l.* habitat or the ripple water components. When the view zenith angle was larger than the illumination zenith angle, M_v was greater than M_i and little to no mutually shaded habitat area was visible, based on the azimuth differences between the imaged objects. Thus, we were able to capture the essence of the mutual-shading effect of the canopy, and the ripple water components.

The authors in Jacob et al. (2013b) quantified the *f*-Ratio of Nonnadir-Viewed Spheroids. First, we considered a single spheroid in the decomposed sub-mixel endmember, trailing vegetation, hyperproductive, georeferenced, canopied, sparsely shaded, spectral data. For the spheroidal case, it is necessary to show whether the *f*-Ratio is still independent of density, as in the case of the nadir-viewing cones [Schowengerdt 1997)]. From the view direction, the spheroid had a projected area $\Gamma_v = \pi R^2 \sec\theta_v'$. However, only the portion $\frac{1}{2}\left(1 + \langle i', v' \rangle\right)$ of the *S. damnosum s.l.* habitat with ripple water components was sunlit. Similarly, the illumination shadow on the ground occupied the habitat area $\pi R^2 \sec\theta_i'$. The compound area of viewed habitat, and ripple water components plus illumination shadow projected onto the background was $\Gamma = \pi R^2 \left[\sec\theta_i' + \sec\theta_v' - O(\theta_i', \theta_v', \phi)\right]$. Thus, we defined the *f*-ratio for the spheroidal and its associated trailing vegetation, canopied, sparsely shaded attributes as

$$F = \frac{\Gamma_c}{\Gamma} = \frac{\frac{1}{2}\left(1 + \langle i', v' \rangle\right)\sec\theta_v'}{\sec\theta_i' + \sec\theta_v' - O(\theta_i', \theta_v', \phi)}$$, where Γ_c was the sunlit area of the *S. damnosum s.l.* riverine habitat, Precambrian rock and ripple water components. We then defined the corresponding ratio $f = \frac{K_c}{1 - K_g}$ for the endmember selection of these QuickBird[TM], sub-mixel emissivities. For our purposes, *n* represented the shadow parameters generated from the decomposed, 5m, *S damnosum s.l.* riverine habitat mixel. If there was no mutual shadowing, there was had $f = F$.

The authors then noted that *n* increased, however, mutual shadowing occurred, and as such, $K_g = e^{-\lambda \pi R^2 \left[\sec\theta_i' + \sec\theta_v' - O(\theta_i, \theta_v, \phi)\right]}$. They then then defined the mutual shadowing proportion M as $M = 1 - \frac{1 - K_g}{\lambda \Gamma}$, which was the fraction of total shadowing cast from the ripple water components that fell onto the *S. damnosum s.l.* habitat instead of the background. The sunlit and viewed habitat surface features were reduced by hiding either from viewing or from illumination. Thus,

153

the *f*-ratio with mutual shadowing was $f = \frac{n\Gamma_c - \sum \Delta_{A_c}}{A(1-K_g)}, \quad = F\frac{1-\sum \Delta_{A_c}/(n\Gamma_c)}{1-M}$, where $\sum \Delta_{A_c}$ was the total decrement from $n\Gamma_c$ to A_c (i.e., the background-projected area of viewed sunlit, riverine habitat surface). We expressed $\sum \Delta_{A_c}$ as three terms: a decrement due to mutual shading in the view direction, plus a decrement due to mutual shading in the sun direction, and a subtraction of those elements shaded in both directions using $\sum \Delta_{A_c} = n\Gamma_v (P_v M_v + P_i M_i - P_o)$, where P_v was the conditional probability that the *S. damnosum s.l.* habitat faced the sun given that it was mutually shaded from view. In this research Pi was the probability that ripple water habitat surface elements faced the viewer given that it was mutually shaded from illumination. Both P_i and P_v were average proportions of the habitat areas projected in the view direction. P_o, the third term, overlapped part of the first two terms, expressed as a fraction of Γ_v. P_o contained three parts derived from the RapidEye[TM]-imaged habitat surface elements: ripple water components and the vegetation canopy geoclassified LULC structure. This collection contributed to the hotspot, due to the spatial correlation of the shadows. Since the probabilities of being hidden in multiple directions were not independent, we were able to substitute $\sum \Delta_{A_c} = n\Gamma_v (P_v M_v + P_i M_i - P_o)$ into $f = \frac{n\Gamma_c - \sum \Delta_{A_c}}{A(1-K_g)}, \quad = F\frac{1-\sum \Delta_{A_c}/(n\Gamma_c)}{1-M}$, which yielded a single expression for $f = F\frac{1-\Gamma_v(P_v M_v + P_i M_i - P_o)/\Gamma_c}{1-M}$. The authors then modeled, P_v, P_i, and P_o. We used all illumination or viewing shadows incorporating M_i and M_v boundaries respectively. In our model, P_v, P_i, and P_o were used to visualize the M_v and M_i boundaries. If viewing and illumination shadows fall strictly below M_v and M_i boundaries, then P_v, the conditional probability that at surface element facing the sun given the mutually shadowed areas will be the ratio of the illuminated portion of the projected surface below the M_v boundary (Schowengerdt 1997).

Correspondingly, in this research, P_i was the conditional probability that the sampled *S. damnosum s.l.* habitat directly faced the viewer given that it was mutually shaded from illumination and was the ratio of the viewed portion of the projected habitat area below the M_i boundary. Note, that M_i was the proportion of mutually-shaded *S. damnsoum* s.l.. habitat surface projected to the direction of illumination, but $P_i M_i \Gamma_v$ was the area of this fraction of the habitat surface with Precambrian rock and ripple water components projected to viewing direction. Proper calculation of this portion of the riverine habitat and its associated attributes involved some projection change. We used P_o as the variable representing the overlap area o, which was also represented as a fraction of Γ_v.

The authors of Jacob et al. (2013) then considered the case in the principal plane. For simplicity, they assumed that all shadows from the *S. damnosum s.l.* habitat, and rippled water components fell below the boundaries M_v and M_i, which were the traces of planes intersecting the spheroid at its center. The angle between the planes of the M_i and the illumination boundary was $\theta_{M_i} = \cos^{-1}(1-2M_i)$. The authors defined θ_{M_v} similarly. At the hotspot, the M_i and M_v boundaries coincided when $P_v = P_i = 1$, $P_o = M_v = M$, and $f = F = 1$. It was then assumed that the viewing zenith angle increased to $\theta_v > \theta_i$. In usual cases when mutual shadowing is

considered, the M_v boundary is higher than the M_i boundary [23, 24]. In the RapidEye^{TM} sensor's view, P_v was the ratio of the *S. damnosum* s.l. habitat's canopied surface area between M_v boundary and the illumination boundary to the whole area under the M_v boundary. That is,

$$P_v = \frac{M_v \Gamma_v - (\Gamma_v - \Gamma_c)}{M_v \Gamma_v}$$

, while P_i was one, and P_o cancelled the M_i term. Then, the equation became

$$f = F \frac{1 - P_v M_v \Gamma_v / \Gamma_c}{1 - M} = F \frac{(1 - M_v)\Gamma_v}{(1 - M)\Gamma_c} = \frac{1 - e^{-\lambda \Gamma_v}}{(1 - K_g)}$$

. This result suggested that when the viewing direction in the principal plane deviated from ($\theta v > \theta i$), the f-ratio will change in the endmember model. When the coverage is very low, the increment of θ_m may be also so small that M_v will be under the M_i boundary. In this case, mutual shadowing can be simply ignored as in Strahler and Jupp, (1985).

The authors in Jacob et al. [2013b] found that when θ_v moved inward on the principal plane but had not reached nadir, the M_i was higher than M_v. Hence, $P_v = 1$, $P_o = M_v$, and

$$P_i = \frac{1 - \cos(\theta_{M_i} - \theta_i' + \theta_v' \cos\phi)}{1 - \cos\theta_{M_i}}$$

. After θv passed the nadir, the M_v boundary went to the opposite side of the spheroid from M_i. In this case, the RapidEye^{TM}, sub-mixel spectral data revealed the horizontal projection of the habitat and its ripple water components at $\phi = \pi/2$. The authors then used *Pi* just as in Strahler and Jupp, [140], with ϕ equal to π, and P_v was the fraction of M_v over the illumination boundary, i.e.,

$$P_v = \begin{cases} \dfrac{1 - \cos(\theta_{M_v} - \theta_v' + \theta_v' \cos\theta)}{1 - \cos\theta_{M_v}}, & (\theta_{M_v} - \theta_v' + \theta_v' \cos\theta) \geq 0 \\ 0, & (\theta_{M_v} - \theta_v' + \theta_v' \cos\theta) < 0 \end{cases}$$

.

Note, the authors reported that when θ_v was between the hotspot and nadir, P_v was always one, so a discontinuity of P_v appeared at the nadir. This discontinuity arose from the assumption that all shadows fell under the M_v boundary. Additionally, in this research, the M_v at $\theta_v = 0$ was the physical intersection of boundaries between the habitat, and the rippled water components, which did not change with viewing geometry, thus, $P_v M_v$ was still continuous at nadir, and equal to P_o. In other words, the formula had a very large viewing zenith, so that $\theta_{M_v} - \theta_i' + \theta_v' \cos\phi) > 0$, $\sum \Delta_{A_c}/(n\Gamma_c) = M$. When M_i and M_v were explanatorialy independent,

$$f = \frac{n\Gamma_c - \sum \Delta_{A_c}}{A(1 - K_g)}, \qquad = F \frac{1 - \sum \Delta_{A_c}/(n\Gamma_c)}{1 - M}$$

.

If all the spatial objects are at the same height, the situation will be very close to the "uniform height case" – that is, mutual shadows will always fall on the lower part of the objects and the object top-viewing effect will be strong [Schowengerdt (1997)]. However, when heights are distributed over a wide range, the top layer of the canopy will play a more important role in

155

determining the BRDF of the canopy than the lower layer (Jacob et al. 2015). Therefore, when the habitat structural heights were quantified by spectral distribution, the BRDF was apparent. The BRDF was determined by the size, shape, and height of the trailing vegetation habitat and the ripple water components in the top layer. Thus, we restricted ourselves to considering a single top layer only, where the range of distribution of height of the sampled *S. damnosum s.l.* riverine habitat and its attributes did not exceed twice the vertical axis of the spheroid. To share the weighting between the spatiotemporally sampled, spectral predictor, covariate, coefficient estimates we used the equation $\beta = \left(1 - \dfrac{h_2 - h_1}{4b}\right)^2$, if $(h_2 - h_1) \leq 4b$. When $(h_2 - h_1) > b$, β is forced to be zero and it is necessary to redefine the layers (Keshava and Mustard 2002). By doing so, both P_v and P_i were calculated as a weighted sum of corresponding terms $P = \beta P_1 + (1 - \beta) P_2$, where P_1 and P_2 were the spectral probabilities associated with the spatial dimensions of the sampled habitat.

The bidirectional reflectance was modeled as a pure phenomenon that resulted in scenes of discrete, three-dimensional objects (i.e., turbid ripple water components) from the *S. damnosum s.l.* riverine habitat being illuminated and viewed from different positions in the hemisphere. The resulting scene was broken down into their canopy fractions, specifically sunlit and shadowed background, as well as scene brightness. Illumination direction was calculated by a linear combination of the canopy fractions and their respective radiance estimates. The shape of the patterns of the diffuse rippled water components were the driving explanatory spectral predictor variables in the model. These *S. damnosum s.l.* riverine habitat, spectral, sub-pixel emissivities conditioned the mixture of sunlit and shaded objects and background data that was observed from multiple viewing directions, thus quantifying all directions of illumination. This mixture, in turn, controlled the brightness in the image. However, since the use of data for quantitative spectral monitoring requires consistent surface reflectance data, we corrected the radiance effects from varying sun sensor target geometries in the multitemporal, RapidEye^{TM} datasets described by the BRDF. In this research, measuring the spread of the corrected results from the desired equal reflectance line provided a measure of the accuracy of our method. After correction, the root mean square reflectance errors were approximately 0.01 in the visible and 0.02 in the near-infrared.

An expression for additional azimuthual variation was also derived from the geometric-optical model. This azimuthual variation differed fundamentally in radiance for each layer of the *S. damnosum s.l.* riverine habitat canopy. It was observed that all non-zero polar angles were most evident in the canopy when vertical and nearly opaque components of the *S.damnosum s.l.* habitat and its neighboring Precambrain rock and ripple water components were illuminated and viewed along polar sun angles. For the variation of the directional reflectance of the multi-scattered diffuse riverine canopy cover, azimuthual view angles and shade-related parameters were quantified using the illuminated area of the imaged habitat (i.e., areas affected by the sun at large angles from the zenith). Our results also indicated that the cause of the azimuthual variation could be traced to solar flux illumination of the vertically-oriented Precambrian rock and rippled water components, as well as the variation of reflectance moderated by azimutually isotropic

sources of flux from sky light and the riverine larval habitat canopy reflectance values. Spectral unmixing yielded abundance estimates for each endmember together summing up to the 100% reflectance measured in the image. A scattergram representing the endmember reference signature of *S. damnosum s.l.* habitat ripple water pixel reflectance values was then generated.

The model developed to forecast hypeproductive, seasonal, *S. damnosum* s.l. aquatic habitats based on this signature which was designated the black rock-rapid (BRR) model. The spectral signature found to be characteristic of the habitat features that formed the basis of the BRR model. The waveband composition data of the signature was 34% red, 11% blue and 55% green

Of the 30 sites along the Sarakawa River in Northern Togo predicted to be larval habitats by the BRR model, all (100%) were found to contain *S. damnosum* s.l. larvae. In contrast, none of the 52 sites not predicted by the QuickBird, BRR model, but deemed to be potential habitat by the entomologist accompanying the verification team contained *S. damnosum* s.l. larvae. Together, these data suggested that the BRR model exhibited a sensitivity and specificity approaching 100% for the prediction of *S. damnosum* s.l., riverine, larval sites in Togo. Water levels in the rivers of West Africa fluctuate substantially between the rainy and dry seasons, potentially producing seasonally active immature sites. A second model was developed to predict such seasonally active sites. There was a complete correspondence between the sites predicted by the the BRR model and this second model suggesting that the BRR model had identified all active and potentially active immature sites in the study area. *S. damnosum* s.s. and *S. sirbanum* are found in savanna ecosystems throughout most of sub-Saharan Africa (www.who.gov).

To test the generality of the Quickbird, BRR model, it was applied to predict *S. damnosum* s.l. riverine sites in Northern Uganda. A total of 25 potential *S. damnosum* s.l. larval sites were predicted. Of the 25 sites predicted to be suitable *S. damnosum* s.l. aquatic habitats by the BRR model, 23 (92%; 95% CI 81–100%) were found to contain *S. damnosum* s.l. larvae. In contrast, just 2/10 (20%; 95% CI 0–45%) sites examined which were not predicted to represent *S. damnosum* s.l., aquatic habitat by the model were found to contain larvae. The BRR model thus exhibited a sensitivity of 80% and a specificity of 92% when applied in Uganda, a performance that was statistically significant (p, 0.0001; Fisher's Exact test). The two sites that were not predicted by the model which nonetheless were found to contain larvae consisted of low hanging streamside vegetation immersed in fast flowing water The mean number of larvae found at the sites predicted by the BRR model (21.91) was significantly greater that the mean number of larvae at the sitesconsisting of immersed overhanging vegetation (4.0; p<0.001, Mann Whitney U test)

Although post-classification and validating iteratively, quantitatively interpolated, explanatively diagnostic, geo-spectrotemporal, geosampled, parameterizable mosquito-related, eco-epidemiological, optimizable, covariates employing ArcGIS object-based classifiers and sub-meter resolution data requires extensive ground truthing to identify regions of high density transmission foci (e.g., positively autocorrelated high density, *Ae egyti,* oviposition sites in newly transitioned, forest-canopy, sparsely shaded, LULC to rice agriculture LULC change seasonal geolocations), these model outputs may be vital for implementing control interventions. Generally, positive autocorrelation occurs in iteratively interpolative, eco-epidemiological,

vector mosquito, larval habitat, data analyses when similar eco-epidemiological, count data, regressed values from spatiotemporal geosampled larval habitats cluster in geographic space, while negative autocorrelation occurs when unlike data sampled values cluster in space (Jacob et al. 2005a) .

Jacob et al. (2013) identified negative spatial autocorrelation (NSA), the tendency for dissimilar neighboring values to cluster on a predictive, *An. gambiae* s.l. map, constructed from mutliple, bio-geophysical multivariate, sub-meter resolution, optimally parameterizable, unbiased, iteratively, quantitatively interpolative, ArcGIS explanatively geoclassified LULC and meteorological covariate coefficients. Unquantified NSA generated from an inverse variance-covariance matrix in AUTOREG geo-visulized misspecifications in a vulnerability forecasting, eco-epidemiological, *An. gambiae* s.l., habitat model. The authors employed an eigenfunction decomposition algorithm based on a modified geographic connectivity matrix to compute the Moran's *I* statistic, to uncover hidden NSA in a dataset of georeferenced *An. gambiae* s.l. habitat explanatory predictor variables geo-spectrotemporally geosampled in Malindi and Kisumu, Kenya.

Moran's *I* Index value and both a Z score and *p*-value were optimally employed to evaluate the significance of an iteratively interpolative, eco-epidemiological dataset of *An. gambiae* s.l. covariates. In general, a Moran's Index value near +1.0 indicates clustering while an index value near -1.0 indicates dispersion (Griffith 2003). Without looking at statistical significance there is no basis for knowing if the observed pattern is just one of many, many possible versions of random in an entomological, geo-spectrotemporal, geospatial, sub-pmter resolution, vulnerability , forecasting, mprobabilisc paradigms (Jacob et al. 2005) . In Jacob et al. (2013) the Moran's *I* was optimnally defined as $I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$ where N was the number of spatial *An. gambiae* s.l. oviposition units indexed by i and j; X was the the immature larval count productivity; $\bar{X}$ was the mean of X; and w_{ij} was an element of a matrix of spatial weights. The expected value of Moran's *I* under the null hypothesis of no spatial autocorrelation inntye model resdiual forecast is $E(I) = \frac{-1}{N-1}$ Its variance equaed $Var(I) = \frac{NS_4 - S_3 S_5}{(N-1)(N-2)(N-3)(\sum_i \sum_j w_{ij})^2} - (E(I))^2$ where $S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2$

$S_2 = \sum_i (\sum_j w_{ij} + \sum_j w_{ji})^2$ $S_3 = \frac{N^{-1} \sum_i (x_i - \bar{x})^4}{(N^{-1} \sum_i (x_i - \bar{x})^2)^2}$ $S_4 = (N^2 - 3N + 3)S_1 - NS_2 + 3(\sum_i \sum_j w_{ij})^2$
$S_5 = (N^2 - N)S_1 - 2NS_2 + 6(\sum_i \sum_j w_{ij})^2$ Values of the range from −1 to +1. Negative values indicate negative spatial autocorrelation and positive values indicate positive spatial autocorrelation. A zero value indicates a random spatial pattern (Griffth 2003). For prise statistical hypothesis testing, the Moran's *I* values was transformed to Z-scores. Z-scores are expressed in terms of standard deviations from their means. Resultantly, these z-scores have a distribution with a mean of 0 and a standard deviation of 1. (Larsen,and Morris. 2000).

The Moran's *I* statistic was eco-epidemiologically, explanatively decomposed into time series, orthogonalized, synthetic, gridded, map patterns. Global tests revealed that $|z_{MC}|$s

generated were less than 1.11 for the presence of latent autocorrelation. The algorithm captured NSA in the *An. gambiae* s.l. habitat data by quantifying all non-normalized, random variables, space-time heterogeneity, and distributional properties of the spatial filters.

Processing of such eco-epidemiological, time series, vulnerability models may permit real-time monitoring and forecasting of immature, *Ae egypti* 0.61m, meter/mixel, spatial resolution data endmembers. The orthogonalized, georeferenceable eco-epidemiological, geolocation and expected mosquito, capture, point, immature, habitat counts from highly productive sites in a newly transitioned, discontinuous, forest-canopied, LULC to irrigated, riceland, agro-ecosystem LULCs (post-tillering, flooded) may then be formally predicted, using sub-meter resolution, satellite data and eco-epidemiological, field measurements of proxy habitat characteristics for identifying significant orthogonalizable, time series predictors of disease transmission at a neighboring, irrigated, study site. An ordinary quantitative interpolator can use the sub-mixel, sub-meter, resolution data along with other spatially continuous explanatory proxy variables geo-spectrotemporally geosampled from productive habitats for targeting other high density foci habitat sites which can help implement larval control strategies in any ecosystem -environment(Gu and Novak 2005, Jacob and Novak 2014) regradeless of mosquito species, abudance and distribution.

For compherensive Yellow fever predictive, model building, a conceptual quantity may be employeable in ArcGIS to quantitate the total productivity of an *Ae egypti* immature, capture point, georeferenceable, seasonally hyperproductive, oviposition sites in a newly, deforested, rice agriculture, canopied LULCs for cartographically and regressively representing the population of emerging female mosquitoes from all habitats. In so doing, total immature productivity can be partitioned into constituent parts from individual georeferenced capture point, seasonal habitats in ArcGIS in an irrigated African, rice-village, agro-ecosystem. In so doing, the effect of time series, larval interventions could be optimally represented by corresponding reductions in the total productivity in ArcGIS from the treated habitats. This model framework has two advantages over most models in literature currently which naivingly construct intervention forecasting paradigms using the assumption is that a general larval population count from one geosampled explanative "habitat" is identical for all habitats. First, although, immature habitat-based eco-epidemiological, vulnerability, endmiec, trasnmission-oriented, forecasting models are generalizable to a certain extent [e.g., geoclassifiable weekly geo-spectrotemporally uncoalesced, sub-meter resolution, near infra-red (NIR), meterological, wavelength radinace], they are implicit regarding both numbers and geolocations of habitats (see Jacob 2009). Second, the models can be specified with cartographically pertinet, empirical data employing geo-spectrotemporally, geospatially, orthogonalized, elucidative, uncoalesced, sub-meter resolution, (e.g., Quickbird 0.61m visible and NIR) sub-mixel, iteratively, quantitatively interpolative, signature estimates of habitat productivity (Jacob and Novak 2014). This new iteratively interpolative, model framework allows experimenters to examine impacts of larval interventions from the perspective of georeferenceable, seasonally geosampled,, seasonal habitats by delving into the complexity associated with landscape juxtapositions of individual habitats using newer remote technologies.

Specifically employing an eigenfunction decomposition algorithm may quantify residual autocorrelation error coefficients in sub-meter resolution, eco-epidemiological, bidirectional, radiance iterative interpolator for addressing the issue of observational correlation among georeferenced, georeferenceable, *Aedes* mosquito habitat, probabilistically regressable, empirical data. paradigms. In the mathematical discipline of linear algebra, eigendecomposition or sometimes spectral decomposition is the factorization of a matrix into a canonical form, whereby the matrix is represented in terms of its eigenvalues and eigenvectors (Griffith 2003). Only diagonalizable matrices can be factorized in this way. A (non-zero) vector v of dimension N is an eigenvector of a square ($N \times N$) matrix A in a vector arthrpod-related, risk model if it satisfies the linear equation $\mathbf{Av} = \lambda\mathbf{v}$ where λ is a scalar, termed the eigenvalue corresponding to v(Jacob et al. 2009). That is, the eigenvectors are the vectors that the linear transformation A merely elongates or shrinks, and the amount that they elongate/shrink by is the eigenvalue. The above equation is called the eigenvalue equation or the eigenvalue problem.This yields an equation for the eigenvalues $p(\lambda) := \det(\mathbf{A} - \lambda\mathbf{I}) = 0$. $p(\lambda)$ isthe characteristic polynomial, and the equation, called the characteristic equation, is an Nth order polynomial equation in the unknown λ. (Griffith 2003) This equation will have N_λ distinct solutions, in a yellow fever forecast, vulnerability model where $1 \leq N_\lambda \leq N$. The set of solutions, that is, the eigenvalues, is called the spectrum of A(Griffth 2003) An experimenetre may factor p as $p(\lambda) - (\lambda - \lambda_1)^{n_1}(\lambda - \lambda_2)^{n_2}\cdots(\lambda - \lambda_k)^{n_k} - 0$. in a datset of regressed geo-epsctrotemrpoally uncoalesced, YFV-related diagnostic, clinical, field or remote geosampled, variables The integer n_i is termed the algebraic multiplicity of eigenvalue λ_i (Anselin 1995). The

$$\sum_{i=1}^{N_\lambda} n_i = N.$$

algebraic multiplicities in the model would then sum to N: For each eigenvalue, λ_i, we have a specific eigenvalue equation $(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{v} = 0.$ In so doing, there will be $1 \leq m_i \leq n_i$ linearly independent solutions to each YFV-related, iteratively interpolative, eigenvalue equation. The m_i solutions are the eigenvectors associated with the eigenvalue $\lambda_{i(\ Griffith\ 2003)}$. The integer m_i is termed the geometric multiplicity of λ (Cliif and Ord 1971)$_i$. It is important to keep in mind that the algebraic multiplicity n_i and geometric multiplicity m_i may or may not be equal in a forecasting, vulnerability, eco-epidemiological, predictive risk model, but the resdiual forecasts will always have $m_i \leq n_i$. The simplest case is of course when $m_i = n_i = 1$ in a yellow fever model. The total number of linearly independent eigenvectors, N_v, may be then calculated by summing the geometric multiplicities

If an experimenetr lets A be a square ($N \times N$) matrix in a yellow fever eco-epidemiological, vulnerability model with N linearly independent diagnostic, clinical, field or remote geo-spectrotemporally iteratively interpolative eigenvectors, $q_i\ (i = 1, \ldots, N).$ Then A can be factorized as $\mathbf{A} = \mathbf{Q\Lambda Q}^{-1}$ where Q is the square ($N \times N$) matrix whose i^{th} column is the eigenvector q_i of A and Λ is the diagonal matrix whose diagonal elements in AUTOREG are the corresponding eigenvalues, *i.e.*, $\Lambda_{ii} = \lambda_i$. Note that only diagonalizable matrices can be factorized in this way. For example, a defective, YFV-related, diagnotsic, time series dependemt, probabilistic matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ cannot be diagonalized. In linear algebra, a defective matrix is a square matrix that does not have a complete basis of eigenvectors, and is therefore not diagonalizable. In particular, an $n \times n$ matrix is defective if and only if it does not have n linearly independent eigenvectors.(Griffth 2003) . The yellow fever regressed spatsil fielter , orthogonalized,

synthetic, eigenvectors q_i $(i = 1, \ldots, N)$ would be normalized, but they need not be. A non-normalized set of eigenvectors, v_i $(i = 1, \ldots, N)$, can also be used as the columns of Q in the model. That can be understood by noting that the magnitude of the eigenvectors in Q gets canceled in the decomposition by the presence of Q^{-1} during the regression.

Taking a 2 × 2 real matrix $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}$ as an example to be orthogonally decomposed into a explanative, time series, diagonal matrix of a yellow fever, heursitically optimizable, eco-epidemiological, sub-meter resolution, endmember dataset of georeferenceable covariate coefficients using a multiplication of a non-singular matrix $\mathbf{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbb{R}^{2 \times 2}$ .Then

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{1} \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$$, for some real diagonal matrix $\begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$.

Shifting $\mathbf{B}$ to the right hand side:would then render $\begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$

The above equation can be decomposed into 2 simultaneous equations:

$$\begin{cases} \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} = \begin{bmatrix} ax \\ cx \end{bmatrix} \\ \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} by \\ dy \end{bmatrix} \end{cases}$$ Factoring out the eigenvalues x and y: $\begin{cases} \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ c \end{bmatrix} = x \begin{bmatrix} a \\ c \end{bmatrix} \\ \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} b \\ d \end{bmatrix} = y \begin{bmatrix} b \\ d \end{bmatrix} \end{cases}$

Letting $\vec{a} = \begin{bmatrix} a \\ c \end{bmatrix}, \vec{b} = \begin{bmatrix} b \\ d \end{bmatrix}$, this could render two elucidative, time series, vector equations: $\begin{cases} A\vec{a} = x\vec{a} \\ A\vec{b} = y\vec{b} \end{cases}$ which may be cartographically represented by a single vector equation involving 2 solutions as eigenvalues: $\mathbf{Au} = \lambda\mathbf{u}$ where λ would represent the two diagnostic YFV-related, geo-epsctrotemporal, diagnostic, explicative, clinical, field or remote geosampled, sub-meter resolution, decomposed eigenvalues x and y, $\mathbf{u}$ which would represent the vectors $\vec{a}$ and $\vec{b}$.

Shifting $\lambda\mathbf{u}$ to the left hand side of the probabilistic, yeallow ferver, time series paradigm and factorizing $\mathbf{u}$ would then render $(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = 0$. Since $\mathbf{B}$ would be non-singular in the explanatorial, resdiual, orthogonlaized, uncoalesced forecasts taregting the hypeproductive *Ae egypti* habitats on a newly trasnitioned, sub-meter resolution, georeferenceable, discontinuous, forest canopied, dense or sparsely shaded, geoclassified, LULC to irrigated, riceland agro-ecosystem, ovipoistion, immature, capture point, seasonal habitat, in an African agro-village complex it is essential that $\mathbf{u}$ is non-zero. Therefore, $(\mathbf{A} - \lambda\mathbf{I}) = 0$ Considering the determinant of $(\mathbf{A} - \lambda\mathbf{I})$, in the vulnerabilit , yellow fever, eco-epidemioloigical, forecasting model would then

reflect $\begin{bmatrix} 1-\lambda & 0 \\ 1 & 3-\lambda \end{bmatrix} = 0$.Thus$(1-\lambda)(3-\lambda) = 0$ where the solutions of the eigenvalues for the matrix $\mathbf{A}$ would be $\lambda = 1$or $\lambda = 3$, and the resulting diagonal matrix from the eigendecomposition of $\mathbf{A}$ would be $\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$..Putting the solutions back into the above simultaneous equations $\begin{cases} \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} a \\ c \end{bmatrix} = 1\begin{bmatrix} a \\ c \end{bmatrix} \\ \begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} b \\ d \end{bmatrix} = 3\begin{bmatrix} b \\ d \end{bmatrix} \end{cases}$. Solving the equations, an expeimenter would then have $a = -2c, a \in \mathbb{R}$and $b = 0, d \in \mathbb{R}$. Thus, the matrix $\mathbf{B}$ required for the optimal eigendecomposition of $\mathbf{A}$ in the yellow fever model would be $\begin{bmatrix} -2c & 0 \\ c & d \end{bmatrix}, [c,d] \in \mathbb{R}$.

$\begin{bmatrix} -2c & 0 \\ c & d \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} -2c & 0 \\ c & d \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}, [c,d] \in \mathbb{R}$ If matrix A can be eigendecomposed and if none of its eigenvalues are zero in the predictive risk model for taregting hyperproductive *Ae egypti* riceland habitats on newly trasnitioned discontinuous, forest-canopied,sub-meter resolution, geoclassified LULCs on rice agriculture , sub-meter resolution, geoclassified LULC, then A would be nonsingular and its inverse would be given by$\mathbf{A}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^{-1}$

Furthermore, because Λ is a diagonal matrix in a robust, geo-spectrotemporal, geosampled, yellow fever, eco-epidemiological, forecast, vulnerability model, its inverse would eb easy to calculate: $\left[\Lambda^{-1}\right]_{ii} = \frac{1}{\lambda_i}$ [1.1]When eigendecomposition is used on a matrix of measured, regressable diagnostic, YFV-related , sub-meter resolution, clinical, field or remote geo-spectrotemporally geosampled, georeferenceable data, the inverse may be less valid when all eigenvalues are unmodified in equation 1.1. This is easily realized since eigenvalues become relatively small when their contribution to the inversion is large (see Griffth 2003). Those YFV eigenvalues near zero or at the "noise" of the measurement system for the diagnostic, explicative, yellow fever, model estimators would have undue influence and would hamper solutions (e.g., remote detection of explanatively significant covariates associated to a transitioned, forest, discontinuously canopied, sub-meter resolution, geo-spectrotemporally geoclassified LULC using the inverse.

Two mitigations have been recently proposed by Jacob et al. 2015 for 1) truncating small/zero *S. damnsoum* s.l. eigenvalues for extending the lowest reliable eigenvalue to those below it in a forecasting, vulnerability, eco-epidemiological, endmember , sub-meter resolution, probabilistic paradigm. The first mitigation method is similar to a sparse sample of the original matrix for removing components that are not considered valuable in the model. However, if the solution or detection process is near the noise level, Jacob et al. (2015) found that truncating may remove components that influence the desired solution. The second mitigation extended the time series *S. damnosum* s.l. synthesized , sub-meter resolution, immature, capture point, residual eigenvalue so that lower values has much less influence over inversion, but still contributes, such that solutions near the noise was found.

In Jacob et al. (2015) the reliable eigenvalue wa sfound by assuming that eigenvalues of extremely similar and low value were a good representation of measurement noise in a 3-D frecasting vulnerability, sub-meter resolution, georeferenceable, S. damnsoum s.l. iterative stochastic interpolator. The authors also found that the eigenvalues are rank-sorted by value, then the reliable eigenvalue can be found by minimization of the Laplacian of the sorted eigenvalues: $\min |\nabla^2 \lambda_s|$ where the eigenvalues are commonly subscripted with an 's' to denote being sorted. The position of the minimization is the lowest reliable eigenvalue (Griffth 2003). In measurement systems for remoetly tergeting, newly trasnitioned, sub-meter resolution, discontinuous, forest canopied, ArcGIS delineated LULC , seasoanlly trasnitioned to a *Ae egypti,* rice agriculture , hyperproductive, immature, capture point, georferenceable, geosampled habitat , the square root of this reliable eigenvalue could be the average noise over the components of the system.

Conventional ANOVA may be inappropriated to simultaneously deal with independent variables of both categorical (e.g., habitat type) and continuous predictors (e.g., vegetation LULC coverage and distance to the nearest house.Jacob et al. (2008). Autocorrelation statistics may be generated using datasets of, geospectrotermporally uncoalesced seasonally prolific, *Ae egypti* signaturizable, sub-meter resolution LULC , paramterizable predictors on newly transitioned deforeseted, rice agriculture, oviposition site in an expanding, irrigated, African riceland agro-ecosystem within a a negative binomial regression matrix with a non-homogenous mean which may reveal that the ouputs generated from the residual forecasts. Commonly, over-Poission variation occurs in vector medical entomological, real-time , sub-meter resolution, eco-epidemiological, vulnerability models especially when targeting forecasts of prolific, georeferenceable, hyperproductive, oviposition sites. Overdispersion occurs in spatial filter hierarchical models commonly due to to outliers (Griffith 2003). Spatial outliers in empericial dataset of vector insect larval habitat, seasonally, geo-spectrotemporally geosampled, fractionalized, time series, iteratively interpolative, endmember data analyses can generate a misspecified model (Jacob et al. 2012, Griffith 2006). Therafter, an expeimenter may employ the spatial filter, orthogonalized, endmember spatial filter, synthetic eigenvectors to determine if the *Aedes* geo-spectrotemporally geosampled, immature, capture point, data represent positive or negative latent autocorrelation in   time series clusters constructed from varying and constant, elucidative, endmember, predictors on newly transitioned, deforested, sparsely or dense shaded, discontinuous, rice agriculture, immature habitat, geoclassified LULC.

In order to quantitate, explanatively, residualized heteroskedaticity(i.e., uncommon variance) among the optimal regressable dataset of diagnostic, YFV, clinical, field or remote geo-spectrotemporally geosampled, georeferenceable paramterizable covariates in the spatial filter, hierarchical, cluster-based, eco-epidemiological models. Heteroskedascity occurs when the standard deviations of a sampled variable, monitored over a specific amount of time is non-constant, (Hosmer and Leneshew 2000). Non–homoskedastic quantiative yellow fever endmember decomposed, iteratively intrepolative, explanative, georeferenceable variables can violate the assumption of common variance in an entomological, real-time, cartographic, logistic/Poisson predictive, geographic, forecast, vulnerability model especially when seasoanlly targeting productive, immature habitat, *Ae egeypti,* eco-epidmiological, sentinal site, capture points on a newly transitioned deforested, irrigated, rice agriculture, sub-meter resolution,

imaged ArcGIS geoclassified LULC. Non-quantification of heteroskedastic parameters in a vector insect larval habitat, endmember distribution model can generate specification error distrinution covariates). Homoskedastic variables can violate the assumption of common variance in an entomological, predictive, geographic, forecast, vulnerability, eco-epidemiological, forecasting vulnerability model targeting productive, immature habiat capture points on a newly transitioned deforested irriagted rice agriculture , capture point. For example, Jacob et al. (2009) employed an error autocovarite matrix and orthogonalized, sub-mter resolution (i.e., panchromatic QuickBird) spatial filter eigenvectors in AUTOREG to spatially adjust georeferenced, predicted prolific, *An. arabiensis* larval habiats in ArcGIS of an irrigated, riceland, agro-village complex in the Mwea Rice Scheme in Central Kenya

In order to mathematically theorectical and optimally operationally logical idea of Gu and Novak (2005) for impelmenting yellow fever control staregeies in a newly deforested , rice agriculture, habiutat in an irrigated African, agro-village environment sub-meter resolution LULC three scenarios of larval control (YF1, YF2, and YF3), can br assumed. For optimally regressively qualitatively quantitating transitional ArcGIS geoclassifiabel, sub-meter resolution, imaged, explanatorial, prolific, uncoalesced, LULC, *Ae egypti,* ovipoistion irrigated, georefernced, ricefield, ploughed habitats, for example, YF1 would represent a situation in which all the immature, capture point, geosampled habitats were identical in contribution to the total productivity P. This scenario is similar to the assumption of Killeen and others[26] except that Gu and Novak (2005) measured productivity as proportional contributions of individual habitats rather than absolute numbers of emerging mosquitoes. Clearly, this difference is trivial in this scenario due to the assumption of identical habitats. For YF1, reduction in the total productivity P would then be a linear function of levels of coverage (C) of habitats under treatment
$$P = 1 - C \quad [1.1]$$

In the next two scenarios, more realistic explicative situations where *Ae. egeypti* immature habitat productivities geo-spectrotemrpoally geosampled on deforested canopied LULC trasnitionial, riceland seasonal habitats would not be not uniform in an irrigated African ago-village ecosystem with some prolific habitats contributing extremely large amounts of emerging adults.This data may be orthogonally devised in ArcGIS cyberenvironmental geodatbase (see Jacob and Novak 2014). For untargeted interventions, YF2 would ideally represent georeferenceable, geocalssifiable, seasonal aquatic habitats which may be randomly chosen for treatment. This scenario would occur when larval control is conducted with little knowledge of habitat productivity. Because of enormous variability in adult productivity observed in the field,[12,27,28] the majority of the total, *Ae egypti* larval productivity in an ArcGIS, simulated, forecasting, vulnerability, deforested, canopied LULC to ricefield ploughed, capture point, geosampled habitat, for instance, may actually originate from a small number of highly prolific habitats on the LULC. Under this circumstance, the random choice of habitats for treatment is likely to miss those prolific habitats unless large proportions of habitats are selected for treatment using newer cartographic gadgeteries (e.g., gridded ArcGIS sub-meter resolution matrices, obeject based classifiers in ENVI, etc) Therefore, the relationship between P and C may be optimally generally describeable using a logistic function in ArcGIS as

$$P = 1/(1 + e^{\alpha+\beta C}), \text{ for } C < 1$$
$$P = 0, \text{ for } C = 1 \quad\quad [1.2]$$

Alternatively, targeted yellow ferver interventions YF3 could represent a scenario where the interventions are targeted toward highly productive immature habitats on a georeferenceable, explanatively geoclassifieable, newly trasnitioned, deforested canopy to rice agriculture, immature, seasonally geo-spectrotemporally geosampled habitat. In so doing, regressive quantization of small proportions of the georfereneced, aquatic habitats may obtain large proportional reductions in the total *Ae egypti,* immature productivity on these LULC change sites. In so doing, the relationship between *P* and *C* can then be approximated using an exponential curve

$$P = e^{\delta C}, \text{ for } C < 1$$
$$P = 0, \text{ for } C = 1 \quad [1.3]$$

where δ is a constant reflecting degrees of aggregation in productivity among the immature geosampled habitats on the newly trasnitioned deforested rice agriculture habitat

The equations in 1.1 to 1.3 described here can be easily modified in ArcGIS to optimally represent a wide spectrum of larval control practices in newly transitioned, georeferenceable, deforested, sparsely or densecanopied, geoclassifiable, sub-resolution, time series, uncoalesced, iteratively interpolative, seasonally tranitioned, riceland LULCs using meterological and other YFV-related, illuminatively diagnostic, clincial, field or remote, geosampled, *Ae egypti*, explanative, geo-spectrotemporal, immature habitat,k sub-meter resolution regressors. In a specific larval control program, the relationship between *P* and *C* should be empirically estimated based on data of both larval and habitat surveys. For instance, the habitat productivity on a deforested, sparsely shaded, discontinuously canopied, rice-flooded, seasonal, georeferenceable, prolific oviposition LULC site can be estimated as a product of estimates of emerging adults or pupal density (D_i) and size (S_i) of habitat *i*. Therefore, for a focal area on a seasoanlly transitioned, *Ae egypti*, irrigated, riceland, African, expanding agro-village, eco-agriculture, signaturizable., iteratively quanatiaively interpolative, orthogonally quantiatively decomposeable, sub-meter resolution,   geoclassifiable, LULC with *n* immature habitats, the constituent contribution (M_i) to the total productivity from the capture point *i* imay be estimated as

$$M_i = \frac{D_i S_i}{\sum_{i=1,\ldots,n} D_i S_i} \times 100\%. \quad [1.4].$$

Because habitats often exhibit seasonal changes after rainfall patterns in tropical Africa, ranking and prioritizing immature, seasonal, geo-spectrotemporally, geosampled, signaturizable, sub-meter resolution, seasaonlly imaged, immature, habitats in an ArcGIS cyberenvironment need to to track the temporal changes in habitat productivity (see Jacob and Novak 2014).

Note that *P*, ranging from 0 to 100%, would be the percent productivity associated with levels of coverage of immature habitats under treatment in a yellow fever, , vulnerability-oriented, seasonal, ArcGIS-derived, robust, agro-ecosystem, irrigated riceland, sub-meter resolution, capture point, eco-epidemioloigical forecasting model. It has been well recognized that the goals of any control intervention should be remotely established based on both mosquito abundances and transmission intensities in African riceland environments (Jacob et al. 2007, Muturi 2007, Mwnagangi et al. 2008, Mururi et al. 2007). To parsimoniously eco-geographically precisely cartographicallyu represent seasonal variability in mosquito abundance

in various irrigated, deforested, sub-meter resolution, uncoalesced, geo-spectrotemporalized, ArcGIS geoclassifiable, georeferenceable, LULC, riceland areas, a parameter γ may be introduced to the empirical eco-epidemiological, optimizable datset of diagnostic geosampled regressands representing the base level of emerging female, *Ae egypti* mosquitoes, per person, per day. γ may reflect local diagnostic, clincial, field or remote characteristics influencing latent, mosquito-related, habitat proliferation of suitability charateristics such as abundance and quality of larval habitats in transitoned, deforested, sparse or densely, discontinuously canopied LULC to rice agriculture geoclassified LULC in ArcGIS-derived, georeferenceable, sub-meter resolution, imaged, area. Therefore, emerging females per person per day under the scenarios of yellow fever larval interventions can be calculated as γ*P*. Two base levels γ = 1 and 5 for the intervention analysis may be further exploited as these values may provide rise to low and intermediate levels of transmission intensities. In practice, parameter γ may not be directly measurable in the new ricefields , but it may be estimated from man-biting rates or entomological inoculation rates(EIR).

Transmission intensity, measure by EIR, is a fundamental predictive explanator of incidence and prevalence of YFV [31–35] Conventionally, EIR is estimated as a product of man biting rate (*ma*) and proportion of sporozoite (*s*) infected mosquitoes as EIR = *mas*. Because only proportions of emerging female mosquitoes that survive the extrinsic incubation period are capable of transmitting the parasite (Novak et al. 2012), assuming that daily mosquito mortality (*d*) is age-independent may substitute *m* with $\gamma P e^{-dT}$ in equation 1.4. Therefore, EIR is calculatable in an expanding, georeferenceable, irrigated African, rice-agriculture ecosystem due to increasing anthropogenic population employing $EIR = \gamma P e^{-dT} as$ .The computated values from the geo-spectrotemporally uncoalesced, non-orthogonalized, iteratively interpolative, sub-meter-resolution, regressable, georeferenceable, optimally parameterizable, seasonally georefernecable, prolific sylvatic, *Ae egypt,i* LULC, ovipoistion site geosampled, covariate coefficients on a newly transitioned, discontinuous, forest-canopied, LULC to riceland irriigted agriculture habitat may be listed seasonally in an ArcGIS geodatabase cyberenvironment. These robustifiable values may be also examined employing newer autorgressive paraidgms in SAS (PROC LOGISTIC). These model estimates can yield monthly EIR values using geosampled *Ae egypti* ,rice agro-village, ecosystem, data variables., for example, for the two levels of γ, respectively which would then be subsequently overlaid in ArcGIS onto georeferenceable, gridded, sub-meter resolution, stratifiable, deforested, canopied, rice agriculture, time series, expanding interface regions These EIR levels may correspond to low and intermediate levels of YFV transmission intensity in a geoclassifiable, georeferenceable, irrigated African, complex ecosystem especially on newly transitioned deforested, dense or sparsely, discontinuosly canopied, LULC along the periphery of the riceland agro-village. An experimenter may extend the analyses to situations of high yellow fever transmission in these LULCs because then the relationships between incidence and prevalence and transmission intensity in a georefernceable, irrigated, riceland agroecosystem, seasoanlly prolific, *Ae egypti* immature habitat may be complicated to quantiate regressively due to inconspciocu convulated, diagnostic, clinical, field or remote, time series variables (e.g., acquired protective immunity). Equation 1.4 may establish a quantifiable relationship between EIR and the total productivity *P*, which in turn amy be a function of the level of coverage of larval habitats *C* based on Equations 1.3. Therefore, it is possible to incorporate control interventions targeting adult mosquitoes such as with insecticide treatment by manipulating mortality rate *d* in Equation 1.4.

Next, an experimenter may examine how sub-meter constructed LULC changes in an ArcGIS cyberenvironment in newly transitioned, deforested canopy into rice agriculture irrigated African, ecosystem immature, capture point, georfernceable, explanative *Ae. aegypti* habitats, and EIR influence both yelloe fever incidence and prevalence. Association between incidence and EIR is affected by several entomological factors including vector competence and host susceptibility. Several studies have shown that not every infectious bite leads to an infection in the susceptible host.[39–41] There are evidences that infection rates tend to decline with age sting an enhanced protective immunity based on repeated exposures.[37] Additionally, various mosquito species may have different vector competence depending on geolocational, seasonal, quanatifiable, environmental attributes. In traditional geo-spectrotemporal, geospatial, vector arthrpod, forecasting, vulnerabilty-orinted, habitat models, this complexity may be incorporated into a parameterizable covariate estimator reflecting the probability of an uninfected person becoming infected due to an infectious bite *b*. Assuming that the outcomes of individual infectious bites on an uninfected host are independent, a probabilistc geospectrotermpoal, sub-meter resolution, autoregressive habitat model may decsribe describe the probability of infection (*I*) as a function of EIR on a ArcGIS, explicatively georferenceable newly trasnitioned, deforested, sparsely or densely canopied, sub-meter resolution, geoclassifiable, LULC to irrigated African, rice agriculture, oviposition, immature, capture point, georeferenceable site employing - $I = [1 - (1 - b)^{EIR}]$[1.5].

If assuming exposure to infectious bites of sylvatic, *Ae egypti* is uniform among humans on transitional, geoclassifiable, ArcGIS-derived, sub-meter resolution LULC form forest canopy to rice agriculture LULC in an irrigated African, riceland ecosystem, the probability of infection *I* would then be equivalent to the proportion of uninfected persons (*N*) who were exposed and then became infected. Therefore, the incidence rate can be approximated as a product of *N* and *I*. Examining the relationships between yellow fever infection rates and the three scenarios of larval control of Gu and Novak 2005 may help remotely trageting prolific *Ae. egypti* aquatic immature these habitat, capture points as decribed in Jacob and Novak (2014).

In situations where larval control interventions, equation 1.5 can be modified to incorporate the amount of predicted seasoanlly geo-spectrotemporally, geosaptially geosampled, immature productivity counts data. By dividing the two groups based on whether the dependent/ respone variable is an explanative, count variable (i.e., non-binarized) in $I = (1 - w)[1 - (1 - b)^{EIR}] + w[1 - (1 - b)^{(1-f)EIR}]$(1.7) may optimally computate iteratively interpolative, georfernceable explanators where *w* is the percent of individual habitat predicted larval contributions, *f* is the percent reduction in exposure based on productive *Ae egypti* oviposition, transitioned, LULC, sub-meter resolution capture point , immature habitat geo-spectrotemporally geosampled paramterizable covariates within a gridded, ArcGIS-derived, geoclassifiable, LULC, change map.

Finally, the established relationship between prevalence (*p*) and EIR may be employed to examine the impact of larval control on YFV prevalence in an expanding irrigated, African, riceland agro-ecosystem due to increasing anthropogenic population. The relationship between prevalence and EIR may be described in ArcGIS cyberenvironment by the following equation in areas of low and intermediate transmission in a trasnitioned forest-canopy LULC to rice-agriculture, eco-georeferenceable, sylvatic, *Ae.aegypti*, oviposition sites without consideration

of acquired protective immunity $p = EIR/(EIR + r)$ [1.8]. where r is recovery rates, as measured by the reciprocal of the infection period. For calculation purposes, EIR and r should have the same unit, either daily or monthly. If, for example, $r = 0.01$ time series, paramterizable covariates influencing YFV endmeic trasnmission may be regressively and cartographically illustrattable in ArcGIS and ArcGIS friendly software packages (e.g.,Python, ENVI). Previous studies on the coastal Kenya[35] and others[45] have suggested much lower recovery rates in medically entomological, vector arthrpod-related immature habitat prognosticating, endemic, trasnmissison models. Generally, adoption of lower recovery rates makes the effect of larval control on prevalence less remarkable than observed (Gu and Novak 2005)

In contrast, targeted interventions YF3 can represent a scenario where control interventions are targeted toward highly productive *Ae. aegypti* habitats on trasnitional landscapes in an irrigated riceland, eco-epidemiological, georeferenceable, study site. Therefore, management of small proportions of aquatic habitats can optimally obtain large proportional reductions in the total productivity. The relationship between *P* and *C* can also be approximated employing an exponential curve

$$P = e^{\delta C}, \text{ for } C < 1$$
$$P = 0, \text{ for } C = 1 \quad [1.9]$$

where δ is a constant reflecting degrees of aggregation in productivity among prolific, immature, geo-spectrotemrporally geospatialized, immature habitats

Equations 1.1–1.3 described here can be easily modified to represent a wide spectrum of *Ae. aegypti* larval control practices in an expanding, irrigated African, riceland agro-ecosystem, village complexes. In a specific larval control program targeting newly transitional, deforested, discontinuously canopied, geocalssifiable, georefernceable, time series, sub-meter resolution, LULCs may quantiate the relationship between *P* and *C* employing empirically estimated data from both larval and habitat surveys. For instance,an *Ae egypti* immature, capture point, previously deforested,sparsely canopied, LULC habitat productivity can be estimated as a product of estimates of emerging adults or pupal density (D_i) and size (S_i) of the rice agriculture, habitat *i*. Therefore, for a focal area with *n* habitats, the constituent contribution (M_i) to the total productivity from the vector arthropod habitat *i* could be estimated as

$$M_i = \frac{D_i S_i}{\sum_{i=1,\ldots,n} D_i S_i} \times 100\%. \quad [1.11]$$

Because habitats often exhibit seasonal changes after rainfall patterns in tropical Africa, ranking and prioritizing habitats need to proceed to track the temporal changes in LULC change habitat productivity using sub-meter resolution, remotely sensed , georeferenceable wavelength, meterological data and an optimizable empirical datast of explicatively diagnostic, clinical, field or remote,geo-spectrotemporally geosampled,optimally parameterizable, uncoalesced covariate coefficient values. Note that *P*, ranging from 0 to 100%, would be the percent productivity associated with levels of coverage of the seasonally prolific, sylvatic, *Ae. aegypti* rice-agriculture habitats on a transitioned, discontinuous, deforested, canopied LULC under treatment.

Operationally the goals of any yellow fever control intervention should be established based on both mosquitoabundances and transmission intensities To represent variability in

mosquito abundance in various LULC areas (e.g., transitioned, forest canopy to ploughed riceland, immature, capture point, georeferenceable, immature habitat plot). By introducing a time series, parameterizable, meteorological covariate γ as the base level of number of emerging female mosquitoes per person per day, γ may aid in reflecting local characters influencing mosquito proliferation such as abundance and quality of larval habitats in a new rice-agriculture expanading LULC, habitat area. Therefore, emerging females per person per day under the scenarios of larval interventions can be calculated as γP.

Transmission intensity, measure by EIR, is a fundamental predictor of incidence and prevalence of vector arthpod related diseses [31–35] Conventionally, EIR is estimated as a product of man biting rate (*ma*) and proportion of sporozoite (*s*) infected mosquitoes as EIR = *mas*. Because only proportions of emerging female mosquitoes that survive the extrinsic incubation period are capable of transmitting the parasite, it is vital to quantaite all geospectrotemporal geosampled, diagnostic, YFV-related, clincial, field or remote georfrenceable explanators. By assuming that daily mosquito mortality (*d*) is age-independent, an experimenter could substitute *m* with γPe^{-dT} in equation1.4. In so doing. EIR could be optimally calculated in a expanding irrigted African, riceland agro-ecosystem employing $EIR = \gamma Pe^{-dT}as$ [1.12].

Precisely regressed, explanatorial, seasonal trasnitioning, LULC estimates may yield monthly EIR values (e.g., 0.55 and 3.71)or annual values (e.g., 6.62 and 33.6), for optimally quantiting two levels of γ, respectively in a projected regressed, georfernceable geo-spectrotemproal, optimaizable, eco-epidemiologiocal, dataset of uncoalesced, iteratively interpolative, sub-meter resolution, YFV-related diagnostic, clinical, field or remote geosampled parameterizable, covariate coefficient, unbiased estimators. These EIR levels may correspond to low and intermediate levels of yellow fever transmission intensity typically found in tropical , irrigated, African, ,expanding agro-village ecosystems. The analyses may be up-graded to quantiate seasonal situations of high transmission because then the relationships between incidence and prevalence and transmission intensity may be stratified based in other variables (e.g., protective immunity) Equation 1.4 establishes a relationship between EIR and the total productivity *P*, which in turn may be a function of the level of coverage of *Ae egypti* , newly trasnitioned deforested canopied, cpature point larval habitats *C* based on Equations 1.3. Therefore, it is possible to incorporate control interventions targeting adult, sylvatic, *Ae.egypti* mosquitoes such as with targeted, prioritized, insecticides treatments by manipulating seasonal mortality rate *d* in Equation 1.4.

Next, an experimenter can examine how LULC changes and seasoanl EIR influence both yellow fever incidence and prevalence. Association between incidence and EIR is affected by several factors including vector competence and host susceptibility (Novak 2012). Several studies have shown that not every infectious bite leads to an infection in the susceptible host.[39–41] There are evidences that infection rates tend to decline with age suggesting an enhanced protective immunity based on repeated exposures.[37] Additionally, various unknown sylvatic deforested-canopy, YFV mosquito species may have different vector competence. In traditional forecasting, yellow fever, vulnerability, probabilistic paradigms, this complexity may be parsimoniously incorporated into a parameter reflecting the probability of an uninfected person becoming infected due to an infectious bite *b* on a transitioned deforested, time series, sparsely or dense, discontinuously canopied, georeferenceable, geo-spectrotemporally geosampled,

geoclassifiable, ArcGIS-derived, explanative, seasonal LULC to rice agriculture, flooded, *Ae egypti,* immature,capture point, sub-meter resolution topographic ArcGIS geo-classifiaction. Assuming that the outcomes of individual infectious bites on an uninfected host are independent, a YFV experimenter can use a binomial model in GEN MOD to describe the probability of infection (*I*) as a function of EIR in $I = [1 - (1 - b)^{EIR}]$[1.5]

In the "Criteria For Assessing Goodness Of Fit" table displayed in tha GEN MOD Output , the value of the deviance in a forecasting YFV vulnerability newly transitioned forest-canopeid LULC to rice agriculture LULC, the residuals (remoetly targeting hypeproductive, sylvatic, *Ae egypti ,* oviposition sites) may be divided by its degrees of freedom is less than 1. A *P*-value is not computed for the deviance; however, a deviance that is approximately equal to its degrees of freedom is a possible indication of a good model model fit. Asymptotic distribution theory applies to binomial data as the number of binomial trials parameter *n* becomes large for each combination of explanatory variables. McCullagh and Nelder (1989) caution against the use of the deviance alone to assess model fit. The model fit for each observation should be assessed by examination of residuals. The OBSTATS option in the MODEL statement produces a table of residuals and other useful statistics for each observation.

In the "Analysis Of Parameter Estimates" GEN MOD tables may be displayed by chi-square values for the explanatory diagnostic, time series, clincial, field or remote geo-spectrotemrpaolly geosampeld, georfernceable variables which may indicate that the parameter values other than the intercept term are all significant. The scale parameter may be set to 1 for the binomialized explanative, immature sub-meter resolution ,forest-canopy to rice –agriculture *Ae egypti* , habitat distribution. An overdispersion analysis may reveal the value of "exterme" (geospatial outiers) yellow fever ,seasonal observations may be indicated here. The preceding table may contain the profile, likelihood, confidence intervals for the explanatory variable sub-meter resolution, geocalssifiable, LULC parameters requested with the LRCI option (see ww.sas.edu). Wald confidence intervals may be also displayed by default. Profile likelihood confidence intervals are considered to be more accurate than Wald intervals (see Aitkin et al. (1989)), which may be applicable to geo-spectrotemporally georefernceable, small datasets of diganostic clinical , field or remote geosampled predictors. Specifying the confidence coefficient with the ALPHA= option in the MODEL statement may allow to see statistical parameter significance in the dataset. The default value of 0.05 may, correspond to 95% confidence limits. Profile likelihood confidence interval (CI) of the elucidative, regressed, georeferenceable, geo-spectotemporally geosampled, explicative,YFV-related seasonally prolifi,c *Ae egypti* parameterizable covariate estimators on transitioned,forest-canopy to rice agriculture LULCs may be optimally determined.

The standard procedure for computing a CI for a parameter in a generalized linear model is by the formula: estimate _ percentile _ SE(estimate), where SE is the standard error (Hosmer and Lemeshew 2002). The percentile is selected according to a desired confidence level and a reference distribution(a t-distribution for   diagnostic clinical, field or remote, sub-meter resolution   LULC regression coefficients in a linear model) otherwise a standard normal distribution. This procedure is commonly referred to as a Wald-type CI. It may work poorly if the seasonal, geo-spectrotemporally geosampled, *Ae egypit* habitat georefernceable, distribution of the parameterizable covariate estimators are markedly skewed or if the standard error is a

poorestimate of the standard deviation of the YFV-related explanative estimator. Since the standard errors in GLM's are basedon asymptotic variances obtained from the information matrix, Wald CI's from an eco-epidemiological, forecast, vulnerability, YF remote sensing LULC deforested canopy LULC to rice agriculture LULC , sub-meter resolution model may perform poorly for small to moderate sample sizes. Profile likelihood CIs don't assume normality of the estimator and appear to perform better for small samples sizes than Wald CI'S (Fotehringham 2002).Nonetheless, for a time series, forecasting, YFV-related , vulnerability, eco-epidemiological, probabilsic paradigm based on an asymptotic approximation – the asymptotic chi-square distribution of the log-likelihood ratio test statistic may be precisely quanatiated

In statistics, a likelihood ratio test is a statistical test used to compare the goodness of fit of two models, one of which (the *null model*) is a special case of the other (the *alternative model*). The test is based on the likelihood ratio, which expresses how many times more likely the data (e.,g, orthogonally decomposed, sub-meter resolution, gridded, YFV-reated, seasoanlly explanative , diagnostic clincial, field or remote, deforested canopy to rice agriculture, geoclassified LULCs are under one *Ae egypti* habitat model than the other. This likelihood ratio, or equivalently its logarithm, can then be used to compute a *p*-value, or compared to a critical value to decide whether to reject a yellow fever specified null model in favour of the alternative model. When the logarithm of the likelihood ratio is used, the statistic is known as a log-likelihood ratio statistic, and the probability distribution of this test statistic, assuming that the null model is true, can be approximated using Wilks' theorem (Cox 1974)

If the seasoanlly related, explanatively diagnostic YFV, clincial, field or remote , deforested canopy to rice agriculture, geoclassified, LULCs quantized distributions of the likelihood ratio correspond to a particular null and alternative hypothesis it can be explicitly used to determine if sub-meter, spatial resolutio,n imaged variables regions can be employed to accept/reject the null hypothesis. In most cases, however, the exact distribution of the likelihood ratio corresponding to specific hypotheses for an LULC forecasting vulbnerability yellow fever, geo-spectrotempora,l probabilistic paradigm is very difficult to determine. A convenient result for such a model may be that as the sample size $n$approaches ∞, the test statistic $-2\log(\Lambda)$for a nested YFV predictive, ricleand irrgated agro-costem model for remote targeting prolific, *Ae egypti* habitat model will be asymptotically χ^2-distributed with degrees of freedom equal to the difference in dimensionality of $\Theta$and Θ_0.This means that for a great variety of hypotheses testing employing a sub-meter resolution, ArcGIS, seasonally explanative, time seriesm risk anlyses for aggregating statistics about trasnitions in peripheral,sub-meter resolution, forest-canopy LULC to riceland agro-ecosystrem, the likelihood ratio may be optimally employed using $\Lambda$for the geospectrotemporal geosampled uncoalesced, itearatively quantatively interpolative. In so doing $-2\log(\Lambda)$ may be compared to the χ^2value corresponding to a desired statistical significance as an approximate statistical test for determining exact sub-meter resolution LULC parameterizable regressors. Employing the Wilks theorem..

Wilks' theorem assumes that the true but unknown values of the estimated parameters are in the interior of the parameter space. This is commonly violated in, for example, random or mixed effects models when one of the variance components is negligible relative to the others. In some

such cases with one variance component essentially zero relative to the others or the models are not properly nested, Pinheiro and Bates showed that the true distribution of this likelihood ratio chi-square statistic could be substantially different from the naive χ^2, often dramatically so.[7] The naive assumptions could give significance probabilities (*p*-values) that are far too large on average in some cases and far too small in other.

In general, to test random effects, Griffith (2005) recommends using restricted maximum likelihood (REML) for optimally quantizing parameterizable, iteratively interpolative, vector arthropod-related, time series, uncoalesced,, probabilistic, covariate coefficients. In statistics, the REML approach is a particular form of maximum likelihood estimation which does not base estimates on a maximum likelihood fit of all the information, but instead uses a likelihood function calculated from a transformed set of data, so that nuisance parameters have no effect (Bartlett. 1937). In the case of variance component estimation, the original geo-spectrotemrpoally geosampled, YFV-related, clinical, field or remote predictor datasets may be replaced by a set of contrasts calculated from the data, and the likelihood function may then be calculable from the probability distribution of these contrasts, according to the forecasting vulnerability model output. In particular, REML is useful as a method for fitting, linear, mixed, entomological models for targeting, sub-meter resolution, LULCs of immature, hyperproductive georeferenceable, geosampled, prolific habitats (Jacob et al. 2013). In contrast to the earlier maximum likelihood estimation, REML can produce unbiased estimates of variance and covariance parameters (Hosmer and Lemeshew 2002).

Simulated tests setting employing one and two random effects variances to zero may help quantiate LULC paramterizable, sub-relaution estimators of rica-agrovillage, eco- agriculture LULC into discontinuous, forest canopy LULC. Fortunately, the simulated p-values with *k* restrictions most closely match a 50-50 mixture of $\chi^2(k)$ and $\chi^2(k-1)$. (With *k* = 1, $\chi^2(0)$ is 0 with probability 1. This means that a good approximation may be $0.5\chi^2(1)$. Pinheiro and Bates (2000) also simulated LULC tests of different fixed effects. In one test of a factor with 4 levels (degrees of freedom = 3), they found that a 50-50 mixture of $\chi^2(3)$ and $\chi^2(4)$ was a good match for actual *p*-values obtained by simulation – and the error in using the naive $\chi^2(3)$" was not too alarming. However, in another test of a factor with 15 levels, they found a reasonable match to $\chi^2(18)$– 4 more degrees of freedom than the 14 that one would get from a naive (inappropriate) application of Wilks' theorem, and the simulated *p*-value was several times the naive $\chi^2(14)$." They authors concluded that for testing fixed effects, it's wise to use simulation. Further the authors provided a "simulate.lme" function in their "nlme" package for S-PLUS and R to support doing that. To be clear, these limitations on Wilks' theorem do not negate any power properties of a particular likelihood ratio test in a predictive, time series, explicative, yellow fever, forecasting vulnerability, sub-meter resolution, probabilsic paradigm only the use of a χ^2 distribution to evaluate its statistical significance.

If assuming exposure to infectious bites is uniform among humans on a newly transitioned riceland, sub-meter resolution geoclassifiable,geo-spectrotemporalnor geo-spatiotemrpoal, eco-epidemiological, optimizable LULC, the probability of infection *I* would be equivalent to the proportion of uninfected persons (*N*) who are exposed and then became infected

in a trasnitional forest canpy, sub-mter resolution *Ae egypti* , riceland agro-ecosystem, immature, capture point, eco-epidemiological, georferenecable, geosampled, habitat. Therefore, the incidence rate in a irrigated rice-agro-village, expanading ecosystem due to increaing anthropogenic populations can be approximated as a product of *N* and *I*. Examining the relationships between infection rates and the three scenarios of *Aedes* larval control may be optimally differentiated linerally and spatially by setting *N* = 100 and *b* = 0.5 in a GLM in a ArcGIS quanatited trasnitioned peripheral forest canopy , sub-meter resolution LULC converted into rice agriculture LULC in an agro- irrigated African ecosystem..

In situations where bednets are used in addition to larval control interventions, equation 1.5 can be modified to incorporate the protection of bednets. This can be obtained by dividing the human hosts into two groups based on whether bednets are used on the transitioned agro-village complex area employing: $I = (1 - w)[1 - (1 - b)^{EIR}] + w[1 - (1 - b)^{(1-f)EIR}]$ [1.6] where *w* is the percent of individuals who slept under bednets, *f* is the percent reduction in exposure protected by bednets. Analyzing a situation where larval control interventions are combined with a bednet program in which half (*w* = 0.5) of the population sleeping under bednets with *f* = 80% may seasonally forecast, hyperproductive, sub-meter resolution, sylvatic, *Ae. aegypti,* , LULC oviposition, capture points in an agro-village,georeferenceable, African, irrigated complex.

Finally, an experimenter may the establish relationship between prevalence (*p*) and EIR to examine the impact of larval control on yellow fever prevalence in an irrigated georefernceable, African, riceland, environment seasonally geo-spectrotemrporally geosampled, LULC In so doing, the relationship between prevalence and EIR can be described by the following equation in ArcGIS for optimally revealing LULC change areas of low and intermediate transmission without consideration of acquired protective immunity $p = EIR/(EIR + r)$ [1.7] where *r* is recovery rates, as measured by the reciprocal of the infection period. For calculation purposes, EIR and *r* should have the same unit, either daily or monthly. For instance, *r* = 0.01 may be employable as an explanatorial georeferenceable , geosampled *Ae egypti* decomposeable, seasonal quantized value which may be incorporated into some yellow fever modeling studies of expanding irrigated African LULC into forest canopied LULC due to increasing anthropogenic populations. Generally, adoption of lower recovery rates makes the effect of larval control on prevalence less remarkable than observed Gu and Novak 2005)

The optimizable, EIR, geometric, endmember, expanded definition of a probability space in a geo-spectrotemporally prolific, geospatialized, ecogeoreferenceable, vulnerabiltiy-oriented, YFV, eco-epidemiological, forecasting, ArcGIS model would be a triple $(\Omega, \mathcal{F}, P)$ consisting of the sample space Ω— an arbitrary non-empty set, the σ-algebra $\mathcal{F} \subseteq 2^{\Omega}$ (also called σ-field) — a set of subsets of Ω, called events, such that: 1) $\mathcal{F}$ contains the sample space: $\Omega \in \mathcal{F}$, 2) $\mathcal{F}$ is closed under complements: if $A \in \mathcal{F}$, then also $(\Omega \setminus A) \in \mathcal{F}$. $\mathcal{F}$ is closed under countable unions: if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$, then also $(\bigcup_{i=1}^{\infty} A_i) \in \mathcal{F}$ (Wasserman 2004). The corollary from De Morgan's law is that $\mathcal{F}$ is also closed under countable intersections: if $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$, then also $(\bigcap_{i=1}^{\infty} A_i) \in \mathcal{F}$ as defined using propositional logic and boolean algebra. De Morgan's laws are a pair of transformation rules that are both valid rules of inference which allow the expression of conjunctions and disjunctions purely in terms of each other via negation. In mathematics and mathematical logic, Boolean algebra is the branch of algebra in which the

values of the variables are the truth values *true* and *false*, usually denoted 1 and 0 respectively. (Neter 1990) Instead of just employing elementary algebra determining where the interpolated values of the riceland oviposition sub-mwter resolution, forest-canopied, discontinuous immature habitat, seasonal sampled, signature, uncoalesced, LULC variables are the main operations, the main operations of Boolean algebra are the conjunction *and* denoted as ∧, the disjunction *or* denoted as ∨, and the negation *not* denoted as ¬. It is thus a formalism for describing logical relations in the same way that ordinary algebra describes numeric relations.

The probability measure $P : \mathcal{F} \to [0,1]$ — a function on $\mathcal{F}$ such that P is countably additive. Hence, if $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{F}$ is a countable collection of explanative, optimizable, pairwise disjoint forecasting, sub-meter resolution, uncoalesced, iteratively interpolative, geo-spectrotemporal, newly transitioned, irrigated riceland agriculture habitat LULC datsets, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ where the measure of entire sample space in the model would be the probabilty zone targeting, hypeproductive, seasonal, *Ae. aegypti,* immature, capture point, georeferenceable, aquatic habitats on ArcGIS derived, discontinuous land use land cover (LULC) polygons. For describing newly trasnsitioned, African, agro-village, riceland, peripheral boundaries into deforested, sparsely or dense, forest canopy ArcGIS geo-classifiable, LULCs as equal to one: $P(\Omega) = 1$.

Figure 1.1 De Morgan's Laws represented with Venn diagrams revealing a geoclassified LULC change area from deforesed sparsely canopied , *Ae egypti* oviposition sites protuding into a irriagted African ,riceland agro-ecosystem, tillering habitat hypothetical ArcGIS 1) Union of two LULC datasets 2) Intersection of two LULC subsets
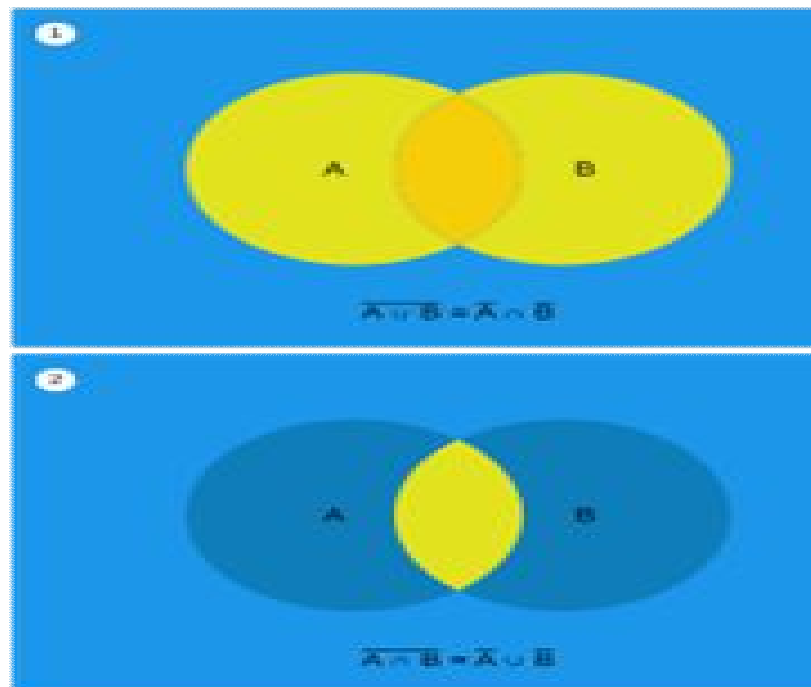
Extending the definition of a probability distribution of one randomized,sub-meter resolution, deforested canopy to rice agriculture LULC, explantively geoclassified variable to the joint probability distribution of two, geo-spectrotemporal, georeferenceable, YFV-related, sub-meter resolution, seasonal, *Ae. aegypti*, capture point, immature habitat,geosampled, random discretely diagnostic, geo-spatiotemporal or geo-spectrotemporal estimatorsX, Y could be resolved, $P(X = x \text{ and } Y = y) = P(Y = y \mid X = x) \cdot P(X = x) = P(X = x \mid Y = y) \cdot P(Y = y)$,where $P(Y = y \mid X = x)$ would be the probability of $Y = y$given that $X = x$ in the vulnerability endmember paradigm. The generalization of the preceding two explanatively, iteratively, quantitatively interpolative, forecasting, variable case would be the joint probability distribution of n optimally, discretely randomized, georeferenceable, diagnostic, time series, YFV-related, clincial, field or remote, geo-spectrotemporally geosampled semi-paramterizable variable dataset $X_1, X_2, \ldots, X_n$which could be subsequently quantitated as :

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1)$$
$$\times P(X_2 = x_2 \mid X_1 = x_1)$$
$$\times P(X_3 = x_3 \mid X_1 = x_1, X_2 = x_2) \times \cdots \times P(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \ldots, X_{n-1} = x_{n-1})$$

The PROBBNRM function can return the probability that an explicative seasonal orthogonally decomposeable, iteratively interpolative seasonally diagnostic, georeferenceable, explicative, clinical, field or remote uncoalesced, YFV, sub-meter reesolution, geoclassifiable, ArcGIS-derivable, LULC observation (X, Y) geosampled in an irrigated African, riceland, agro-ecosystem complex with expanding peripheral boundaries into discontinuous, neighboring, sparsely or dense covered, forest-canopied, seasonal LULC   from a standardized bivariate normal distribution with mean 0, variance 1, and a correlation coefficient **r**, is less than or equal to (**x**, **y**). That is,a vulnerability, probabilistic, linear regression, of the parameterizable YFV estimators in the geo-spatrotemporal or geo-spectrotemporal, eco-epidemiologically geosampled, empirical, optimizable dataset will return the probability that X$\leq$**x** and Y$\leq$**y** got determining seasonal diagnostic predictors associated with irrigated,rice agriculture, ArcGIS geoclassiable,time series LULCs. The following equation describes the PROBBNRM function, where **u** and **v** could optimally regressively represent a dataset of geo-spectrotemporally, uncoalesced, geosampled, sub-meter resolution, geoclassifiable, sylvatic, *Ae-egypti* ,LULC, probabilistic,eco-epidemiologically randomized, sub-meter resolution, endmember irriagted riceland agro-village variables **x** and **y**, respectively:

$$\text{PROBBNRM}(x, y, r) = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{x} \int_{-\infty}^{y} \exp\left[-\frac{u^2 - 2ruv + v^2}{2(1-r^2)}\right] dv\, du$$

A PROBBNRM function YFV model may apply the chain rule of probability to the forecasts targeting prolific, georeferenceable, *Ae egypti* oviposition, capture points LULC sites on sub-meter resolution, ArcGIS-derived predictive simulated maps of sparesely or dense shaded,discontinuous, forest canopy to rice-irrigated, geoclassified, agro-village LULC. In probability theory, the chain rule (also called the general product rule) permits the calculation of any member of the joint distribution of a set of random variables using only conditional

probabilities(Hazewinkle 2001). An experimenter may optimally consider an indexed set of geo-spectrotemporally geosampled, sub-meter resolution, georeferenceable, geoclassified, diagnostic, YFV-related clinical, field or remote-specified, explanatively paramterizable, data variables $A_1, \dots, A_n$. To find the value of this member of the joint distribution, an experimenter could apply the definition of conditional probability to obtain: $P(A_n, \dots, A_1) = P(A_n | A_{n-1}, \dots, A_1) \cdot P(A_{n-1}, \dots, A_1)$. Repeating this process with each final term would create the product: $P\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} P\left(A_k \,\middle|\, \bigcap_{j=1}^{k-1} A_j\right)$. With four elucidatively geoclassifiable, seasonal, geoclassifed, georeferenceable, *Ae egypti,* time series, decomposeable, forest deforested, transitioned rice agriculture, partially or dense, discontinuously canopied, decomposed, iteratively quantatively interpolative, sub-meter resolution, uncoalesced, LULC variables, the chain rule would produce conditional probabilities: $P(A_4, A_3, A_2, A_1) = P(A_4 | A_3, A_2, A_1) \cdot P(A_3 | A_2, A_1) \cdot P(A_2 | A_1) \cdot P(A_1)$.

Since these orthogonally quantitatively decomposed, geoclassified, georeferenceable, sub-meter resolution, explicative, time series, seasoanl geoclassified, LULC variables are probabilities, there would in a robust, time series, explicative, two-variable, forecasting vulnerability modelling case scenario, n $\sum_i \sum_j P(X = x_i \text{ and } Y = y_j) = 1,$ which could subsequently generalize for n discrete, explanatively diagnostic, regressed, georeferenceable, geosampled, time series, geo-spectrotemporalized, illuminative, endmember, randomized, predictive variables $X_1, X_2, \dots, X_n$ to $\sum_i \sum_j \cdots \sum_k P(X_1 = x_{1i}, X_2 = x_{2j}, \dots, X_n = x_{nk}) = 1.$ The joint probability density function $f_{X,Y}(x, y)$ for the continuous, YFV-related, diagnostic, clinical, field or remotely regressively randomized variables would then be equal to: $f_{X,Y}(x, y) = f_{Y|X}(y | x) f_X(x) = f_{X|Y}(x | y) f_Y(y)$ …where $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$ which would optimally render the conditional distributions of Y given $X = x$ and of X given $Y = y$ respectively. Consequently, $f_X(x)$ and $f_Y(y)$ would render the marginal distributions for X and Y respectively. Again, since these are probability distributions, the final model residual output would reveal s $\int_x \int_y f_{X,Y}(x, y)\, dy\, dx = 1.$ The "mixed joint density" may be also optimally defined where one randomized, time series, explicative, clinical, field or remote geo-spectrotemporally geosampled, yellow fever, regression variable where X is continuous and the other random variable Y is discrete, or vice versa, as: $f_{X,Y}(x, y) = f_{X|Y}(x | y) P(Y = y) = P(Y = y | X = x) f_X(x)$.

Parsimonioulsy regressively quantitating a sub-meter resolution, geo-spectrotemporal, ecogeoreferenceable, geo-spatiotemporal, geosampled, uncoalesced, iteratively interpolatable, geoclassifiable, discontinuous, *Ae egypti*-specified, sparsely shaded, deforesed, partially canopied, georeferenceable, LULC polygon protuding into a irrigated, rice agriculture, agro-ecosystem village in an ArcGIS geo-database cyberenvironment may reveal prolific, immature, *Ae egypti* eco-epidemiological, capture points and their explicative, correlation coefficients In so doing, the extent of at least two explanatively diagnostic, illuminative, time series, explanative, geo-spectrotemporal, clinical, field or remotely randomizable, eco-epidmiological, *Ae egypti,* geo-spectrotemporally geosampled datset of explanatively uncoalesced, geoclassifiable,

orthogonally elucidative, eco-epidemiological, seasonal, sub-meter resolution, YFV-related, LULC variables may be found to be linearly related. The residualized regressands may reveal a robust, bivariate distribution from the forecasting vulnerability, yellow fever, probabilistic paradigm.

In generalizable, entomological,seasonal, geocalssifiable, LULC, vector arthropod-related, diagnostic endmember ,signaturizable, forecasting vulnerability, sub-meter resolution, geo-spectrotemporally uncoalesced, iterative, interpolative datasets the multivariate normal distribution or multivariate Gaussian distribution, is a generalization of the one-dimensional (univariate) normal distribution to higher dimensions. One possible definition is that a random vector in these probabilisc paradigms has a k-variate normally distributed if every linear combination of its k components has a univariate normal distribution (see Jacob et al. 2005, Griffith 2005).In mathematics, probability, and statistics, a multivariate random variable or random vector is a list of mathematical variables each of whose value is unknown, either because the value has not yet occurred or because there is imperfect knowledge of its value (Kendrick, 1981). The individual yellow fevel-related, newly transitioned, geoclassified, discontinuous canopied, forest LULC to irrigated rice agriculture, uncoalesced,sub-meter resolution variables in a random vector may be grouped together because there may be correlations among them (e.g., different properties of an individual statistical wavelength, transmittance frequency of a *Ae egypti* immature, georeferenceable, oviposition, capture point, pre-flooded habitat unit). For example, while a given georeferenceable, prolific, seasonal deforested , sparsely canopied, transitioned LULC to a riceland pre-harvested, capture point,immature habitat has a specific emittance, the representation of any waveband from within a sub-meter resolution sensor would be a random vector. Normally each element of a random vector is a real discret integer .Random vectors are often used as the underlying implementation of various types of aggregate random variables, (e.g., a random matrix, random tree, random sequence, stochastic process, etc. (Hosmer and Lemeshew 2002) More formally, a multivariate, explanative, diagnostic, forecasting vulnerability, sylvatic, YFV-related eco-epidemiological, iteratively interpolative, geo-spectrotemporalized, endemic, transmission-orinted, predictive, random variable is a column vector $\mathbf{X} = (X_1, ..., X_n)^T$ (or its transpose, which is a row vector) whose components are scalar-valued, eco-epidemiological, LULC trasnitioned (forest discontinuous canopy to rice-agriculture), discrete, random variables on the same probability space $(\Omega, \mathcal{F}, P)$, where Ω is the sample space, $\mathcal{F}$ is the sigma-algebra (i.e., the collection of all geo-sampling, immature,capture point, georeferenceable habitat events in an irrigated, African, riceland, agro-village ecosystem), and P is the probability measure (a function returning each event's probability). The logic behinds this model processing derives mainly from the multivariate central limit theorem.

In probability theory, the central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent, explanative, diagnostic, random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution (Rice 1995). To illustrate what this means in a geo-spectrotemporal, yellow fever, forecasting vulnerability, sub-meter resolution, LULC-related, ArcGIS, probabilistic paradigm,suppose that a diagnostic YFV clinical, field or remote sample is obtained containing a large number of *Ae egypti* capture point, seasonal irrigated, African, riceland, agro-village, immature habitat quantitated

observations. Importantly, each georeferenceable, geosampled, immature habitat covariate observation would randomly be generated in a way that does not depend on the explicatively orthogonally quantitative, orthogonally decomposed regression values of the other sub-meter observations. In so doing, the arithmetic average of the observed, forest canopy to rice agriculture, quantizable, sub-meter resolution, parameterizable, LULC, geo-spectrotemporal uncoalesced, wavelength, transmittance, frequency variables would be computable. If this procedure is performed in PROC REG, for example, many times, then the computed values of the average will be distributed according to the normal distribution (commonly known as a "bell curve") based on the central limit theorem.

The central limit theorem has a number of variants. In its common form, the random variables must be identically distributed. In variants, convergence of the mean to the normal distribution also occurs for non-identical distributions or for non-independent observations, given that they comply with certain conditions (see Hosmer and Lemeshew 2002). In more general entomological modelling usage, a central limit theorem is any of a set of weak-convergence theorems in probability theory that can be employed to express a sum of many yellow fever , diagnostic, *Ae egypti* ,independent and identically distributed (i.i.d.) randomized, uncoalesced, geo-spectrotemporal, sub-meter resolution, LULC, explanative variables, or alternatively, random variables with specific types of dependence, will tend to be distributed according to one of a small set of attractor distributions. When the variance of the i.i.d., geo-spectrotemporal, optimally regressand ,time series, YFV-related, diagnostic, empirically uncoalesced sub-meter resolution, explanatorial,optimizable,eco-epidemiological datasets of georeferenceable, orthogonally, quantitatively, decomposeable, clincial, field or remote variables is finite, the attractor distribution would be the normal distribution. In contrast, the sum of a number of explicative, i.i.d., time series, YFV-related, randomized, elucidative variables with power law tail distributions would decrease as $|x|^{-\alpha-1}$ where $0 < \alpha < 2$ (and therefore having infinite variance) will tend to an alpha-stable distribution with stability parameter (or index of stability) of α as the number of geo-spectrotemporally geosampled, optimizable, parameterizable, LULC, explanatorial, georferenceable estimator, *Ae egypti* capture point, decomposeable dataset time series, discontinuous irrigated African riceland agro-ecosystem LULCs grows due to anthropogenic ecological pressures.

In probability theory, a distribution is said to be stable (or a random variable is said to be stable) if a linear combination of two independent copies of a random sample has the same distribution, up to specific geolocation and set of scale parameters. The stable distribution family is also sometimes referred to as the Lévy alpha-stable distribution (Hosmer and Lemeshew 2002) Of the parameters defining the family, most attention has been focused on the stability parameter, α Stable distributions have $0 < \alpha \leq 2$, with the upper bound corresponding to the normal distribution, and α = 1 to the Cauchy distribution. The Cauchy distribution $f(x; x_0, \gamma)$ is the distribution of the X-intercept of a ray issuing from (x_0, γ) with a uniformly distributed angle which is used in statistics as the canonical example of a "pathological" distribution since both its mean and its variance are undefinable (Johnston et al. 1994). The Cauchy distribution does not have finite moments of order greater than or equal to one; only fractional absolute moments exist

Theses distributions may be undefined variance for α < 2 in a vulnerability forecasting ,sub-meter resolution, ArcGIS –derived discontinuous forest-canopy to irrigated, African riceland, expanding, agro-ecosystem, eco-epidemiological, explicative, forecasting vulnerability, LULC model outputs and an undefined mean for α ≤ 1. The importance of the diagnostic, optimizable, yellow fever, eco-epidemiological, clincial, field or remote geo-spectrotemporally, georeferenceable, parameterizble estimators and the *Ae egypti,* oviposition, capture point, geo-spectrotemporally geosampled, habitat variables may tabulate stable probability distributions since t they would be "attractors" for properly normed sums of i.i.d, time series, YFV-related, randomizable variables. The normal distribution defines a family of stable distributions (Hosmer and Lemeshew 2002). By the classical central limit theorem the properly normed sum of a set of geo-spectrotemporally geosampled, YF, ,disgnostic, eco-epidemiological, randomized variables, each with finite variance, will tend towards a normal distribution as the number of trasnitional discontinuous forest canopied LULC variables to riceland irrigated agriculture, oviposition sites increases. Without the finite variance assumption the limit may be a stable distribution.

A non-degenerate, sub-meter resolution, ArcGIS–derived, discontinuous forest-canopy to irrigated, African, riceland agro-ecosystem, LULC, seasonal, geo-spectrotemporal, forecast, vulnerability, sub-meter resolution, *Ae egypti* oviposition capture point, model distribution may be a stable distribution if it satisfies the following property: Let X_1 and X_2 be independent copies of a random diagnostic, eco-epidemiological, geo-spectrotemporal or geo-spatiotemrpoal geosampled,, forecast, vulnerability, YFV-related, clincial, field or remote, time series sub-meter resolution, LULC, variable X. Then X would be said to be stable if for any constants a > 0 and b > 0 the randomized, YFV variable $aX_1 + bX_2$ has the same distribution as cX + d for some constants c > 0 and d. The distribution would be strictly stable if this holds with d = $0^{[\wedge}$. Since the yellow fever data may express a normal distribution, the Cauchy distribution, and the Lévy distribution would have similar properties, in the forecasting vulnerabilty, probabilistic paradigm since it follows that both distributions would be special cases of stable distributions

In probability theory and statistics, the Lévy distribution, named after Paul Lévy, is a continuous probability distribution for a non-negative random variable(Balakrishnan, and Nevrozov, 2003). An endmeic, trasnmission-oriented, sylvatic, YF, forecast, vulnerability, sub-meter resolution, geoclassified LULC, geo-spatiotemporal or geo-spectrotemporal, *Ae. aegypti,* oviposition, model may be optimally expressed using a Levy distribution as $\mathscr{F}_k \left[P_N(k) \right](x) = \mathscr{F}_k \left[\exp\left(-N\,|k|^{\beta}\right) \right](x).$ In the medical entomological, remotely sensed, time series, model $\mathscr{F}$ would be the Fourier transform of the probability $P_N(k)$ for N-step addition of random discontinuous, forest-canopied, or riceland, agro-irrigation, sub-mter resolution, LULC variables. Lévy showed that $\beta \in (0,\,2)$ for $P(x)$ to be nonnegative.(Hosmer and Lemeshew 2002). The Lévy distribution has infinite variance and sometimes infinite mean (Cressie 1993). The case $\beta = 1$ may render a Cauchy distribution, whilst $\beta = 2$ may render a normal distribution in a geo-spectrotemporal or geo-spectrotemporal, eco-epidemiological, YF, forecast, vulnerability, sub-meter resolution, signature paradigm for targeting seasonal *Ae. aegypti,* oviposition, capture point, hypeprproductive foci. The Lévy distribution for a African, agro-irrgation-related, predictive, ecosystem, YF, risk model for targeting *Ae aegypti,* immature, seasonal, prolific habitats on newly transitioned inhomogeneous, sub-meter resolution, geoclassified LULCs is implementable in the Wolfram Language as LevyDistribution[*mu, sigma*].

Figure 6. Distributions form a hypeothetical yelloe fever related regressable, four-parameter family of continuous probability distributions parametrized by prolific *Ae egypti* habitat geo location and scale parameters μ and *c*, respectively wher two shape parameters β and α, correspondi to measures of asymmetry and concentration, respectively
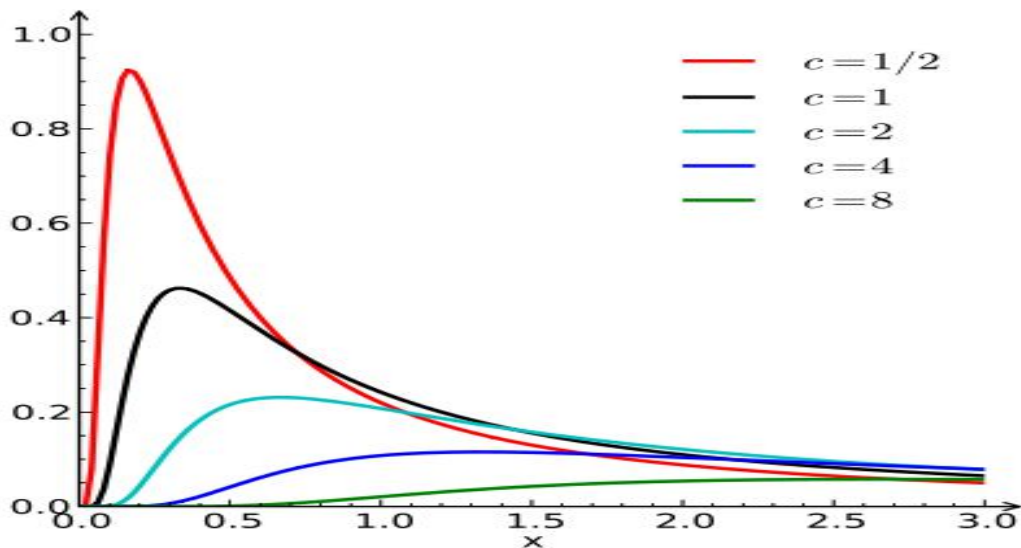
The marginal probabilities in a georferenceable, geosampled, geo-spectyrotemporalized, YFV-related, diagnostic , clinical, field or remote, sub-meter resolution, iteratively intrepoltaive, LULC transitioned, discontinuous, deforested-canopy to irrigated African rice agro-village, *Ae egypti,* immature, aquatic habitats would then be quanatiated by

$$P(x_1) = \int_{-\infty}^{\infty} P(x_1, x_2)\, dx_2 = \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-(x_1-\mu_1)^2/(2\sigma_1^2)} \quad \text{and} \quad P(x_2) = \int_{-\infty}^{\infty} P(x_1, x_2)\, dx_1 = \frac{1}{\sigma_2 \sqrt{2\pi}} e^{-(x_2-\mu_2)^2/(2\sigma_2^2)}$$

Let Z_1 and Z_2 be two independent, diagnostically explanative, time series, clinical, field or remote geosampled, normal variates with means $\mu_i = 0$ and $\sigma_i^2 = 1$ for $i = 1$, 2. Then the variables a_1 and a_2 defined would be normal bivariates where the unit variance and correlation coefficient would be optimally quantitable as ρ: $a_1 = \sqrt{\frac{1+\rho}{2}}\, z_1 + \sqrt{\frac{1-\rho}{2}}\, z_2$ and $a_2 = \sqrt{\frac{1+\rho}{2}}\, z_1 - \sqrt{\frac{1-\rho}{2}}\, z_2$. To derive the bivariate normal probability function, let X_1 and X_2 be normally and independently distributed variates with mean 0 and variance 1(Hazewinkle 2002). Subsequently, an experimenter optimally could define $Y_1 = \mu_1 + \sigma_{11} X_1 + \sigma_{12} X_2$ and $Y_2 = \mu_2 + \sigma_{21} X_1 + \sigma_{22} X_2$ in the for summarizing the probabistic uncertainities in the residual forecasts, targeting ,prolific *Ae geypti* capture, point aquatic habitats on discontinuous, deforested, sparsely or dense canopied, irrigated, riceland tillered, sub-meter resolution, image habitats, for example. In so doing, the variates Y_1 and Y_2 would then be themselves normally distributed with means μ_1 and μ_2, variances $\sigma_1^2 \sigma_{11}^2 + \sigma_{12}^2$ or $\sigma_2^2 = \sigma_{21}^2 + \sigma_{22}^2$, and covariance $V_{12} \equiv \sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}$. The covariance matrix in the forecasting, YFV , vulnerability, eco-epidemioloigical, probabilistic paraidigm could then be optimally defined by $V_{ij} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$, where using $\rho \equiv \frac{V_{12}}{\sigma_1\sigma_2} = \frac{\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}}{\sigma_1\sigma_2}$. PROC CORR statements.

Now, the joint probability density function for x_1 and x_2 is

$$f(x_1, x_2)\, dx_1\, dx_2 = \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2}\, dx_1\, dx_2,$$

but from ($\diamondsuit$) and ($\diamondsuit$), an experimenter would have

$$\begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

in a lineraized, elucidative, eco-epidemiological, sub-meter resolution, residual dataset of regressed, explicative, time series, YFV clinical, field or remote regressed forecasts.

The underlying, ideal, distributional assumptions for each geospectrotemporally geosampled, georefernceable diagnostic, clinical, field or remote variable are usually different from each other. An individual variable might be best modeled as a *t* distribution or as a Poisson process. The correlation of the various variables are very important to estimate as well. A joint estimation of a set of geoclassifiable, sub-meter resolution LULC variables would make it possible to estimate a correlation structure but would restrict the modeling to single, simple multivariate distribution (for example, the normal). Even with a simple multivariate distribution, the joint estimation would be computationally difficult and would have to deal with issues of missing data.

By using the MODEL procedure ERRORMODEL statement, an experimenter could can combine and simulate from sub-meter resolution LULC, *Ae egypti*, forecasting vulnerability paradigms of different distributions. The covariance matrix for the combined model may be optimally constructed by using the copula induced by the multivariate YFV, normalized distribution. A copula is a function that couples joint distributions to their marginal distributions (Hosmer and Lemeshew 2002).-

In probability theory and statistics, a copula is a multivariate probability distribution for which the marginal probability distribution of each variable (e.g., empirically geo-spectrotemrpoally geosampled, georferenceable, Yellow fever related, diagnostic clincial, field or remote, sub-mter resolution, geoclassifiable decomposed *Ae, egypti* oviposition LULC site on a trasnitioned sparsely canopied, deforested, agro-village , irrigation scheme polgonized paramterizable covariate) is uniform. Copulas are used to describe the dependence between random variables(Wassermen 2004).

Sklar's Theorem states that any multivariate joint distribution can be written in terms of univariate marginal distribution functions and a copula which describes the dependence structure between variables. Copulas are popular in high-dimensional statistical applications as they allow one to easily model and estimate the distribution of random vectors by estimating marginals and copulae separately. There are many parametric copula families available, which usually have parameters that control the strength of dependence

Consider a random vector $(X_1, X_2, \ldots, X_d)$ for determining high larval desity count of a positively autocorrelated geospatialized, ArcGIS-derived, georeferenceable, cluster of immature *Ae egypti,* aquatic habita,t capture point, transitional, deforested to rice-irriagted,ArcGIS-derived polygon. Suppose its marginals are continuous, (i.e. the marginal $F_i(x) = \mathbb{P}[X_i \leq x]$ and has optimizable functions. By applying the probability integral transform to each geoclassifiable

geospectrotemporal uncoalesced, iteratively interpolative, elucidative, decomposed, sub-mter resolution, explanative, seasonal, trasnitioning disocntinuous deforested canopy to rice agriculture LULC components, the random vector in the forecast vulnerability paradigm parameterized dataset would reveal $(U_1, U_2, \ldots, U_d) = (F_1(X_1), F_2(X_2), \ldots, F_d(X_d))$ which would optimally render robust uniformly distributed marginals.In so doing, the copula of $(X_1, X_2, \ldots, X_d)$ would be optimally defined as the joint cumulative distribution function of $(U_1, U_2, \ldots, U_d)$: $C(u_1, u_2, \ldots, u_d) = \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2, \ldots, U_d \leq u_d]$. The copula C would contain all information on the dependence structure between the LULC components of $(X_1, X_2, \ldots, X_d)$ whereas the marginal cumulative distribution functions F_i would contain all information on the marginal distributions of the diagnostic, clinical, field or remote –specified, *Ae egypti* aqutic laral habitat capture point, georferenceable, regressors.

The importance of the above methodology is that the reverse of these steps can be used to generate pseudo-random samples from general classes of multivariate YFV-related explicative probability distributions. That is, given a procedure to generate a sample $(U_1, U_2, \ldots, U_d)$ from the copula distribution, the required sample may be optimally constructed in SAS as $(X_1, X_2, \ldots, X_d) = \left(F_1^{-1}(U_1), F_2^{-1}(U_2), \ldots, F_d^{-1}(U_d) \right)$ [Eqn.1.1]The inverses F_i^{-1} may be unproblematic as the F_i may be assumed to be continuous in the residual forecasts targeting the prolic *Ae egypti* habitats on recently trannstioned , sparsely deforeseted .LULCs along irriagted African ricland agro-ecosystem boundaries. Equation 1.1 for the copula function can be rewritten to correspond to: $C(u_1, u_2, \ldots, u_d) = \mathbb{P}[X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2), \ldots, X_d \leq F_d^{-1}(u_d)]$.

In probabilistic terms, $C : [0,1]^d \rightarrow [0,1]$ is a *d*-dimensional copula if C is a joint cumulative distribution function of a *d*-dimensional random vector on the unit cube $[0,1]^d$ with uniform marginals (Hazewinkle 2002) In analytic terms, $C : [0,1]^d \rightarrow [0,1]$ is a *d*-dimensional copula if1) $C(u_1, \ldots, u_{i-1}, 0, u_{i+1}, \ldots, u_d) = 0$, the copula is zero if one of the arguments is zero, 2) $C(1, \ldots, 1, u, 1, \ldots, 1) = u$, the copula is equal to *u* if one argument is *u* and all others 1;and, 3) C is *d*-non-decreasing, (i.e., for each hyperrectangle $B = \prod_{i=1}^{d} [x_i, y_i] \subseteq [0,1]^d$ the C-volume $\int_B dC(u) = \sum_{\mathbf{z} \in \times_{i=1}^{d} \{x_i, y_i\}} (-1)^{N(\mathbf{z})} C(\mathbf{z}) \geq 0,$ of B is non-negative: where the $N(\mathbf{z}) = \#\{k : z_k = x_k\}$. For instance, in the bivariate case, $C : [0,1] \times [0,1] \rightarrow [0,1]$ employed to tabulate a normalized distribution forexplicative, an empirical dataset of optimally regressed yellow fever, geo-spectrotemporally geosampled, LULC, sub-meter reoslution, diagnostic, clinical, field or remote may express a bivariate copula if $C(0,u) = C(u,0) = 0$, $C(1,u) = C(u,1) = u$ and $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ for all $0 < u_1 < u_2 < 1$ and $0 \leq v_1 \leq v_2 \leq 1$.

Sklar's theorem, provides the theoretical foundation for the application of copulas (wasserman 2004). Sklar's theorem states that every multivariate, cumulative distribution function{ e.g., a non-normalized plot of optimally regressed, YFV-related, diagnostic , clinical, field or remote geo-spectrotemporally geosampled, *Ae.egypti* hypeorductive oviposition site immature seasoanl productivity rates on a georeferenceable, transitioned deforested, sparsely

shaded, discontinuous canopied LULC to ricland ploughing capture point LULC could be tabulated emplying $H(x_1, \ldots, x_d) = \mathbb{P}[X_1 \leq x_1, \ldots, X_d \leq x_d]$ of a random vector $(X_1, X_2, \ldots, X_d)$ which may be expressed in terms of its marginals $F_i(x) = \mathbb{P}[X_i \leq x]$ and a copula C. Indeed: $H(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$ (Wasserman 2004).

The concept of the cumulative distribution function makes an explicit appearance in statistical analysis in two (similar) ways. Cumulative frequency analysis is the analysis of the frequency of occurrence of values of a phenomenon (immature counts of a prolific georferenceable, sylvatic, *Ae egypti* , ovipoition, immature habitats on a newly trasnitioned, discontinuous, forest-canopied, LULC to riceland flooded capture point)less than a reference value. The empirical distribution function is a formal direct estimate of the cumulative distribution function for which simple statistical properties can be derived and which can form the basis of various statistical hypothesis tests(Hosmer and Lemeshew 200) . Such tests can assess whether there is evidence against a sample of geoclassified orthogonally decomposed datset of eco-epidemiological, YFV-rleated clinical, field or remote specified, georeferenceable seasonal, LULC, submeter resolution, data parameters. having arisen from a given normlized distribution, or evidence against two samples of data having arisen from the same, immature, aquatic, habitat, population distribution.The Kolmogorov–Smirnov test is based on cumulative distribution functions and can be employed to test to see whether two empirical LULC distributions are different based on geosampled, geo-spectrotemporal, fractionalized regressors or whether an empirical distribution is different from an ideal distribution. The closely related Kuiper's test is useful if the domain of the distribution is cyclic in a foreasting vulnerability, YFV eco-epidemiogical model as in day of the week. For instance Kuiper's test might be used to see if the number of precipitation events varies during the year or if a newly transitioned, sparsely shaded, discontinuous deforested, geoclassifiable canopied, LULC to irrigated African riceland, ploughing in gridded, ArcGIS, stratifiable polygons vary by day of the week or day of the month.

In case that the multivariate LULC distribution has a density f, it holds further that $f(x_1, \ldots x_d) = c(F_1(x_1), \ldots F_d(x_d)) \cdot f_1(x_1) \cdot \ldots \cdot f_d(x_d)$, where c is the density of the copula in the foreasting, yellow fever, vulnerability model. .The theorem also states that, given H, the copula is unique on $\mathrm{Ran}(F_1) \times \cdots \times \mathrm{Ran}(F_d)$, which is the cartesian product of the ranges of the marginal continuous functions.. This implies that the copula in a YFV risk model delineating prolific, *Ae egypti* ,aquatic habitats on a trasnitioned, forest canopied, geoclassifiable LULC to rice agriculture LULC may be unique if the marginals F_i are continuous in the analyses. The converse is also true: given a copula $C : [0,1]^d \rightarrow [0,1]$ and margins $F_i(x)$ then $C(F_1(x_1), \ldots, F_d(x_d))$ would define a *d*-dimensional, geo-spectrotemporal, YFV-related, sub-meter resolution., explanative, diagnostic, geoclassifiable, LULC cumulative distribution function.

As long as $\begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{vmatrix} \neq 0,$ this can be inverted to give $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{bmatrix} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21}} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix} \begin{bmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{bmatrix}.$ The joint probability distribution in a forecasting vulnerability, yellow fever, LULC model can be expressed either in

terms of a joint cumulative distribution function or in terms of a joint probability density function (in the case of continuous diagnostic, geo-spectrotemporally uncoalesced, iteratively interpolative, endmember variables) or joint probability mass function (e.g., in the case of discrete, dignaostic YFV, clincial, field or remote geos-spectrotemporally, geoeeferenceable, geosampled, sub-meter resolution, geo-predictive variables). These in turn can be used to find two other types of seasonal,optimizable, YFV distributions:whereby, the marginal distributionrendering the probabilities for any one of the explicative, decomposed, LULC iteratively interpolative signaturized explanative variables in the regressed datset with no reference to any specific ranges of diagnostic values for the other quanatiated clinical, field or remote variables, and the conditional probability distribution giving the probabilities for any subset of the variables conditional on particular values of the remaining variables in theyellow fever model parameter estimator dataset.

A necessary (but, in general, not sufficient) condition for statistical independence of two explanative, geo-spectrotemporally, diagnostically geosampled, time series, geo-predictive, yellow fever, variables in forecasting eco-epidemiological, sub-meter resolution, *Ae egypti* geoclassifiable, LULC, vulnerability model is that they be statistically uncorrelated; that is, their covariance is zero. Therefore,optimally the covariance matrix R of the components of a white noise vector w with n elements must be an n by n diagonal matrix, where each diagonal element R_{ii} is the variance of component w_i; and the correlation matrix must be the n by n identity matrix. In linear algebra, the identity matrix, or sometimes ambiguously called a unit matrix, of size n is the n × n square matrix with ones on the main diagonal and zeros elsewhere which may be denoted by I_n, or simply by I if the size is immaterial or can be trivially determined by the context.(Neter 1990).

In probability theory and statistics, a covariance matrix (also known as dispersion matrix or variance–covariance matrix) is a matrix whose element in the i, j position is the covariance between the i [th] and j [th] elements of a random vector (Wasserman 2004). A randomized geo-spectrotemrpoal, geospatialized, georeferenceable, sub-mter resolution, uncoalesced, sylvatic, YFV, clinical, field or remote geosampled vector would be a random variable with multiple dimensions. Each element of the vector would then be a scalar random variable in a forecasting vulnerability, yellow fever, risk model. Each element in the model then would have either a finite number of *observed* empirical values or a finite or infinite number of *potential* values. The potential values for targeting prolific *Ae egypti* immature capture point, georeferenceable, seasonal,immature habitats on transitional ,discontinuous, forest-canopy to rice-agriculture habitats in an African, irrigated village complex may be specified by the joint probability distribution in SAS/GIS.

Intuitively, the covariance matrix would generalize the notion of covariance to multiple dimensions in the sylvatic, YFV-related LULC forecasting vulnerability model. As an example, let's considere two, explicatively quantiated, *Ae egypti,* LULC mathematical vectors $X = [x_1, x_2]^T$ and $Y = [y_1, y_2]^T$. There are four covariances to consider: x_1 with y_1, x_1 with y_2, x_2 with y_1, and x_2 with y_2. These variances cannot be summarized in a scalar format. Of course, a 2×2 matrix is the most natural choice to describe the covariance: the first row containing the

covariances of x_1 with y_1 and y_2, and the second row containing the covariances of x_2 with y_1 and y_2 in the yellow fever, forecasting vulnerability model.

Because the covariance of the i th random variable with itself is simply that random variable's variance, each element on the principal diagonal of the covariance matrix is just the variance of each of the elements in the vector (Wassermen 2004). Because $\mathrm{Covar}(x_i, x_j) = \mathrm{Covar}(x_j, x_i)$, every covariance matrix in a geo-spectrotemporal, geospatial, yellow fever, forecasting vulnerability, LULC regression paraidigm for remotely optimally targetinf Ae egypti riceland irriaged Ae egypti paddies in an deforested forecst canopy LULC would be symmetric. In addition, every covariance matrix would positive semi-definite.

In linear algebra, a symmetric $n \times n$ real matrix M is said to be positive definite if the scalar $z^{\mathbf{T}} M z$ is positive for every non-zero column vector z of n real numbers (e.g., geo-spectrotemrporal georfernceable, uncoalesced , diagnostic, sub-metre resolution YFV-related geosampled, clincial, field or remote specified Ae egypti, orthogonally decomposed discretere interger values) . Here $z^{\mathbf{T}}$ denotes the transpose of z (Hazewinkle 2001). In mathematics, a Hermitian matrix (or self-adjoint matrix) is a square matrix with complex entries that is equal to its own conjugate transpose—that is, the element in the i-th row and j-th column is equal to the complex conjugate of the element in the j-th row and i-th column, for all indices i and j: $a_{ij} = \overline{a_{ji}}$ More generally, an $n \times n$ Hermitian matrix M is said to be positive definite if the scalar $z^* M z$ is real and positive for all non-zero column vectors z of n complex numbers. Here z^* denotes the conjugate transpose of z.

In mathematics, the conjugate transpose or Hermitian transpose of an m-by-n matrix A with complex entries is the n-by-m matrix A^* obtained from A by taking the transpose and then taking the complex conjugate of each entry (i.e., negating their imaginary parts but not their real parts). The conjugate transpose is formally defined b $(A^*)_{ij} = \overline{A_{ji}}$ where the subscripts denote the i,j-th entry, for $1 \le i \le n$ and $1 \le j \le m$, and the overbar denotes a scalar complex conjugate. (e.g., the complex conjugate of $a + bi$, where a and b are real geospectrotemrpoally georefernceable YFV-related, explicative, clincial, field or remote specified rice agroculture, parameterizable covariates on a geoclassifiable, *Ae egypti* forest canopied, LULC r, is $a - bi$.)This definition can also be written as $A^* = (\overline{A})^{\mathrm{T}} = \overline{A^{\mathrm{T}}}$ where $A^{\mathbf{T}}$ denotes the transpose and $\overline{A}$ denotes the matrix with complex conjugated entries.A Hermitian conjugate, bedaggered matrix, adjoint matrix or transjugate. The conjugate transpose of a matrix A can be denoted by any of these symbols: A^* or $A^{\mathbf{H}}$, commonly used in linear algebra $A^{\dagger}$ (sometimes pronounced as "A dagger"), universally used in quantum mechanics. A^+, although this symbol is more commonly used for the Moore–Penrose pseudoinverse (Hazewinkle 2002)

In mathematics, and in particular linear algebra, a pseudoinverse A^+ of a matrix A is a generalization of the inverse matrix (Wasserman 2004) The most widely known type of matrix pseudoinverse is the Moore–Penrose pseudoinverse. When referring to a matrix, the term pseudoinverse, without further specification, is often used to indicate the Moore–Penrose pseudoinverse. The term generalized inverse is sometimes used as a synonym for pseudoinverse.A common use of the pseudoinverse is to compute a 'best fit' (least squares)

solution to a system of linear equations that lacks a unique solution Another use is to find the minimum (Euclidean) norm solution to a system of linear equations with multiple solutions. The pseudoinverse facilitates the statement and proof of results in linear algebra.The pseudoinverse is defined and unique for all matrices whose entries are real or complex numbers. It can be computed using the singular value decomposition (SVD)

In linear algebra, the SVD is a factorization of a real or complex matrix . It is the generalization of the eigendecomposition of a positive semidefinite normal matrix (for example, a symmetric geo-spectrotemporalized, geospatial, YFV-related disgnostic , clinical, field or remote weighted matrix with positive eigenvalues) to any $m \times n$ matrix via an extension of polar decomposition. It has many useful applications in signal processing and statistics.

The polar decomposition of a square complex matrix A is a matrix decomposition of the form $A = UP$ where U is a unitary matrix and P is a positive-semidefinite Hermitian matrix. Intuitively, the polar decomposition separates A into a component that stretches the space along a set of orthogonal axes, represented by P, and a rotation (with possible reflection) represented by U. The decomposition of the complex conjugate of A is given by $\overline{A} = \overline{UP}$. This decomposition always exists; and so long as A is invertible, it is unique, with P positive-definite. Note that $\det A = \det U \det P = re^{i\theta}$ gives the corresponding polar decomposition of the determinant of A, since $\det P = r = |\det A|$ and $\det U - e^{i\theta}$. The matrix P is always unique, even if A is singular, and given by $P = \sqrt{A^*A}$ where A^* denotes the conjugate transpose of A. This expression is meaningful since a positive-semidefinite Hermitian matrix has a unique positive-semidefinite square root. If A is invertible, then the matrix U is given by $U = AP^{-1}$. In terms of the singular value decomposition of A, $A = W \Sigma V^*$, one has $P = V\Sigma V^* U = WV^*$ confirming that P is positive-definite and U is unitary. Thus, the existence of the SVD is equivalent to the existence of polar decomposition.One can also decompose A in the form $A = P'U$ Here U is the same as before and P' is given by $P' = UPU^{-1} = \sqrt{AA^*} = W\Sigma W^*$. This is known as the left polar decomposition, whereas the previous decomposition is known as the right polar decomposition. Left polar decomposition is also known as reverse polar decomposition

The matrix A is normal if and only if $P' = P$. Then $U\Sigma = \Sigma U$, and it is possible to diagonalise U with a unitary similarity matrix S that commutes with Σ, giving $S U S* = \Phi^{-1}$, where Φ is a diagonal unitary matrix of phases $e^{i\varphi}$. Putting $Q = V S^*$, one can then re-write the polar decomposition as $A = (Q\Phi Q^*)(Q\Sigma Q^*)$, so A then thus also has a spectral decomposition $A = Q\Lambda Q^*$ with complex eigenvalues such that $\Lambda\Lambda^* = \Sigma^2$ and a unitary matrix of complex eigenvectors Q. In the mathematical discipline of linear algebra, eigendecomposition or sometimes spectral decomposition is the factorization of a matrix into a canonical form, whereby the matrix is represented in terms of its eigenvalues and eigenvectors.

In mathematics, particularly linear algebra and functional analysis, the spectral theorem is any of a number (e.g., a regressed,optimizable, eco-epidemiological, African riceland,geosampled, YFV, diagnostic, clincial, field and remote parameterizable covariate) of results about linear operators or matrices. In broad terms, the spectral theorem provides conditions under which an operator or a matrix can be diagonalized (that is, represented as a diagonal matrix in some basis). Intuitively, diagonal matrices are computationally quite

manageable, so it is of interest to see whether an arbitrary matrix can be diagonalized. The concept of diagonalization is relatively straightforward for operators on finite-dimensional vector spaces but requires some modification for operators on infinite-dimensional spaces. In general, the spectral theorem identifies a class of linear operators that can be modeled by multiplication operators, which are as simple as one can hope to find.

In more abstract language, the spectral theorem is a statement about commutative C*-algebras. See also spectral theory for a historical perspective.Examples of operators to which the spectral theorem applies are self-adjoint   operators or more generally normal operators on Hilbert spaces.   In mathematics, the Pythagorean theorem, also known as Pythagoras' theorem, is a fundamental relation in Euclidean geometry among the three sides of a right triangle. It states that the square of the hypotenuse (the side opposite the right angle) is equal to the sum of the squares of the other two sides. The theorem can be written as an equation relating the lengths of the sides a, b and c, often called the "Pythagorean equation":$]a^2 + b^2 = c^2$,where c represents the length of the hypotenuse and a and b the lengths of the triangle's other two sides.

We begin with the abstract characterization of C*-algebras given in the 1943 paper by Gelfand and Naimark.A C*-algebra, A, is a Banach algebra over the field of complex numbers, (e.g., geo-spectrotemporal datset of, yellow fever, geosampled, georeferenceable, diagnostic, clinical, field or remote, *Ae egypti* forest canopy LULC map $* : A \rightarrow A$. An experimenter, $x*$ for the riceland, agro-ecosystem, geo-spectrotemporal, LULC image of an element x of A. The forecast, vulnerabilty , YFV map C * has the following properties: It is an involution, for every x in A $x^{**} = (x^*)^* = x$.For all x, y in A:$(x + y)^* = x^* + y^*$,$(xy)^* = y^* x^*$For every geosampled λ in C and every x in $A$$(\lambda x)^* = \overline{\lambda} x^*$·For all x in A: $\|x^* x\| = \|x\| \|x^*\|$·The first three identities say that A is a C*-algebra. The last identity is the C* identity and is equivalent to: $\|xx^*\| = \|x\|^2$,which is the B*-identity.

In mathematics, especially functional analysis, a Banach algebra, named after Stefan Banach, is an associative algebra A over the real or complex numbers (or over a non-Archimedean complete normed field) that at the same time is also a Banach space, i.e. normed and complete. The algebra multiplication and the Banach space norm are required to be related by the following inequality:$\forall x, y \in A : \|xy\| \le \|x\| \|y\|$(i.e., the norm of the product is less than or equal to the product of the norms). This ensures that the multiplication operation is continuous. A Banach space is a vector space X over the field **R** of real numbers, or over the field **C** of complex numbers, which is equipped with a norm and which is complete with respect to that norm, that is to say, for every Cauchy sequence $\{x_n\}$ in X, there exists an element x in X such that $\lim_{n \to \infty} x_n = x$,or equivalently:$\lim_{n \to \infty} \|x_n - x\|_X = 0$·The vector space structure allows one to relate the behavior of Cauchy sequences to that of converging series of vectors. A normed space X is a Banach space if and only if each absolutely convergent series in X converges,[2] Completeness of a normed space is preserved if the given norm is replaced by an equivalent one.All norms on a finite-dimensional vector space are equivalent. Every finite-dimensional normed space over **R** or **C** is a Banach space.
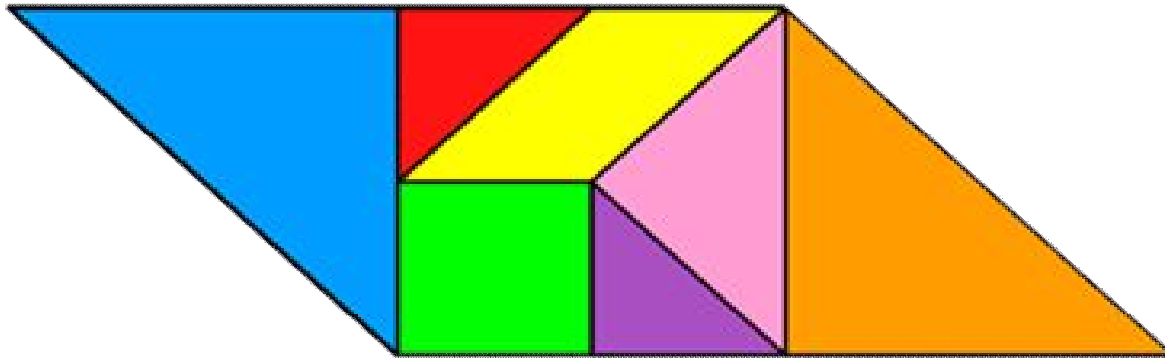
The C*-identity is a very strong requirement. For example, together with the spectral radius formula, it implies that the C*-norm is uniquely determined by the algebraic structure: $\|x\|^2 = \|x^*x\| = \sup\{|\lambda| : x^*x - \lambda 1 \text{ is not invertible}\}$. A bounded linear map, $\pi : A \to B$, between C*-algebras A and B is called a *-homomorphism if For x and y in A $\pi(xy) = \pi(x)\pi(y)$. For x in A $\pi(x^*) = \pi(x)^*$ In the case of C*-algebras, any * homomorphism π between C*-algebras is contractive, i.e. bounded with norm ≤ 1. Furthermore, an injective *-homomorphism between C*-algebras is isometric. These are consequences of the C*-identity. A bijective *-homomorphism π is called a C*-isomorphism, in which case A and B are said to be isomorphic.

The mathematical concept of a Hilbert space, generalizes the notion of Euclidean space. It extends the methods of vector algebra and calculus from the two-dimensional Euclidean plane and 3-dimensional (D) space to spaces with any finite or infinite number of dimensions. *-algebras (pronounced "C-star") are an area of research in functional analysis, a branch of mathematics. A C*-algebra is a complex algebra A of continuous linear operators on a complex Hilbert space with two additional properties:1)A is a topologically closed set in the norm topology of operators.2) A is closed under the operation of taking adjoints of operators. A Hilbert space is an abstract vector space possessing the structure of an inner product that allows length and angle to be measured. Furthermore, Hilbert spaces are complete: there are enough limits in the space to allow the techniques of calculus to be used.

Exact analogs of the Pythagorean theorem and parallelogram law may hold in a Hilbert space when constructing a geo-spectrotemporal, geospatial, forecasting, vulnerability YFV-related eco-epidemiological, sub-meter resolution, geoclassifiable, *Ae egypti* LULC. At a deeper level, perpendicular projection onto a subspace (the analog of "dropping the altitude" of a trianglized LULC geoclassified by irigation riceland) plays a significant role in optimization problems and other aspects of the theory. An element of a Hilbert space can be uniquely specified by its coordinates with respect to a set of coordinate axes (an orthonormal basis), in analogy with Cartesian coordinates in the plane.

In mathematics, the simplest form of the parallelogram law (also called the parallelogram identity) belongs to elementary geometry. It states that the sum of the squares of the lengths of the four sides of a parallelogram equals the sum of the squares of the lengths of the two diagonals. Using the notation in the diagram on the right, the sides are (*AB*), (*BC*), (*CD*), (*DA*). But since in Euclidean geometry a parallelogram necessarily has opposite sides equal, or (*AB*) = (*CD*) and (*BC*) = (*DA*), the law can be stated as, $2(AB)^2 + 2(BC)^2 = (AC)^2 + (BD)^2$ If the parallelogram is a rectangle, the two diagonals are of equal lengths (*AC*) = (*BD*) so, $2(AB)^2 - 2(BC)^2 - 2(AC)^2$ and the statement reduces to the Pythagorean theorem. For the general quadrilateral forecast vulnerability model with four sides not necessarily equal, $(AB)^2 + (BC)^2 + (CD)^2 + (DA)^2 = (AC)^2 + (BD)^2 + 4x^2$, where x is the length of the line segment joining the midpoints of the diagonals. It can be seen from the diagram that, for a parallelogram, $x = 0$, and the general formula simplifies to the parallelogram law.

Applying parallelogram law **for geo-classifying agro-irrigated gri-startified, African, riceland, agro-villag,e complex ecosystem areas and peripheral, forest-canopy, oviposition,** *Ae. egypti* **LULC bas**

-

When that set of axes is countably infinite, this means that the Hilbert space can also usefully be thought of in terms of infinite sequences that are square-summable One of the most familiar examples of a Hilbert space is the Euclidean space consisting of three-dimensional vectors, denoted by $\mathbb{R}^3$, and equipped with the dot productThe dot product of two vectors A = [A_1, A_2, ..., A_n] and B = [B_1, B_2, ..., B_n] is defined as:[1]

$$\mathbf{A} \cdot \mathbf{B} = \sum_{i=1}^{n} A_i B_i = A_1 B_1 + A_2 B_2 + \cdots + A_n B_n$$

where Σ denotes summation notation and *n* is the dimension of the vector space.

The polar decomposition of any bounded linear operator *A* between complex Hilbert spaces is a canonical factorization as the product of a partial isometry and a non-negative operator.The polar decomposition for matrices generalizes as follows: if *A* is a bounded linear operator then there is a unique factorization of *A* as a product *A = UP* where *U* is a partial isometry, *P* is a non-negative self-adjoint operator and the initial space of *U* is the closure of the range of *P*.The operator *U* must be weakened to a partial isometry, rather than unitary, because of the following issues. If *A* is the one-sided shift on l^2(N), then |*A*| = {*A*A*}$^{1/2}$ = *I*. So if *A = U* |*A*|, *U* must be *A*, which is not unitary.The existence of a polar decomposition is a consequence of Douglas' lemma:

If *A*, *B* are bounded operators on a Hilbert space *H*, and *A*A* ≤ *B*B*, then there exists a contraction *C* such that *A = CB*. Furthermore, *C* is unique if *Ker*(*B**) ⊂ *Ker*(*C*).The operator *C*

can be defined by *C(Bh) := Ah* for all *h* in *H*, extended by continuity to the closure of *Ran(B)*, and by zero on the orthogonal complement to all of *H*. The lemma then follows since *A*A ≤ B*B* implies *Ker(A)* ⊂ *Ker(B)*.

In particular. If *A*A = B*B*, then *C* is a partial isometry, which is unique if *Ker(B*)* ⊂ *Ker(C)*. In general, for any bounded operator *A*, $A^*A = (A^*A)^{\frac{1}{2}}(A^*A)^{\frac{1}{2}}$, where $(A^*A)^{\frac{1}{2}}$ is the unique positive square root of *A*A* given by the usual functional calculus. So by the lemma, we have $A = U(A^*A)^{\frac{1}{2}}$ for some partial isometry *U*, which is unique if *Ker(A*)* ⊂ *Ker(U)*. Take *P* to be $(A^*A)^{\frac{1}{2}}$ and one obtains the polar decomposition *A = UP*. Notice that an analogous argument can be used to show *A = P'U'*, where *P'* is positive and *U'* a partial isometry.

When *H* is finite-dimensional, *U* can be extended to a unitary operator; this is not true in general (see example above). Alternatively, the polar decomposition can be shown using the operator version of singular value decomposition. By property of the continuous functional calculus, |*A*| is in the C*-algebra generated by *A*. A similar but weaker statement holds for the partial isometry: *U* is in the von Neumann algebra generated by *A*. If *A* is invertible, the polar part *U* will be in the C*-algebra as well.

Formally, the singular value decomposition of an $m \times n$ real or complex matrix $\mathbf{M}$ is a factorization of the form $\mathbf{U\Sigma V^*}$, where U is an $m \times m$ real or complex unitary matrix, $\mathbf{\Sigma}$ is a $m \times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal, and $\mathbf{V}$ is an $n \times n$ real or complex unitary matrix. The diagonal entries σ_i of $\mathbf{\Sigma}$ are known as the singular values of $\mathbf{M}$. The columns of $\mathbf{U}$ and the columns of $\mathbf{V}$ are called the left-singular vectors and right-singular vectors of $\mathbf{M}$, respectively.

In the following discussion, the following conventions are adopted. K will denote one of the fields of real or complex numbers, denoted $\mathbb{R}, \mathbb{C}$, respectively. The vector space of $m \times n$ matrices over K is denoted by $M(m, n; K)$. For $A \in M(m, n; K)$, $A^{\mathbf{T}}$ and A^* denote the transpose and Hermitian transpose (also called conjugate transpose) respectively. If $K = \mathbb{R}$, then $A^* = A^{\mathbf{T}}$. For $A \in M(m, n; K)$, then $\text{im}(A)$ denotes the range (image) of A (the space spanned by the column vectors of A) and $\text{ker}(A)$ denotes the kernel (null space) of A. Finally, for any positive integer n, $I_n \in M(n, n; K)$ denotes the $n \times n$ identity matrix

Geovizualization of the SVD of a two-dimensional, real shearing matrix In mathematics, a shear matrix or transvection is an elementary matrix that represents the addition of a multiple of one row or column to another. Such a matrix may be derived by taking the identity matrix and replacing one of the zero elements with a non-zero value. A typical shear matrix is shown below:

$$S = \begin{pmatrix} 1 & 0 & 0 & \lambda & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The name shear reflects the fact that the matrix represents a shear transformation. Geometrically, such a transformation takes pairs of points in a linear space, that are purely axially separated along the axis whose row in the matrix contains the shear element, and effectively replaces those pairs by pairs whose separation is no longer purely axial but has two vector components. Thus, the shear axis is always an eigenvector of *S*.

A shear parallel to the *x* axis results in $x' = x + \lambda y$ and $y' = y$. In matrix form: $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$. Similarly, a shear parallel to the *y* axis has $x' = x$ and $y' = y + \lambda x$. In matrix form: $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$. Clearly the determinant will always be 1, as no matter where the shear element is placed, it will be a member of a skew-diagonal that also contains zero elements (as all skew-diagonals have length at least two) hence its product will remain zero and won't contribute to the determinant. Thus every shear matrix has an inverse, and the inverse is simply a shear matrix with the shear element negated, representing a shear transformation in the opposite direction. In fact, this is part of an easily derived more general result: if *S* is a shear matrix with shear element λ, then S^n is a shear matrix whose shear element is simply $n\lambda$. Hence, raising a shear matrix to a power *n* multiplies its shear factor by *n*.

The negative definite, positive semi-definite, and negative semi-definite matrices are defined in the same way, except that 0's are allowed, i.e. the expression $z^{\mathrm{T}} M z$ or $z^* M z$ is required to be always negative, non-negative, and non-positive, respectively. Positive definite matrices are closely related to positive-definite symmetric bilinear forms (or sesquilinear forms in the complex case), and to inner products of vector spac

In particular, if in addition to being independent every geo-spectrotemporally geosampled YFV diagnostic, eco-epidemiological, time series, diagnostic, clinical, field or remote, geo-predictive *Ae egypti* LULC variable in *w* also has a normal distribution with zero mean and the same variance σ^2, *w* would be a Gaussian white noise vector. In that case, the joint distribution of *w* would be a multivariate normal distribution; the independence between the variables then implies that the distribution has spherical symmetry in *n*-dimensional space. Spherical symmetry refers to any spherical object that can be divided through the center and produce two equal halves Therefore, any orthogonal transformation of the vector will result in a Gaussian white random vector. In particular, under most types of discrete Fourier transform, such as FFT and Hartley, the transform *W* of *w* will be a Gaussian white noise vector, too in the vulnerability paradigm; that is, the *n* Fourier coefficients of *w* will be independent Gaussian variables with zero mean and the same variance σ^2.

The power spectrum *P* of a random vector *w* can be optimally defined in a YFV gridded, sub-meter resolution model, as the expected value of the squared modulus of each coefficient of its Fourier transform *W*, that is, $P_i = E(|W_i|^2)$. Under that definition, a Gaussian white noise vector will have a perfectly flat power spectrum, with $P_i = \sigma^2$ for all *i*.

If *w* is a white random vector, in a seasonal, hyperproductive, ovipsoition, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition ,

191

discontinuous, forest-canopied, African riceland model.but not a Gaussian one, its Fourier coefficients W_i will not be completely independent of each other; although for large n and common probability distributions the dependencies would be very subtle, and their pairwise correlations may be assumed to be zero. Often the weaker condition "statistically uncorrelated" is used in the definition of white noise, instead of "statistically independent"(Hosmer and Lemeshew 2002). However, some of the commonly expected properties of white noise (such as flat power spectrum) may not hold for this weaker version in an eco-epidemiological, sub-meter resolution, gridded, YFV-related, geoclassifiable LULC model. Under this assumption, the stricter version can be referred to explicitly as independent white noise vector.

An example of a random vector that is "Gaussian white noise" in a geo-spectrotemporal, YFV-related, sylvatic, *Ae egypti* irrigated ricland LULC model in the weak but not in the strong sense could be optimally resolved as $x=[x_1,x_2]$ where x_1 is a normal random variable with zero mean, and x_2 is equal to $+x_1$ or to $-x_1$, with equal probability. These two variables would be uncorrelated and individually normally distributed, but they may not be jointly normally distributed and may not be independent. If x is rotated by 45 degrees, its two components in the YFVforecast vulnerability model may still be uncorrelated, but their distribution will no longer be normal.

In some situations a experimenter may relax the definition in a yellow fever model a by allowing each component of a white random vector w to have non-zero expected value μ. In image processing especially, where samples are typically restricted to positive values, one often takes μ to be one half of the maximum sample value (Jensen 2005). In that case, the Fourier coefficient W_0 corresponding to the zero-frequency component (essentially, the average of the w_i) in the model will also have a non-zero expected value $\mu\sqrt{n}$; and the power spectrum P will be flat only over the non-zero frequencies.

In order to define the notion of "white noise" in the theory of continuous-time signals, a experimeneter must replace the concept of a "random vector" by a continuous-time random signal; that is, a random process that generates a function w of a real-valued parameter t for robustly quantitating an empriical dataset of geo-spectrotemrpoally geosampled, georferenceable, explicative, diagnostic, sub-meter resolution, diagnostic, clinical, field or remote dataSuch a process is said to be white noise in the strongest sense if the value $w(t)$ for any time t is a random variable that is statistically independent of its entire history before t. A weaker definition may only require independence between the geosampled values $w(t_1)$ and $w(t_2)$ at every pair of distinct times t_1 and t_2. An even weaker definition requires only that such pairs $w(t_1)$ and $w(t_2)$ be uncorrelated (Rao 1972).

However, a precise definition of these concepts is not trivial, because some quantities that are finite sums in the finite discrete case must be replaced by integrals that may not converge in a YFV forecasting *Ae egypti* irrigtion, riceland *Ae egypti* model. Indeed, the set of all possible instances of a geoclassified expanding African riceland agro-ecosystem village complex signal w may no longer have a finite-dimensional space $\mathbb{R}^n$, but an infinite-dimensional function space. In mathematics, a function space is a set of functions of a given kind from a set X to a set Y. It is

called a space because in many applications it is a topological space (including metric spaces), a vector space, or both. Namely, if Y is a field, functions have inherent vector structure with two operations of pointwise addition and multiplication to a scalar.(Griffith 2003) Moreover, by any definition a white noise signal w in a yellow fever seasonal model for targeting prolific, georeferenceable, sylvatic *Ae egypti* ,riceland, capture point, seasonal, immature habitats would have to be essentially discontinuous at every point; therefore even the simplest operations on w, like integration over a finite interval would require advanced mathematical machinery.

Some authors require each value $w(t)$ to be a real-valued random variable with some finite variance σ^2. Then the covariance $E(w(t_1) \cdot w(t_2))$ between the values at two times t_1 and t_2 is well-defined: it is zero if the times are distinct, and σ^2 if they are equal. However, by this definition, the integral $W_{[a,a+r]} = \int_a^{a+r} w(t)\, dt$ over any interval with positive width r would be zero. This property would render the concept inadequate as a model of physical "white noise" signals.Therefore, most authors define the signal w indirectly by specifying non-zero values for the integrals of $w(t)$ and $|w(t)|^2$ over any interval $[a, a + r]$, as a function of its width r. In this approach, however, the value of $w(t)$ at an isolated time cannot be defined as a real-valued random variable. Also the covariance $E(w(t_1) \cdot w(t_2))$ becomes infinite when $t_1 = t_2$; and the autocorrelation function $R(t_1, t_2)$ must be defined as $N\delta(t_1 - t_2)$, where N is some real constant and δ is Dirac's "function".In this approach, one usually specifies that the integral W_I of $w(t)$ over an interval $I = [a, b]$ is a real random variable with normal distribution, zero mean, and variance $(b - a)\sigma^2$; and also that the covariance $E(W_I \cdot W_J)$ of the integrals W_I, W_J is $r\sigma^2$, where r is the width of the intersection $I \cap J$ of the two intervals I, J. This model is called a Gaussian white noise signal (or process).

In statistics and econometrics one often assumes that an observed series of data values is the sum of a series of values generated by a deterministic linear process, depending on certain independent (explanatory) variables, and on a series of random noise values. Then regression analysis is used to infer the parameters of the model process from the observed data, e.g. by ordinary least squares, and to test the null hypothesis that each of the parameters is zero against the alternative hypothesis that it is non-zero. Hypothesis testing typically assumes that the noise values are mutually uncorrelated with zero mean and the same Gaussian probability distribution — in other words, that the noise is white. If there is non-zero correlation between the noise values underlying different observations then the estimated model parameters are still unbiased, but estimates of their uncertainties (such as confidence intervals) will be biased (not accurate on average). This is also true if the noise is heteroskedastic — that is, if it has different variances for different data points.

Alternatively, in the subset of regression analysis known as time series analysis there are often no explanatory variables other than the past values of the variable being modeled (the dependent variable). In this case the noise process is often modeled as a moving average process, in which the current value of the dependent variable depends on current and past values of a sequential white noise process.

These two ideas are crucial in applications such as channel estimation and channel equalization in communications and audio. These concepts are also used in data compression.In particular, by a suitable linear transformation (a coloring transformation), a white random vector can be used to produce a "non-white" random vector (that is, a list of random variables) whose elements have a prescribed covariance matrix. Conversely, a random vector with known covariance matrix can be transformed into a white random vector by a suitable whitening transformation.

Geospectrally eigen- decomposable, sub-meter resolution, gridded data [QuickBird, visible and near infra-red(NIR), remotely sensed,   sub-mixel (i.e., mixed pixel), endmember (i.e., reference biosignature) fractions of incident radiation reflected, transmitted and absorbed by prolific, georeferenced, entomological, sylvatic,capture point, YF , *Ae aegypti,* larval habitats is crucial in implementing control strategies in various ecosystems (e.g., urbanizing, pre-flooded, tillering riceland, discontinuous, deforested, sparsely shaded, forest canopy LULCs). However, non-quantitation of probabilistic bidirectional internal and external geometric error radiance uncertainties derived in an eco-epidemiological, gridded sub-meter resolution, endmember, wavelength-oriented, regression equation constructed from log-transformed discrete, categorical and continuous, bionomic, seasonal, hypeprproductive, *Ae. aegypti,* time series, clinical, field and radiance-related, non-diagnosed, heterogeneous explanators (e.g., daily precipitation count and measure of drop sizes from rain gauges defined in disdrometers, quantized, uncoalesced, frequencies of LULC sites measured in nanometers) heteroskedastic (i.e.,uncommon variance) independently, probabilistic, non-normalities may be a limiting factor in forecasting vulnerability paradigms. Injudiciously, endemic, YF, linear regression metamodels and their concomitant experimental designs in literature assume a univariate (not multivariate) simulation response and white noise. By definition, white noise is normally (Gaussian), independently (implying no common random numbers), and identically (constant variance) distributed with zero mean (valid metamodel)It is important to recognize the source of the internal and external error and whether it is systematic (predictable) or nonsystematic (random) in a medially entomological, vector arthropod, forecasting, vulnerability, eco-epidemiological model.

Many tests of the constancy of vector arthropod-related geo-spectrotemporally geosampled, empirical datasets of georeferenceable, regression coefficients have been proposed in various paradigms in SAS. If the regressed diagnostic, YF, time series clinical, field or remote, geosampled non-parameterizable, sub-meter resolution, uncoalesced, iteratively interpola LULC coefficients are suspected of discrete LULC changes a Chow (1960) test or the test proposed by Quandt (1960) may be appropriate. If the coefficients are suspected of changing smoothly through time, (e.g., are generated by some process, another class of tests can be used. For these tests an ARIMA process is used as a proxy for the true generating process. Tests for coefficients suspected of following specific ARIMA processes have been proposed in the literature. A simple white noise process generates the random coefficients model, which can be tested using the Lagrange multi-plier test of Breusch and Pagan (1979). For coefficients suspected of following a random walk, tests have been proposed by Brown, Durbin, and Evans (1975), Garbade (1977), Pagan and Tanaka (1979), LaMotte and McWhorter (1978), and a series of tests have been proposed and compared by Harvey and Phillips (1976). Cooley and Prescott (1976) introduced a

model where the coefficients follow an ARIMA (0, 1,1) process, and proposed a likelihood ratio test.

There are four principal assumptions which justify the usage of linear regression models for purposes of eco-epidemiological, inferences or geo-prediction for YFV surveillance on geoclassifiable, time series expanding geoclassifiable LULCs:(i) linearity and additivity of the relationship between dependent and independent variables: (a) The expected value of dependent variable (e.g., prevalence of yellow fever in an African, agro-village, complex escalating LULC ecosystem) is a straight-line function of each geo-spectrotemporally, iteratively interpolative, explicatively, heuristically optimal, diagnostic, time series endmember, dependent, diagnostic, clinical, field or remote, uncoalesced, geosampled, signaturized, independent variable, holding the others fixed. (b) The slope of that line does not depend on the values of the other probabilistic, LULC explicative, YFV variables. (c) The effects of different diagnostic, independent variables on the expected value of the dependent variable (e.g., prevalence of yellow fever) are additive.(ii) statistical independence of the errors (in particular, no correlation between consecutive errors in the empirical, georefernceable, yellow fever ,time series regressors). (iii) homoscedasticity (constant variance) of the errors including; (a) versus time (e.g., remotely sensed, unmixed,clustered, yellow fever,iteratable data) (b) versus the LULC geo-predictions of prolific, seasonal, sylvatic *Ae egeypti* , capture point, immature, seasonal, georferenceable habitats (c) versus any independent geosampled, diagnostic, clinical, field or remote, optimizable variable;and, (iv) quanatiation of time series normality of the error distribution. If any of these assumptions is violated (i.e., if there are non-quantized, nonlinear, explanativerelationships between YFV-related, time series dependent and independent, geoclassified, diagnostic, LULC variables in a forecasting vulnerability model or the errors exhibit correlation, heteroscedasticity, or multicolineraity, then the forecasts, confidence intervals, and scientific insights yielded by a yellow ferver regression geo-predictive,eco-epidemiological model may be (at best) inefficient or (at worst) seriously biased or misleading.

In inferencial forecasted datasets rendered from linearization of seasonally geo-spectrotemporally geosampled, empiricalized, vector arthropod-related, time series, sub-meter resolution, uncoalesced LULC and other endmember explanators [3 dimensional (D) catchment slope coefficients) associated with seasonally hyperproductive, vector arthropod-related], multicollinearity is a common phenomenon. Collinearity occurs when two or more explanatorial variables in a multiple regression model are highly correlated, meaning that one can be linearly forecasted from the others with a substantial degree of accuracy. In this situation the coefficient estimates of the multiple regression may change erratically in response to small changes in the model or the data. Multicollinearity will not reduce the explanative, geo-predictive power or reliability of a seasonal, YFV-related, *Ae. egypti,* optimally sub-meter resolution, geoclassifiable,georeferenceable, LULC, eco-epidemiological model as a whole, at least within the sample dataset; it only would affect calculations regarding individual regressors. That is, a multiple regression forecasting, vulnerability model with correlated geo-spectrotemporal, uncoalesced, endmember, *Ae egypti* LULC, submeter resolution geo-predictors can indicate how well the entire bundle of geo-predictors render the outcome variable, but it may not give valid results about any individual regressors, or about which explanators are redundant with respect to others.

In case of perfect multicollinearity the design matrix is singular and therefore cannot be inverted in most statistical packages. The design matrix is defined to be a matrix X such that the jth column of the ith row of X represents the value (e.g, weigthage of a forest canopy LULC visible band) of the jth variable associated with the ith object (Hosmer and Lemeshew 2002).A geo-spectrotemporal, YFV-related, regression model which would be a linear combination of the georeferenceable, explanatory variables may therefore be represented via matrix multiplication as $y - X\beta$ where X is the design matrix, β is a vector of the model's coefficients (one for each geosampled, clinical, field or remote variable, and y is the vector of geo-predicted model outputs for each object (e.g., geoclassifiable discontinuous , georeferenceable riceland pre-harvesting, flooded , immature capture point habitat).

Under these circumstances, for quantitating a generalizable YFV-related explanative, seasonal endmember, sub-meter resolution. sylvatic, *Ae. egypti*, linearized, LULC ec-epidemiological, geo-spectrotemporal, geospatialized,geo-predicti ve model $y = X\beta + \epsilon$, the ordinary least-squares estimator $\hat{\beta}_{OLS} = (X^{\top}X)^{-1}X^{\top}y$ does not exist.Note that in statements of the assumptions underlying regression analyses of most medically important, seasonally, georeferenceable, geo-spectrotemporally uncoalesced, hyperproductive, vector, arthropod-related, immature, capture point, habitat endmember, sub-meter resolution, iteratively interpolative, signaturizeable, orthogonalized, data feature attribute, an ordinary least squares, may reveal the absence of multicollinearity . However, an exact , non-stochastic, linearizable relationship qualitatively quantized among the YFV-related LULC,weighted regressors (e.g., geo-spectrotemporally uncoalesced, georeferenced, time series, diagnostic, clinical, field or remote geosampled, *Ae egypti,* fractionalized, radiance endmember, wavelength , visble and NIR frequencies from a georferenced forecast canopy habitat protruding onto a tillering ricland agro-ecosystem immature habitat).

Two variables in a YFV, forecat model are perfectly collinear if there is an exact linear relationship between them. For example, X_1 and X_2 are perfectly collinear in a forecasting, iteratively interpolative, seasonally explanative, georeferenceable, geo-spectrotemporal, geospatialized, yellow fever vulnerability paradigm if there exists parameters λ_0 and λ_1 such that, for all the signaturizable, diagnostic, uncoalesced, *Ae egypti,* illuminative, LULC explicative observations *I,*then a model may berobustly constructed in a statsistical software package (SAS or R) employing $X_{2i} = \lambda_0 + \lambda_1 X_{1i}$.[Eqn;1.1] and the robustness of the rendered probabilsitic non-normalities may be qualitatively quantitated with parsimony and precisison.

Perfect multicollinearity exists in a geo-spectrotemporal, geospatial, YFV-related *Ae. egypti* LULC , forecasting, vulnerability model if, for example, as in equation 1.1 , the correlation between two diagnostic geo-spectrotemporalized, endmember, clinical, field or remote, diagnostic, independent variables is equal to 1 or −1. More commonly, the issue of multicollinearity arises when there is an approximate linear relationship among two or more independent variables (Rao 1972).Mathematically, a dataset of geo-spectrotemporally geosampled,georeferenceable, seasonal, YFV-related, geoclassifiable, *Ae. egypti*, explicatively uncoalesced, parameterized LULC, capture point, immature habitat, eco-epidemiological dataset is perfectly multicollinear if there exist one or more exact linear relationships among some of the

variables. For example, an experimenter could optimally have $\lambda_0 - \lambda_1 X_{1i} + \lambda_2 X_{2i} + \cdots + \lambda_k X_{ki} = 0$ holding for all the diagnostic, time series explanative, sylvatic, YFV-related, diagnostic clinical, field or remote geosampled observations i, where λ_j are constants and X_{ji} is the ith observation on the jth explanatory endmember, iteratively interpolatable, geo-predictive variable. In so doing, explorations may be optimally conducted where one issue caused by multicollinearity in the yellow fever, forecasting, vulnerability model is invasively examined by attempting to obtain estimates for the parameters of the multiple regression equation $Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \varepsilon_i$.

The ordinary least squares estimates involve inverting the matrix $X^T X$

$$X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{k1} \\ \vdots & \vdots & & \vdots \\ 1 & X_{1N} & \cdots & X_{kN} \end{bmatrix}.$$

where (Hosmer and Lemeshew 2002) In statistics, OLS or linear least squares is a method for estimating the unknown parameters in a linear regression model, with the goal of minimizing the differences between the observed responses in some arbitrary dataset (e.g., geo-spectrotermporally uncoalesced, sub-meter resolution, yellow fevere georferenceable, geosampled, *Ae egypti*, geoclassifiable sparsely shaded, discontinuously canopied, forest canopied, explanative LULC, neigbouring an agro-ecosystem riceland ploughing, capture point, immature habitat) and the responses predicted by the linear approximation of the data . Geo-visually this may be efficiently optimally delineated cartographically in an ArcGIS cyberenvironment whiah can render the sum of the vertical distances between each explanative capture point in the set and the corresponding point on the regression line. Importanttly, the smaller the differences in the OLS model the better the model fits the data). The resulting diagnostic, clinical, field or remote estimator can be expressed by a simple formula, especially in the case of a single regressor on the right-hand side of the geopredictive, geo-spectrotemporal, endmember equation..

The OLS estimator is consistent when the regressors are exogenous and there is no perfect multicollinearity, and optimal in the class of linear unbiased estimators when the errors are homoscedastic and serially uncorrelated. Under these conditions, the method of OLS can provide minimum-variance mean-unbiased estimation for a geo-spectrotemporal, geospatial, yellow fever-related, eco-epidemiological, sylvatic, *Ae egypti*, uncoalesced, sub-meter resolution, sub-mixel, iteratievly , quantaitively interpolative, LULC, forecasting vulnerability paradigmwhen the errors have finite variances. Consider estimation of $g(\theta)$ based on a dataset of uncoalesd sub-meter resolution, YFV-related, LULC explanators $X_1, X_2, \ldots, X_n$ I and are if identically independently distributed from some member of a family of densities (e.g., uncoalesed geoclassifiable meterological data varaiables $p_\theta, \theta \in \Omega$, where Ω is the parameter space. An unbiased, explanative, LULC estimator $\delta(X_1, X_2, \ldots, X_n)$ of $g(\theta)$ would then be unbiased estimator or minimum-variance unbiased estimator *UMVUE* if $\forall \theta \in \Omega$, $\operatorname{var}(\delta(X_1, X_2, \ldots, X_n)) \leq \operatorname{var}(\tilde{\delta}(X_1, X_2, \ldots, X_n))$ for any unbiased georferenceable estimator in the forecasting vulnerability model residual forecast.

If an unbiased estimator of $g(\theta)$ exists, in the yellow fever, time series eco-epidemiological LULC model then an expeimenter can can prove there is an essentially unique MVUE in the

resdiualized YFV forecasts (targets of hyperproductive clustering, *Ae egypti* , eco-epidemioloigical, capture points). Using the Rao–Blackwell theorem an expeimemter can also prove that determining the MVUE in aYFV-related forecasting, vulnerability paradigm is simply a matter of finding a complete sufficient statistic for the family $p_\theta, \theta \in \Omega$ and conditioning any unbiased estimator on it. An estimator δ(*X*) is an observable random variable (i.e. a statistic) used for precisely estimating some *unobservable* quantity(geospatial heterosekedascitic outlier) . For example, an yellow fever experimeneter may be unable to observe the average weight of all sylvatic,*Ae egypti*, geoclassifiable, geo-spectrotermporal, uncoalesced, iteratively interpolated forest canopied, discontinuous, sparsely shaded, LULCs adjacent to a tillering riceland agro-village comapled, captiure point, immatire habitatt) in X, (a georeferenceable, agro-ecosystem, African, ricleand agro-ecosystem) but the expeimenter may observe the regreession weights of a random sample of 4 of them. The average weight of those 4—the "sample average"—may however not be used as an estimator of the unobservable "population average". If an experimenter allow linerations to occur in PROC REG and then the endmember data is exported into PROC VAROGRAM for interpolation, then the endmember estimators may be misspecified.

A sufficient statistic *T*(*X*) may be a statistic calculated from non-normalized uncoalesced, geo-spectrotermporialized, sub-meter resolution data *X* to estimate some parameter θ for which it is true that no other statistic which can be calculable, from sylvatic, *Ae egypti* LULC data X may provide any additional information about θ forecasting vulnerability. The output (e.g., targets of hyperproductive,georferenceable, geolocations of *Ae egypti* habitats on seasonlly transitioned , sparsely shaded, discontinuously canopied, *Ae egypti* forest-canopy, geoclassifiable LULCs may be be optimally defined as an observable georferenceabloe, explicative, random variable such that the conditional probability distribution of all observable data *X* given *T*(*X*) does not depend on the unobservable parameter θ, such as the mean or standard deviation of the whole population from which the data *X* was geosampled.. In the most frequently cited examples, the "unobservable" quantities are commonly LULC estimators that can parametrize a known family of probability distributions according to which the data (e.g., diagnostic time series, clinical, fielsd or remote YFV paramterizable covariates) are distributed. In other words, a sufficient statistic *T(X)* for a parameter in a robust YFV-related explicative *Ae egypti,* LULC vulnerability, paraidigm, θ is a robustifiable statistic for geopredicting yellow fever prevalance if the conditional distribution of the geo-spectrotemporal, geo-spatializable,diagnostic, clincial, field or remote specified, geo-spectrotemporally geosampled regressors and the conditional distribution of the data *X,* given *T*(*X*), does not depend on the parameter θ.

The mean squared error of an geo-spectrotemporally orthogonally explicatively decomposeable, sub-meter resolution, yellow fever, iteratively interpolative, explicative sub-meter resolution, regression LULC estimator is the expected value of the square of its deviation from the unobservable quantity being estimated. One case of Rao–Blackwell theorem states: The mean squared error of the Rao–Blackwell estimator does not exceed that of the original estimator.In other words $\mathrm{E}((\delta_1(X) - \theta)^2) \leq \mathrm{E}((\delta(X) - \theta)^2)$. In statistics, the Rao–Blackwell theorem, sometimes referred to as the Rao–Blackwell–Kolmogorov theorem, is a result which characterizes the transformation of an arbitrarily crude estimator into an estimator that is optimal

by the mean-squared-error criterion or any of a variety of similar criteria.The Rao–Blackwell theorem states that if $g(X)$ is any kind of estimator of a parameter θ, then the conditional expectation of $g(X)$ given $T(X)$, where T is a sufficient statistic, is typically a better estimator of θ, and is never worse. It may be possible that a yellow fever experimneter utilize a regression froemweork in PROC REG so as to easily construct a very crude estimator $g(X)$, and then evaluate that conditional expected value to achive an optimal munbiased Ae egypti LULC estimator that can orthogonally , cartographoically represent a sub-meter resoltuioon, uncoalesed endmember signature estimator that is in various senses optimal for mtragetimng hyupeporductiev *Ae egypti* habitats in an African riseland agro-ecosysem,

The essential tools of the proof besides the definition provided in literature arebased on the the law of total expectation and the fact that for any random variable Y, $E(Y^2)$ in an eco-epidemiological, predictive, sylvatic *Ae. aegypti,* YFV–related LULC model residualizable explantorial, paramterizable covariate estimator dataset cannot be less than $[E(Y)]^2$. That inequality may be logically ascertained employing Jensen's inequality, although it may also be shown in PROC LOGISTIC to follow instantly in a YFV related forecasting sub-meter resolution, linear decomposition algorith that a *Ae egypti* , eco-epidemioloigicasl, capture point, georefrenecable, geo-spectrotemporal, geospatialized, model that the

$$0 \leq \mathrm{Var}(Y) = \mathrm{E}((Y - \mathrm{E}(Y))^2) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2.$$

In probability theory, a convex function applied to the expected value of a geospectrotemrpaolly geosampled, *Ae egypti,* LULC sub-meter resolution, forecast vulnerabiliy model randomized variable would be always less than or equal to the expected value of the convex function of the random variable. This result,is due to Jensen's inequality, which underlies many important inequalities (including, for instance, the arithmetic–geometric mean inequality and Hölder's inequality).

Exponential growth is a special case of convexity. Exponential growth narrowly means "increasing at a rate *proportional* to the current value", while convex growth generally means "increasing at an increasing rate (but not necessarily proportionally to current value)" (Grifftyh 2003)For a real convex function φ, geospectrotermpaolly geosampled, georeferenceable, YFV-related geoclassifiable *Ae egypti* LULCs x_1, x_2, ..., x_n in its domain, and positive weights a_i,

Jensen's inequality can be stated as: $\varphi\left(\dfrac{\sum a_i x_i}{\sum a_i}\right) \leq \dfrac{\sum a_i \varphi(x_i)}{\sum a_i}$ [Eqn 1.2]and the inequality

is reverseable if φ is concave, which is] $\varphi\left(\dfrac{\sum a_i x_i}{\sum a_i}\right) \geq \dfrac{\sum a_i \varphi(x_i)}{\sum a_i}.$ [Eqn 1.3]. Equality holds if and only if $x_1 = x_2 = \cdots = x_n$or φ is linear (Rao 1972) As a particular case, if the weights in a geo-spectrotermporally uncoalesced, YFV-related, geoclassifiable, *Ae egypti*, sub-meter resolution , discontinuously canopied, sparsely shaded, forest-canopy LULC neigboring an African agro-system ricleand irrigated LULC a_i are all equal, then (equation 1.21) and equation

1.3 become $\varphi\left(\dfrac{\sum x_i}{n}\right) \geq \dfrac{\sum \varphi(x_i)}{n}$ [Eqn 1.4] and $\varphi\left(\dfrac{\sum x_i}{n}\right) \leq \dfrac{\sum \varphi(x_i)}{n}$ [Eqn 1.5].

For instance, the function $\log(x)$ would be concave in a precisoion, yellow fever, forecasting, vulnerability paradigm so substituting $\varphi(x) = \log(x)$ in the previous regression

formula (Equn 1.5) establish the (logarithm of the) familiar arithmetic mean-geometric mean inequality:
$$\log\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) \geq \frac{\sum_{i=1}^{n} \log(x_i)}{n} \quad \text{or} \quad \frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdots x_n}$$

The more general version of the Rao–Blackwell theorem speaks of the "expected loss" or risk function: $\mathbb{E}(L(\delta_1(X))) \leq \mathbb{E}(L(\delta(X)))$ where the "loss function" L may be any convex function. For the proof of the more general version, Jensen's inequality cannot be dispensed with. The improved estimator is unbiased if and only if the original estimator is unbiased, as may be seen at once by using the law of total expectation. The theorem holds regardless of whether biased or unbiased estimators are used.

The theorem seems very weak: it says only that the Rao–Blackwell estimator is no worse th In mathematics, Jensen's inequality, named after the Danish mathematician Johan Jensen, relates the value of a convex function of an integral to the integral of the convex function. It was proven by Jensen in 1906.[1] Given its generality, the inequality appears in many forms depending on the context, some of which are presented below. In its simplest form the inequality states that the convex transformation of a mean is less than or equal to the mean applied after convex transformation; it is a simple corollary that the opposite is true of concave transformations.

Jensen's inequality can generalizes the statement that the secant line of a convex function lies *above* the graph of the function, which may be usable in an Jensen's inequality quantaifiaction for two geoclassifiable explanative, georfernecable, geo-spectrotemporal, fractionalized, endmember, forecast-oriented, YFV-related, sub-meter resolution, non-orthogonalized, geoclasisified eco-epidmeiolgioical, uncoalesced, wavelength, frequency points: the secant line consists of weighted means of the convex function, $tf(x_1) + (1-t)f(x_2)$ while the graph of the function is the convex function of the weighted means, $f(tx_1 + (1-t)x_2)$. In the context of probability theory, itthe forecasting vulnerability paradigm residual forecast (i.e., targets of hyperproductive, sylvatic *Ae egypti* , immature, capture points, on geoclassified LULCs generally stated in the following form: if X is a randomizable explanatorial unmixed YFV-related regressable variable and φ is a convex function, then an the original estimator. In practice, however, the improvement is often enormous. Further, by the Lehmann–Scheffé theorem, an unbiased estimator that is a function of a complete, sufficient statistic is the UMVUE estimator. Put formally, suppose $\delta(X_1, X_2, \ldots, X_n)$ is unbiased for $g(\theta)$, and that T is a complete sufficient statistic for the family ofdensities.Then $\eta(X_1, X_2, \ldots, X_n) = \mathbb{E}(\delta(X_1, X_2, \ldots, X_n)|T)$ is the MVUE for $g(\theta)$. A Bayesian analog is a Bayes estimator, particularly with minimum mean square error (MMSE). Under the additional assumption that the errors be normally distributed, OLS is the maximum likelihood estimator. OLS is used in economics (econometrics), political science and electrical engineering (control theory and signal processing), among many areas of applications in Yellow fever modelling . The Multi-fractional order estimator is an expanded version of OLS.

The multi-fractional order estimator (MFOE) is a straightforward, practical, and flexible alternative to the Kalman filter (KF) for tracking targets(seasonal hyperproductive *Ae aegypti* oviposition geolocations in expanding African ricelands). The MFOE may be focused strictly on simple and pragmatic fundamentals along with the integrity of mathematical forecast

modeling. Like the KF, the MFOE is based on the least squares method (LSM) and the orthogonality principle at the center of Kalman's derivation (Cressie 1993). Optimizing, the MFOE may yield better accuracy than the KF in a YF model and subsequent algorithms such as the extended KF and the interacting multiple model (IMM). MFOE is an expanded form of the LSM, which effectively includes the KF and OLS(Hosmer and Lemeshew 2002). as subsets (special cases).. The MFOE offers two major advances for YF target modelling seasonal hypeproductive foci on newly transitioned : (1) minimizing the mean squared error (MSE) with fractions of estimated coefficients (useful in target tracking)[1][2] and (2) describing the effect of deterministic OLS processing of statistical inputs (of value in econometrics)[

Consider equally time spaced noisy measurement samples of a target trajectory described

$$y_n = \sum_{j=1}^{J} c_j n^{j-1} + \eta_n = x_n + \eta_n$$

by where n represents both the time samples and the index in a YF , eco-epidemiological, oviposition, African, riceland, discontinuous, forest-canopy, vulnereability, *Ae aegypti,* capture point, forecast,LULC model. The polynomial describing the trajectory may be ay a degree J-1; and η_n may be zero mean, stationary, white noise (not necessarily Gaussian) with variance σ_n^2. Estimating x(t) at time τ with the MFOE may then

$$\hat{x}(\tau) = \sum_{n=1}^{N} y_n w_n(\tau)$$

describe where the hat (^) denotes an estimate, N which may be the number of ovispoition hyperproductive samples in the data window. In the YF model τ would be the time of the desired estimate, and the data weights would be measured as

$$w_n(\tau) = \sum_{m} U_{mn} T_m(\tau) f_m$$

The U_{mn} would be orthogonal polynomial LULC, sub-mter resolution.coefficient estimators. $T_m(\tau)$, a function may detai iprojects and estimate the polynomial coefficient c_m from the desired estimation time τ. The MFOE parameter $0 \le f_m \le 1$ may apply a fraction of the projected coefficien YF estimate.

The combined terms $U_{mn} T_m$ effectively constitute a novel set of expansion functions with coefficients f_m(Cressie 1993). The MFOE in the YF model may be optimized at time τ as a function of the f_ms for given geo-spatiotemporal or geo-spectrotemporal, LULC measurement noise, target (hyperproductive *Ae aegypti* riceland habitat) dynamics, and non-recursive sliding data window size, N. However, for all $f_m = 1$, the MFOE nay reduce and hence would be equivalent to the KF in the absence of process noise, and to the standard polynomial LSM in the model output.

As in the case of coefficients in conventional series expansions, the f_ms typically decrease monotonically as higher order terms are included to match complex target trajectories. For example, in the f_ms monotonically decreased in the MFOE from $f_1 = 1$ to $f_5 \gtrsim 0$, where $f_m = 0$ for m ≧ 6. The MFOE may consist of five point, 5th order processing of composite real (but altered for declassification). A window of only 5 data points(geosampled seasonal, hypeproductive *Ae. aegypti,* ovipsoition sites in an African, riceland, expanding environment) provided excellent maneuver following; whereas, 5th order processing included fractions of higher order terms to better approximate the complex maneuvering target trajectory. The MFOE

overcomes the long-ago rejection of terms higher than 3rd order because, taken at full value (i.e., $f_m s = 1$), estimator variances increase exponentially with linear order increases. (This is elucidated below in the section "Application of the FOE".)

As described inthe MFOE can be written more efficiently as $\hat{x} = <\psi, \omega_m>$ where the estimator weights $w_n(\tau)$ of order m are components of the estimating vector $\omega_m(\tau)$. By definition $\hat{x} \doteq \hat{x}(\tau)$ and $\omega_m \doteq \omega_m(\tau)$. The angle brackets and comma $<,>$ denote the inner product, and the data vector ψ comprises noisy measurement samples y_n. Perhaps the most useful MFOE tracking estimator is the simple fractional order estimator (FOE) where $f_1 = f_2 = 1$ and $f_m = 0$ for all $m > 3$, leaving only $0 \leq f_3 \leq 1$. This is effectively an FOE of fractional order $2 + f_3$, which linear interpolates between the 2nd and 3rd order estimators describedin)as $w_{2 \mid f_3} = (1 - f_3)\omega_2 + f_3\omega_3 = \omega_2 + f_3(\omega_3 - \omega_2) = \omega_2 + f_3\nu_3$ where the scalar fraction f_3 is the linear interpolation factor, the vector $\nu_3 = \omega_3 - \omega_2 = v_3 T_3$, and v_3 (which comprises the components v_{3n}) is the vector estimator of the 3rd polynomial coefficient $c_3 \equiv \frac{a\Delta^2}{2}$ (a is acceleration and Δ is the sample period). The vector ν_3 is the acceleration estimator from ω_3.

The mean-square error (MSE) from the FOE applied to an target (e.g., seasonal, hyperproductive, African, *Ae. aegypti*, oviposition, LULC, sub-meter, resolution, capture point, riceland, is $MSE = \sigma_\eta^2(|\omega_2|^2 + f_3^2|\nu_3|^2) + |c_3 T_3(1 - f_3)|^2$, where for any vector θ, $|\theta|^2 \doteq <\theta, \theta>$. The first term on the right of the equal sign is the FOE target habitat location estimator variance $\sigma_\eta^2(|\omega_2|^2 + f_3^2|\nu_3|^2)$ composed of the 2nd order location estimator variance and part of the variance from the 3rd order paramter estimator as determined by the interpolation factor squared f_3^2. The second term is the bias squared $[c_3 T_3(1 - f_3)]^2$ from the 2nd order target habitat location estimator as a function of acceleration in c_3. Setting the derivative of the MSE with respect to f_3 equal to zero and solving yields the optimal f_3: $f_{3,opt} \doteq f_{3,opt}(\tau) = \frac{(c_3 T_3)^2}{(c_3 T_3)^2 + \sigma_\eta^2|\nu_3|^2} = \frac{c_3^2}{c_3^2 + \sigma_\eta^2|v_3|^2} = \frac{\rho_3^2}{\rho_3^2 + |v_3|^2}$ where $\rho_3 \equiv \frac{c_3}{\sigma_\eta} = \frac{a\Delta^2}{2\sigma_\eta}$, as defined in an eco-epidemiological, forecast-oriented, YF , eco-georeferenceable, vulnerability, vulnerability, endmember, sub-meter resolution, LULC model. The optimal FOE is then very simply $w_{2+f_{3,opt}} = \omega_2 + f_{3,opt}\nu_3 = \omega_2 + v_3 T_3 f_{3,opt} = \omega_2 + v_3 T_3 \frac{\rho_3^2}{\rho_3^2 + |v_3|^2}$ Substituting the optimal FOE into the MSE yields the minimum: $MSE_{min} = \sigma_\eta^2(|\omega_2|^2 - f_{3,opt}|\nu_3|^2)$ Although not obvious, the MSE_{min} may includs the bias squared. The variance in the FOE MSE is the quadratic interpolation between the 2nd and the 3rd order location estimator variances as a function of $f_{3,opt}^2$. Whereas, the MSE_{min} is the linear interpolation between the same 2nd and the 3rd order location estimator variances as a function of $f_{3,opt}$. The bias squared accounts for the difference(Griffith 2003)..

Since a target's future immature habitat, capture point, riceland, *Ae egypti* habitat location is generally of more interest than where it is or has been, consider one-step prediction. Normalized with respect to measurement noise variance, the MSE for equally spaced samples reduces for the predicted, capture point, oviposition, LULC, position to $MSE = \frac{1}{N} + \frac{3(N+1)}{N(N-1)} + f_3^2 \frac{5(N+1)(N+2)}{N((N-1)(N-2)} + \rho_3^2 \left[\frac{(N+1)(N-2)}{6} \right]^2 (1-f_3)^2$ where N is the number of habitat samples in the non-recursive sliding data window. Note that the first term on the right of the equal sign is the variance from estimating the first coefficient (position); the second term is the variance from estimating the 2nd coefficie nt (velocity); and the 3rd term with $f_3 = 1$ is the variance from estimating the 3rd coefficient (which includes acceleration). This pattern continues for higher order terms. Further, the sum of the variances from estimating the first two coefficients is $\frac{4N+2}{N(N-1)}$). Adding the variance from estimating the 3rd coefficient yields $\frac{9N^2 + 9N + 6}{N(N-1)(N-2)}$.

Estimator variances obviously increase exponentially with unit order increases. In the absence of process noise, the KF yields variances equivalent to these. (A derivation of the variance from a 1st degree polynomial corresponding to $f_3 = \rho_3 = 0$ for the generalized case of arbitrary estimation time and sample times is given in reference. In addition, establishing a multi-dimensional tracking gate at the predicted position may easily be aided with the simple approximation of the error function in an eco-epidemiological, forecast-oriented, YF , eco-georeferenceable, vulnerability, endmember, sub-meter resolution LULC model.

Tuning the KF consists of a trade-off between measurement noise and process noise to minimize the estimation error. The KF process noise serves two roles: First, its covariance is sized to account for the maximum expected target acceleration. Second, process noise covariance establishes an effective recursive , LULC geoclassified,YF data window (analogous to the non-recursive sliding data window), described by Brookner as the Kalman filter memory.

Contrary to process noise covariance as a single independent parameter in the KF serving two roles, the FOE has the advantage of two separate independent eco-epidemiological, forecast-oriented, YF, eco-georeferenceable, vulnerability, endmember, sub-meter resolution, LULC model parameters: one for acceleration and the other for sizing the sliding data window. Therefore, as opposed to being limited to just two tuning parameters (process and measurement noises) as is the KF, the FOE includes three independent tuning parameters: measurement noise variance, the assumed maximum deterministic target acceleration (for simplicity both target acceleration and measurement noise are included in the ratio of the single parameter ρ_3), and the number of *Ae,aegypti* ovispoition samples in the data window.

A YF model MSE_{min} curve can yield the RMSE (square root of the MSE). On the other hand, choosing a vaable n yields the second curve As shown the optimal FOE is essentially a 3rd order non-recursive capture point, LULC estimator which may yieldsless than, for example, 4% RMSE improvement over an optimal FOE in the case of no acceleration. However, in the case of maximum acceleration the optimal MSE form a remotely sensed, riceland, African, capture

point, eco-epidemiology, YF model markedly volatile and can have large error spikes that can confuse an arborvirologist, medical entomology, or expeimenter one spike exceeding the optimal MSE for worst case. Obviously, higher values of YF immature habitat N produce larger error spikes.Since trackers encounter greatest difficulties and often lose track during target capture point, *Ae aegypti,* forecasts., the much smoother MSE_{min}transition of the optimal FOE has a major advantage over larger YF data siumations windows.

One consequence of a high degree of multicollinearity is that, even if the matrix $X^{\top}X$is invertible, a computer algorithm may be unsuccessful in obtaining an approximate inverse, and if it does obtain one it may be numerically inaccurate. But even in the presence of an accurate $X^{\top}X$matrix, the following consequences arise.

In the presence of multicollinearity in a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC, oviposition , African, riceland model, the estimate of one variable's impact on the discontinuous, forest-canopied, dependent variable Ywhile controlling for the others tends to be less precise than if predictors were uncorrelated with one another. The usual interpretation of a sub-meter resolution, explanatorial, time series, regression coefficient is that it provides an estimate of the effect of a one unit change in an independent variable, X_1, holding the other variables constant. If $X_1$is highly correlated with another independent variable, X_2, in the given eco-epidemiological, LULC, riceland African, orthogonal, forest-canopied, discontinuous, orthogonal, grid-stratified, geosampled dataset, then we have a set of observations for which $X_1$and X_2have a particular linear stochastic relationship. Currently a set of YF geosampled observations for which all LULC changes in riceland $X_1$are independent of changes in inhomogeneous, forest-canopy X_2, so we have an imprecise, forecasted, LULC estimate of the effect of independent changes in X_1.

In some sense, the collinear variables contain the same information about the dependent variable. If nominally "different" measures actually quantify the same phenomenon then they are redundant. Alternatively, if the YF variables are accorded different names and perhaps employ different numeric measurement scales but are highly correlated with each other, then they suffer from redundancy.One of the features of multicollinearity is that the standard errors of the affected coefficients tend to be large. In that case, the test of the hypothesis that the coefficient is equal to zero may lead to a failure to reject a false null hypothesis of no effect of the explanator, a type II error.

In statistical hypothesis testing, a type I error is the incorrect rejection of a true null hypothesis (a "false positive"), while a type II error is the failure to reject a false null hypothesis (a "false negative"). More simply stated, a type I error is detecting an effect that is not present, while a type II error is failing to detect an effect that is present. The terms "type I error" and "type II error" are often used interchangeably with the general notion of false positives and false negatives in binary classification, such as medical testing, but narrowly speaking refer specifically to statistical hypothesis testing in the Neyman–Pearson framework In statistics, the Neyman–Pearson lemma, , states that when performing a hypothesis test between two simple hypotheses H_0: $\theta = \theta_0$ and H_1: $\theta = \theta_1$, the likelihood-ratio test which rejects H_0 in favour of H_1

when $\Lambda(x) = \dfrac{L(\theta_0 \mid x)}{L(\theta_1 \mid x)} \leq \eta$ where $P(\Lambda(X) \leq \eta \mid H_0) = \alpha$ is the most powerful test at significance level α for a threshold η. If the test is most powerful for all $\theta_1 \in \Theta_1$, it is said to be uniformly most powerful (UMP) for alternatives in the set Θ_1.

Each of the two competing models, the null model and the alternative model, is separately fitted to the data and the log-likelihood recorded. The test statistic (often denoted by *D*) is twice the log of the likelihoods ratio, *i.e.*, it is twice the difference in the log-likelihoods:

$$D = -2 \ln \left(\frac{\text{likelihood for null model}}{\text{likelihood for alternative model}} \right) = 2 \ln \left(\frac{\text{likelihood for alternative model}}{\text{likelihood for null model}} \right)$$
$$= 2 \times [\ln(\text{likelihood for alternative model}) - \ln(\text{likelihood for null model})]$$

The model with more parameters (here *alternative*) will always fit at least as well; that is: They have a greater or equal log-likelihood, than the model with less parameters (here *null*). Whether it fits significantly better and should thus be preferred is determined by deriving the probability or *p*-value of the difference *D*. Where the null hypothesis represents a special case of the alternative hypothesis, the probability distribution of the test statistic is approximately a chi-squared distribution with degrees of freedom equal to $df_{alt} - df_{null}$.[10] Symbols df_{alt} and df_{null} represent the number of free parameters of models *alternative* and *null*, respectively.

Here is an example of use. If the YF null model has 1 *Ae aegypti,* parameter and a log-likelihood of −8024 and the alternative model has 3 parameters and a log-likelihood of −8012, then the probability of this difference is that of chi-squared value of $2 \times (-8012 - (-8024)) = 24$ with $3 - 1 = 2$ degrees of freedom, and is equal to 6×10^{-6}. Certain assumptions[6] must be met for the statistic to follow a chi-squared distribution, and often empirical *p*-values are computed.

The likelihood-ratio test requires nested models – models in which the more complex one can be transformed into the simpler model by imposing a set of constraints on the parameters. If the models are not nested, then a generalization of the likelihood-ratio test can usually be used instead: the relative likelihood.

Suppose that the ML estimate for *θ* is $\hat{\theta}$ in a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model.Relative plausibilities of other *θ* values may be then found by comparing the likelihood of those other values with the likelihood of $\hat{\theta}$. The relative likelihood of *θ* is defined[3][4] as $\mathcal{L}(\theta|x)/\mathcal{L}(\hat{\theta}|x)$. A 10% likelihood region for *θ* is $\{\theta : \mathcal{L}(\theta|x)/\mathcal{L}(\hat{\theta}|x) \geq 0.10\}$ and more generally, a *p*% likelihood region for *θ* is defined[3][4] to be $\{\theta : \mathcal{L}(\theta|x)/\mathcal{L}(\hat{\theta}|x) \geq p/100\}$. If *θ* is a single real parameter, a *p*% likelihood region will typically comprise an interval of real values. In that case, the region is called a likelihood interval.

Likelihood intervals can be compared to confidence intervals in a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model especially Likelihood ratio tests based on Wilks's theorem, which says that 2 times log(likelihood ratio) of nested hypotheses is approximately chi-square under certain commonly met criteria. Most importantly, the parameters of the larger YF model fixed to produce the smaller model must lie in the interior of the parameter space; one or more cannot be on a boundary. In such cases, the approximating chi-square distribution has degrees of freedom equal to the number of parameters in the larger model fixed to produce the smaller model. The indexing percentage for likelihood intervals is quite different from confidence intervals: If θ is a single real parameter, then under certain conditions, a 14.7% likelihood interval for θ will be the same as a 95% confidence interval.

Ideally, statistical software (e.g., SAS, R) automatically can provide charts and statistics that test whether regression assumptions in a are satisfied for any given eco-epidemiological, LULC<, YF forecast vulnerability, eco-epidemiological model. Unfortunately, many software packages do not provide such output by default (i.e., additional menu commands must be executed or code must be written) and some (such as Excel's built-in regression add-in) offer only limited options. RegressIt does provide such output and in graphic detail. Example of outputs from YF explicatively, geopredictive paradigms that violate all of the assumptions previously mentioned is littered in literature. In most of these models the naïve experimeneter has accepted the diagnostic, clinical ,field or remote LULC forecasts due to s large value of R-squares. R-squared is a statistical measure of how close the data are to the fitted regression line. It is also known as the coefficient of determination, or the coefficient of multiple determination for multiple regression. The Coefficient of Determination Interpretation is the accuracy of the predictor of the independent variable on the dependent variable value. Multiple regression analysis expresses a relationship between a set of predictor variables and a single criterion variable by the multiple correlation R, multiple coefficient of determination R2, and a set of standard partial regression weights β , β , etc

A eco-epidemiological, robust YF , time series, forecasting, vulnerability, *Ae egypti,* African riceland, habitat, ovispoition, LULC model that satisfies regression assumptions at least reasonably well by the usage of robust, nonlinear explanative, log-transformations of an empiricial geosampled, geo-spectrotempora, eco-epidemiologicall datset of diagnostic, clinical, field or remote explanatory variables doe not currently exist. In literature normalized quantile plots from diagnostic, vector arthropod-related medical entomological geo-predictive models are rare.

In statistics, a Q–Q plot ("Q" stands for *quantile*) is a probability plot, which is a graphical method for comparing two probability distributions by plotting their quantiles against each other. First, the set of intervals for the quantiles is chosen. A point (x, y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate). Thus the line is a parametric curve with the parameter which is the (number of the) interval for the quantile.If the two distributions being compared are similar, the points in the Q–Q plot will approximately lie on the line $y = x$. If the distributions are linearly related, the points in the Q–Q plot will approximately lie on a line, but

not necessarily on the line *y* = *x*. Q–Q plots can also be used as a graphical means of estimating parameters in a location-scale family of distributions.

A Q–Q plot is used to compare the shapes of distributions, providing a graphical view of how properties such as location, scale, and skewness are similar or different in the two distributions. Q–Q plots can be used to compare collections of data, or theoretical distributions. The use of Q–Q plots to compare two samples of a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland.data can be viewed as a non-parametric approach to comparing their underlying distributions. A Q–Q plot is generally a more powerful approach to do this than the common technique of comparing histograms of the two samples, but requires more skill to interpret(Anselin 1995).. Q–Q plots are commonly used to compare a data set to a theoretical model. This can provide an assessment of "goodness of fit" that is graphical, rather than reducing to a numerical summary. Q–Q plots are also used to compare two theoretical distributions to each other.[4] Since Q–Q plots compare distributions, there is no need for the values to be observed as pairs, as in a scatter plot, or even for the numbers of values in the two groups being compared to be equal.

A scatter plot (also called a scatter graph, scatter chart, scattergram, or scatter diagram) is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data. If the points are color-coded you can increase the number of displayed variables to three. The data is displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis. A scatter plot can be used either when one continuous variable that is under the control of the experimenter and the other depends on it or when both continuous variables are independent. If a parameter estimator exists in a a seasonal, hyperproductive, YF eco-epidemiological, *Ae. aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African, riceland model that is systematically incremented and/or decremented by the other, it is called the control parameter or independent variable and is customarily plotted along the horizontal axis. The measured or dependent variable is customarily plotted along the vertical axis. If no dependent variable exists, either type of variable can be plotted on either axis and a scatter plot will illustrate only the degree of correlation (not causation) between two variables.

A scatter plot can suggest various kinds of correlations between variables with a certain confidence interval. For example, weight and height, weight would be on y axis and height would be on the x axis. Correlations may be positive (rising), negative (falling), or null (uncorrelated). If the pattern of dots slopes from lower left to upper right, it indicates a positive correlation between the variables being studied. If the pattern of dots slopes from upper left to lower right, it indicates a negative correlation. A line of best fit (alternatively called 'trendline') can be drawn in order to study the relationship between the variables. An equation for the correlation between the variables can be determined by established best-fit procedures. For a linear correlation, the best-fit procedure is known as linear regression and is guaranteed to generate a correct solution in a finite time. No universal best-fit procedure is guaranteed to generate a correct solution for arbitrary relationships. A scatter plot may be very useful when an arbovirologist, medical entomologist or YF experiment wishs to see how two comparable a

seasonal, hyperproductive, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland datasets agree with each other. In this case, an identity line, *i.e.*, a *y=x* line, or an 1:1 line, is often drawn as a reference. The more the two data sets agree, the more the scatters tend to concentrate in the vicinity of the identity line; if the two data sets are numerically identical, the scatters fall on the identity line exactly.

One of the most powerful aspects of a scatter plot, however, is its ability to show nonlinear relationships between variables. The ability to do this can be enhanced by adding a smooth line such as LOESS. Further, if the a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model data are represented by a mixture model of simple relationships, these relationships will be visually evident as superimposed patterns.

The scatter diagram is one of the seven basic tools of quality control(Griffith 2003). Scatter charts can be built in the form of bubble, marker, or/and line charts.[6] For a set of data variables (dimensions) X_1, X_2, ... , X_k, the scatter plot matrix shows all the pairwise scatter plots of the variables on a single view with multiple scatterplots in a matrix format. For k variables, in a a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model,the scatterplot matrix will contain k rows and k columns. A plot located on the intersection of i-th row and j-th column is a plot of variables X_i versus X_j. This means that each row and column is one dimension, and each cell plots a scatterplot of two dimensions.

The Tukey-lambda PPCC plot is used to suggest an appropriate distribution. One should follow-up with PPCC and probability plots of the appropriate alternatives In probability theory and statistics, the beta distribution is a family of continuous probability distributions defined on the interval [0, 1] parametrized by two positive shape parameters, denoted by α and β, that appear as exponents of the random variable and control the shape of the distribution. The beta distribution has been applied to model the behavior of random variables limited to intervals of finite length in a wide variety of disciplines. For example, it has been used as a statistical description of allele frequencies in population genetics; time allocation in project management / control systems; sunshine data; variability of soil properties proportions of the minerals in rocks in stratigraphy and heterogeneity in the probability of HIV transmission.

The probability density function (pdf) of the beta distribution, for $0 \leq x \leq 1$, and shape parameters α, β > 0, is a power function of the variable *x* and of its reflection (1−*x*) as follows:

$$f(x; \alpha, \beta) = \text{constant} \cdot x^{\alpha-1}(1-x)^{\beta-1}$$

$$= \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}$$

$$= \frac{1}{\mathrm{B}(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

where $\Gamma(z)$ is the gamma function. The beta function, $\mathbb{B}$, is a normalization constant to ensure that the total probability integrates to 1. In the above equations x is a realization—an observed value that actually occurred—of a random process X. This definition includes both ends $x = 0$ and $x = 1$, which is consistent with definitions for other continuous distributions supported on a bounded interval which are special cases of the beta distribution, for example the arcsine distribution, and consistent with several authors, like N. L. Johnson and S. Kotz. However, several other authors, including W. Feller, choose to exclude the ends $x = 0$ and $x = 1$, (such that the two ends are not actually part of the density function) and consider instead $0 < x < 1$.

Statistical tests can be divided in two large groups based on the distribution of the variables to be investigated (Précsényi et al., 2000). Parametric tests, as it is in their name, estimate a parameter of the investigated population. They assume that the distribution of the investigated variable (or the error) is normal. The power of parametric tests is high (also small differences can be detected), but they have several assumptions and usually they can be used only on variables measured in a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti*, forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model.on ratio or interval scale. In contrary, non-parametric tests do not estimate a parameter. They do not require normality, but in case of some of them it is assumed that the distribution has a particular shape (e.g. symmetric). Non-parametric tests have fewer assumptions and can be used also on variables measured on nominal or ordinal scale(Griffth 2003).. They have usually lower power then their parametric counterpart, and many, especially the more complex parametric tests, do not have a non-parametric counterpart.

Violations of independence are potentially very serious in time series regression models: serial correlation in the errors (i.e., correlation between consecutive errors or errors separated by some other number of periods) means that there is room for improvement in the model, and extreme serial correlation is often a symptom of a badly mis-specified model. Serial correlation (also known as autocorrelation") is sometimes a byproduct of a violation of the linearity assumption, as in the case of a simple (i.e., straight) trend line fitted to data which are growing exponentially over time which may render mis-specified itaretively interpolated sub-mter resolution, *Ae. aegypti*, ovispoition, LULCs in a African, riceland, agro-ecosystem, forecast, vulnerability model.

Estimation of conditional choice probabilities in a binary choice, medical entomological, vector, arthropod-related, YFV, forecast, vulnerability model under uncertainty is considered weak due to nonparametric restrictions on expectations and the error distribution. The estimation method follows a two-stage strategy: the first stage estimates expectations employing realizations of elucidative, geo-spectrotemporal, diagnostic, clinical, field or remote-specified, explanatory, futuristically-oriented,, YFV-related descriptor (prevalanace rate) and their second stage, forecasted conditional choice probabilities employing choice parameterizable, covariate, coefficient data and their regression expectations estimates . Although under given conditions the two-stage nonparametric kernel estimator is uniformly, strongly consistent and asymptotically normal, geospectrotemporally geospatialized, explucatively diagnostic, georeferenced, time series, log-transmformed clinical data(e.g., annual adult, *Ae. egypti*, biting rates), field covariates (e.g., GPS ground cordinates of a hyperendemic, gridded, startified, land cover cluster centroid and its bidirectional wavelength remotely sensed frequencies (e.g., uncoalesced, iteratively

interpolatable, visible and near infra-red, transmittance) can render precision forecasts, primarily due to conflicting measurement coefficient values in the negative log-likelihood computation. Hence, the chi-square distributions are commonly erroneous in these geo-predictive eco-epidemiological, time series,probabilistic paraigidms.

The usual approach in   discrete choice problem solving when optimally   quantizing explanatorial, geo-spectrotemporal, uncoalesced,geospatialized,   sub-meter resolution, uncoalesced, wavelength frequency and clinically diagnostic, YFV-related, vector, forecastable, vulnerability analysis for targeting, immature, *Ae egypti* seasonal, georeferenceable, prolific, immature habitats   has maintained strong assumptions in regard to probability distributions based on a fixed set parameter expectations. For example,see Khormi (2013) and Gerade (2004). The disadvantage of these parametric assumptionsis the danger of incorrect specification, which may distort inference on preferences or choice probabilities.

Logistic regression was optimally employed in Reyes-Villanueva and, Rodríguez-Pérez 2004 for quantitating the likelihood of occurrence of a non-gonoactive *Aedes aegypti* female, previously fed human blood, with relation to body size and collection method.This study was conducted in Monterrey, Mexico, between 1994 and 1996. Ten samplings of 60 mosquitoes of *Ae. aegypti* females were carried out in three dengue endemic areas: six of biting females, two of emerging mosquitoes, and two of indoor resting females. Gravid females, as well as those with blood in the gut were removed. Mosquitoes were taken to the laboratory and engorged on human blood. After 48 hours, ovaries were dissected to register whether they were gonoactive or non-gonoactive. Wing-length in mm was an indicator for body size. The logistic regression model was used to assess the likelihood of non-gonoactivity, as a binary variable, in relation to wing-length and collection method.Of the 600 females, 164 (27%) remained non-gonoactive, with a wing-length range of 1.9-3.2 mm, almost equal to that of all females (1.8-3.3 mm). The logistic regression model showed a significant likelihood of a female remaining non-gonoactive (Y=1). The collection method did not influence the binary response, but there was an inverse relationship between non-gonoactivity and wing-length.

Logistic regression was a useful tool to estimate the likelihood for an engorged female to remain non-gonoactive in Reyes-Villanueva and Rodríguez-Pérez (2004). The forecasts revealed necessity for a second blood meal is present in any female, but small mosquitoes are more likely to bite again within a 2-day interval, in order to attain egg maturation. However, when building a logistic regression, *Ae.egypti,* time series, eco-epidemiological, forecast vulnerability model,a experimeneter may assume that the logit of the outcome variable is a linear combination of the independent variables. This involves two aspects, as the model dichotomous framework are dealing with the two sides of a logistic regression equation in the YFV-related vulnerability analyses. First, consider the link function of the outcome geosampled , diagnostic, explicative, time series, geo-variable on the left hand side of the equation. A experimenter may assume that the logit function is the correct function to use for quantitating an empirical geomapled datset of geo-spectrotemporally paramterizable, diagnostic, YFV, clinical, field an remote specified covariates. Secondly, on the right hand side of the equation, an experimenter may assume that all the relevant variables,that are not included should not be in the model, and the logit function is a linear combination of the explanative geo-predictors. It could happen that the logit function (i.e., link function) is not the correct choice or the relationship

between the logit of outcome variable and the independent variables is not linear. In either case, a specification error would occur. The misspecification of the link function is usually not too severe compared with other alternative link function choices such as probit (based on the normal distribution) (Griffith 2003). In practice, a yellow fever experimenter should be concerned with whether the Logistic binary forecasting vulnerability model has all the relevant geo-predictors and if the linear combination of them is sufficient.

Logistic regression models are commonly used to study the association between a binary response variable and an exposure variable. Besides the exposure of interest, other covariates are frequently included in the fitted model in order to control for their effects on outcome. Unfortunately, misspecification of the main exposure variable and the other covariates is not uncommon, and this can adversely affect tests of the association between the exposure and response. We allow the term "misspecification" to cover a broad range of modeling errors including measurement errors, discretizing continuous explanatory variables, and completely excluding covariates from the model.

Suppose that x denotes the explanatory variable of interest, and z denotes a vector of other explanatory variables in a a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model.For simplicity of presentation, we shall hereafter refer to x and z as the exposure variable andcovariates, respectively. When the outcome of interest is binary, logistic regression models are commonly used to study the association between exposure and response. If Y denotes the binary response, then the relationshipbetween exposure and response is modeledas:logit Pr(Y = lIx,z) = 0 + ax + ,'z (1) where 0, a, and , are unknown parameters. The hypothesis of no association between exposure and outcomecan be expressed as: Ho:a = 0 Given n independent and identically distributed observations of the form (Yi, xi, zi), i = 1,2, .. ,n, this hypothesis can be assessed using the likelihood score testof a - 0, say Q (x,z), which is asymptotically equivalentto tests of a = 0 based on the maximum likelihoodestimator of a and on the likelihood ratio statistic (1).

Alternatively, Manski (1988b, 1991) introduced a two-stage semiparametric method applicable if expectations are fulfilled and are conditioned only on variables (i.e., geo-spectrotemporal, kriged, proxy, *Ae egypti,* forest-canopy, LULC, sub-meter resolution, unmixed biosignatures) observed by the researcher. The first stage estimates expectations nonparametrically using realizations of agents' futures, and the second stage estimates preference parameters employing choice data and estimated expectations under a parametric restriction on the error distribution. The strong consistency of the two-stage estimator was shown in Manski (1991). The rate of convergence and the asymptotic distribution of the estimator was studied by Ahn and Manski (1993). In this semiparametric method, conditional choice probabilities are inferred from a triple: estimated expectations, two-stage estimation of the preference parameters, and a parametrically specified error distribution. This two-stage semiparametric method may be promising for elucidating geo-predictive explanators of hypeproductive*, Ae egypti* aquatic larval habitats on geoclassified, seasonal geoclassifiable LULCs in that it does not prespecify the functional form of expectations and separates inference on expectations from inference on preferences. Also, it substantially simplifies what would be a complicated finite-horizon dynamic choice problem; it begins with the observation (e.g., geosampled georeferenced, *Ae. egypti,*

riceland, agro-village complex, sub-meter resolution, geoclassified LULC centroid prolific oviposition, capture point) and the assumption that any dynamic choice problem has an observationally equivalent static representation. As such solving a backward recursion problem in a forecast vulnerability, georeferenceable, geo-spectrotemporal, geosampled, *Ae. egypti* which is very often computationally intractable.

Recursion is the process of repeating items in a self-similar way(Cressie 1993). For instance, when the surfaces of two seasonally geoclassifiable, georeferenceable, *Ae.egypti* LULCs are exactly parallel with each other, the nested images that occur may be a form of infinite recursion, The most common application of recursion is in mathematics and computer science, in which it refers to a method of defining functions in which the function being defined is applied within its own definition. Specifically, this defines an infinite number of instances (function values), using a finite expression that for some instances may refer to other instances, but in such a way that no loop or infinite chain of references can occur. The term is also used more generally to describe a process of repeating objects in a self-similar way.

Finite subdivision rules are a geometric form of recursion, which can be used to create fractal-like images. A subdivision rule can start with a collection of seasonally geoclassifiable, georeferenceable, sub-meter resolution, *Ae.aegypti* LULCs polygons labelled by finitely many labels (e.g., discontinuous, forest canopy, riceland, agro-ecosystem, post tillering) , and then each polygon can be subdivided into smaller labelled polygons (e.g., sparsely shaded, trailing vegetation) in a way that depends only on the labels of the original polygon. This process can be iterated in SAS or R. The standard `middle thirds' technique for creating the Cantor set is a subdivision rule, as is barycentric subdivision.

In mathematics, the Cantor set is a set of points lying on a single line segment that has a number of remarkable and deep properties. Through consideration of this set, Cantor and others helped lay the foundations of modern point-set topology. Although Cantor himself defined the set in a general, abstract way, the most common modern construction is the Cantor ternary set, built by removing the middle thirds of a line segment. Cantor himself only mentioned the ternary construction in passing, as an example of a more general idea, that of a perfect set that is nowhere dense.

The Cantor ternary set $\mathcal{C}$ is created by deleting the open middle third from each of a set of line segments repeatedly. One starts by deleting the open middle third $(1/3, 2/3)$ from the interval $[0, 1]$, leaving two line segments: $[0, 1/3] \cup [2/3, 1]$. Next, the open middle third of each of these remaining segments is deleted, leaving four line segments: $[0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$. This process is continued ad infinitum, where the *n*th set is

$$C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3} \right) \text{ and } C_0 = [0,1].$$

A Cantor ternary set contains all points in the interval [0, 1] that are not deleted at any step in this infinite processwher $\mathcal{C} := \lim_{n=1}^{\infty} C_n = \bigcap_{n=1}^{\infty} C_n.$

An explicit closed formula for the Cantor set may be $\mathcal{C} = \bigcap_{n=1}^{\infty} \bigcap_{k=0}^{3^{n-1}-1} \left(\left[0, \frac{3k+1}{3^n}\right] \cup \left[\frac{3k+2}{3^n}, 1\right] \right)$ or $\mathcal{C} = [0,1] \setminus \bigcup_{n=1}^{\infty} \bigcup_{k=0}^{3^{n-1}-1} \left(\frac{3k+1}{3^n}, \frac{3k+2}{3^n} \right).$ for a a seasonal, hyperproductive, YF eco-epidemiological, *Ae.aegypti,* forecast, vulnerability, sub-meter resolution, LULC ,oviposition , discontinuous, forest-canopied, African riceland model. This process of removing middle thirds is a simple example of a finite subdivision rule. In mathematics, a finite subdivision rule is a recursive way of dividing a polygon or other two-dimensional shape into smaller and smaller pie.

To construct a Canto function, an arbovirologist, medical entomologist or YF expeimenter, may consider a seasonal hyperproductive, capture points in the [0, 1] interval in terms of base 3 (or ternary) notation. Recall that some points admit more than one representation in this notation, as for example $^1/_3$, that can be written as 0.1_3 but also as $0.022222..._3$, and $^2/_3$, that can be written as 0.2_3 but also as $0.12222..._3$. This alternative recurring representation of a number with a terminating numeral occurs in any positional system. When an arbovirologist, medical entomologist or YF expeirmenter remove the middle third, this contains the numbers with ternary numerals of the form $0.1xxxxx..._3$ in aYF model where $xxxxx..._3$ is strictly between $00000..._3$ and $22222..._3$. So the numbers remaining after the first step consist of

- Numbers of the form $0.0xxxxx..._3$
- $^1/_3 = 0.1_3 = 0.022222..._3$
- $^2/_3 = 0.122222..._3 = 0.2_3$
- Numbers of the form $0.2xxxxx..._3$.

This can be summarized by saying that those numbers that admit a ternary representation such that the first digit after the decimal point is not 1 are the ones remaining after the first step.

The second step removes numbers of the form $0.01xxxx..._3$ and $0.21xxxx..._3$, and (with appropriate care for the endpoints) it can be concluded that the remaining numbers are those with a ternary numeral where neither of the first *two* digits is 1. Continuing in this way, for a number

213

not to be excluded at step n, it must have a ternary representation of a YF model whose nth digit is not 1. For a number to be in the Cantor set, it must not be excluded at any step, it must admit a numeral representation consisting entirely of 0s and 2s. It is worth emphasising that numbers like 1, $1/3 = 0.1_3$ and $7/9 = 0.21_3$ are in the Cantor set, as they have ternary numerals consisting entirely of 0s and 2s: $1 = 0.2222..._3$, $1/3 = 0.022222..._3$ and $7/9 = 0.2022222..._3$. So while a number in $\mathcal{C}$ may have either a terminating or a recurring ternary numeral, one of its representations will consist entirely of 0s and 2s.

The function from $\mathcal{C}$ to [0,1] is defined by taking the numeral that does consist entirely of 0s and 2s, replacing all the 2s by 1s, and interpreting the sequence as a binary representation of a real number. In a formula, $f\left(\sum_{k=1}^{\infty} a_k 3^{-k}\right) = \sum_{k=1}^{\infty} \frac{a_k}{2} 2^{-k}.$ (Cressie 1993) For any number y in [0,1], its binary representation can be translated into a ternary representation of a number x in $\mathcal{C}$ by replacing all the 1s by 2s. With this, $f(x) = y$ so that y is in the range of f in any eco-epidemiological, YF , geo-spectrotemrpoal or geo-spatiotemrpoal, *Ae. aegypti* , forecast, vulnerability, LULC model For example, if $y = 3/5 = 0.100110011001..._2$, an arbovirologist , medical entomologist or other expiementer couls write $x = 0.200220022002..._3 = 7/10.$ Consequently, f would be surjective; however, f would not be injective in the African riceland YF model— interestingly enough, the model values for which $f(x)$ coincides with would be those that would be at opposing ends of one of the middle thirds removed. For exmaple, $7/9 = 0.2022222..._3$ and $8/9 = 0.2200000..._3$ so $f(7/9) = 0.101111..._2 = 0.11_2 = f(8/9)$.

So there are as many points in the Cantor set as there are in [0, 1], and the Cantor set is uncountable (see Cantor's diagonal argument). However, the set of endpoints of the removed intervals is countable, so there must be uncountably many numbers in the Cantor set which are not interval endpoints. As noted above, one example of such a number is ¼, which can be written as $0.02020202020..._3$ in ternary notation.

The Cantor set contains as many points as the interval from which it is taken, yet itself contains no interval of nonzero length. The irrational numbers have the same property, but the Cantor set has the additional property of being closed, so it is not even dense in any interval, unlike the irrational numbers which are dense in every interval.It has been conjectured that all algebraic irrational numbers are normal. Since members of the Cantor set are not normal, this would imply that all members of the Cantor set are either rational or transcendental.

The Cantor set for a YF, ricland , African ecosystem, *Ae. aegypti*, sub-meter resolution, LULC,forest-canopy risk model would be of the prototype of a fractal. It would be self-similar, because it would be equal to two copies of itself, if each copy is shrunk by a factor of 3 and translated. More precisely, there may be two functions, the left and right self-similarity transformations, $f_L(x) = x/3$ and $f_R(x) = (2+x)/3$, which leave the Cantor set invariant tabulated from a YF forecast vulnerability model up to homeomorphism: $f_L(\mathcal{C}) \cong f_R(\mathcal{C}) \cong \mathcal{C}$.

Jacob et al. (2014) considered estimation of conditional choice probabilities without estimating preference malaria, mosquito, vector, *An. arabiensis* sub-meter resolution, QuickBird, waveband dataset of geo-spectrotemporal uncoalesced parameters (i.e., the only parametric

component of the entomological, forecast, vulnerability model) under weak nonparametric restrictions on expectations and the error distribution.The analysis was then two-stage nonparametric: The first stage estimates expectations nonparametrically, and the second stage estimatsd conditional choice probabilities nonparametrically.

In Jacob et al. (2014) the authors spectrally decomposed a sub-meter spatial resolution, riceland, agro-village complex, paddy preparation, *An. arabiensis* capture point, immature, habitat, mixed pixel (mixel) for iteratively interpolating a hyperproductive,immature habitat, uncoalesced signature in a neighboring riceland agro-ecosystem, study site, georferenceable, environment. Initially,the authors constructed a regression model which revealed that paddy preparation *An. arabiensis* habitats were the most productive based on spatiotemporal, geo-spectrotemporal, field-geosampled count data. Individual mixel, geospectral, wavelength, frequency,emissivity radiance estimates from the QuickBird visible and near-infra-red (NIR) data of a paddy preparation *An. arabiensis* habitat were then extracted by using a Li-Strahler geometric-optical model. The model used three scene components: sunlit canopy (C), sunlit background (G) and shadow (T) generated from the image. The G, C, T components' classes were estimated using ENVI, an object-based classification algorithm. ENVI bundles together a number of scientific algorithms for image processing a lot of which are contained in automated, wizard-based approach that walks users through complex tasks (http://www.harrisgeospatial.com/ProductsandSolutions/GeospatialProducts.aspx)

In ENVI®, the Digital Number (DN) of the mixel in every QuickBird band was viewed using the z-profile from a spectral library. After making an atmospheric correction from the image for the study site, the DN was converted into ground reflectance. At the "hotspot," (e.g., a paddy preparation, *An. arabiensis* capture point georferenced immature habitat) when illumination and viewing positions coincide, shadows were hidden behind sparsely canopied, plant crowns and the scene appeared bright. *As* the viewing position diverged from that of illumination, the shadows behind the crowns were progressively revealed and the scene darkened. Because, in general, the shadows of the habitat were not circular, the amount of shadow revealed was a function of both the zenith and azimuth angles by which the viewing and illumination positions diverged, rather than a simple phase angle between them. This effect creates an assymetric hotspot, in which the shape of the habitat hotspot was related to the shape of the plant crowns in the QuickBird scene. At large zenith angles, mutual shadowing of crowns becomes an important factor(Jensen 2005).. Illumination shadows will tend to fall on other crowns, rather than the background, and will preferentially shadow the lower portions of adjacent crowns. Further, these shadows were obscured since adjacent crowns tended to obscure the lower portions of other crowns. This effect produces a "bowl-shaped" bidirectional reflectance distribution function (BRDF) in which the scene brightness increases at the function's edges. Deriving formulas describing the hotspot and mutual shadowing effects and presents examples showed how the shape of the BRDF was dependent on the shape of the *An. arabiensis* capture point habitat crowns, their density and their their brightness relative to the background. A convex geometrical model was also used for endmember validation of the spectrally decomposed paddy preparation habitat. An ordinary kriged-based interpolation was performed in ArcGIS Geostatistical Analyst^{TM} employing the reference signature generated from the unmixing models. Linear unbiased predictors and variance estimates were derived of all productive paddy preparation *An. arabiensis* habitats in the study site based on the extracted

decomposed mixel endmember reflectance estimates. Spectral unmixing tools may be used to decompose QuickBird visible and NIR mixel reflectance of a productive *Ae egypti* capture poin ,georferenced, immature habitat. Thereafter, an ordinary interpolator can use the sub-mixel data along with other spatially continuous explanatory variables geosampled from productive immature *Aedes* habitats for targeting other high density foci habitat sites which can help implement larval control strategies in a riceland expanding environment.

In the literature, the 'two-stage' approach is often used due to the presence of a nuisance parameter or nuisance function in the equation of interest when risk modeling, uncoalesced, sub-meter resolution, geo-spectrotemporal, explicatively geospatialized, medical entomological, geopredictive paraidigms. See, for example, Ahn and Powell (1993) Heckman (1976), Powell (1987) or Robinson (1988). This two-stage method is appealing since it is conceptually simple and easily applicable. In the first stage, one estimates the nuisance parameter (or function); in the second stage, one proceeds to the equation of interest with the first-stage estimate replacing the unknown nuisance parameter (or function).

From a technical perspective, the present two-stage method in an explicatively robustifiable, geo-spectrotemporal, sylvatic, *Ae egypti* LULC, forecast vulnerability, geospatial model for inference on conditional choice probabilities would involve nonparametric estimation of conditional mean functions in both the first and second stages, commonly qualitatively quantiated in SAS/GIS. The estimation method in the entomological model would be 'nonparametric' because neither the functional form of the mean regression nor the error distribution woulod be prespecified. Basically, the nonparametric method can be understood by the following 'local average' idea (see Eubank, 1988; Hardle, 1990). If the mean regression function is believed to be smooth, information about the function at a georferenceable *Ae. egypti* larval habitat, geoclassifiable LULC, capture point, should be contained in the data near the point. Then, a local average of the signature data of the point can be used to construct an estimator of the function at the point.In the context of kernel estimation, a researcher may investigate two main properties of the two-stage yellow fever vulnerability, diagnostic, sub-meter resolution LULC polgonized parameter estimator in ArcGIS: uniform rate of strong convergence and pointwise asymptotic normality of the *Ae. egypti* geo-spectrotemporally geosampled, seasonal covariates. In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable(Hosmer and Lemeshew 2002) Since both the first and the second stages in the algorithm would be nonparametric, one can easily guess that the convergence rate (either uniform or pointwise) would be slower especially when the n is the sample size of 10,000 YFV clinical, field and remote datatset variables, and are dependent on the dimensions of the forecasted arguments.Jacob et al. (2015) determined that, if kernels and bandwidth sequences for both stages of a black fly *S. damnosoum* s.l., vector of onchcercisis are chosen appropriately, the pointwise convergence rate for asymptotic normality would be equal to the minimum of the convergence rates of the first-stage estimate and of the hypothetical estimate presuming the first-stage regression is known. A lower bound on the uniform convergence rate was obtained by the authors of Jacob et al. (2014) , which was slower than the minimum of the rates of convergence of the first stage and thegeopredicted, targeted hypeproductive trailing vegation, sparsely canopied, immature *S. damnosum* s.l. immature estimates.

Residual, aympoptically normalized, elucidative, diagnostic, regression plots from previously constructed time series, malaria, mosquito, vector, *An. arabiensi,s* aquatic larval habitat, PROC REG models has revealed that errors in variance uncertainty estimation can substantially alter numerical predictions of a model by inflating the value of test statistic thereby, increasing the chance of a Type I error – incorrect rejection of the null hypothesis, H_0: no spatial autocorrelation [Jacob et al. 2009]. Spatial autocorrelation is the correlation among values of a single variable strictly attributable to their relatively close geolocational positions on a two-dimensional surface, introducing a deviation from the independent observations assumption of classical statistics [Griffith 2003]. The complexity of techniques for optimally conducting a yellow fever regression estimation analyses of seasonally,geosampled, geoclassifiable, explicative, LULC assessment ranges from simple correlations between different types to species abundance and distribution of vector mosquito arthropod species (Mwangangi et al. 2007, Kalluri et al. 2007, Patz 2000).

The autoregressive conditional variance is important in mapping vector arthropods*, as it is used in immature habitat prediction and confidence intervals, tests of hypotheses and for estimating, geo-spatiotemporal, geospectral prediction error in a vunerability model framework(See Griffith 2005, Jacob et al. 2005). Nuisance parameters are often variances, but there are exceptions:for example, in an optimizable dataset of errors-in-variables model parameter estimates,the unknown true habitat location of each georeferenceable observation is a nuisance parameter (see Jacob et al. 2009). Stochastic models have been generated with non-linear nuisance parameters for examining the interrelationship between mosquito productivity and oviposition of gravid mosquitoes (Jacob et al. 2010, Jacob et al. 2013). By designing a model that explicitly features non-stationary behavior of regressed, YFV, mosquito vector, LULC, aquatic, immature habitat data, a hierarchy of conditional variance components may be linked by applying Bayes theorem in PROC MCMC.

PROC MCMC uses a random walk Metropolis algorithm to obtain posterior samples. The algorithm is simple but practical, and it can be used to obtain random samples from any arbitrarily complicated target distribution (e.g., prolific, georeferenced, *Ae. egypti,* capture point immature habitat) of any dimension that is known up to a normalizing constant. The Bayesian procedures use a special case of the Metropolis algorithm called the Gibbs sampler to obtain posterior samplers (Cressie 1993).

The Gibbs sampler, is a special case of the Metropolis sampler in which the proposal distributions exactly match the posterior conditional distributions and proposals are accepted 100% of the time(Geman and Geman 1984. Gelfand et al. 1990). Gibbs sampling would require decomposing the joint posterior distribution into full conditional distributions for each geo-spectrotemporal uncoalesced, sylvatic, YFV-related, elucidative or non-elucidative, explanatorily diagnostic, normalized, clinical, field or remote geosampled, parameter estiamtors in the forecasting vulnerability model and then sample from them. The sampler may be efficient when sub-meter resolution LULC wavelength, sub-meter resolution, emissivity parameters are not highly dependent on each other and the full conditional distributions are easy to sample from. Unfortunately while deriving the conditional distributions in a regression YFV model it may be

relatively easy to find an efficient way to sample from quantitable, frequency conditional distributions.

Suppose $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ is the parameter vector, $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood, and $\pi(\boldsymbol{\theta})$ is the prior distribution in a robustifiable, forecasting, sylvatic, YFV-related, vulnerability , Bayesain regression. The full posterior conditional distribution of $\pi(\theta_i|\theta_j, i \neq j, \mathbf{y})$ would then be proportional to the joint posterior density; that is, $\pi(\theta_i|\theta_j, i \neq j, \mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ For example, the one-dimensional conditional distribution of θ_1 given $\theta_j = \theta_j^*, 2 \leq j \leq k$, may be then computed as the following in PROC MCMC $\pi(\theta_1|\theta_j = \theta_j^*, 2 \leq j \leq k, \mathbf{y}) = p(\mathbf{y}|(\boldsymbol{\theta} = (\theta_1, \theta_2^*, \ldots, \theta_k^*)')\pi(\boldsymbol{\theta} = (\theta_1, \theta_2^*, \ldots, \theta_k^*)')$ The Gibbs sampler in the YFV, forecasting vulnerability model would optimally works as follows: 1) Set $t = 0$, and choose an arbitrary initial value of $\boldsymbol{\theta}^{(0)} = \{\theta_1^{(0)}, \ldots, \theta_k^{(0)}\}$,(e.g., yellow fever prevalance) 2)Generate each geosampled forest-canopy, geoclassified, seasonal LULC component of $\boldsymbol{\theta}$:3) Draw $\theta_1^{(t+1)}$ from $\pi(\theta_1|\theta_2^{(t)}, \ldots, \theta_k^{(t)}, \mathbf{y})$,4) draw $\theta_2^{(t+1)}$ from $\pi(\theta_2|\theta_1^{(t+1)}, \theta_3^{(t)}, \ldots, \theta_k^{(t)}, \mathbf{y})$ ...4) Draw $\theta_k^{(t+1)}$ from $\pi(\theta_k|\theta_1^{(t+1)}, \ldots, \theta_{k-1}^{(t+1)}, \mathbf{y})$ and finally 5) Set $t = t + 1$. If $t < T$, the number of desired seasonal, forecasted YFV-related samples would be non-specified.

Suppose an expeimneter wants to obtain T samples from a univariate, YFV Bayesianized regression distribution with probability density function $f(\theta|\mathbf{y})$. Suppose θ^t is the tth sample from f. To employ the Metropolis algorithm for qualitatively quantating geospectrotemporally uncoalesced sub-meter resolution, diagnostic, *Ae egypti*-related clinical, field or remote geosampled parameterizable covariate coefficients an order must be prescribed in order have an initial value θ^0 and a symmetric proposal density $q(\theta^{t+1}|\theta^t)$ in the risk model. For the $(t+1)$th iteration, the algorithm would optimally generate a sample from $q(\cdot|\cdot)$ based on the quantized sample θ^t, and would make a decision to either accept or reject the new sample. If the new YFV sample is accepted, the algorithm would repeat itself by starting at the new sample. If the new sample is rejected, the algorithm would start at the current point and then repeat in geospace. Since the algorithm is self-repeating, it can be carried out as long as required in order to facilitate the LULC radiance frequency transmittance endmember quantitation. In practice,a reseracher must decide the total number of YFV-related diagnostic clinical, field or remote samples needed in advance and stop the sampler after that many iterations has been completed.

Suppose $q(\theta_{new}|\theta^t)$ is a symmetric distribution in a exploratory diagnostic clinical, field or remote geo-spectrotemrpoally geosampled, YFV-related, sylvatic, *Ae. egypti* forecast, vulnerability model. The proposal distribution should be an easy distribution from which to sample, and it must be such that $q(\theta_{new}|\theta^t) = q(\theta^t|\theta_{new})$, meaning that the likelihood of jumping to θ_{new} from θ^t would be the same as the likelihood of jumping back to θ^t from θ_{new}. The most common choice of the proposal distribution is the normal distribution $N(\theta^t, \sigma)$ with a fixed σ (www.sas.edu). The Metropolis algorithm for qualitatively quantitating an optimizable dataset of uncoalesced, sub-meter resolution, YFV-related, LULC, wavelength irradiance frequency estimators can be summarized as follows: 1) Set $t = 0$ and choose a starting point θ^0 which an arbitrary capture point (e.g., prolific geosampled *Ae egypti* immature, seasonal habitat) as long

as $f(\theta^0|y) > 0$. 2) Generate a new sample, θ_{new}, by using the proposal distribution $q(\cdot|\theta^t)$. 3) Calculate the following quantity:

$$r = \min\left\{\frac{f(\theta_{new}|y)}{f(\theta^t|y)}, 1\right\}$$

4) Sample u from the uniform distribution $U(0,1)$., 5) Set $\theta^{t+1} = \theta_{new}$ if $u < r$; otherwise set $\theta^{t+1} = \theta^t$: and, 6) Set $t = t + 1$. If $t < T$, the number of desired YFV-related samples would return to step 2.

Note that the number of iteration keeps increasing in a robustifiable, explanative, diagnostic, sylvatic, YFV-related, clinical, field or remote geo-spectrotemporally geosampled, probabilistic, LULC-oriented, forecast, vulnerability model regardless of whether a proposed sample is accepted. This algorithm can metaheuristically optimally define a chain of randomized geo-spectrotemporally geosampled, YFV-related, parameterizable, explanative, seasonal, uncoalesced, sub-meter resolution, *Ae egypti*, LULC-oriented, covariates whose distribution may converge to the desired distribution $f(\theta|y)$, and so from some point forward, the chain of samples is a sample from the distribution of interest. In Markov chain terminology, this distribution is called the stationary distribution of the chain, and in Bayesian statistics, it is the posterior distribution of the model parameters. The random-walk Metropolis algorithm is used in PROC MCMC.

Fortunatley, a yellow fever regresssion probabilsic paradigm would not be limited to a symmetric random-walk proposal distribution for establishing a valid sampling algorithm. A more general form, the Metropolis-Hastings (MH) algorithm, was proposed by Hastings (1970). The MH algorithm uses an asymmetric proposal distribution: $q(\theta_{new}|\theta^t) \neq q(\theta^t|\theta_{new})$. The difference in its implementation comes in calculating the ratio of densities:

$$r = \min\left\{\frac{f(\theta_{new}|y)q(\theta^t|\theta_{new})}{f(\theta^t|y)q(\theta_{new}|\theta^t)}, 1\right\}$$

The extension of the Metropolis algorithm to a higher-dimensional θ is straightforward. Suppose $\theta = (\theta_1, \theta_2, \cdots, \theta_k)^t$ is a geosampled, clincially diagnostic, sylvatic, YFV-related parameterizble covariate of intetrest. To start the Metropolis algorithm, an experimenter would select an initial value for each θ_k and use a multivariate version of proposal distribution $q(\cdot|\cdot)$, such as a multivariate normal distribution, to select a k-dimensional new parameter. Other steps would remain the same as those previously described, and this Markov chain eventually would converge to the target distribution of $f(\theta|y)$ (forecasted, georeferenced, seasonally hyperproductive, sylvatic, *Ae egypti*, immature habitat, capture point)

Another type of Metropolis algorithm is the "independence" sampler. It is called the independence sampler because the proposal distribution in the algorithm does not depend on the current point as it does with the random-walk Metropolis algorithm. For this sampler to work well in a YFV-related, forecast, vulnerability, wavelength model, the experimeter would have to have a proposal distribution that mimics the target distribution (hypeprodutcive, *Ae egypti,* oviposition site) and have the acceptance rate be as high as possible. In order to do so, the experimenter would have to set $t = 0$. Initially there would be a requirement to choose a starting point θ^0. This can be an arbitrary point as long as $f(\theta^0|y) > 0$. Thereafter a new sample, θ_{new}, may be genereated by using the proposal distribution $q(\cdot)$. The proposal distribution does not depend on the current value of θ^t (Cressie 1993). Then b calculating the following quantity:

$r = \min\left\{ \dfrac{f(\theta_{new}|\mathbf{y})/q(\theta_{new})}{f(\theta^t|\mathbf{y})/q(\theta^t)}, 1 \right\}$, the sample u from the uniform YFV distribution would be precisely $U(0,1)$. By setting $\theta^{t+1} = \theta_{new}$ if $u < r$; or otherwise set $\theta^{t+1} = \theta^t$. All the geo-spectrotemporally geosampled, georeferenced, YFV-related, explanative, diagnostic, clinical, field or remote geosampled, time series data would then be efficiently quantized. Finally a set $t = t + 1$ would be robustly rendered. If $t < T$, is rendered this discretre inter value would be the number of desired samples in the *Ae. egypti* stcochastic or deterministic, uncoalesced, sub-meter resolution, LULC biosignature, iterative interpolator.

An ideal proposal density in a geo-spectrotemporal, geospatialized, sylvatic, YFV-related, forecasting, vulnerability model should have thicker tails than those of the target distribution. This requirement sometimes can be difficult to satisfy especially in cases where you do not know what the target posterior distributions are like In addition, this sampler does not produce independent samples as the name seems to imply, and sample chains from independence samplers can get stuck in the tails of the posterior distribution if the proposal distribution is not chosen carefully. The independence sampler is used in PROC MCMC(www.sas.edu).

For the actual implementation details of the Metropolis algorithm in PROC MCMC for constructing a robust, YFV, forecasting regression, *Ae egypti*, LULC sub-meter resolution,vulnerability model, the blocking of the parameters and tuning of the covariance matrices must be highlighted. One key factor in achieving high efficiency of a Metropolis-based Markov chain in a YFV simulation model may be finding a good proposal distribution for each block of parameters (i.e.,tuning). The tuning phase consists of a number of loops.

In SAS, the minimum number of loops is controlled by the option MINTUNE=, with a default value of 2 (www.sas.edu). The option MAXTUNE= controls the maximum number of tuning loops, with a default value of 24. Each loop lasts for NTU= iterations, where by default NTU= 500. At the end of every loop, PROC MCMC would examine the acceptance probabilities of the regressed, sylvatic, YFV-related *Ae.aegypti*, geoclassified, LULC explanators for each block. The acceptance probability is the percentage of NTU= proposals that have been accepted(www.sas.eud). If the probability falls within the acceptance tolerance range (the current configuration of c/Σ or p for the forecasting vulnerability model would be kept. Otherwise, the geo-spectrotemporally geosampled landscape parameters are modified before the next tuning loop.

A good proposal distribution should resemble the actual posterior distribution of the YFV-related georeferenceable, *Ae.aegypti*, geoclassified, LULC parameters. Large sample theory states that the posterior distribution of the parameters approaches a multivariate normal distribution (see Gelman et al.; 2004, and Schervish; 1995,). That is why a normal proposal distribution would work well in practice in a forecast, YFV-related vulnerability model. The default proposal distribution in PROC MCMC is the normal distribution: $q_j(\theta_{new}|\theta^t) = MVN(\theta_{new}|\theta^t, c^2\Sigma)$. In addition, Roberts and Rosenthal (2001) empirically demonstrated that an acceptance rate between 0.15 and 0.5 is at least 80% efficient, so there is really no need to fine-tune the algorithms to reach acceptance probability that is within small tolerance of the optimal YFV-related, geoclassified, LULC paramterized covariate coefficient

values (e.g, targeted forest-canopy capture point, immature, hypeproductive, georefenced *Ae.aegypti* oviposition sites). PROC MCMC works with a probability range, determined by the PROC options TARGACCEPT ±ACCEPTTOL (www.sas.edu).

The default value of TARGACCEPT would then be a function of the number of sylvatic, *Ae.egypti*, geoclassified, LULC sub-mter resolution explanators in the YFV model. The default value of ACCEPTTOL is 0.075 (www.sas.edu) If the observed acceptance rate in a given tuning loop is less than the lower bound of the range, the scale is reduceable; if the observed acceptance rate is greater than the upper bound of the range, the scale is increased (Gelmans 1996). During the tuning phase, a scale parameter in the normal distribution is adjusted as a function of the observed acceptance rate and the target acceptance rate(www.sas.edu). The following updating scheme imay be used in PROC MCMC for constructing a robust, stochastic, yellow fever, forecast eco-epidemiological, diagnostic, clinical, field or remote geosampled, georefenced, risk model [1]:

$$c_{new} = \frac{c_{cur} \cdot \Phi^{-1}(p_{opt}/2)}{\Phi^{-1}(p_{cur}/2)}$$

for optimally targeting prolific, seasonal, *Ae egeypti* habitats where c_{cur} is the current scale, p_{cur} is the current acceptance rate, p_{opt} is the optimal acceptance probability. Alternative, an experimeneter may choose a multivariate *t*-distribution as the proposal distribution for quantitating a dataset of geo-spectrotemporally optimizable, YFV-related *Ae.aegypti*, geoclassified, LULC, sub-meter resolution, geosampled, explanators It may be a good distribution to use if the posterior distribution has thick tails and a *t*-distribution which may improve the mixing of the Markov chain during burn in periods

By default, PROC MCMC would assume that all the geospatialized YFV-related explanatorial observations in the dataset are independent, and as such the logarithm of the posterior density could be calculatable as follows:

$$\log(p(\theta|y)) = \log(\pi(\theta)) + \sum_{i=1}^{n} \log(f(y_i|\theta))$$

where θ is a parameter or a vector of parameters. The term $\log(\pi(\theta))$ is the sum of the log of the prior densities specified in the PRIOR and HYPERPRIOR statements(www.sas.edu). The term $\log(f(y_i|\theta))$ would then be the log likelihood specified in the MODEL statement. The MODEL statement specifies the log likelihood for a single observation in the dataset(www.sas.edu).

The statements in PROC MCMC are in many ways like DATA step statements; PROC MCMC evaluates every statement in order for each observation. The procedure cumulatively would add the log likelihood for each diagnostic, YFV-related, clinical, field or remote geosampled, sylvatic, *Aea.egypti*, geoclassified, LULC observation. Statements between the BEGINNODATA and ENDNODATA statements are evaluated only at the first and the last observation (www.sas.edu). At the last sylvatic, YFV observation, the log of the prior and hyperprior distributions may be added to the sum of the log likelihood to obtain the log of the posterior distribution.

With multiple PARMS statements (e.g., multiple blocks of YFV parameters), PROC MCMC updates each block of explanatorial, LULC sub-meter resolution, uncoalesced, geo-spectrotemporalized, geospatial parameters while holding the others constants. The procedure still steps through all of the programming statements to calculate the log of the posterior

distribution, given the current or the proposed wavelength, frequency values of the updating block of landscape parameters. In other words, the procedure would not calculate the conditional distribution explicitly for each block of YFV-related *Ae.egypti*, geoclassified, LULC explanators. Instead the Bayesain paradigm would use the full joint distribution in the Metropolis step for every block update. If a seasonal YFV-related model dependent data—that is, $\log(f(\mathbf{y}|\boldsymbol{\theta})) \neq \sum_i \log(f(y_i|\boldsymbol{\theta}))$— the PROC option JOINTMODEL may be appropriate

PROC MCMC will assume that the input geo-spectrotemporally geosampled YFV-related georeferenceable explicative *Ae.egypti*, geoclassified, LULC explanators are independent then the joint log likelihood would be the sum of individual log-likelihood functions. By specifying the log likelihood of one YFV-related observation in the MODEL statement. PROC MCMC would be then able to evaluate that function for each successive observation in the geospatialized explanatorial dataset and cumulatively sum them up. If the YFV-related observations are not explicity independent of each other, this summation would produce the incorrect log likelihood.

There are two ways to analyzed geo-spectrotemporally uncoalesced, YFV-related , *Ae.egypti*, geoclassified, LULC, signature model dependent data. An experimenter can either use the DATA step LAG function or use the PROC option JOINTMODEL. The LAG function returns values of a variable from a queue (www.sas.edu). As PROC MCMC steps through a YFV—related geo-spectroterporally uncoalesced dataset, the LAG function would queues each dataset eco-epidemiological, irradiance variable. If the log likelihood for a diagnostic, clinical, field or remote geosampled georcferenceable observation x_i depends only on observations 1 to i in the dataset, an experimenter can use this SAS function to construct the log-likelihood function for each YFV observation. Note that the LAG function enable access observations from different rows, but the log-likelihood function in the MODEL statement must be generic enough that it applies to all the geosampled YFV observations.

A second option is to create arrays, store all relevant variables in the arrays, and construct the joint log likelihood for the entire dataset instead of for each geo-spectrotemporally geosampled, YFV-related, sylvatic, *Ae.egypti*, geoclassified, LULC observation. Following is a simple example that illustrates the usage of this option.

```
/* allocate the YFV-related Ae.egypti, geoclassified, LULC explanators sample size. */
data exi;
  call streaminit(17);
  do ind = 1 to 100;
    y = rand("normal", 2.3, 1);
    output;
  end;
run;
```

The log-likelihood function for each YFV-related georeferenceable, explicative, time series signature observation would then be quantited in the forecasting vulnerability, geo-spectrotemrpoal, geospatial model as follows: $\log(f(y_i|\mu,\sigma)) = \log(\phi(y_i;\mu,\text{var}=\sigma^2))$ The joint

log-likelihood function would be as follows: $\log(f(\mathbf{y}|\mu,\sigma)) = \sum_i \log(\phi(y_i; \mu, \text{var} = \sigma^2))$ The following statements may fit a simple model with an unknown mean (*mu*) in PROC MCMC, with the variance in the likelihood assumed known. The MODEL statement may indicate a normal likelihood for each YFV-related *Ae.egypti*, geoclassified, LULC observation *y*.

```
proc mcmc data=exi seed=7 outpost=p1;
   parm mu;
   prior mu ~ normal(0, sd=10);
   model y ~ normal(mu, sd=1);
run;
```

The following statements specify the log-likelihood function for an entire dataset of geo-spectrotemporally geospatialized, metaheursitically optmizable ,YFV-related diagnostic, clinical, field or remote–geosampled ,*Ae.egypti*, geoclassified, LULC explanatorial, covariate coefficiuents values.

```
data a;
run;

proc mcmc data=a seed=7 outpost=p2 jointmodel;
   array data[1] / nosymbols;
   begincnst;
     rc = read_array("exi", data, "y");
     n = dim(data, 1);
   endcnst;

   parm mu;
   prior mu ~ normal(0, sd=10);
   ll = 0;
   do i = 1 to n;
     ll = ll + lpdfnorm(data[i], mu, 1);
   end;
   model general(ll);
run;
```

The JOINTMODEL option can indicate that the function used in the MODEL statement calculated for the log likelihood for the entire geo-spectrotemporally geosampled YFV dataset, rather than just for one observation. Given this option, the procedure would no longer step through the input data during the simulation. Consequently, an experimenter would not be able use any YFV-related dataset explanatorial, georefernceable, *Ae.egypti*, geoclassifiable, LULC explanators to construct the log-likelihood function. Instead, an experimenter can store the geosampled geo-spectrotemporalized, uncoalesced wavelength, sub-meter resolutionm signature dataset in arrays . In so doing, arrays instead of dataset variables would be employeable to calculate the log likelihood in the forecasting vulnerability paradigm.

The ARRAY statement allocates a temporary array(www.sas.edu) The READ_ARRAY function selects the *y* variable from the *exi* data set and stores it in the *data* array. In the programming statements, an experimenter may use a DO loop to construct the joint log likelihood in a geo-spectrotemporal, YFV-related, geoclassified, *Ae.egypti*, LULC sub-meter resolution, explanative, forecast, vulnerability model. In so doing, the expression in the YFV model in the GENERAL function would takes the value of the joint log-likelihood for all the geosampled geo-spectrotemporal, uncoalesced, probabilistically regressed, iteratively interpolative, fractionalized, iterative, interpolative, endmember, biosignature data.

To work with a new density that is not listed in the section Standard Distributions, a researcher can use the GENERAL and DGENERAL functions. The letter "D" stands for discrete (www.sas.edu). The new, diagnostic, clinical, field and remote geosampled, YFV-related *Ae.egypti*, geoclassified, LULC, normalized distributions would have to be specified on the logarithm scale. Suppose that a researcher wants to use the inverse-beta distribution:

$$p(\alpha|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)+\Gamma(b)} \cdot \alpha^{(a-1)} \cdot (1+\alpha)^{-(a+b)}$$

The following statements in PROC MCMC could then define the density of the YFV-related, sylvatic, *Ae.egypti*, LULC geoclassifiable, explanators on its log scale as:

    a = 3; b = 5;
    const = lgamma(a + b) - lgamma(a) - lgamma(b);
    lp = const + (a - 1) * log(alpha) - (a + b) * log(1 + alpha);
    prior alpha ~ general(lp);

The symbol *lp* is the expression for the log of an inverse-beta (a = 3, b = 5). The function general(lp) will assigns the YFV-related forecasted distribution to *alpha*. Note that the constant term, *const*, can be omitted as the Markov simulation requires only the log of the density kernel.

Let $(x_1, x_2, \ldots, x_n)$ be an independent and identically distributed sample drawn from some distribution with an unknown density f. If an experimenter is interested in estimating the shape of this function f. Its kernel density estimator is $\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$, where $K(\bullet)$ is the kernel — a non-negative function that integrates to one and has mean zero — and $h > 0$ is a smoothing parameter called the bandwidth. A kernel with subscript h is called the scaled kernel and defined as $K_h(x) = 1/h\ K(x/h)$ (Griffith 2003). Intuitively, one wants to choose h as small as the data allow, however there is always a trade-off between the bias of the estimator and its variance; more on the choice of bandwidth below.

A range of kernel functions are commonly used: uniform, triangular, biweight, triweight, Epanechnikov, normal, and others. The Epanechnikov kernel is optimal in a mean square error sense (Cressie 1993) though the loss of efficiency is small for the kernels listed previously,[4] and due to its convenient mathematical properties, the normal kernel is often used, which means $K(x) = \phi(x)$, where ϕ is the standard normal density function.

The construction of a kernel density estimate finds interpretations in fields outside of density estimation (Rosenblatt,1956). For example, in thermodynamics, this is equivalent to the

amount of heat generated when heat kernels (the fundamental solution to the heat equation) are placed at each data point locations x_i. Similar methods are used to construct discrete Laplace operators on point clouds for manifold learning In mathematics, the discrete Laplace operator is an analog of the continuous Laplace operator, defined so that it has meaning on a graph or a discrete grid. For the case of a finite-dimensional graph (having a finite number of edges and vertices), the discrete Laplace operator is more commonly called the Laplacian matrix.

When you use the GENERAL function in the MODEL statement, specify the dependent variable (e.g., YFV-related explanatorial, dataset ofgeo-spectrotemporal, georefernceable, *Ae.egypti*, geoclassifiable, LULC forest-canopy) on the left of the regression equation$(\sim)$. The log-likelihood function would take the YFV-related explanatorial, dependent variable into account; hence there is no need to explicitly state the dependent variable in the MODEL statement. However, in the PRIOR statements, a researcher may need to explicitly state the diagnostic, clinical, field or remote geosampled LULC sub-meter resolution defined parameter names with the GENERAL and DGENERAL functions.

A researcher can specify any YFV-related, optimizable, explanatorial, dataset of georefernceable, sylvatic, *Ae.egypti*-related, geoclassifiable, LULC explanative, normalized, distribution function by using the GENERAL and DGENERAL functions as long as they are programmable with SAS statements. When the function is used in the PRIOR statements, the reseacher must supply initial values. This can be done in either the PARMS statement (PARMS Statement) or within the BEGINCNST and ENDCNST statements (e.g., BEGINCNST/ENDCNST Statement). It is important to remember that PROC MCMC does not verify that the GENERAL function you specify is a valid distribution—that is, an integrable density. You must use the function with caution.

Given a time series of geo-spectrotemporally geosampled, uncoalesced, YFV-related, explanatorial, dataset of georefernceable, *Ae.egypti*, geoclassifiable, LULC, sub-meter resolution, data X_t where t is an integer index and the X_t are geosampled covariate coefficient values, then an ARMA(p',q) model may be given by: $\left(1-\sum_{i=1}^{p'}\alpha_i L^i\right) X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$ where L is the lag operator, the α_i are the parameters of the autoregressive part of the model, the θ_i are the parameters of the moving average part and the ε_t are error terms. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.Assume now that the polynomial $\left(1-\sum_{i=1}^{p'}\alpha_i L^i\right)$ has a unitary root of multiplicity d. Then the YFV-related,time series, weighted, equation can be rewritten as: $\left(1-\sum_{i=1}^{p'}\alpha_i L^i\right) = \left(1-\sum_{i=1}^{p'-d}\phi_i L^i\right)(1-L)^d$. in PROC ARIMA. An ARIMA(p,d,q) process would express this polynomial factorization property with $p=p'-d$, which would be given by: $\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$. Thus this YFV-related, probabilistic, forecasting vulnerabil;ity model construction can be thought as a particular case of an ARMA($p+d,q$)

process having the autoregressive polynomial with *d* unit roots. (For this reason, every ARIMA model with *d*>0 is not wide sense stationary.)The model can be generalized as

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d X_t = \delta + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t$$

follows. In so doing, the residual forecast would define an ARIMA(*p*,*d*,*q*) process with drift $\delta/(1-\Sigma\varphi_i)$.

The explicit identification of the factorization of the autoregression geoclassifiable, YFV-related dataset of georefernceable, *Ae.egypti*, LULC oriented, explanatorial polynomial into factors as mentioned previously can be extended to other cases, by firstly applying the moving average polynomial and secondly including other special factors. For example, having a factor $(1 - L^s)$ in a yelow fever model is one way of including a non-stationary seasonality of period *s* into the model; this factor has the effect of re-expressing the geos-epctrotemporally uncoalesced, sub-meter resolution, log-transmformed, wavelength, frequency, transmittance data as changes from *s* periods ago. Another example is the factor $(1 - \sqrt{3}L + L^2)$, which includes a (non-stationary) seasonality of period 2. The effect of the first type of factor would allow each season's geoclassifiable LULC uncoalesced explanative, wavelength value to drift separately over time, whereas with the second type values for adjacent seasons move together.

Identification and specification of appropriate factors in an ARIMA model can be an important step in a geo-spectrotemporal, YFV-related dataset of explanatorial, georefernceable, *Ae.egypti*, geoclassifiable, uncoalesced LULC model estiamators as it can allow a reduction in the overall number of parameters to be estimated, while allowing the imposition on the model of types of behaviour that logic and experience suggest should be there.The ARIMA model can be viewed as a "cascade" of two models. The first is non-stationary:$Y_t = (1 - L)^d X_t$while the second is wide-sense stationary:$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)Y_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t .$ (Griffith 003) As such, geo-spectrotemporally uncoalesced, sub-mteer resolution, YFV-related explanatorial, geospatialized, wavelength, frequency datasets of uncoalesced iteratively interpolative, georefernceable, *Ae.egypti*, geoclassifiable, diagnostic, clinical and LULC explanators forecasts can be made for the process Y_t, using a generalization of the method of autoregressive forecasting.

Some well-known special cases can arise naturally in a geo-spectrotemrpoal *Ae.egypti* LULC ARIMA-related YFV model. For example:An ARIMA(0,1,0) model (or I(1) model) is given by $X_t = X_{t-1} + \varepsilon_t$— which is simply a random walk. A random walk is a mathematical formalization of a path that consists of a succession of random steps (Cressie 1993).An ARIMA(0,1,0) forecast vulnerability YFV-related paradigm may be generated with a constant, given by $X_t = c + X_{t-1} + \varepsilon_t$— which is a random walk with drift. If the random walk model predicts that the value at time "t" will equal the last period's value plus a constant, or drift (α), and a white noise term (ε_t), then the process is random walk with a drift. The resdiual forecast would not revert to a long-run mean and has variance dependent on time. An geospectrotemrpoal, geosampled, ARIMA(0,2,2) model may be given by $X_t = X_{t-1} + X_{t-2} + (\alpha + \beta - 2)\varepsilon_{t-1} + (1 - \alpha)\varepsilon_{t-2} + \varepsilon_t$— in PROC ARIMA which is equivalent to Holt's linear method with additive errors. Holt (1957) extended simple exponential

smoothing to allow forecasting of data with a trend which involves a forecast equation and two smoothing equations (one for the level and one for the trend).

Although the exponential smoothing methods have been around since the 1950s, there has not been a well-developed vector medical entomological modelling framework for incorporating stochastic immature habitat targeting sub-meter resolution LULC models, likelihood calculation, prediction intervals and procedures for model selection. Some important steps toward this framework for other disciplines were established by Gardner (1985), and Ord, Koehler, and Snyder (1997). Earlier work in establishing prediction intervals for exponential smoothing methods appeared in Chatfield and Yar (1991), Ord et al. (1997) and Koehler, Snyder, and Ord (2001).

The work of Brown (1959) and Gardner (1985) led to the use of exponential smoothing in automatic forecasting (e.g., Stellwagen & Goodrich, 1999). Makridakis, Wheelwright, and Hyndman (1998) advocate the methods in the taxonomy proposed by Pegels (1969) and extended by Gardner (1985)However, for developing a more general class of methods with a uniform approach to calculation of prediction intervals, maximum likelihood estimation and the exact calculation of a YFV-related, optimizable, explanatorial, dataset of georefernceable, *Ae.egypti*-related, geoclassifiable, LULC explanative, normalized, distribution model selection criteria such as Akaike's Information Criterion AIC).

. To determine the order of a non-seasonal ARIMA model, is (AIC) which is written as:$AIC= -2\log(L)+2(p+q+k+1)$, where L is the likelihood of the data, $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$, p is the order of the autoregressive part and q is the order of the moving average part.The corrected AIC for ARIMA,geo-spectrotemporal, geospatial, YFV-related datasets of georefernceable, explanatorial, *Ae.egypti*, geoclassifiable, sub-meter resolution, uncoalesced, iteratively interpolative, LULC explanatorial models can be written as:$AICc= AIC+ (2(p+q+k+1)(p+q+k+2))/(T-p-q-k-2)$The Bayesian Information Criterion (BIC) can be written as:$BIC= AIC+(\log(T)-2)(p+q+k+1)$(Gelman 1996). The objective is to minimize the AIC, AICc or BIC values for a good model

Commonly, having obtained the joint conditional distribution of all of the unknown, random, explanatorial, geo-spectrotemporal, uncoalesced, geo-predictive, YFV-related, explicatively diagnostic clinical and geoclassifiable LULC explanators given an optimizable datset of known geosampled, larval habitat, georeferenceable, parametrizable covariates, by applying Bayes theorem, nuisance variables may be marginalized to obtain the conditional distribution for determining ecological parameters associated with georeferenceable, prolific, *Ae. aegypti,* geospatialized, sub-meter resolution, LULC, aquatic, habitat data. However, even though this generalized treatment of the conditional variance can generate an autoregressive error model, the residual estimates may not be able to spatially target prolific *Ae. egypti* aquatic habitats based on larval/pupal productivity. Treatments of vector mosquito auatic habitat perturbations should be based on surveillance of larvae in the most productive areas of an ecosystem [Gu and Novak 2005].

Accurate diagnostic, autoregressive forecasts of *Ae egypti* aquatic habitat geolocations requires an absolute relative error estimator to identify prolific habitats for developing habitat-

based intervention nmodels for implementing Integrated Vector Management(IVM). Integrated Vector Management is a decision-making process for the management of vector populations, so as to reduce or interrupt transmission of vector-borne diseases (http://www.ivmproject.net); the strategy is based on the premise that effective control is not the sole preserve of the health sector but of various public and private agencies, including communities. Salient attributes of IVM including methods based on knowledge of co-ecofactors influencing local vector biology. For example, hyperendemic, yellow fever transmission and morbidity statistics analyzed in an ArcGIS cyberenvironment may use a range of interventions, (e.g., insectide treatment of *Ae. egypti,* immature, capture point, prolific habitat) often in combination and synergistically for combating uncertainty plagued, (e.g., heteroskedastic independent variables), probabilistic, seasonal, explicative, LULC-oriented, regression variables One of the key assumptions of regression is that the variance of the errors is constant (i.e., homoscedastic) observations. (Hosmer and Lemeshew 2002). Standard estimation methods are inefficient when the errors are heteroscedastic or have non-constant variance (Rao 1972). Typically, vector, seasonal, arthropod-related, diagnostic regression, residuals are plotted to assess assumptions. In so doing, precise collaboration model scenarios employing within health sector, explanative geosampled variables and with other public and private sector covariates and their impact on dependent variables (e.g., monthly prevalance yellow ferver statistics) representing vector data may be efficiently mapped. Also an experimenter can create robust, eco-epidemiological, forecast-oriented, vulnerability, canopy-forested, *Ae. egpti,* LULCs, to quantiate georefernceable, illuminative, prolific, riceland complex,geo-classified LULC protusions that may reveal relationships between expanding riceland irrigated, agro-ecosystem polygons and immature productivity in ArcGIS. Linking tabular data in SAS with geospatial, immature habitat, canopy forested, decomposeable, *Ae. egypti,* LULC data in ArcGIS queries may also aid in quantitating geospatial proximities of the LULCs which then may be log-transformed regressors as part of the risk analysis. Driving forces in an IVM for yellow fever eradication could include the need to overcome challenges experienced with conventional single-intervention approaches to targeted vector control in an ArcGIS cyberenvironments and in other software packages (C ++) for promoting multi-sectorial approaches to human health.

IVM pilot studies, especially for the control of malaria have been successful in sub-Saharan Africa. Castro and colleagues (2004) described a successful intervention in Dar es Salaam, Tanzania, from the late 1980's through the 1990's, with elements of IVM included in the strategy. Gilroy et al. (1945) reported decreases in malaria incidence and sporozoite prevalence rates in Nigeria using a variety of environmental management techniques, including source reduction and drainage. Keiser et al. (2005) conducted a meta-analysis on EM studies globally and concluded that EM can have a significant impact on clinical malaria in Kenya, if EM is appropriate to the eco-epidemiological setting. EM was also used with available larvicides to control malaria in the coastal city of Mombasa. Schliessmann et al. (1973) employed a combination of water drainage techniques and larviciding to reduce the number of malaria cases by 98% from 1969 to 1970 in a coastal flood plain of Haiti. In India, Sharma et al. (1986) reported over 95% reduction in malaria incidence over a four-year period for communities receiving a combination of water-source reduction activities and biological control in larval habitats. Collectively, these studies suggest that multiple vector control strategies (i.e., IVM) may be beneficial for yellow fever eradication.

Importantly, an ArcGIS cyberenvironment IVM approach for YFV vector control may take into account available health infrastructures and resources in the surrounding eco-epidemiological, riceland, agro-ecosystem study sites and integrate all available and effective measures, whether chemical, or biological. In so doing, IVM for yellow fever control for a riceland, agroecosystem, village–complex, eco-geographical, eco-hydrological, orthogonlized, gridded, time series, geoclassified, LULC in an ArcGIS cyberenvironment would effectively model coordination of the control activities of all LULC sectors (e.g., bordering forest canopied LULC) that would have an impact on the vector borne disease, including health, water, solid waste management, housing and agriculture.

For remote identification of vector, medical entomological, georeferenced datasets of mosquito larval habitats the first step is often to construct a discrete tessellation of the region(Jacob et al. 2007). In applications where complex geometries do not need to be represented such as urban habitats, regular orthogonal grids are constructed in ArcGIS and overlaid onto satellite images. However, some vector mosquito aquatic habitats are rarely uniform in space or character (e.g., a sparsely canopied, *Ae. egypti,* tillering riceland seasonal hypeproductive ovipoistion site), An orthogonal grid overlaid on satellite data of Riceland agro-village complex areas may fail to capturephysical or man-made structures, (i.e paddies, canals, berms) at these habitats. Unlike an orthogonal grid, digitizing each habitat converts a polygon into a grid cell, which can conform to rice-land, immature seasonal geosampled habitat boundaries.

Jacob et al. (2006) illustrated the application of a random sampling methodology,comparing an orthogonal and a digitized grid for assessment of riceland habitats in the Mwea rice scheme in Kenya. A land cover map was initially procured in Erdas *Imagine* V8.7® using QuickBird visible and near infra –red (NIR) data acquired July 2005, for three agro-village complexes, Karima, Kururi, and Kangichiri within the Mwea Rice Scheme, Kenya. Information from visible and NIR channels can distinguish between high and low mosquito producing rice fields (Washino and Woods 1994). QuickBird multispectral products provides 4 discrete non-overlapping spectral bands covering a range from 0.45 μm to 0.72 μm. The QuickBird products were delivered as radiometrically corrected image mixed pixels (mixels). The projection used for all of the spatial datasets is the UTM Zone 37S datum WGS-84 projection. After geographic registration alignment of the satellite's Universal Transverse Mercator(UTM) coordinates with known reference objects.

An orthogonal grid was overlaid on the images. In the digitized dataset, each georferenceable habitat was traced in Arc Info 9.1®. All habitats in each study site rice-village complex were stratified based on levels of rice stage. The QuickBird 0.61 m visible and NIR data was classified using the Iterative Self-Organizing DataAnalysis Technique (ISODATA) supervised routine inERDAS *Imagine* V8.7®. A supervised classification cana ssign spectral signatures automatically generated by the ISODATA algorithm (www.esri.com) for remote identification ofmosquito aquatic habitats [Jacob et al. 2011].

The grid cell size in the orthogonal grid was calculated by the mean size of the paddy in ArcInfo 9.1®. The area of the orthogonal grids that were overlaid on the QuickBird visible and NIR data which were 64.3 m × 64.3 m by polyon size.Unique identifiers were placed in each grid cell unit. The grids extended out to a 1 km distance from the external boundary of each study site providing a 1 km radial area. A digitized grid tracing each rice land *An. arabiensis* aquatic habitat was generated in Arc Info 9.1® for the 3 study sites. Unique identifiers were placed in each grid cell . The grid extended out to a 1 km distancefrom the external boundary of each study site.The orthogonal grid and the digitized gridded image were'screened' to determine fit and to provide an indication ofthe geolocation of riceland *An. arabiensis* aquatic habitats..Due to the variation in the number of dips collected per habitat based on the size, the authors based their results on thenumber of larvae collected per dip. The entomological variable was total rice land anopheline larvae.

An ANOVA test was performed to compare the differences in larval abundance between different paddy categories in each study site. An independent sample t-test was used to compare differences in larval abundance between paddies and canals as well as between vegetated and non-vegetated canals. Robust standard errors were used because data were collected at both the orthogonal grid cell and digitized grid cell level. An alpha level of 0.05 was used to indicate significance. All data management and calculationswere performed using SAS 9.1.3® (SAS inc. Carey,NC, USA).

The digitized grid captured each riceland *An. arabiensis* aquatic habitat in the study sites while the orthogonal grid was unable to identify any aquatic habitats in any of the three sites.The digitized grid cell data was used determine abundance of riceland *Anopheles* larvae in the paddy and canal habitats in each study site. The abundance of 1st instar larvae/dip collected in Rurumi and Kangichiri study sites was 0.99 and 1.95, respectively and significantly lower than 4.81 in the Kiuria study site ($F = 5.16$, df 2, 751, $p < 0.01$).Similarly, the abundance of 2nd instar larvae differed significantly among villages with that of 0.66 in the Rurumi study site being significantly lower than 1.09 or 2.11 in the Kangichiri and Kiuria study sites, respectively ($F = 3.79$, df2, 751, $p < 0.05$). The abundance of 3rd and 4th instar larvaeas well as that of pupae did not differ significantly among villages ($F = 1.64$, 0.97 and 1.04, df 2, 75, $p > 0.05$).The abundance of riceland *An. arabiensis* larvae/20 dips collected in the paddy and canal habitats at the 3 study sites. In the Kangichiri study site,the difference in the abundance of pupae and 1st, 2nd and 3rd instar larvae collected in paddy and canal habitats was not significant ($p > 0.05$) while that of 4th instar larvae was significantly higher in the paddy habitats than in the canals ($F = 5.19$, df 1, 179, $p < 0.05$). In the Kiuria studysite, significantly higher abundance of 3rd instar larvae were collected in the canals ($F = 4.68$, df 1, 179, $p < 0.05$) while the other immature stages did not differ significantly between canal and paddy habitats. In the Rurumi study site, paddy habitats had significantly higher abundanceof 1st and 2nd instar larvae compared with the canals($F = 5.60$ and 3.94, df 1, 188, $p < 0.05$) but the otherimmature stages did not vary significantly between paddyand canal habitats.

Alternate wet/dry (intermittent) griided, rice-irrigation geoclassifiable, eco-epidemiological, predictive maps combined with other vector control methods in an ArcGIS cyberenvironment 3-D module may reveal effective methods in controlling prolific, seasonal, georferenceable, larval habitats in interface regions between eco-geographically expanding

riceland agroecoystem, partially and dense canopied, sylvatic, *Ae egypti,* prolific, habitat regions and their surrounding forested canopied LULCs. In so doing, a remotely regressively, quantitative, targeted IVM may be implemented in riceland agroecosystems and a more economic usage of rice-irrigation, ,ecohydrological, ecogeographic-related, geo-spatiotemporal, explanatorial, georefereceable, explanative LULCs may be also determined thereby reducing farmers' costs.

An IVM approach conceived in an ArcGIS cyberenvironment for yellow fever however must be evidence-based. Thus, an essential feature of IVM for yellow fever control in ArcGIS should optimally be based on the ability to generate geo-localized ecobiological, eco-geographical, ecohydrological,YFV-related, orthogonally decomposed, iteratively interpolated geo-spectrotemrpoally geosampled. proxy biosignature, gridded, sub-mter resolution(e.g., panchromatic QuickBird 0,61m data) explanatorial geoclassifiable, LULC data. IVM integrates all available resources to achieve a maximum impact on vector-borne disease (http://www.ivmproject.net/). In so doing, an IVM strategy in an ArcGIS cyberenvironment server may utilize multitemporal, geo-spectrotemporally geosampled, optimal sub-resolution satellite imagery for constructing robustifable, forecasting, seasonal, vulnerabilty, eco-epidemiological, ecogeographical, ecohydrological, probabilsitic, predictive, risk maps for remotely targeting YFV, transmission-related, geoclassifiable LULC, explanatorial, orthogonally decomposed, endmember, wavelength, reflectance, emissivity, transmittance data, for example. Cartographic, vulnearbility, time series, risk descriptions of seasonally georefernceable, explanatorial, geopredictive, geosampled points in the transmission cycle (e.g., positively autocorrelated, explanative, georeferenceable clusters of prolific, *Ae. egypti*, canopy sparsely, forested LULC, immature habitats) at the periphery of a georeferenceable, riceland irrigated, seasonal agroecosystem, geoclassified LULC, may then act synergistically with field level strategies (e.g., indoor residual targeted insecticide spraying) in a cyberenvironment simulatiuon model to improve upon single-tool interventions. Thus, rather than relying upon blanket solutions, vector control managers – together with managers in health, rice agriculture, water resources, and land use can examine the local eco-setting of seasonal, YFV-related, disease, transmission LULC zones(e.g., mesoendmeic, hypoendemic) and seasonal, explanatorial, topographic, sub-meter resolution, reflectance, emissivity, transmittance of decomposed explicative, LULC, patterns employing ArcGIS tactics and thereafter, systematically design and qauis-periodically update strategies (e.g., geo-spectrotemporal, iteratively interpolating, unmixed shade, canopied forestland proxy biosignature of a prolific riceleand, *Ae.egypti,* LULC endmembers based on geosampled, immature, seasonal, count data) and actionizable IVM plans. New scientific knowledge about the bioecology and behavior of seasonal, YFV mosquitoes and their hosts and natural seasonal predators on geoclassified forest canopy and ricland agroecosystem LULCs can improve the precision of vector control methods in general, in an ArcGIS cyberenvironmnent and would permit better precision targeting of prolific, seasonal, immature habitats in a customized IVM, ecogeographical, ecohydrological framework.

In particular, advances in newe,r ArcGIS, geospatial, cartographic, analytical tools can contribute to more precise, forecastable, iteratively interpolative, endmember endemic, wavelength reflectance, emissivity, transmittance, time series dependent, eco-epidemiological, risk mapping of the distribution of vector mosquito species and their ecogeographic, LULC-related, ecohydrological, larval habitats areas, and other disease transmission covariate,

parameterizable explanatorial, coefficient estimator values. For example, ET Surface is a set of tools for ArcGIS that enable users to create surfaces and perform surface analysis without the need of 3D or Spatial Analyst extensions. The only requirement would be an ArcGIS license (ArcView, ArcEditor or ArcInfo). The functionality of ET Surface is available in three different ways for explanatorily geoclassifying, geopredictive, geoclassifiable, eco-epidemiological, LULC, forecast, risk mapping, orthogonally decomposed, canopied, *Ae. egypti* proxy LULC habitat biosignature, unmixed data;1)Via the user friendly wizard type interface 2)Via a set of tools for Arc Toolbox which can be used in the Model Builder, at Command Line or in Python scripts and 3) in NET customizations (www.esri.com). In so doing, explanatorily geo-spectrotemporally, iteratively interpolatable, seasonal ecogeographic, echohydrological, geoclassified, LULC, surface endmembers (e.g., forest, canopied, riceland perpiherial) may be constructed in TIN from points, polylines or polygons of the georeferenced, hyperproductive, seasonal, YFV-related, larval habitats and their geo-spatiotemporally, spectrotemporally geosampled, parameterizable, covariate, estimator, coefficient values. The output can be an ESRI TIN or PolygonZ TIN. For ESRI TIN features can be triangulated as Mass LULC points or Hard breaklines. For PolygonsZ,, TIN georeferenceable, elucidatively, optimally decomposeable, YFV-related, eco-epidemiological, seasonal, data feature attributes may be triangulated as Mass points (e.g., prolific, irrigated riceland, *Ae. egypti* habitats). The TIN features may then be triangulated as Mass points employing Contours to Raster tools, an inverse distance weighted matrix or Kernel Density. In so doing, seasonally bordering, riceland agroecosystem and sparsely shaded, forest canopied LULC, *Ae. egypti* habitat, explanatorily orthogonally decomposed fractionalized endmember, vulnerability, forecastable, risk, mapping ecogeographic, echohydrological variables maybe robustly empirically quantitable employing TIN surface analysis which would identify and quantitatively regressively represent productive estimate all Slope, Aspect, Visibility, Volume, Cut/Fill variables. A Raster Surface Analysis can then diagnosis Slope, Aspect, Hillshade, Viewshed, Volume, Cut/Fill, iteratively interpolatable, seasonal, *Ae. egypti* habitats, Contours Raster Calculator performs complex mathematical operations on rasters employing Raster Distance Analysis, Euclidean Distance, Direction and Allocation, Weighted Voronoi (Thiessen) allocation, Cost Distance and Allocation (www.esri.com). Echohydrological, explanatorial, geo-spatiotemporal, operationizable, ArcGIS functions may then include Flow Direction, Flow Accumulation, Derive Streams, Watershed, Fill depressions, of partially canopied geoclassfiable, LULCs related to prolific, seasonal, *Ae. egypti* habitats. ET TerrainViewer can generate 3-D geo-visualizationa of expanding riceland agroecosystem LULC surfaces into forested canopied LULC employing Polygon Z TINs and rasters as well as shapefiles. Areal photography can be draped on the LULC surfaces as well as Digitized 3-D features or graphics with elevation extracted from ESRI TIN, Raster, PolygonZ TIN may reveal seasonal variation in these habitats.

The analytical tools within an ArcGIS cyberenvironment may be used to guide targeted control efforts, improve cost-effectiveness and minimize unwanted ecosystem disruption or damage (Jacob et al. 2014, Jacob et. al. 2007, Griffith 2005, Wood et al. 1991a, Wood and Washino 1994) for implementing IVM. Technical advances in knowledge about larval control and ArcGIS have made the use of biological insecticides more practical and feasible today than in the past. Remote sensing and GIS are powerful tools for larval surveillance, predicting potential outbreaks and targeting intervention programs (CDC, 2006).

Fortunately, remote sensing techniques for seasonal, YVR-related, time series, eco-epidemiological, forecast, regression-related, probabilistically optimizable, risk mapping and other time series dependent, endemic, transmission-oriented, explanatorial, LULC, riceland agroecosystem and forest canopied, explanatorial, geo-classifiable, parameterizable diagnostic, clinical, field and remote, geosampled covariate, estimator, coefficient estimates has evolved considerably during the past two decades for implementing IVM. Currently, time-series, optimally forecasting, risk analysis and geospatial trend analysis encompass a range of technologies and approaches, including digital mapping vulnerability analysis of remotely sensed sub-meter resolution, waveband imagery, spatial statistics, ecological niche modeling, the use of global positioning systems, and others (Jacob et al. 2010, Jacob et al. 2007 et al. 2000; Griffith 2005, Hay 2000). Common, time series-related, vector, explanatorial, entomological-related, ecogeographic, ecohydrological data analyses include LULC overlay analysis of thematic data and geospatial intersection, buffer generation, neighborhood analysis, vector-based, grid generation, network analysis and geo-morphological, terrain-related, surface and elevational, 3-D, eco-epidemiological, modeling. Singh (1989) and Coppin and Bauer (1996) summarized eleven different LULC change detection algorithms that were found to be documented in the literature. These include: 1. Mono-temporal change delineation. 2. Delta or post-classification comparisons. 3. Multi-dimensional temporal feature geospace analysis. 4. Composite analysis. 5. Image differencing. 6. Multitemporal linear data transformation. 7. Change vector analysis. 8. Image regression. 9. Multitemporal biomass index 10. Background subtraction and 11 Image rationing. As such, currently the complexity of explanatorial, diagnostic, LULC techniques in ArcGIS cyberenvironments range from employing correlations between simple techniques that link satellite-derived, seasonal, field distribution traits (e.g., georeferenced, *Ae. egypti*, larval habitat coordinates overalid onto Google Earth^{TM} maps) to explanatorily, iteratively interpolated, decomposed, sub-meter resolution, fractionalized, endmember, geospectral biosignatures extracted from different ecogeographic, ecohydrological LULC cover types and proxy remotely sensed indicators [normalized vegetation difference index (NDVI)] for species interpolation and abundance mapping in ArcGIS (e.g., see Jacob et al. 2013, Jacob et al. 2012).

ArcGIS layers of environmental information (such as topography, climate, and vegetation), eco-epidemiological and geospatial, explanatorial, time series, risk-related, orthogonal, LULC stratification can also be analyzed for regressively qualitatively, hierarchically, explanatively quantitatable, seasonal, geosampled, ecogeographic, ecohydrological, YVR-related, georeferenced,endemic, transmission-oriented, geolocational, explanatorial, orthogonally parameterizable, covariate, estimator, coefficient values. This approach has been successfully used in the case of Chagas disease and for vectors of leishmaniasis, forexample, (see Peterson et al. 2002, 2004, 2004) filarsis(Jacob et al. 2006), malaria (Jacob et al. 2005, Jacob et al. 2009, Jacob et 2011), Eastern Equine Encephalitis Virus (EEEV) (Jacob et al. 2010b), West Nile Virus (WNV)(Jacob et al. 2010ab) and most recently for onchocerciasis ("river blindness')(Jacob et al. 2012, Jacob et al. 2013). ArcGIS has many digitizing tools that can seasonally cartographically display similar non-linear, explanatorily, empirically, probabilsitically regressed, georefernced, uncoalesced, YFV-related, data, feature attributes geo-classified by themes and strategically archived them in a time series-dependent geodatabase format. These ecogeographic, ecohydrological, seasonal, explicative, time series,

orthogonalized geodata (e.g., seasonal, geosampled, *Ae. egypti,*immature habitat observations) may then be combined with riceland agroecosystem, geoclassified, georeferenced, LULC-related, infrastructure data (e.g., roads, rivers, village boundaries) in a seamless, stand-alone Windows-based application. While a variety of numerical techniques may be employeable to create robust, geo-predictive, eco-epidemiological,YFV-related, riceland agroecosystem, explanatorial, LULC-oriented, seasonal, forecast,vulnerability maps in time and space from sub-resolution satellite data, only those that models that possess statistical operational logistics (e.g., residual autocorrelation risk maps of prolific, seasonal shaded, forest canopied, *Ae. egypti* habitat, ground corrdinates) can aid in understanding the biophysical explanatorial, clustering processes related to *Aedes* and other YFV-related mosquitoes which may provide meaningful eco-epidemiological LULC information for developing and implementing an IVM in a expanding riceleand agro-ecosystem. For instance, since seasonal YFV has three transmission cycles geo-spectrotemporally geospatially and seasonally associated with varying geoclassfiable, georeferenceable, LULCs: jungle (sylvatic), intermediate (savannah) and urban, swaths, optimal, image acquisitions of seasonal, geosampled, explanatorial, YFV-related, LULC reflectance, emissivty, transmittance wavelength, sub-meter resolution, orthogonally decomposable, georeferencable data, feature attributes, may be optimized in as ArcGIS cyberenvironment by constructing Thessian polygons which may be subsequently regressively remotely quantitatively analyzed for ecogeographically representing LULC zones and their boundaries in a riceland agroecosystem, geoclassified LULC. Further, since the jungle (sylvatic) cycle involves transmission of the virus between non-human primates (e.g., monkeys) and mosquito species found in forest-canopied LULCs, remote strategies should be specifically ecogeographically, ecohydrologically, seasonally, optimally determined for these geoclassifiable LULC data in the cyberenvironment.

Attacking multiple, georeferencable, geo-spectrotemporally geosampled, explanatorial, LULC-related, seasonal weighted, explicative regressors in the transmission cycle may act synergistic ally and improve upon current YFV-related interventions by enabling larval habitat source reduction.For example, environmental management of YFV on seasonally geoclassified LULCs can complement insecticide-treated bednets (ITNs) because of a potential secondary mode of action that both control strategies share. In addition to increasing vector mortality, ITNs reduce the rate at which female *Aedes* mosquitoes geolocate human hosts for blood feeding, thereby extending their gonotrophic cycle (Novak 2012).Similarly, while reducing adult vector emergence and abundance, source reduction of larval habitats may prolong the cycle duration by extending delays in geolocating, optimal, seasonally prolific, *Ae. aegypti* oviposition sites. Treatments of vector mosquito auatic habitat perturbations should be based on surveillance of larvae in the most productive areas of an ecosystem(Gu and Novak 2005).

However, source reduction of seasonally prolific, *Ae. aegypti* larval habitats may only operate through a secondary mode of action when larval habitat density is below a critical threshold. Regardless, an ArcGIS seasonal, YFV-related, control program (e.g., targeted insectide treatment) can be effective even when larval habitats are limited. In ArcGIS. an experimenter may also precisely cartographically illustrate, forecasted regression resdiuals for revealing covariates trageting georferenceable, geolocations seasonal hypeproductive *Ae. aegypti* paramterizable covariates whiles parsimonioulsy demonstrating how a, prolific, seasonal,

YFV-related, larval habitat source reduction technique is better suited to human populations in higher urbanizing geo-spectrotemporally geoclassifiable, explanatorial, LULCs (e.g., expanding riceland agroecosystems into geoclassifieable sparsely shaded forest canopy) in the presence of insecticide resistance, or when the insecticidal properties of ITNs are depleted.Importantly, since the *Ae. aegypti* mosquito spreads YFV when it bites non-infected humans, ArcGIS can establish the environmental conditions for an epidemic which may occur on or within a specific vicinity of an eco-geographically explicatively remotely delineateable seasonal, rice agriculture, geoclassifiable LULC (e.g.,post tillering). A biting frequency distribution may be georeferenced in ArcGIS. As such, by travel patterns of infected humans or infected *Ae. aegypti* mosquitoes, an YFV epidemic spread from non-urbanized, explanatorial, seasonal transitioning, LULC habitats into a more urbanized LULC habitat especially in areas of low yellow fever prevalence (e.g., turbid meandering riverine pathways), may be optimally qualitatively regressively quantized and subsequently seasonally mapped in ArcGIS.

Remotely sensed, anthropogenic, explanative, time series, explanatively geoclassifiable, LULC, dependent, geo-spectrotemporally, geosampled, environmental, eco-epidemiological drivers that especially affect YFV- risk (positively autocorrelated cluster of georeferenced riceland *Ae. aegypti,* immature habitats) can include encroachment, site destruction or regressands associated with wildlife, particularly through logging and road building or, changes in the distribution and availability of surface waters. A 3-dimensional (D) digital terrain model constructed in ArcGIS may identify process thresholds to predict the extent of a stable gully network in a 5 km^2 catchment in an eco-epidemiological, agro-village, georeferenceable, riceland ecosystem, study site, for example. The model, may optimally explicatively forecast georeferenceable, LULC controls on channel networks and interpret these in terms of a critical shear stress for channel incision (τ_c) modelling applied by saturation overland flow. The ecohydrological ArcGIS model may compare the shear stress applied by Hortonian overland flow to that applied by saturation overland flow in a riceland model delineating agro-village complex, time seris expansions into neighboring tilled land cover and its precise protusion(i.e.,deforestation) levels into forest-canopy, geoclassified, georeferenced, LULC polygons. In soil science, Horton overland flow describes the tendency of water to flow horizontally across geoclassifiable LULC surfaces when rainfall has exceeded infiltration capacity and depression storage capacity (Hortian 1933).

Discrete, seasonal, explanative, *Ae. aegypti* eco-epidemiological, time series, geo-spectrotemporally geosampled,immature, empirical, habitat data may also be thought of as thematic data in ArcGIS. As such, georeferenceable, iteratively interpolated, uncoalesced, capture point, proxy biosignature data feature attributes may be eco-geographically represented in an optimizable dataset of archived seasonal, expanding, explanatorial, geoclassfiable, LULC time series, eco-epidemiolgical, geo-predictive, risk maps. Points, lines, or polgonized areas in an urbanizing, riceland, georeferenceable, agro-ecosystem environment may be optimally delineated in ArcGIS [Jacob et al. 2007, Jacob et al. 2008]. Eco-epidemiological, regression-related, probablistic, explanatorily, metaheuristically optimizable,LULC forecast risk, mapping of uncontrolled urbanization may reveal accidental introduction of pathogens from neighboring, forest-canopied, goclassifiable LULCs into a riceland, agroecosystem, village complex.

For raster geo-representation of a discontinuous explanative, surface such as those seasonal LULCs associated with meandering, infrequently sparsely canopied, stream diversional tributaries (e.g., irrigation dam), agricultural LULC, explanatorial, time series changes , geo-spectrotemporally explicatively geoclassified in ArcGIS, fractionalized endmember deposition with subsequent sub-mixel (i.e., mixed pixel) iterative interpolation (e.g., co-kriging) in ArcGIS may reveal seasonal, hyperproductive, georeferenceable, sylvatic, *Ae. aegypti* capture point, immature habitats. Discrete, *Ae. aegypti,* endemic, categorical or continuous explanative, orthogonally decomposeable, spatial filtered data, may represent georeferenced objects in both the feature and raster data storage systems in_3D_Analyst as well (see http://webhelp.esri.com/arcgisdesktop/).

ArcGIS technologies in a cyberenvironment offer a unique way to collect and manage multiple, explicative, bio-ecological, seasonal geo-spectrotemporally geosampled, explanatorial, georeferenceable, diagnostic, clinical, field or remote, endemic, *Ae. aegypti,,* transmission-oriented, explantorial, regression variables. For example, developing and implementing streamlined data collections, aggregations and reporting methodologies employing Personal Dital Assitant (PDA) handheld computers equipped with differentially coirrected global positioning systems (DGPS) can provide eco-geographically explanatively detailed, georeferenceable, *Ae. aegypti,* immature habitats and geoclassifiable, elucidative, LULC information when conducting overhead surveys in Riceland, irrigated, agro-village ecosystems.

Differential Global Positioning System (DGPS) is an enhancement to Global Positioning System that provides improved geolocation accuracy, from the 15-meter nominal GPS accuracy to about 0.178m cm in case of the best implementations for vector arthropod-related habitat monitoring (Jacob et al. 200imaging tec).DGPS uses a network of fixed, ground-based reference stations to broadcast the difference between the positions indicated by the GPS satellite systems and the known fixed positions. These stations broadcast the difference between the measured satellite pseudoranges and actual (internally computed) pseudoranges, and receiver stations may correct their pseudoranges by the same amount. The digital correction signal is typically broadcast locally over ground-based transmitters of shorter range.

A similar system that transmits corrections from orbiting satellites instead of ground-based transmitters is called a Wide-Area DGPS (WADGPS) or Satellite Based Augmentation System. Augmentation of a global navigation satellite system (GNSS) is a method of improving the navigation system's attributes, such as accuracy, reliability, and availability, through the integration of external information into the calculation process. A satellite-based augmentation system (SBAS) is a system that supports wide-area or regional augmentation through the use of additional satellite-broadcast messages. Such systems are commonly composed of multiple ground stations, located at accurately-surveyed points (e.g., a prolific seasoanal *Ae. egypti*, forest canopy , sparsely canopied, habitat geosampled along a riceland tillered habitat boundary). There are many such systems in place and they are generally named or described based on how the GNSS sensor receives the external information. Some systems transmit additional information about sources of error (e.g., clock drift, ephemeris, or ionospheric delay), while others provide direct measurements of how much the signal was off in the past. Further, a third group can provide additional vehicle information to be integrated in the calculation process (e.g., determining exact geolocation of a kriged seasoanl*, Ae. egypti*, forest canopied, oviposition

sites). The ground stations take measurements of one or more of the GNSS satellites, the satellite signals, or other environmental factors which may impact the signal received by the user (www.nas.gov).. Using these measurements, information messages are created and sent to one or more satellites for broadcast to the end users. SBAS is sometimes synonymous with WADGPS, wide-area DGPS Post-processing may be used in Differential GPS to obtain precise positions of geo-predicted, unknown, remote, YFV –related, riceland and forest-canopy, georeferenced LULC points by relating them to known points such as survey markers.

Remotely targeted larval interventions as conceived in an ArcGIS cyberenvironment has great potential to maximize limited resources for accurate ecogeographic robust geo-predictions of vector arthropod immature , hypeproductive , ovipoistion, capture points. For instance, Jacob and Novak (2014) developed a new model Perersonl Digital Assistant (PDA), regression-based, sub-meter resolution [i.e.,QuickBird visible and near-infra-red (NIR) 0.61m wavebands] orthogonally gridded, autoregressive framework to evaluate larval interventions employing entomological inoculation rates (EIR), incidence, and prevalence of malaria in a riceland agroecosystem village complex in Mwea, Kenya. The authors determined the feasibility of using a PDA –related, seasonal, ArcGIS cyberenvironment as a mobile field data collection system by monitoring, and mapping multiple, malaria, mosquito vector, *Anopheline arabiensis s.s.*, aquatic, larval, habitat, Thessian polygons and other field-geosampled georeferenceable, parameterizable, covariate, estimator, time series, dependent, decomposed, coefficient values. The system employed QuickBird raster imagery displayed on a Trimble Recon X 400 MHz Intel PXA255 Xscale CPU®. The mobile mapping platform was employed to identify specific explanatorial, LULC geolocations of treated and untreated, seasonally, prolific, *An. arabiensis s.s.* aquatic larval habitats in the riceland complex. As data pertaining to prolific geosampled, *An. arabiensis s.s.* larval habitats were entered into the PDA, all treated and untreated rice paddies within a 2 km buffer of the agro-village, riceland-complex, study site were viewed in real-time and managed in the cyberenvironment.

To account for the seasonal variability in immature productivity of the *An. arabiensis s.s.* related, riceland, agro-ecosystem, geoclassified,seasonal, LULC habitats in Jacob and Novak (2014), the authors conceived a conceptual quantity, (i.e., total immature productivity). This explanatorial geopredictive dependent variable consisted of proportional, field geosampled, contributions as aggregated in the PDA-ArcGIS cyberenvironment of each individual, georefernced, larval habitat, seasonal count and other georeferenceable, regressable, feature data attributes (e.g., Depth of habitat, Daily precipitation rates). The impact of various scenarios of larval interventions on the empirically, geo-spectrotemporally geosampled, riceland, parasitological indicators of malaria transmission were then examined from the perspective of the immature habitat geolocations within the cyberenvironment employing a linear model. The objectives of the modeling effort were twofold: 1) to evaluate to what extent larval control should be undertaken to achieve specified goals in reducing incidence and prevalence of malaria in the riceland agroecosystem, eco-epidemiological, study site and, 2) to put into perspective how bioecological, geo-spatiotemporal, geosampled surveys of larval populations on specific seasonally geoclassified LULCs (e.g., post-harvesting) and prolific, aquatic habitats could assist in designing intervention programs.

The PDA models in the ArcGIS cyberenvironment generated had important messages with respect to planning of larval interventions. First, management of all aquatic habitats was deemed unnecessary for obtaining specified objectives of reductions in incidence and prevalence of malaria as in Gu and Novak (2005). This reduction in habitats, for geosampling, for implementing control strategies the authors in Jacob and Novak (2014) deemed especially important as the seasonally geo-spatiotemporally, geosampled, georeferenced *An. arabiensis* s.s., aquatic, larval habitats were not uniform in productivity at the rice agroecosystem village, complex, study site. Second, given the enormous variability observed in anopheline productivity between the habitats, the authors concluded that optimal intervention could be achieved by remotely targeting control efforts toward specific geoclassifiable, seasonally hyperproductive, capture point, immature habitats. In many situations, the majority of female mosquitoes in a focal area are likely to emerge from prolific habitats, which might account only for a small proportion of habitats (Novak 2012). Third, for targeted interventions, the authors in Jacob and Novak (2014) surmised it was crucial to survey seasonally, georeferenceable, prolific, aquatic larval habitats to robutstly forecast their seasonal larval productivity within a PDA-friendly, ArcGIS cyberenvironment. Targeted YFV-related interventions may require field and remote seasonal regressive quantification of georeferenceable, larval habitat productivity based on on-going, geosampled, mosquito data collections, (e.g. larval density and surface size of habitats) on a specific LULC (e.g., sparsely canopied, flooded, riceland habitat) for effectively implementing control strategies in an ArcGIS cyberenvironment.

While proper identification of mosquito species and knowledge of their bionomics focuses on control efforts, remote inspection of georeferenced, *Aedes,* larval habitats and/or weekly trapping for adult mosquitoes can ensure knowledge of the mosquito population in a given geoclassified LULC in an ArcGIS cyberenvironment (Jacob et al. 2011). Accurate, georeferenceable, entomological collection data is crucial for understanding mosquito biogeography, ecology, and the impact of environmental changes, as well as for species distribution landscape modeling, planning mosquito surveys, and for determining disease risk [www.cdc.gov]. A mobile collection system combined with newer ArcGIS software may help prioritize health interventions based on empirical, seasonal datasets of georeferenceable, geo-spectrotemporal, geo-spatiotemporal, empirically explanatorial, diagnostic, clinical, field or remote-specified, eco-epidemiological,optimizable predictors (e.g., parameterizable georeferenceable, sylvatic, *Ae. aegypti*-related, endemic, transmission-oriented, geoclassified, post-harvested, LULC, ricleand agro-ecosystem, estimator coefficients) in a more standardized regressionable fashion(e.g., Possionized probability distributions) .

Developing and implementing streamlined data collection, aggregation and reporting methodologies, employing a PDA-GIS-DGPS-remote sensing cyberenvironment can provide precise eco-geographically detailed, real-time, elucidative, LULC information which can lower overall seasonal, *Ae. aegypti* -related, larval habitat, treatment costs. For example, a bidirectional PDA-GIS-DGPS-remote sensing, web-based reporting system using a broadband satellite access point and the Internet can provide optimally efficient and timely amounts of relevant field-level information of immatures on geoclassified riceland agroecosystem LULCs. This is important, as seasonal, riceland, *Ae. aegypti*-related, treatment and management information (risk vulnerability maps of hyperproductive, flooded, seasonal habitats) would have the most value when it is quick, clear and easy to understand, and relevant to decisions that need to be made immediately.

An adaptable modular information surveillance system in the cyberinfrastrure can help insure that the right decisions are made to reduce the parasite, the vector or both in geo-classfiable, explanative, seasonal, riceland agroecosystem, LULC areas. This system can also support dissemination of information such as geosampled data about field activities and costs as they are compiled to provide an expedited understanding of how much is being spent in seasonal, *Ae egypti,* immature, habitat treatment. The cyberenvironment may also determine whether the intended outcomes (i.e., insectide trtreatement of predicted prolific larval habitats) are being realized. The cyberenvironment would allow both field managers and health ministries timely and optimal information to make field operational adjustments and maintain the most efficient and economically feasible pressure on the immature habitat, *Ae. egypti,* capture point populations. For example, once geoclassifiable explicative, LULC patterns and seasonal quantizable, explanatory correlations of geo-spatiotemporal, geosampled, hyperproductive, *Ae.egypti,* aquatic habitats are mapped in ArcGIS, field and remote, management practices can be modified to optimize applications of fertilizers and pesticides yielding lower overall costs and minimizing environmental impacts caused by excessive insecticide applications. In addition, a PDA-GIS-DGPS-remote sensing, cybernvironment, web-based, reporting system can provide a module that is dedicated to measuring the economic status of a riceland community and clusters of communities, in order to monitor and measure the LULC impacts of yellow fever and the economic benefits of control interventions. These LULC optimizations may be a critical factor for yellow fever eradication.

Unfortunately, it is very difficult to control or eliminate *Ae aegypti* or any seasonal, YFV-related, mosquito on any seasonally related, eco-geographical, ecohydrological, explanatively geoclassifiable, elucidative, orthogonally decomposed, proxy LULC, geo-spectrotemporal sub-meter resolution, endmember, biosignature, reflectance, emissivity wavelength transmittance dataset in any ArcGIS cyberenvironment especially based if the pameterizable covarites are stochastically and/or deterministically explanatorily iteratively interpolated, geo-spatializable, unmixed, feature, attribute value as the vector can make adequate adaptations to changing environments. The vector is highly remotely resilient to homogenous landscapes. Immature *Aedes* have the ability to rapidly bounce back to initial numbers after disturbances resulting from natural phenomena (e.g., droughts) or human interventions (e.g., mosquito control measures)(see Novak 2012). One such eco-adaptation of YF viruses is the ability of the *Aedes* eggs to withstand desiccation on specific LULCs (e.g., riceland-irrigation). People not only provide the mosquitoes with blood meals but also water-holding containers in and around the home for drinking, washing and cooking necessary for the mosquito to complete their development (www.dengue.gov.) The *Aedes* mosquito lays her eggs on the sides of containers above the water and eggs hatch into larvae when the eggs are submerged (Novak 1992). Thus, the *Ae. egypti* mosquito population in a riceland agroecosystem transitional LULC can be maintained as a result of eggs hatching following rainfall or by the man-made addition of water to containers harboring eggs. During the dry seasons productive containers with eggs limit the mosquito population to only those containers that are filled artificially with water primarily for home use or for home-grown plants in container (www.cdc.gov). YFV may be episodically transmitted by *Ae. aegypti* as well as by other, YFV-related mosquitoes even during the dry season since the habitats may originate in both rural and densely settled, urbanized, riceland non-homogenous, geoclassifed, seasonally explanatorial, partially canopied, LULC areas and

forest canopied LULCs. These expanding LULC seasonal habitats may be remotely captured in an ArcGIS cyberenvironment.

Aedes mosquiotes have been geo-spatiotemporally/spectrotemrporally geosampled on many different, multitemporal, explanatorial, ecogeographical, ecohydrological, geo-classifiable LULCs. Intensive sampling for eggs of *Aedes vexans* (Meigen) in upland wet-prairie habitats revealed that three species of plant, *Echinochloa crusgalli* (barnyard grass), *Polygonum pennsylvanicum* (smartweed), and *Cyperus esculentus* (yellow nutgrass), commonly occurring in contiguous U.S.A. and southern Canada and dense mats of vegetative LULC detritus, could be used as remotely sensed indicators of oviposition sites (Novak 2012). Where a slope was steep, specific elevations captured in a DTM indicated oviposition site preferences. Plant species and elevation were both indicators of areas where soil moisture was attractive to gravid females. Soil moistures between 55% and 69% were found optimal for oviposition.

Natural breeding habitats of *Aedes aegypti* in the Caribbean region were reviewed by Chadee (1993) by conducting larval surveys in Trinidad, Puerto Rico, and the U.S. Virgin Islands. Twelve types of natural LULC habitats were recorded: rock holes (9.7%), calabashes (2.4%), tree holes (19.5%), leaf axils (4.8%), bamboo joints (14.9%), papaya stumps (7.3%), coconut shells (4.8%), bromeliads (7.3%), ground pools (14.9%), coral rock holes (9.7%), crab holes (2.4%), and conch shells (7.3%), of which the coconut shell and calabash habitats were new to the Caribbean. The countries having the highest prevalence of natural habitats were Trinidad. Puerto Rico, and Jamaica, with 9 types (22.0%), 7 types (17.0%), and 6 types (14.6%), respectively. In Gargan II et al. (1988), floodwater *Aedine* mosquito eggs were recovered from soil samples taken from grassland, ecohydrological, LULC, depressions, called pans, in the Orange Free State Province of South Africa. Asedge, *Mariscus congestus* (Vahl) C. B. C1., was a useful indicator of *Aedes* (*Ochlerotatus*) *juppi* Mcintosh oviposition areas. No transovarial transmission of virus was demonstrated by *Ae. juppi* females reared from the eggs and allowed to feed shortly after eclosion on hamsters. No virus was recovered from 557 pools of 5,425 adult *Ae. juppi* that were collected as eggs and reared to the adult stage in the laboratory. Rift Valley fever virus replicated to high titres in experimentally infected *Ae. juppi* females, but horizontal transmission experiments proved inconclusive.

Historically, between 1959 and 1962, an explosive epidemic of yellow fever occurred in the Blue Nile region of the southern Sudan and then in the Omo River Valley of Ethiopia (Lindrec et al., 1968). The epidemic is estimated to have caused 30,000 fatalities. Yellow fever was again reported from east of Lake Margarita in Ethiopia during 1966. The occurrence of epidemic yellow fever in Ethiopia stimulated immunity surveys in 5 localities of the arid northern districts of Kenya specifically in two communities; Lokitaung Township and Marsabit Mountain (see Hendersen et al. 1968). Approximately 14% of the sera collected at Lokitaung and Marsabit Mountain were yellow-fever-immune. Although the immunity detected at Lokitaung could have resulted from a southward extension of the Ethiopian epidemic, the immunity detected in children and adults at Marsabit Mountain seemed likely to have resulted from YFV transmission within the vicinity of that township although this was not proven. However, the finding of yellow fever immunity in both regions was of sufficient importance to necessitate further serological surveys. Eco-epidemiological studies were then initiated at the sites to locate possible vectors and reservoir hosts of YFV. The initial collection of mosquitoes found that

Aedes (Stegomyia) africanus was common in eco-agricultural, explanatively geoclassified, LULC areas. The commonest mosquito in the forest collections was *Mansonla (Mansonioides) africana* (Theobald) and at forest edge LULC *Aedes (Finlaya) embuensi*. Additionally, the collections at three areas of the forest edge yielded 1,042 mosquitoes of the *Aedes (Aedimorphus) dentatus* group. One additional isolate came from *Culex zombaensis* (Theobald).

Further, at the forest, ecogeographical, cartographically delineated, LULC edge and in the areas of gum trees, a total of 30 *Aedes (Stegomyia) aegypti* (Linnaeus) were collected. On two occasions *Ae.aegypti* was found biting man indoors. Apart from *Ae aegypti,* the only *Stegomyia* collected was one specimen of *Aedes (Stegomyia) keniensis* van Someren. Although *Aedes* and *Haemagogus* mosquito species are relevant for the transmission of YVR, only *Ae aegypti* is particularly important because of its adapted ecobiology to the human domestic environment (www.cdc.gov/yellowfever/transmission/index.html).

Although in Hendersen et al. (1968) and other research thereafter, yellow fever mosquito vectors have been identified on varying geoclassifiable, explanatorial, ecogeographical, ecohydrological LULCs, no possible mechanisms of endemic, disease transmission have been remotely elucidated based on empirically seasonally, regressed, endemic, explanatorial, transmission-oriented, geo-spatiotemporal, geo-spectrotemporal, diagnostic, clinical, field geosampled, endmember reflectance emissivity, wavelength transmisstance, involving uncoalesced, iteratively interpolated parameterizable, covariate, estimator, coefficient values. It is evident that considerable additional information is necessary before YFV transmission activity can be resolved in a seasonal LULC in an ArcGIS cyberenvironment.

One of the limitations overlooked by Hendersen et al. (1968) and others may have been the non-detection of seasonally, explanatorial YFV–related, ecogeographical, ecohydrological orthogonally decomposeable, partially canopied, endmember, sub-mixel,unmixed riceland, agro-ecosystem, proxy biosignatures representing LULC reflectance emissivity, transmisstance, wavelength, seasonal, explanatorily iteratively interpolative, parameterizable, covariate coefficient, estimator values in ArcGIS. This data may be optimally employable in elucidative, time series dependent, eco-epidemiological, forecast-oriented, regression-based, risk-related, hierarchical, SAS trees and other programs (Python, C++) for accurate explanatorial, ecogeographical, ecohydrological, explicative, LULC, predictions of case distribution data and other seasonally hyperproductive, sparsely or dense canopied, endmember, YFV-related, endemic, transmission-oriented, seasonal covariates of LULC dynamics(geolocations of con needed to implement integrated vector management (IVM). In epidemic areas malaria control programmescould make significant inroads in morbidity and mortality who.gov]. Recent analyses by Jacob and Novak 2014, and Gu and Novak 2005 , suggest that spatial statistics and remote sensing data can develop enand implement larval control strategies and play an important role in future control programmes suh as IVM, especially under hyperendmeic circumstances .

Typical time series, ArcGIS applications of simulated riceland and forest canopied, LULCs and   vector, arthrpod-related, immature, explanative, field geo-spectrotemporally/spatiotemporally, larval habitat, seasonal data involve either images from sub-

resolution, optical, senor systems, or to a lesser degree, active radar sensors such as RADARSAT. These systems have proven to be satisfactory for many ecobiological ecohydrological, medical entomological, immature habitat, forecasting, heursitically optimizable, explanative, forecasting vulnerability, model applications, such as seasonal, eco-epidemiological, ecogeographical, risk mapping, georeferenced, explanatorial, riceland and forest, canopied LULCs into broad classes and, in some biomes, for estimating above ground canopied, ricland biomass, imature habitats. Another remote system, the canopy volume method (CVM) could employ the ability of a waveform-recording sensor (SLICER) to directly measure the 3- D explanatorial, georeferenceable, optimally quantized distribution of a forested canopy structure of a cluster of prolific, geo-spectrotemporally geosampled, seasonal,*Ae. egypti* habitats. Using Lidar data, Lefsky and colleagues (1999b), were able to treat geoclassifiable, forest canopy LULC as a matrix of voxels (i.e., 3-D pixels), each of which was orthogonally defined as containing canopy or not, either in the brightly or dimly sunlit portion of the canopy. This information was then employed to describe the quantitative and qualitative differences in partially canopied, seasonal LULC, discontinuous, empirically, explanatively decomposed, optimally regressed geospectral, uncoalesced endmembers where primate habitats are known to occur. Commonly, the YFV is transmitted by mosquitoes from monkeys to humans when humans are visiting or working in the jungle (www.cdc.gov). The forest YFV lives in mosquito species (*Aedes africanus*, *Haemagogus* sp, and others) that breed in tree holes (http://www.who.int/csr/resources/publications/yellowfev). Further, since sylvatic vectors are present at high density, such as in the savanna LULC zone of Africa, humans may serve as the principal host in epidemic transmission (www.who.int/csr/resources/publications/yellowfev).

It may be remotely pertinent to acquire and geospectrally/geospatially differentiate, seasonal, riceland, YFV-related, agroecosystem and forest canopied, iteratively interpolative, geoclassfiiable, endemic, transmission-oriented, time series, LULC, explanatorial, diagnostic, clinical, field and remote geospecified ecogeographic, ecohydrologic, larval habitat data. To data these processes have been imaged only using coarse resolution. Landsat Multi-Spectral Scanner (MSS) data and TM data was employed by Prince and Yang 2000 to monitor geocalssfiied, deforestation LULC changes in savanna vegetation between 1972 and 1989 in the South Luangwa National Park region, Eastern Zambia. Each pixel in an MSS scene represented a 68 m x 82 m ground area, while each pixel in a TM scene represented a 30m x 30m ground area (except in the case of the far-infrared band 6 which used a larger 120m x 120m pixel). Land-cover types in the region were mapped in ArcGIS and major changes in LULC from 1972 to 1989 were detected from the satellite data, change vulnerbility, risk maps. Forest canopy LULC cover was estimated for woodland vegetation from the MSS data employing a linear relationship between woody LULC cover and red reflectance captured from a neighboring riceland environment. The changing forest canopy cover LULC estimated from MSS data agreed with those measured from geo-spatiotemporal expanding riceland, geoclassifiable LULC (r=0.94). Woody, LULC cover canopy, changed significantly in the region from 1972 to 1989 and revealed strong spatial LULC patterns of deforestation in *Colophospermum mopane* woodland on alluvial soils and vegetation regrowth of valley miombo vegetation and geoclassified, rice woodland LULCs. This information on the geospatial patterns of canopy LULC cover change from 1972 to 1989 suggested certain criteria that any causative process must satisfy, and it provided a baseline to manage the natural resources in the region. Thus, the forest canopy and

riceland LULC cover estimated in ArcGIS from MSS and other high resolution sensor data provides important explanatorial, LULC reflectance, emissivity transmittance wavelength data for robustly constructing, biophysical and climatic process, eco-epidemiological, forecasting, geo-spectrotemporally optimizable, proxy biosignature, deforestation, risk models for estimating the seasonal, impact of rice vegetation LULC structure on forest canopy vegetation and climate processes.

High spatial resolution 3-band airborne images (G, R and NIR) have also been used to map savannah vegetation LULCs for qualitatively remotely quantitating the up scaling of transpiration measurements made over the tree-bush-shrub of Serowe area in Botswana. These images were acquired with digital TETRACAM camera mounted on a small aircraft to collect data in 30, 60 and 100 cm geospatial pixel resolutions. The airborne imaging results were then compared with 1m, pan-sharpened, multi-spectral, IKONO images. Two object-oriented, ecogeographic, ecohydrologic, endmember classification techniques corresponding to eCognition and Feature Analyst software packages were used in data processing. eCognition is a computer object based classification method for geo-spectrotemporally geospatially extracting information from images using a hierarchy of image objects (i.e., groups of homogenous pixels) as opposed to traditional pixel processing methods (see Jensen 2005). In the particular application of forecast, eco-epidemiological, geoclassified, optimizable LULC, risk mapping individual savannah tree species, the eCognition showed to be more accurate and reliable than Feature Analyst. The geo-spectrotemproal characteristics of the TETRACAM images were similar to IKONOS satellite images, while the geospatial characteristics were much better in TETRACAM images than in IKONOS images. This was optimally reflected in the substantial accuracy calculations obtained by quantiating the difference between the airborne and satellite data of the same resolution both quantitatively and qualitatively (i.e., seasonal, endmember LULC, geoclassifiable, orthogonally, explicatively decomposed eigenvector maps). The overall study demonstrated that the higher the geospatial resolution, the higher the number of tree species properly identified with regard to the species type and sparsely canopied topographic pattern, hence the higher the accuracy of the remotely sensed, LULC-related, explanatorial empirical, ecogeographic, regressable, observational, georferenceable predictors. The authors surmised that the more fully formed individual tree crowns were the better the remote distinction among their geospectral explanatorial, unmixed proxy biosignatures, especially where tree species shed and regained their leaves with season, thus affecting the stable green color associated with their endmember reflectance (Kimani 2007). Since explanatorial, remotely sensed, expositively geopredictive, time series co-factors can lead to better ecohydrologic, LULC savannah, orthogonally geoclassification then the chances of overestimating and/or underestimating seasonal, remotely geo-spectrotemporally geosampled, explanatorial, ecogeographic and other uncoalesced, geoclassifiable LULC, data feature attributes on regressively qualitatively, quantizable, irrigated, YFV-related, riceland agroecosystem, immature habitats, for example, may be minimal.

Interestingly, in Africa, an intermediate (savannah) cycle exists that involves YFV transmission from mosquitoes to humans living or working in jungle geo-spectrotemporally geoclassfied, LULC bordered, seasonal areas. In this cycle, the YFV can be transmitted from monkey to human or from human to human via mosquitoes. Meanwhile the urban LULC cycle involves transmission of the virus between humans and urban mosquitoes, primarily *Ae. aegypti.*

The virus is usually brought to the urban LULC setting by a viremic human who is infected in the jungle or savannah (www.infectionlandscapes.org/2011/07/yellow-fever.htm). Thus, geo-spectrotemporal, geospatial, non-quantitation and/or non-decomposition, of seasonally explanative, LULC reflectance, emissivity, ecogeographic, ecohydrologic, sub-meter, resolution, wavelength transmittance may hamper robustly forecasting surveillance, orthogonally, gridded, LULC-related changes associated with seasonal, YFV-related, endemic, transmission-oriented, explanatorial geopredictive, diagnostic, clinical, field or remote, geo-spectrotemporally geosampled, regressed, geospatialized, paramterizable covariate, estimator, coefficient outputs in a vulnerability, forecasting, eco-epidemiological, risk map when implementing IVM.

Since, seasonal, explanatorial, forecasting geoclassifiable, geo-spectrotemporal, endemic transmission LULC patterns of YFV would involve remotely explanatively defining multiple ecogeographic and ecohydrologic, polgonized boundaries of disease transmission zones, (e.g., riceland agroecosystem and forest canopy)an ArcGIS cyberenvironment may be important to accurately remotely capture all these data in real time for implementing an IVM. For instance, seasonal, explanatorial, georefernceable, YFV-related, geoclassified, endemic transmission zones (e.g., hyperendemic forested canopied LULC zones) may be optimally highlighted in an ArcGIS cyberenvironment constructed from seasonal, eco-epidemiological, time series dependent, ecohydrologic, forecasting, vulnerability, endemic, risk maps where the virus is continuously present and geosampled on an enzootic basis. The model may include probabilistically regressed multivariate, endmember, sub-mixel, unmixed, proxy, biosignature-related, geopredictive, explanatorial, uncoalesced variables ecogeographically representing partially canopied, forested, LULC zones areas where the YFV circulates between mosquitoes and monkeys or chimpanzees. An intermediate or emergence geo-spectrotemporal LULC change zone may however be needed to generate an accurate, geo-spatiotemporal, explanatorial, YFV, real-time, geopredictive,uncertainty-oriented, probablistic, eco-epidemiological, risk model in an ArcGIS cyberenvironment where for example, a partially canopied, agro-village complex, ecosystem, LULC, riceland area next to or outside of a geoclassified, forest- canopied LULC may be eco-epidemiologically, geoclassified as an endemic transmission zone. This remote area may include georefernceable geolocations of domestic human activity such as farmland, and herding LULC areas. By remotely capturing all time series dependent, geo-spectrotemporally geosampled, seasonal, LULC zones in an interventional,YFV-related, riceland agro-ecosystem, eco-epidemiological, study site and their boundaries, an updated seasonal, endemic, transmission-oriented, geopredictive, eco-epidemiological, risk model may accurately ecogeographically predict, seasonal, YFV increases in a ArcGIS-related cyberenvironment.

By generating robust, remotely-sensed, explanatorial, geopredictive, temporal, forecasting, vulnerability, seasonal, eco-epidemiological, ecohydrologic, risk maps the potential for human-to-human transmission in agro-village and forest canopy, geoclassfiable LULC zones as ecogeographically orthogonally defined in an ArcGIS cyberenvironment may be efficiently gridded and mapped especially when YFV -infected mosquitoes from the endemic zones lay their eggs in fields or savannahs outside the dense canopy, forested, LULC area. These environmental, ArcGIS-derived, optimized, time series dependent, explanatorial, geopredictive, eco-epidemiological, vulnerability maps may forecast the LULC areas at risk based on seasonal

geolocations of *Aedes* mosquito and wild monkey populations, for example. These risk maps may reveal seasonal, explanatorial, geoclassfiable, elucidative, LULC sites (e.g., dicontunous canopy gaps in geoclassiable forest LULC) with specific, overlaid, georeferenceable, breeding areas (e.g. dump sites) and seasonal, meteorological, sub-meter resolution, proxy, uncoalesced, geo-spectrotemporal, elucidative, parameterizable covariate, estimator, coefficient values such as evaporation demand for determining length of time for standing water presence to allow mosquitoes to complete their life cycle. Since the humans working or living in the fields on savannah LULC can become infected when they are bitten by the infected mosquitoes, early alerts may be then provided to at-risk agro-village communities using time series, iteratively interpolative, eco-epidemiological, risk maps constructed in an ArcGIS cyberenvironment. In some instances full scale evacuation may be mandatory. A high, risk, geoclassified, LULC area is an area where there is potential for an epidemic to because a human infected with YFV has been bitten by the sylvatic, *Ae aegypti* mosquito (www.cdc.gov). Since explanatorial, time series dependent, geo-spectrotemporally unmixed, geo-spatialized, geoclassifiable, biophysical, YFV, ecogeographic, geoclassified, LULC attributes are geo-characteristically remotely distributed on different explanatorial, geolocations during varying seasonal sample frames, qualitatively regressively quantitating meteorological, geo-morphological, (elevation, slope coefficients), and/or vegation-related parameterizable covariate, estimator, coefficient values employing satellite remote sensing in ArcGIS, Different LULCs change classes for intensive ecogeographic, ecohydrologic, remote sampling. Sattelite sensed data from weather radar are an important source of information in the analysis of the weather situations (Jensen 2005).

These data usually have a higher geospatial temporal resolution than conventional observations. They can provide LULC information on the conditions on the earth's surface and the atmosphere which may be important data for constructing robust models for cartographically delineating seasonal, soil moisture changes, for instance. On Anguilla,West Indies, two distinct populations have been identified; a domestic immature *Ae, egypti*, population with in large domestic undergroundwater storage cisterns and old asphalt drums, and a rock hole population. The latter habitat was first cited by Belkin and Heinetnann (1976) from a collectionmade during August 1966. The rock holes containing immature occuredat distances up to more than 1 km from the nearest human habitation (Parker et al. 1983). *Ae. aegypti* is closely associated with humans and human habitation. The female is primarily an indoor day-biter that feeds almost exclusively on humans and exploits artificial containers as sites to deposit her eggs (Focks and Alexander 2006, Halstead 2008.) Yellow fever mosquito aquatic, larval habitats include old tires, flower pots, aluminum cans, bird baths, rain gutters, tree holes, and many other items that can hold even small amounts of water Christophers SR..1960) Many *Aedes* mosquitoes will lay their eggs in damp soil; the eggs will hatch when flood waters cover (www.cdc.gov).

Fortunately, remote sensing techniques allow eco-epidemiological research teams to gain knowledge of meteorological, geomorphological and vegetation-related, geo-spectrotemporalized objects without being in direct contact with them (Hay 2000). Clouds, precipitation, atmospheric gases like water vapor and the earth and ocean surfaces are meteorological objects. All elevational and vegetation LULC objects absorb, scatter and emit electromagnetic radiation which can be utilized in passive and active remote sensing techniques in ArcGIS for risk, eco-epidemiological forecast mapping, YFV-related LULC-oriented, seasonal, prolific, georeferenceable, larval habitats. The geographic distribution of *Ae. aegypti* is

considered to, in part, be limited by cold temperatures: the low-latitude areas equatorward of the average 10°C winter isotherms in the northern and southern hemispheres approximate the climatic boundary for establishment of the mosquito (Richardson et al. 2011,Yang et al. 2009). Eggs of *Ae. aegypti* can be transported over long distances in artificial containers through human activities, including to areas outside the established range, but the innate climate tolerance precludes establishment in colder areas at middle and high latitudes and at high elevations at lower latitudes. The highest previously published elevation records for *Ae. aegypti* in the Americas are 1,630 m for México (Ibanez-Bernal and S. Nuevo 1987)and 2,200 m for Colombia. (Suarez MF, Nelson MJ. 1981). Moreover, the work by Herrera-Basto et al. (1992)(reported a dengue outbreak in the Méxican city of Taxco, located at 1,700 m, but the collected mosquitoes were only reported as *Aedes* species. Although these mosquitoes likely included *Ae. aegypti*, they also may have included specimens of another container-inhabiting mosquito *Ae. epactius*, which occurs at high elevation in México (Heinemann SJ and Belkin JN 1977). An important question is whether ongoing climate warming potentially could lead to the elevation ceiling for sylvatic, *Ae. aegypti* moving up to an extent where currently unaffected high-elevation cities are threatened by mosquito vector proliferation and establishment of local dengue virus transmission cycles. Although the mosquito has been studied intensely in dengue-endemic areas at the core of its geographic range, virtually nothing is known of its natural history at the cool margins of its range.

There may be a dearth of field information deriveable from ngeo-spectrotemrpoally uncoalesced remotely identifiable explanative specific climatic, elevation and vegation-related conditions under which *Ae. aegypti* is capable of establishment at the cool margins of its range. This dearth may (1) enhance the development of robust forecasting eco-epidemiological models for quantiating the distribution and abundance of the mosquito at the cool margins of its range and (2) assess the potential for climate seasonal shifts that may lead to LULC changes in the geographic distribution of the mosquito, which could place additional human populations at risk for exposure to the arbovirus vector. The latter includes not only high-elevation urban areas but also high-latitude pasture fields.

Studies of remotely derived, climate parameters,slope coefficients and optiamlly quantiziable vegetation-related, geoclassifiable, LULC regression covariates and associations with and *Ae. aegypti* may be complicated by the dependence of the mosquito on humans, especially its preference for human blood and its use of artificial containers as larval development sites. Socioeconomic conditions and human behavior (e.g., water storage practices or use of air conditioning or mosquito screening to prevent intrusion of mosquitoes into homes) may confound basic LULC associations between climate parameters and mosquito abundance at an agro-ecosystem ricland eco-epidemiological, study site. It may be therefore, especially improtnat to study geoclassifiable LULC associations between climate, vegetation ,elevation and geoerferenceable, geosampled, *Ae. aegypti* along transects that include high variability in socioeconomic conditions, such as transects extending from a forested sparsely canopied centroid to a riceland agrovillage complex centroid agro-village, for example. To minimize the potential confounding effects of varying conditions, a research effort may be focused on an elevation and climate gradient within an eco-epidemiological study site, where the targeted communities may include neighborhoods of comparable LULC status.

Fortunately, sub-meter resolution, remote sensing satellites provide continuous measurements of the earth and its environment, and offer a synoptic monitoring capability. As such, a remote ArcGIS cyberenvironment may provide continuous real-time sub-meter resolution, satellite measurements of geo-spatiotemporal, geo-spectrotemporal, geosampled, YFV–related, explanatorial, georeferencable, orthogonalized, LULC explicative, time series, gridded, uncoalesced, data feature attributes and other time series, eco-epidemiological, observational, transitioning illuminative, optimizable predictors (e.g., seasonal shifting riceland agroecosystem, geos-pectrotemporal, geo-classified LULC bordering a forest canopied, sub-mter resolution, optimally parameterizable, LULC covariate, estimator) for determination of optimal geo-spatiotemporal intervention time frames for implementing IVM strategies. These explicit robustifiable, explanative, time series, ground measurements may have distinct advantages over other remote cyberenvironmental, geospatial, surveillance tools in ArcGIS since they can be collected repeatedly and automatically. Generally, once an area has been imaged and then, ecogeographically, geoclassified, explicative, quantizable, regressive geospatial associations between observational, sub-meter resolution, LULC, geopredictive ecohydrologic, explanatorial, time series dependent, risk-related, explanatorial, georeferenceable variables (e.g., vector, pathogen, and reservoir host occurrence, abundance and distribution), and multiple orthogonally operationizable, probabilistically regressive, remotely extracted geosampled, paramterizable covariate coefficient values can be accurately quantitated using the LULC analysis capabilities of ArcGIS. Thus, sub-meter resolution, remote sensing data can be an important tool for developing and understanding seasonal, climatic and geomorphological, terrain-related, explanative, riceland agroecosystem, orthogonally decomposeable, time series, explanatorial, LULC processes for quantitating riceland agro-ecosystem, LULC intrusion into forest canopied LULCs while regressively and cartographically delineating endemic YFV-related, disease transmission potential for effectively implementing IVM.The study may reveal that the endemic transmission potential of YFV is very sensitive to climate changes on the periphery of geoclassifiable, georeferenceable, riceland agroecosystem, geoclassfiiable, sub-meter resolution, optimally delineated, explicative, multitemporal LULCs and forest canopied LULC encroaching endemic areas in ArcGIS. The health impact may then be optimally determined to be most pronounced in populations living in the less economically developed riceland temperate areas in which YFV endemicity is low or absent.

An experimenter may require knowledge of advanced, invasive, geospatial, rater-based analytical tools in a ArcGIS cyberenvironment for accurately eco-epidemiologically, forecast, risk mapping georefernceable, seasonal-geosampled, YFV-related, meterological, ecogeographic, ecohydrologic, explanatorial, parameterizable covariate estimator coefficient values on geo-spectrally orthogonally decomposed, riceland agroecosystem and forest canopied LULCs. For example, ArcGIS® Spatial Analyst, an optional extension to ArcGIS Desktop (e.g., ArcInfo®, ArcEditor™, and ArcView®) can provide powerful tools for optimally conducting geomorphological, terrain-related, meteorologically, comprehensive, seasonal, vector, YFV-related, endemic, transmission-oriented, raster-based, explanatorily biophysical,elevational, geospatial, eco-epidemiological, forecast, risk analysis. Within ArcGIS Spatial Analyst, experimenters can employ a wide range of seasonally, georeferncable, vector data formats for combining and qualitatively regressively quantitating, empirically geo-spectrotemporally geosampled, explanatorial, disgnostic, clinical, field and remote datasets with weather data while interpreting new data, and performing complex raster operations such as 3-D terrain trend

analysis, surface networking analysis, statistical analysis, and others. For instance, Jacob et al. (2013) employed ArcGIS Spatial Analyst for constructing an integrated environment within ArcGIS Desktop for performing a raster analysis on an empirical, time series-related, georeferenced, eco-epidemiological, explanatorial, geo-spectrotemporally geosampled, agro-village tributary, riverine, larval habitat, ecohydrological, georefernced empirical dataset of field and remote operationizable, probabilistically regressable, data feature attributes of a black fly vector of onchocerciasis *Similium damnosum* s.l.. Onchocerciasis known as river blindness and Robles disease, is a parasitic disease caused by infection by *Onchocerca volvulus*, a nematode (roundworm) which is second in the world only to trachoma as an infectious cause of blindness (www.who.gov).

In Jacob et al. (2013) a geospectral, geo-spatiotemporal, endmember,sub-mixel iteratively interpolative, frcationalized endmember, biosignature from QuickBird sub-meter satellite data (e.g., 0.61m pixel spatial resolution) was qualitatively extracted.The sub-meter, geospatial, resolution data was obtained from Digital Globe Inc., (Longmont, CO, USA). The satellite image and data of the eco-epidemiologcal, georeferenced, riverine, vector, arthropod, immature habitat, capture point were acquired on July 15, 2010, roughly at the mid-point of the rainy season. The data contained 25 km^2 polygons of the geoclassifiable, LULC at the eco-epidmeiological,study site. QuickBird products provide four discrete non-overlapping spectral bands in the 0.45 to 0.72µm range covering the visible and near NIR electromagnetic spectrum (EMS) with an 11-bit collected information depth (www.digitalglobe.com).

The QuickBird sensor has coverage of 16.5-19 km in the across-track direction. In addition, the along-track and across-track capabilities provide a good stereo-geometry and a high revisit frequency of 1-3.5 days (www.digitalglobe.com). Basic Imagery products are geo-radiometrically corrected and sensor corrected (Jensen 2005). The QuickBird sensor correction blends all mixels from all detectors into the synthetic array to form a single, high quality, sub-meter, geospatial, resolution image. In Jacob et al. (2013) the QuickBird image data were delivered as pan-sharpened composite products in IR colors. The clearest cloud-free images available of the contiguous sub-areas of the eco-epidemiological, riverine, study site along the river and tributaries were used to identify georeferencable, LULC and other explanatorial, ecogeographic, ecohydrologic, geospatial features associated with prolific, partial shade canopied, *S. damnosum* s.l. habitats.

The model developed to predict a georefernceable, capture point, trailing vegetation, *S. damnosum* s.l., riverine, agro-village, turbid water, larval habitats based on the endmember biosignature was designated the black rock-rapid (BRR) model. An optimal habitat representive 0.61m pixel was initially extracted employing object-based classifiers in ENVI. The mixel data included fast flowing water passing over a substrate of Precambrian rock. To develop the BRR model, individual pixel (0.61 m^2 per pixel) endmember, geospectral, ecogeographic, ecohydrologic, explanatorial, decomposed, reflectance estimates in the QuickBird images were imdividually extracted from a georeferenced, field-validated, prolific, eco-epidemiological, *S. damnosum* s.l. capture point, immature habitat. This procedure allowed for the creation of a robust, proxy biosignature based on of a unit of an actual confirmed seasonal habitat. The model used three scene components: sunlit canopy (C), sunlit background (G) and shadow (T) generated from the QuickBird image, to determine the sub-pixel endmember, derivative spectra

associated with the known habitats. The C, G and T explanative, time series geo-spectrotemporal, geosaptial, component classes were estimated using the ENVI software package (Exelis Visual Information Solutions, Boulder, CO) which employed the object-based classification algorithm for separating the prolific, shade,canopied, non-homogenous, larval, habitat-related, endmember continuous, quantizable, regresssive, date feature attributes. Explanatorial, non-parametric, heuristically optimizable, sub-mixel explanative, georferenceable estimators from the derivative spectra and the geometric-optical model were then used to construct the Boolean model that subsequently rendered a geospectral, endmember, reference, proxy biosignature in an ArcGIS geodatabase. The waveband composition data of the QuickBird signature was 34% red, 11% blue and 55% green. In QuickBird composites fully canopied, vegetation LULC is portrayed in red as the blue band is not used and the remaining bands are shifted (i.e., visible green to the blue color gun, red to the green gun, and the NIR band to the red gun)(www.digitalglobe.com).These unique identifiers of the agro-village complex, trailing vegetation, tubid water, discontinuously canopied, riverine habitat, geospectral biosignatures were then employed to forecast prolific, georefernced, larval habitats along unsurveyed rivers in both Togo and Uganda.

To assess the BRR model's ability to forecast the, prolific, georefernced, riverine, tributary, agro-village, complex, larval habitat, densly shade, canopied sites that was temporarily active under varying flow or flooding conditions, a second model was developed. The strategic approach taken was to overlay a Digital Elevation Model (DEM) with sub-meter resolution signals characteristic of Precambrian rock plus white water, or Precambrian rock alone. The DEM is a simple tool to geolocate differences in elevation and other catchment variables that would show areas where such fast flowing water could occur during different riverine meandering flow conditions (Jensen 2005). Crosskey (1960) revealed that explanatorial, orthogonally decomposeable, shade canopied, hyperproductive, *S. damnosum* s.l., larval habitats can be affected by non-temporally and temporally dependent, explicative, ecohydrological, seasonal attributes (e.g., Precambrian rocks, floating vegetation, turbid waters). In order to accomplish the iteration in ArcGIS, Jacob et al. (2013), employed PCI Geomatics software (PCI Geomatics, Toronto, Canada), which supported an automatic overlay of any interpolated wet and dry Precambrian rock, proxy biosignature along the river course. This endmember analysis revealed the geolocations of both active habitats (i.e. those with water flowing over Precambrian rock) as well as sites that might become active under increased flow or flooding conditions. Typical uses of ArcGIS Spatial Analyst for seasonal, vector, entomological-related, explanative, endemic, transmission-oriented, eco-epidemiological, risk-based, explanatorial, data analyses include DEM generation, larval habitat, population density, 3-D mapping, surface runoff modeling, cluster modeling, vegetation LULC, targeted mapping and eco-epidemiological study site geolocational probability exploratory risk,   analysis (see Jacob et al. 2013b, Jacob et al. 2012,Jacob et al. 2011c).

Of the 30 sites along the Sarakawa River in Northern Togo forecasted to be prolific, larval habitats by the BRR model, all (100%) were found to contain *S. damnosum* s.l. larvae. In contrast, none of the 52 sites not predicted by the BRR model, but deemed to be potential habitat by the medical entomologist accompanying the verification team contained larvae. Together, these data suggested that the BRR model exhibited a sensitivity and specificity approaching 100% for the prediction of *S. damnosum* s.l. riverine larval sites in Togo.

To test the generality of the BRR model, it was applied to predict, *S. damnosum* s.l., riverine, larval habitat, capture point, sites in Northern Uganda. A total of 25 potential *S. damnosum* s.l. larval habitat sites were predicted. Of the 25 sites predicted to be suitable habitats by the BRR model, 23 (92%; 95% CI 81–100%) were found to contain *S. damnosum* s.l. larvae. In contrast, just 2/10 (20%; 95% CI 0–45%) sites examined which were not optimally forecasted to represent *S. damnosum* s.l. aquatic habitats by the model were found to contain larvae. The BRR model thus exhibited a sensitivity of 80% and a specificity of 92% when applied in Uganda, a performance that was statistically significant ($p<0.0001$; Fisher's Exact test). The two sites that were not predicted by the model were found to contain larvae consisted of low hanging, streamside, geoclassfied, geo-spectrotemporally geosampled, georeferenceable, LULC, vegetation immersed in fast flowing water. The mean number of larvae found at the sites predicted by the BRR model (21.91) was significantly greater that the mean number of larvae at the sites consisting of immersed , trailing vegetation (4.0; $p<0.001$, Mann Whitney U test). Based on literature thus, ArcGIS Spatial Analyst software's strong integration with the ArcGIS Desktop geoprocessing environment may allow research experimenters to create and model YFV-related IVM employing orthogonally decomposeable, geo-spatiotemporal, georeferenced, field-geosampled, explanatory count data as a response count variables   and sub-meter, geospatially uncoalesced, satellite imagery and remotely-sensed geo-statistical, eco-epidemiological, forecasting, risk-related, field and remote independent estimators in a count variable (i.e., Poisson) probability framework.

Poisson models are commonly employed in vector arthropod, eco-epidemiological risk analyses. In such count variable regression equations, the dependent variable is usually prevalence data or habitat productivity. Possionized variables tend to render more realistic outcomes (e.g., non-inflated pseudo R^2 values) than a binary logistic paradigm. The log transformations in a dichotomous, non-linear, regression framework imposes that values above zero to be categorized as only 1. Hencforth all values above zero would be coded homogenously. Explanatorial quantizable multivariate differences in exogenous or endogenous, categorical or contimnuous, observational, YFV-related, explanative predictors could not be differentiated in liner regression space. Possion variables employ actual count variables as regressors but with the imposition that the mean in the model is equal to the variancxe. Unfortunately may vector arthropod-related models commonly violate this assumption, thereby rendering overdispersed, eco-epidemiological, georferenceable forecasts. Thus negative binomial regression models with a gamma distributed mean is employed to compensate for the over-Possion variation in these models, since in these paradigms the standard deviation is equal to the mean. Thus, extreme observations(e.g., geospatial outliers) may be accounted for. Outliers in geo-spectrotemrpoal, geospatial, vector arthropod-related medical entomological, forecasting, regression models can genereate heteroskedascitic or multicollinear variables which will then subsequently render misspecifications (e.g., covariates that are not associated with hyperproductive, seasonal riceland agro-ecosystem, *Ae. egypti,* seasonal georeferenceable, capture points)

Alternatively, an experimenter may choose to perform a non-linear estimation procedure. Many algorithms currently exist in ArcGIS that can quanatite clusteriung tendencies in geo-spectrotemporal, geosampled, empirical datset of georeferenceable *Ae. egypti* aquatic larval habitats in a ricleand agro-ecosystem environment. These non-linear models include autocorrelation using Moran's *i*,(see Moran 1950, Griffth 2003) autoregressive geospatially

weighted matrices (see Cliff and Ord 1972, Anselin 1995,) , Bayesain generalizable hierarchical paradigms (Gilks 1996)), stochastic/deterministic variograms and other iterative explicative geospatial interpolators (see Cressie 1993). Users of ArcGIS Spatial Analyst include those who need to build complex site geo-locational, analytical, eco-epidemiologial, time series dependent, ecogeographic, ecohydrologic, vector, arthropod-related, risk models as well as those users who are interested in conducting terrain and visibility modeling or who want to perform density mapping, overlay, distance analysis, or interpolation (Jensen 2005).

Because of complexities and limitations in remotely estimating, georeferenceable geopredictive, explanative, LULC variables through time series dependent, remote sensing data, proxy, time series, explanative, graphical indicator, uncoalesced biosignature variables such as decomposed proxy vegetation indices [e.g., Normalized Difference Vegetation Index (NDVI)] that measure the abundance, spatial extent, and dynamics of vegetation-related were created (e.g., see Tucker 1991, Huete 1994). The biophysical temporally dependent and/or non-independent, explanatorial, graphical indicator may be employed as a surrogate indicator of climate variability in seasonal, medical entomological-related, eco-epidemiological studies (Jacob et al. 2006, Hay 2000). A widely used remote, proxy, observational,time series, explanatorial, ecogeographic, predictor variable in vector mosquito is the Normalized Difference Vegetation Index (NDVI), (Jacob et al. 2010b, Brown et al. 2008, Cooke et al. 2006, Kunkel et al. 2005, Ward et al. 2004, Backenson et al. 2002, Brownsteinco 2002, Linthicum 1987).The NDVI, which exploits the strong contrast in the reflectance of vegetation in the red and NIR wavelengths, is a commonly employed index to study vegetation dynamics (Jensen 2005). Importantly, healthy vegetation absorbs blue- and red-light energy to fuel photosynthesis and creates chlorophyll. Photosynthetically regressively quantitated plant specific chlorophyll associated with a seasonal, georefernced, vector, entomological-related, aquatic, larval habitats on an explanatorial geo-classfied LULC in an ArcGIS cyberenvironment can determine an optimal geolocational site which will reflect more NIR energy than an unhealthy plant. Thus, analyzing forested, ecohydrologic, endmember –related, riceland, LULC spectrums associated to a diffuse wavelength sub-meter resolution, bi-directional, seasonal, YFV-related, explanatorial, regression-related, forecasting, risk model employing absorption and reflection indices extracted from visible and NIR transmittance emissivities may provide frequency information about levels of sparse or dense canopy forest health, deforestation levels and its relation to immature, mosquito productivity.

Since partial, discontinuous canopy, vegetation, LULC dynamics are influenced by variations in climate, strong quantitative correlations between vegetation indices, meterological variables and seasonal, YVR-related, orthogonally explanatively, decomposeable, explanatorial, ecogeographic, ecohydrologic, parameterizable, covariate estimator coefficients may be optimally revealed in a dynamic, seasonal, ArcGIS-generated, NDVI-related, forecasting, eco-epidemiological, geo-spatiotemporal, partially canopied, risk-related, forecasting vulnerability map. For instance, Brown et al. (2008) employed time series dependent, remotely-sensed, vegetation indices to discriminate among multiple West Nile Virus (WNV) mosquito, aquatic, larval, immature habitats within a densely populated urban environment in New Haven, Connecticut USA. ASTER-derived vegetation indices were identified for 16 sites where adult WNV mosquitoes were trapped. A canonical correlation analyses was then employed to determine if a significant relationship existed between NDVI, disease/water stress index and

distance to water and four local WNV competent vectors (*Culex pipiens, Culex restuans, Culex salinarius, and Aedes vexans*). Their model determined that a significant relationship existed between explanatorial, diagnostic, clinically geo-spectrotemporally geosampled, geopredictive, covariate, parameterizable, explicative estimator coefficients and prolific geosampled mosquito habitats (0.93, P=0.03). The final, eco-epidemiological, decomposeable, time series, dependent model forecasts explained 86% of the variance in the environmental and WNV mosquito measures.

Although Brown was able to show geospatial seasonal segregation between presumed enzootic vectors *Cx. pipiens* and *Cx. restuans* and the presumed, bridge vectors, *Cx. salinarius* and *Ae. vexans,* employing relatively coarse ASTER spatial resolution [e.g., 30 x 30m in the short wave infrared (SWIR)], data, a sub-meter resolution [e.g. QuickBird visible and NIR at 0.61m pixel resolution] may have geo-visualized the individual capture point, georferenceable, endemic, transmission-oriented, WNV, geo-spectrotemporally geosampled, decomposeable, iteratively interpolative, geosampled, mosquito data more accurately. In Jacob et al. (2013)seasonal explanatorial, geopredictive, LULC, characteristics were qualitatively, remotely quantitated employing QuickBird visible and NIR satellite data products in an ArcGIS cyberenvironment for geospatially, geopredicting, malaria case distribution Uganda. Initially, case, as counts, were used as a response variable in a Poisson probability model framework for quantifying datasets of explicative, district-level covariates (i.e., meteorological data, LULC densities and distribution of health centers, etc.) geo-spectrotemporally geosampled from 2006 to 2010 in Uganda. Results from both a Poisson and a negative binomial (i.e., a Poisson random variable with a gamma distrusted mean) revealed that the paramterized covariates rendered from the model were significant, but furnished virtually no predictive power. Inclusion of indicator variables denoting the time sequence and the district location spatial structure was then articulated with Thiessen polygons which also failed to reveal meaningful covariates. Thereafter, an Autoregressive Integrated Moving Average (ARIMA) model was constructed which revealed a conspicuous but not very prominent, first-order temporal, autoregressive structure in the individual district-level time-series dependent data. A random effects term was then specified using monthly time-series dependent data. This specification included a district-specific intercept term that was a random deviation from the overall intercept term which was based on a draw from a normal frequency distribution. The random effects specification revealed a non-constant mean across the districts. This random intercept represented the combined effect of all omitted covariates that caused districts to be more prone to the malaria prevalence than other districts. Additionally, inclusion of a random intercept assumed random heterogeneity in the districts' propensity or, underlying risk of malaria prevalence which persisted throughout the entire duration of the time sequence under study. This random effects term displayed no spatial autocorrelation, and failed to closely conform to a bell-shaped curve. In statistics, a random effect(s) model, also called a variance components model, is a kind of hierarchical linear model (Hosmer and Lemeshew 2002). Multilevel models (also hierarchical linear models, nested models, mixed models, random coefficient, random-effects models, random parameter models, or split-plot designs) are statistical models of parameters that vary at more than one level (Cressie 1993). The model's variance, however, implied a substantial variability in the prevalence of malaria across districts. The estimated model contained considerable overdispersion (i.e., excess Poisson variability): quasi-likelihood scale = 76.565. The following

equation was then employed to forecast the expected value of the prevalence of malaria at the district-level: prevalence = exp[-3.1876 + (random effect)i] .

QuickBird imagery products may be combined with other elucidative, forecasting, geopredictive, eco-epidemiological, risk, model information (e.g., seasonal, autoregressively, probabilistically quantitated, geo-spectrotemporally geo-spatialized, sparse or dense partially canopied, rice –agriculture, geo-classifiable LULC predictors along a forest line), for advanced photogrammetric processing and exploitation of georefernced, explicative, YFV-related, geopredictive, seasonal LULC transitions and proxy, remotely-sensed, explanatorial, partially canopied,biophysical-related field-geosampled, geo-spatiotemporal attributes (e.g., NDVI parameters). These remotely synthesized explanatorily optimally paramterized, covariate, estimators and their respected sub-meter, resolution, reflectance, emissiviy, transmittance coefficient, ecohydrologic, indicator , regressable values, could be included in a parsimoniously orthogonally constructed time series dependent, explanative, endemic, transmission-oriented, eco-epidemiological, forecasting, vulnerability model framework along with other multivariate, QuickBird-derived, explicatively diagnostic, time series, clinical, field and remote-specified, eco-geographic, partially canopied, georeferenceable, data feature attributes, (e.g., demographic population estimators) for accurately forecasting seasonal, YFV-related, field-verifiable case distribution.

QuickBird gridded imagery products have been applied to image to identify explanatorial, seasonal, georefernced, entomological,-related, partially canopied, ecogeographic, ecohydrologic, geo-spectrotemporal, geospatial, larval habitat data. For instance, Jacob et al. (2008) employed gridded multitemporal, QuickBird visible NIR waveband frequency data to optimally identify seasonal, malaria-related, aquatic, larval, aquatic habitats of *Anopheles gambiae* s.l. in Gulu, Uganda. Habitats of *Culex quinquefacistus,* a major mosquito vector of WNV and filarsis were also optimally identified. Endemic transmission-oriented, explanatorial, time series dependent, vulnerability, ecobiological, explanatorily iteratively quantitatively iteratively interpolatable, field-operationizable, sparse or dense, partially canopied, forecasting, vulnerability, eco-epidemiological, risk maps for the Gulu study site were created from QuickBird visible and NIR band data employing differentially corrected ground positioning system (DGPS) coordinates in ArcGIS 9.3$^®$ (ESRI, 380 New York Street, Redlands, CA 92373-8100, USA). A 150 m × 150m orthogonal, spatial filter, residual algorithmic, digitized, grid matrix was overlaid onto the sub-resolution image of the Gulu study site in ArcGIS with the purpose of providing a technique for quantitatively probabilistically estimating explanatorial, decomposeable, geospatial, LULC change employing georefernceable, eco-epidemiological, quantitative, elucidative, time series, geo-spectrotemporal regressors for illuminatively representing multiple, empirical, geo-spatiotemporally, geosampled, prolific, mosquito, aquatic, larval habitat, parmeterizable, explicative, covariate, estimator, time series, dependent coefficients.

The eco-epidemiological, study area was stratified based on the level of drainage and planning present within each digitized gridded, QuickBird, cell. A grid cell was classified as well-drained if functional (e.g. clear of debris or vegetation at the time of observation) engineered drainage systems were present and no standing water was visible, or if the grid cell was eco-geographically geolocated on a slope and no standing water was visible. A

georeferenceable, grid cell was geoclassified as poorly drained if it was geolocated in a depression or valley and had either no drainage systems, or the drainage systems were blocked with debris or vegetation. A unique identifier was then assigned to each grid cell.

Thereafter, a DEM of the eco-epidemiological, study area was downloaded seamlessly from United States Geological Survey (USGS, March 17th, 2007). The use of the DEM for determining seasonal explanatorial, ecogeographic, orthogonalized, LULC parameters for establishing a robust, endemic, transmission-oriented, eco-epidemiological, ecogeographic, ecohydrologic, geopredictive, risk model and has been proven by different studies on the ecology of malaria vectors including *An. gambiae* s.l along with the impact of canopied landscape on their populations and malaria transmission (Jacob et al. 2005, Jacob et al. 2007). The DEM was constructed based on a contour map of 1:50,000. The range of the elevation in the DEM had a minimum value of 996 m with a maximum value of 1,132 m. The slope of the georeferenced, *An. gambiae* s.l., aquatic, larval habitats was then found to be 0.171%. The slope of the *Cx. quinquefacistus* was 0.006%. There was a significant positive correlation for prolific, *Cx quinquefacistus,* aquatic, larval habitat, count and slope (0.24) whilst for *An. gambiae,* aquatic, larval habitat, immature count and slope there was a negative correlation (-0.23) for a local model based on the parameterizable explicative covariate, estimator Distance to stream.

The purpose of DEM construction for probabilistically regressively qualitatively quantitating seasonal, georeferenceable, ArcGIS-derived, sylvatic, YFV-related, eco-epidemiological, forecastable, risk-related, ecogeographic, ecohydrologic, mapping variables would be then to extract 3-D, topographic geomorphological, reflectance, emissity wavelength, transmittance of explanatorial, bidirectional, optimally georeferenceable, explanatively parameterizable, covariate estimator, regression coefficient values that are associated with time series, LULC formations, such as elevation, flow accumulation, flow direction and stream order that may be then geo-spectrotemporally geospatially associated with georeferenceable, endemic (ie., hyperendenic, mesoendemic), transmission-oriented, objects (e.g., georeferenced, prolific, shade-canopied, larval habitat, based on field-geosampled, time series, count data). Jacob et al. (2008) employed a wetness index or topographic index to represent soil surface moisture content at the Gulu, eco-epidemiological, study site. It was calculated in ArcGIS as ln(A/Tan B) where *A* was the upslope contributing area and TanB was the local slope. Parameters A and Tan B were then remotely qualitatively derived using a multiple flow-direction algorithm in ArcGIS to geostatistically and cartographically quantitate the explanatorial, geomorphological time series dependent, ecogeographic, ecohydrologic, geoclassified, YFV-related, geoclassifiable, LULC, regressed, agro-village riceland, covariate coefficients.

The Stream Raster Grid was then generated in ArcGIS. The advantage of using flow distance-to-stream rather than simple distance-to-stream is that flow distance takes flow direction and landscape profile into consideration (www.esri.com). Euclidian distance-to-nearest ecohydrological explicative, georeferenceable, georefernceable bodies were thereafter calculated by Distance from a grid cell to a stream grid cell which was defined geo-statistically employing a Stream Raster Grid. Flow distance-to-stream may affect availability of the aquatic habitat and is commonly calculated as the distance from a grid cell moving downstream to a stream grid cell defined by the Stream Raster grid (Jacob et al. 2008). The Terrain Analysis using DEM (TauDEM) in an ArcGIS cyberenvironment was then employed to retrieve the

explanatorial, georeferenceable, terrain-realted, iteratively quantitatively interpolatable, exegetically geomorphological, eco-epidemiological, endemic, transmission-oriented, regression-related, geo-spectrotemporally dependent, bi-directional, ecogeographic, ecohydrologic, sub-meter resolution, geoclassified, LULC, emissity transmittance, wavelength, parameterized covariate, radiance estimators. A 3-D model of the eco-epidemiological study area was then constructed based on the DEM using ArcScene extension of ArcGIS. The 3-D GIS platform has been advancing rapidly due to recent development of internet high-speed networks, knowledge-based databases, sensor network technology, information technology (IT), and availability of high-resolution space- and air-borne images, Web-based platforms, such as Google Earth, Virtual Earth and Sensor Map(www.esri.com).

Importantly, a web-based, joint, multivariate georferenceable, geodatabase within an ArcGIS cyberenvironment server can ensure timely deliverables employing multitemporal, QuickBird wavelength, reflectance, emissity transmittance, geo-spectrotemporally geosampled, ecogeographic, ecohydrologic, explanative, time series, LULC data within a PDA (Jacob and Novak 2014). This platform could lend itself to a larger public health capability for implementing a YFV-related IVM parsimoniously as it would have the ability to scale-in, time series dependent, eco-epidemiological, real-time, geo-spatiotemporally geosampled, probabilistically regresseable, explanatively orthogonally decomposable, sub-meter resolution,thereby, endmember data providing an easily accessible tool linking data from handheld to desktop to web. This database could extend current, seasonal, field, geosampling temporal, YFV-related protocols by providing the ability to transfer bio-geographical time series information from either a centralized repository web-hosted, database or directly between research experimenters. Data-based explanative, geo-spectrotemporal, geolocational intelligence aggregated and analyzed within an ArcGIS cyberenvironment server can provide effective dissemination of time series, YFV-related information which may lead to implementing an IVM. For instance, in Jacob et al. (2009), employed an uncoalesced dataset of iteratively interpolative, QuickBird visible and NIR, wavelength reflectance, emissity transmittance polygons to identify hyperproductive, georeferenceable, malaria-related, discontinuous, shade, sparsely canopied, prolific, georeferenceable, aquatic larval habitats of *Anopheles arabensis* in a riceland, agro-ecosystem village-complex, interventional, eco-epidemiological, study site complex in the Mwea Rice Scheme, Kenya. The system utilized sub-meter resolution, QuickBird raster imagery in the visible and NIR data spectrums with a PDA, (i.e., a Trimble Recon X 400MHz Intel PXA255 Xscale CPU®). A polygon layer outlining each georeferenced, *An. arabiensis,* aquatic, larval habitat was then created by digitizing the QuickBird imagery in GeoGrid®.

A GEO Grid Implementation for 3D-GIS was recently implemented in National Applied Research Laboratories (NARL), in Taiwan**.** The GEO Grid framework was based on grid technology, remote sensed data, and geographic information. The approach was initiated by synergy of NARL's core competence on environment monitoring and disaster reduction techniques. The constructed Geogrid included high-resolution satellite image processing, virtual reality, geovisualization and disaster mitigation technology along with the advanced cyberinfrastructure environments established within NARL. The framework was constituted by three layers, (i.e. application module, service interface, and computing/data/sensor grids). A prototype platform entitled 3D-GIS Taiwan was then constructed employing 2m resolution FORMOSAT-2 data and a 5m, geospatial, geo-spectrotemporal DEM. These explicative,

geomorphological, ecohydrologic, bidirectional, reflectance, emissity transmittance, sub-meter resolution, uncoalesced, wavelength, endmember, data variables were displayed in 3D stereo visualization and in web pages for the island of Taiwan. The approach emphasized the synergy of multi-discipline with cross-field cooperation for geosciences' application which presently has become a benchmark for implementation of Global Earth Observation System of Systems (GEOSS).

In Jacob et al. (2008), the authors overlaid an orthogonalized, stratified, algorithmic, 2km grid over the QuickBird visible and NIR data of a eco-epidemiological study site in a geo-spectrotemporally geoclassified, riceland agroecosystem, orthogonalized, georeferenced, LULC village complex. Each explanatorial cell within the gridded 0.61m matrix contained an riceland, endemic, explanatorial, LULC cycle (e.g., flooding, post harvesting etc), georeferenced, geo-spectrotemporal, geosampled, uncoalesced, data, feature, ecogeographic, attribute value, as well as multiple *An. arabiensis s.s.,* immature habitat explicatively, georeferenceable, geolocational coordinates which was joined relationally to other satellite geodatabases in ArcGIS. The geospatial geolocation of each georeferenced cell was implicitly contained within the ordering of the algorithmic sequential probabilistic, estimation matrix. The georeferenced, aquatic, larval habitats were then geo-spectrotemporally regressively characterized in relation to the explanorial, time series dependent, georeferenceable, ecobiological explanative attributes geosampled of a neighboring georeferenced, prolific habitat at the riceland agro-ecosystem, eco-epidemiological, study site in ArcGIS. Each georeferenced, seasonal, explanatorial, *An. arabiensis s.s.,* aquatic habitat/polygon was assigned a unique identifier. Field attribute tables were then linked to the polygons. The polygons were used to define the sampling frame, which extended to include a 1 km buffer from the external boundary of the interventional, agro-village, ecosystem complex, eco-epidemiological, study site.

Thereafter, a time series explanatorial, ecogeographic, geopredictive, LULC, endemic, transmission-oriented, classification was performed employing seasonal, rice cycle transitions and time series, disgnostic, clinical, field and remote, geo-spectrotemporally geosampled, *An. arabiensis s.s.,*larval habitat, regressed data variables in ArcGIS. Overall accuracy and class-specific user and producer accuracies were calculated for each of the resultant land cover bi-directional, reflectance, emissity transmittance, explanative, covariate classes. The producer accuracy is a measure of the omission error and indicates the percentage of pixels of a given land cover type that are correctly classified (Jensen 2005). Conversely, the user accuracy is a measure of the commission errors and indicates the probability that a pixel classified into a given class actually represents that class on the ground. These validation techniques in ArcGIS were calculated by dividing the number of pixels of the ith class correctly classified by the total number of mixels classified as the ith class. For each eco-epidemiological, riceland, agro-village,study site, metaheuristically optimizable, LULC risk mapping region, stratified sampling formulas were applied to estimate the probabilistic error matrix cell proportions and consequently, the estimates of overall and class-specific user's and producer's accuracy as well. There are two primary components of error in thematic maps such as land cover maps; position error and thematic error (www.esri.com).

Next, the overall accuracy of the eco-epidemiologically, empirically, qualitatively, ecogeographically, regressively quantitated, riceland agroecosystem, sub-meter resolution,

geoclassifiable LULC, irradiance ecogeographic, ecohydrologic, emissity transmittance, orthogonally quantitatively, decomposed wavelength, biosignature, frequency, data, orthogonalized, feature attributes was calculated by dividing the number of pixels correctly geoclassified (i.e. the sum of the diagonal axis of the matrix) by the total number of mixels included in the evaluation process. The authors noted that the multitemporal, multivariate, geo-spectrotemporal, QuickBird visible and NIR image datasets containing the extracted, digitized, georeferenceable, larval habitat boundaries, with unique identifiers displayed amongst the geosampled, georeferenced habitats within a 1 km buffer of the eco-epidemiological study site had the highest forecasted immature counts. Thereafter, an object-oriented classification in ENVI technology separated the geo-spatiotemporally/geo-spectrotemporally geosampled, prolific, georefernced habitats by larval/pupal productivity in *n*-dimensional, feature attribute, endmember geospace. ENVI combines advanced geo-spectral image processing and proven geospatial image analysis technology with a modern, user-friendly interface for generating optimal remote sensing models (http://www.exelisvis.com).Once all the multivariate, multitemporal, explicatively remotely diagnostic, quantitative, explanatorial, clinical, field and remote-geosampled geopredictive, endemic, transmission-oriented, ecogeographic, ecohydrologic, predictor variables were extracted, a time series regression analyses was performed on the mixel data which revealed significantly higher *An. arabiensis* larval/pupal counts in the tillering LULC stage of rice development.

For the probabilistic, YFV-related, eco-epidemiological, geopredictive, explanatorial, ecogeographic, ecohydrologic, metaheuristically optimizable, explicative, risk model construction, the producer's accuracy related the probability that a QuickBird-derived, ArcGIS, time series dependent, LULC class was correctly mapped and measured employing the errors of omission (1 - producer's accuracy). In contrast, the user's accuracy indicated the probability that a geo-spectrotemporally geosampled, endemic, transmission-oriented, explanatorial, eco-epidemiological, LULC mapping variable actually matched the field geosampled, information from the quantitated georeferenced, bidirectional, reflectance, emissity transmittance, explicatively interpretative, decomposed, wavelength, time series, weighted, parameterizable, covariate, estimator dataset by measuring the error of commission (1- user's accuracy). Accuracy results were then computed by weighting the endmember proportions of each geo-spatiotemporal, seasonal, endemic, transmission-oriented, georeferenced, YFV-related, geoclassified LULC at the Gulu eco-epidemiological study site, against total LULC geoclassified area used in the seasonal geosampling frame. Thereafter, the overall and producer's accuracy was optimally estimated using post-stratified formulas. Post-stratified estimators uses the known pixel totals for each LULC class for treating the sample as a stratified random sample of mixels in that class (Jensen 2005). The overall accuracy of the LULC classification in the QuickBird image was then determined employing a Kappa statistic. Cohen's kappa coefficient is a statistical measure of inter-rater agreement or inter-annotator agreement for qualitative categorical items (see Jensen 2005).

.
Thus, based on Jacob et. al. (2012) and other recent contributions to literature (Heuvelink and Griffith 2010, Jacob et al. 2013), here we assumed that compilation of accurate, seasonal, eco-epidemiological, ecogeographic, diagnostic, time series, geospectrotemporal. YFV-related, explicative, clinical, and QuickBird endemic, transmission-oriented, operationizable, field

georeferenced, reflectance, emissity transmittance, endmember, orthogonally explanatively decomposed, wavelenght data may allow probabilistically quantitating random effects term estimates in an ArcGIS-related, time series dependent, explanatorial, LULC-oriented, geopredictive, regression-related, eco-epidemiological, risk model for precisely forecasting case distribution. Conventional space-time geostatistics in ArcGIS may need to be extended with methods for estimating and quantifying georeferenceable, seasonal, explanatorial, YFV-related, geo-spectrotemporal, variation for quantitatively and iteratively,optimally interpolating geo-spectrotemporally uncoalesced, parameterizable, seasonal geosampled, eco-epidemiological, covariates for optimally quantitating stochastic/deterministic simulating time series dependent, endemic, transmission-oriented, diagnostically explanative, clinical, field and remote LULC, parameterizable, covariate, signature estimator, coefficient values.

Fortunately, current developments in quantitative geography in ArcGIS have given rise to many spatial indices using Local Indicators of Spatial Association (LISA), such as local Moran's *I* (*Ii*) and local Geary's *C* (*Ci*) (Anselin, 1995; Sokal et al., 1998), as well as Local Spatial Autocorrelation (LSA) statistics and Getis-Ord statistics, *Gi* and *Gi** (Getis and Ord, 1992) for precisely regressively qualitatively quantitating, seasonal explanatorial, geopredictive, georeferencable, YFV-related, diagnostic, clinical, field and remote-specified, uncoalesced, endmember, endemic, transmission-oriented, sub-meter resolution, uncoalesced, reflectance, emissity transmittance, wavelength-oriented, orthogonally parameterizable, covariate estimator coefficients. These geostatistics may be optimally employed to summarize probabilistic, uncertainty-oriented, latent autocorrelation in seasonal geo-spectrotemporally geosampled, ecogeographic, ecohydrologic, endemic, YFV-related, transmission-oriented, mosquito datasets, for example, for adjusting and displaying hyperproductive, LULC zones in ArcGIS, field operational, YFV-related, illuminative, time series dependent, vulnerability-oriented, probabilistic, geospatial, forecasting, eco-epidemiological, risk maps. Temporal and latent autocorrelation violates standard statistical techniques when stochastically/deterministically, iteratively interpolated, vulnerability modeling georeferenced medical entomological, endemic, transmission-oriented, time series, explicatively dependent, geopredictive, eco-epidemiological, geospatialized variables by altering asymptotically normalized residual variance and probability estimates (see Jacob et al. 2006, Jacob et al. 2007, Jacob et al., 2009). The variance standardized or asymptotically standardized residual matrix built employing a dataset of metaheuristically optimizble, georeferenceable,diagnostic, sylvatic, YFV-related, clinical, field or remote categorical or continuous, paramterizable covariate may also be qualitatively quantaited in PROC CALIS.

Residual-based experiments represent an important area of research in vector, medical entomological-related, time series-dependent, eco-epidemiological, forecasting, optimizable, risk-related, model analyses for diagnostic checking any non-normality in proposed reflectance, emissivity, transmittance wavelength, uncoalesced frequency, sub-meter resolution datasets of empirically regressable, geopredictive, fractionalized, endmember, orthogonally, explicatively decomposed, paradigm ecogeographic, ecohydrologic, time series dependent, heuristically parameterizable covariate, estimator, coefficient values. Such latent specification tests are covered in many econometrics and statistics textbooks and remain of interest in ongoing medical entomological research such as for qualitatively, quantitating geo-spectrotemporally geosampled, empirically probabilistically regresseable, YFV-related, time series-dependent, LULC-related,

explanatorial,orthogonally parameterizable, covariate, estimator, coefficient estimated values. Similarly, residual-based, explanatorial, probablistic, regression-related, orthogonally parameterizable, forecastble estimators (often referred to as two stage estimators) may be applied for generating robust, estimator hierarchical significance (e.g., 95% credibility for rejecting the null hypothesis that the geosampled geo-spectrotemporal, geospatial, *Ae egypti* data are positively aggregated (i.e, autoccorelated) along forest canopy and rice agriculture geoclassfied LULC polygons) .

Usually, the asymptotic distribution of residual-based, probabilistic, medical entomological-related, time series dependent, geo-predictive, explanatorial, non-normal, explicatively residualizable, eco-epidemiological, diagnostic georeferenceable forecasts is derived on a case-by case basis employing a particular model specification and/or some stringent assumptions about the statistic and/or the first-stage parameter estimators employed in the regression. The key assumption for deriving the limiting distribution of a latent explanatorial erroneous regression-related, geo-spectrotemporally geosampled, empirically, qualitatively quantitated, LULC-related, explanatorial, parameterizable, hierarchical, covariate estimator radiance, sub-meter resolution, emissity transmittance, wavelength endmember may be based on the unquantitated geospectrally orthogonally decomposed data smoothness conditions (e.g., differentiability). In such circumstances, optimally determining whether the limiting distribution of a regressively qualitatively quantitated, time series-related, YFV-related, optimally rendered linearly/geospatially regressed, explanatory probabilty statistic may be affected by the geosampled, diagnostic, time series, probabilistic clinical, field and/or remote specified, regressor or not, may be based on whether or not the empiricalized, YFV-related, operationizable, dependent, endemic, transmission-oriented, eco-epidemiological, forecasting, uncertainty-related, risk model dataset has a non-zero derivative with respect to the estimator (see for instance, Pierce, 1982, and Randles, 1982). Conversely an experimenter may rely on an invariance condition in a regression-based model framework for trivially satisfying any uncorrelated unquantitated, residually forecasted derivatives. Especially in cases where smoothness conditions do not hold or are non-trivial to establish (as, for instance, is the case for many rank-based YFV-derived, linear models), an overhead approach offers a useful and unifying alternative. Our alternative and novel approach for deriving the asymptotic distribution of the residual-based, time series forecasted, explanatorial, geopredictive statistics, or, more generally, the limiting distribution of statistics that involve estimated nuisance parameters was based on probabilistically regressively quantitative, iteratively interpolative, geo-spatiotemporal autocorrelation, uncertainty-related, YFV-related, diagnostic, clinical, field or remote geosampled coefficient estimates.

Here, we limit ourselves to considering strictly increasing, continuous singular, seasonally geo-spectrotemporally geosampled, explanatorial diagnostic, clinical, field or remote, empirically regressed, YFV-related, ecogeographic, ecohydrologic, elucidative, time series, explicative regressor functions. There exists quite a few of these model types described in the literature for other disciplines. The best known are: Minkowski's fragefunktion and Riesz and Sz.-Nágy's function. Minkowski's function was proved to be singular in 1932 by Denjoy, who had previously in 1915 established the existence of nonconstant singular functions with a complicated construction. Riesz and Sz.-Nágy presented their function in the–now classic– textbook *Functional Analysis* of 1952 [6, pp. 48–49]. This function, also known as de Rahm's

function or Lebesgue's function, had already been previously described and studied by other authors: (e.g., Billingsley [11, pp. 35–37] abd Hewitt and Stromberg [13, pp. 278–282].

Interestingly, in previous, explanatorial time series dependent, entomological-related, endemic, transmission-oriented, eco-epidemiological, vulnerability analyses, ordinary Least Square(OLS) and autocorrelation coefficients were optimally employed for parametrically estimating unknown time series dependent, geo-spectrotemporal, geosampeled, geo-spatialized, covariate, estimator coefficients in autoregression eco-epidemiological, probabilistic models constructed from explanatorial, empirical datasets of metaheursitically elucidative, georeferenceable, probabilistic, clinical, field and remote specified, uncoalesced, iteratively interpolative, reflectance coarse and low resolution, fractionalized, endmember, emissivity transmittance, wavelength, frequency estimators. To optimize explicative, distance-based, latent autocorrelation, error coefficients amongst clustering georeferenceable, aquatic, larval habitats and other explanatorial, ecohydrologic, predictor variables of *An. arabiensis*. Jacob et al. (2008) applied OLS and the $Gi(d)$ statistic in ArcGIS and found a significant cluster in a rice-village complex (Z score > 3.70, p < 0.05) with the clustering of habitats highest at a distance of 400m from the agro-rice village complex LULC. Commonly, non-binary weights are allowed in $Gi(d)$ and $G*i(d)$ statistics, and the correlations between nearby values of the statistics are derived and verified by simulation (Getis and Ord, 1992).When the endemic, transmission-oriented, eco-epidemiological, time series dependent, forecasting, vulnerability analysis was conducted in the neighboring riceland agro-ecosystem village complex two significant clusters were noted (Z score > 3.70, p < 0.05)—but only up to a maximum distance of 400m for a northern operationizable, georeferenceable LULC cluster and 150 m for a southern LULC cluster.

LISA may delineate geostatistically significant explicative, YFV geo-spectrotemporal, geospatialized, georeferenceable geolocations of non-stationary or eco-epidemiological, YFV-related, seasonal "hotspots" in geo-spatiotemporally-related, satellite mapped, endemic, transmission-oriented, ecohydrologic, ecogeographic, risk patterns of seasonal, explanatorial, geopredictive variables on specific LULC where global spatial autocorrelation is absent. Unfortunately, in the presence of global spatial autocorrelation, the distributional properties of LISA statistics are less certain (see Tiefelsdorf and Boots, 1997). In such circumstances, significance testing is problematic. Nevertheless, LISA (Ii, and Ci) are proportional to I and $C,$ which may be remotely regressively qualitatively quantiated for determining the eco-epidemiological, time series dependent, relative contribution of specific, georeferenced, empirical, sylvatic, YFV-related, LULC-oriented, explanatorial, endemic, transmission-oriented, geopredictive, variables based on regressively quantized, global autocorrelation, elucidative components, a property not shared by LISA (see Jacob et al., 2005). Consequently, LISA can be optimally employed for invasive, seasonal, time series, probabilistically related, YFV-oriented, endemic, eco-epidemiolgical, sub-meter resolution, vulnerability analysis of georeferenceable, explanatorial, geoclassifiable LULC-dependent, eco-epidemiological, clinical, field or remote geo-spectrotemporal endemic patterns generated from geosampled, clinical, field and remote-specified, geopredictive, empirically geospatially regressed variables exhibiting latent autocorrelation. Other autocorrelation statistics that may be optimally employed to regressively quantitate spatial interdependencies in seasonal, entomological-related, georeferenced datasets of explanatorial, empirical, geopredictive, endemic, transmission-oriented, time series dependent, observational, YFV-related, geo-spectrotemporally geosampled, eco-geographic variables for

conducting a risk-based time series dependent, data analyses include: Geary's c (global differences); the cross product statistic [GAMMA]; Ripley's K (cumulative pairs over distance); the geospatial autoregressive parameters [rho] and [lambda]; Getis and Ord's [G.sub.i] and [G*.sub.i] (local clustering); Anselin's[l.sub.i] and [C.sub.i]; and Matheron's 1/[LAMDA] (inverse of the semivariogram;i.e., the correlogram).

Thereafter, spatial statistics were used to address the issue of observational, bias amongst time series dependent, empirical ecobiological datasets of diagnostic, clinical, field and remote-geospecified multitemporal, multivariate, seasonal, YVR-related, georeferenced, LULC co-factors and other endemic, transmission-oriented, orthogonally parameterizable, covariate estimator, decomposed, sub-meter resolution, geo-spectrotemporally wavelength, endmember coefficient, uncoalesced values at the eco-epidemiological study site. Specifically we employed an eigenfunction decomposition algorithm for qualitatively quantitating latent autocorrelation error coefficients in a geopredictive, explicative, YFV-related, explanatorial, LULC, risk-related, eco-epidemiological, probabilistic, regression-based model framework. Conventional ANOVA may be inappropriate to simultaneously deal with independent variables of both categorical (e.g., LULC type) and continuous remotely sensed, explanatorial, canopied geopredictors (e.g., NDVI coverage) for accurately geospatially targeting, seasonal, YFV-related ,endemic transmission zones (Jacob et al. 2007, Jacob et al. 2006, Jacob et al 2005b) .

Interaction terms were also constructed with the georeferenced, explanatorial, empirical, time series dependent, YFV-related, geo-classified, endemic, LULC, transmission-oriented, geo-spectrotemporal geospatial, data feature attribute, sub-meter resolution, wavelength reflectance, emissity, transmittance variables and orthogonalized, spatial filter,synthetic eigenvectors for quantitating varying regression coefficients for detecting residual autocorrelation coefficients in the ArcGIS-related, geopredictive, eco-epidemiological, ecohydrologic, ecogeographic, risk model. In the context of geospatial regression analysis, several methods can be employed to control for the statistical effects of dependencies among remotely geo-spectrotemporally geosampled, medical entomological-related, empirical datasets of georeferencable observations employing a parametric (e.g., maximum likelihood or Bayesian) framework. However, more recent spatial filtering approaches focus on non-parametrically removing geospatial autocorrelation for accurate, parameterizable covariate, estimator hierarchical, statistical significance, biosignature-related estimation of orthogonally decomposed, geopredictive, entomological-related, endemic, transmission-oriented, endmember, covariate, risk-related, time series dependent, probabilistic, parameterizable, covariate estimator datasets (Jacob et al. 2013b, Jacob et al. 2012b, Griffith 2005).

One advantage of the non-parametric eigenvector filtering procedures for aiding in ecogeographically forecasting robustifiable, explanatorial, time series dependent, endemic, transmission-oriented, YFV case distribution data using LULC-related, sub-meter resolution, partially canopied, reflectance, emissity transmittance, wavelength, covariate, parameter estimator coefficients and other seasonal-geosampled data is that they do not require restrictive and unjustified distributional assumptions such as in non-regressive, parametric, estimation procedures. These procedures normally require explicitly geo-specifying the distributional characteristics of the underlying, endemic, transmission-oriented, optimally forecasting, eco-

epidemiological, risk model (e.g., time series-related, explanatorial, meteorological-based, geoclassified, YFR-related, LULC, geopredictive, uncertainty-oriented, residual map) in the spatial probablistic domain. Commonly, the non-parametric, spatial filtering, eco-epidemiological approach uses a specific subset of eigenvectors from a transformed spatial link matrix to capture dependencies among the disturbances of a spatial regression model; however, we assumed that the optimal subset for filtering a seasonal, endemic, transmission-oriented, explanatorial, YFV-related, time series dependent, explicatively geopredictive, eco-epidemiological, vulnerability model may be an objective function that minimizes spatial autocorrelation rather than maximizes model fit.

As such, we employed the eigenvector filtering approach promoted by Griffith (2003) and Getis and Griffith (2002) by focusing on specification of a mean response to regressively force geospatially dependent, time series-related, probabilistic, geopredictive, endemic, transmission-oriented, endmember, YFV-related, decomposed, biosignature sub-meter resolution, wavelenght, endmember, covariate coefficient values of an auto-model to zero. Our assumption was that a geo-spatiotemporal, eco-epidemiological, auto-model generated from a dataset of empirically explanatorily regressed, sub-meter, operationized, optimizable, clinical, field and remote-geosampled time series-related, multivariate, endemic, transmission-oriented, YFV-related, geopredictive, seasonal covariate, parameter estimator, coefficient values and their probability density mass/functions could contain a linear combination of the dependent variables value at a nearby LULC geolocation (e.g., agro-village complex). In probability theory, a probability density function (PDF), or density of a continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value(Cressie 1993). The probability of a seasonal, sylvatic, YFV-related, orthogonally decomposed, LULC, endmember wavelength, random variable falling within a particular range of covariate, parameterizable estimator, coefficient values was given by the integral of the variable's density over a range—that is, it was given by the area under the density function but above the horizontal axis and between the lowest and greatest values of the LULC geopredictive, range in ArcGIS.

In probability theory and statistics, a probability mass function (pmf) is a function that gives the probability that a discrete random variable is exactly equal to some value (see Hosmer and Lemeshew 2000). The pmf is often the primary means of defining a discrete probability distribution, and as such functions exist for either scalar or multivariate, multitemporal, seasonal, explanatorial, geopredictive, vector, arthropod-related, endemic, transmission-oriented, probabilistically regressable, time series, ecohydrologic, ecogeographic random variables whose domain is discrete (see Jacob et al. 2005b). In medical entomological, geopredictive, explanatorial, eco-epidemiological, endmember, decomposed, biosignature-oriented, wavelenght, risk models, a pmf is a function that gives the probability that a discrete random variable is exactly equal to some value. Probability density function is most commonly associated with absolutely continuous univariate distributions (Cressie 1993).

A random entomological time series variable X has density f_X, where f_X is a non-negative Lebesgue-integrable function, if: $\Pr[a \le X \le b] = \int_a^b f_X(x)\,dx.$ (Hosmer and Lemeshew 2002). Hence, if F_X is the cumulative distribution function of a regressed georeferenced, explanatorial, time

series dependent, sylvatic, YFV-related, geospatial, data feature attribute, reflectance, emissity, transmittance variables[i.e.,X],then: $F_X(x) = \int_{-\infty}^{x} f_X(u)\,du$ and (if f_X is continuous at x) $f_X(x) = \frac{d}{dx} F_X(x)$ Intuitively, in a time series dependent, sub-meter resolution, wavelength, orthogonally decomposed, emissivity transmittance, YFV-related eco-epidemiological forecasting, risk model, $f_X(x)\,dx$ would be the probability of X falling within the infinitesimal interval $[x, x + dx]$.

Importantly, when there is a natural order among the hypotheses x in a geo-spatiotemporally, geosampled, entomological-related, forecasting, eco-epidemiological, remotely geosampled, ecohydrologic, ecogeographic, risk model, it may be convenient to assign numerical, endemic, transmission-oriented, georeferenced, explanatorial, time series dependent, endemic, YFV-related, transmission-oriented, geospatial, data feature attribute, reflectance, emissity, transmittance sub-meter resolution, wavelength, covariate coefficient values to n-tuples in case of a discrete, multivariate,multitemporal, random variable for geo-spatiotemporally quantifying other explanatorial, clinical, field or remote geosampled values not in the image of X.For instance, Jacob et al. (2013) built a robust, regression model where f_X was defined for an empirical-geosampled dataset of explanatorial, *S. damnosum* s.l., riverine, larval habitat, hyperendemic, transmission-oriented, ecohydrologic, ecogeographic probablistically time series, regressed data values where $f_X(x)$ was 0 for all $x \notin X(S)$. Since the QuickBird decomposed image of X (i.e., georefernced, eco- epidemiological, capture point at the riverine study site) was countable, the pmf $f_X(x)$ was zero for all but a few geo-spectrotemporally geosampled, prolific, shade canopied, georeferenced, riverine larval habitat values of x. The discontinuity of the larval habitat pmfs was then calculated and the value was then determined to be related to the fact that the cumulative distribution function (CDF) of a discrete random variable (e.g., larval habitat distance measurement from the georeferenced capture point) was also discontinuous in the empirical, geosampled, georeferenced, covariate, parameter estimator dataset. In probability theory and statistics, the CDF, or just distribution function describes the probability that a real-valued random variable X with a given probability distribution will be found at a value less than or equal to x (see Hosmer and Lemeshew 2002). Explanatorial, YFV-related, time series dependent, geospectrotemporal regressors were differentiable in our ArcGIS-related, geopredictive, endemic, transmission-oriented, risk model where the derivative was zero, just as the PMF was zero at all the geosampled, explanatorial, YFV-related, seasonal, LULC points.

Unfortunately, the standard methods for time series regression analyses of clustered, seasonal, vector, entomological-related, georeferencable, operationizable, multivariate, multitemporal, LULC-related, georeferencable, data models relating observational bioecological-geosampled, ecogeographic, explanatorial, clinical, remote and field, covariate, parameter estimator, orthogonalized coefficients to prolific, endemic, transmission risk zones do not include qualitative quantitation of residualizable, geospatial, intra-cluster, probability-oriented, error correlation effects. Generally, this correlation comes from two sources: (1) the design of the random effects and their assumed covariance from the multiple levels within the regression model; and, (2) the correlation structure of the probablistic residuals. Unfortunately, inconspicuous errors in georefernced, residual, intra-cluster, probabilistic, regressed correlation estimates can overstate precision in explanatorily, forecasted, seasonal-geosampled,

entomological-related, larval, habitat, endemic, transmission-oriented, data feature attributes regardless how they are treated (e.g., independent, autoregressive, Toeplitz, etc.). For instance, in Jacob et al. (2012b), the geolocations of multiple, multivariate, multitemproal,prolific, shade canopied, riverine-based, S. *damnosum* s.l., larval, ecosystem habitats geosampled from 2 pre-established eco-epidemiological sites in Togo were identified and recorded from July 2009 to June 2010. Initially, the data was aggregated into PROC GEN MOD. An agglomerative, hierarchical, residual, explanatorial, cluster-based, endemic, transmission-oriented, risk-based, eco-epidemiological, hierarchical analysis was then performed in ArcGIS. The geosampled, clustered, time series-related, geopredictive, endemic, transmission-oriented, empirically regressed,probabilistic, data, feature attributes was then analyzed for statistical correlations using Monthly Biting Rates (MBR) which is a rate to calculate the bites an individual could potentially receive by *Onchocerca*-infected simuliids. MBR = number of black flies caught multiplicated by the number of days in month divided through the number of catching days per month while the Annual Biting Rate (ABR) is the sum of the 12 MBRs (see Crooskey 1960).

Euclidean distance, parametric,geo-spectrotemporal measurements and geomorphological terrain-related statistics were then generated in ArcGIS. A digital overlay was performed also in ArcGIS employing the georeferenced ground coordinates of high and low *Similium* density clusters stratified by ABR. This data was then overlain onto multi-temporal, multivariate, sub-meter, pixel resolution satellite data (i.e., QuickBird 0.61m visible and NIR wavebands). Orthogonalized, spatial filter, synthetic eigenvectors were then generated in ArcGIS. Univariate and non-linear regression-based models (i.e., Logistic, Poisson and negative binomial) were constructed in PROC REG to determine probability distributions and to identify statistically significant georeferencable linear, covariate, parameterizable estimator coefficients values from the geosampled emprically regressed decomposed endmember, ecohydrologic, ecogeographic wavelenght emissivity transmittance datasets. An LULC was performed using the sub-meter resolution data.

Thereafter, Durbin-Watson test statistics were employed to test the null hypothesis that the geo-spectrotemporal, regression-based, endemic, transmission-oriented, geospatialized, explanatorial endmember, decomposed, biosignature-oriented, sub-meter resolution, wavelength residuals were not autocorrelated against the alternative that the residuals followed an autoregressive process in AUTOREG. The endemic, transmission-oriented, geopredictive, residually forecasted, explanatorial, clinical, field and remote-specified derivatives revealed both time series-dependent geospatially structured and unstructured error effects in the high and low ABR-stratified, georeferenced, geospatial clusters. The analyses also revealed that the parameter estimators, Levels of turbidity and Presence of Precambrian rocks were statistically significant for the seasonal high-ABR-stratified clusters, while the Euclideanized estimators Distance between habitats and Floating vegetation were important for the low-ABR-stratified cluster.

Importantly, Jacob et al. (2012, Jacob et al. 2010, Jacob et al. 2009) proved that varying and constant, decomposed, endmember wavelenght, time series dependent, endemic, transmission-oriented, regression models that have ArcGIS-generated geospatialized, explanatorial regressors derived from QuickBird satellite imagery can be parsimoniously derived employing a robust, intra-cluster, eco-epidemiological, residual, diagnostic test, and an eigendecomposition spatial filter algorithm in SAS (i.e. (AUTOREG) for accurate estimation of

latent autocorrelation uncertainty-oriented affects and other predictive, non-normal probabilities (i.e., unquantiated, heteroskedastic residuals in an empirical dataset of georeferenced, vector, arthropod-related, geo-spectral, endmember, decomposed, biosignature, parameter estimators). The asymptotic distribution of the resulting residually adjusted, intra-cluster geopredictive, erroroneous, autocorrelation coefficients may thus be established on seasonal, YFV-related, ArcGIS, eco-epidemiological,probablistic, ecohydrologic, ecogeographic regression-related, vulnearbility, risk maps using parasitological measurement indicators while estimates of the normalized asymptotic variance can lead to the construction of approximate confidence intervals for when geo-spatiotemporally remotely targeting productive endemic transmission zones based on specific explanatorial, LULCs and field-geosampled, time series count data.

We employed the misspecification interpretation of geospatial autocorrelation in AUTOREG, which assumed that the residual, geopredictive, explanatorial, error correlation in a dataset of seasonal-geosampled, empirically, orthogonally regressable, QuickBird-derived, explanatorial, sylvatic, YFV-related, LULC, time series dependent, covariate, wavelength, emissivity transmittance,frequency, ricland, African, forest-canopied, discontinuous, *Ae aegypti* regressors   and other remotely sensed, endemic, transmission-oriented, endmember, decomposed, fractionalized, biosignature-related, parameterizable estimators were induced by missing exogenous variables. Our assumption was that these data may have been geospatially correlated. Autocorrelation is characterized by a correlation in a signal among nearby locations in geospace (Cressie 1993). The aim of our spatial filtering analyses was then to control for latent autocorrelation coefficients in the empirical dataset of geo-spatiotemporal, geosampled, time series-dependent, georeferenced, explanatorial, clinical, field and remote specified, YFV-related, geopredictive, endemic, transmission-oriented, ecogeographic, operationalizable, ecohydrologic variables with a set of proxy variables, rather than to identify a global autocorrelation parameter for regressively quantizing a geospatially dependent process. Spatial autocorrelation is more complex than one-dimensional autocorrelation because spatial correlation is multi-dimensional (i.e. 2 or 3 dimensions of space) and multi-directional (Griffith 2003). We then employed Global statistics for summarizing the   standard error and other uncertainty parameters estimates from the regressed dataset of YFV-related, clinical, field and remote specified, explanatorial georeferenced, data feature attributes for geospatially assessing the autoregressive error at a single geosampled LULC site at the Gulu eco-epidemiological, study site.

Importantly, among our assumptions was that uncertainty-oriented latent, autocorrelation coefficients and other non-normalized, probability, error propagation may be qualitatively quantitated in the latent, probabilistic residualized,orthogonalized, derivative of eigenfunction decomposition algorithms.We assumed employing spatial filter orthogonal eigenvectors in an empirical dataset of time series-related, regressable, explanatorial, clinical, field and remote-sampled, endemic, transmission-oriented, geo-spatialized reflectance, wavelength, emissivity transmittance, covariate, parameterizable estimator, decomposed coefficient values   would generate robust residual autovariance terms and mean squared geopredictive error estimates for accurately forecasting YFV-related, case distributions at the Gulu eco-epidemiological, study site parsimoniously.   Conversely we assumed that violations of normality would compromise the estimation of coefficients and the calculation of confidence intervals in an explanatorial, geopredictive, time series dependent, regression-based, endemic, transmission-oriented, eco-

epidemiological, YFV-related, emissivity fractionalized, endmember, orthogonally decomposed, biosignature-oriented, ecohydrologic, ecogeographic, sub-meter, wavelength, probablistically regressed, eco-epidemiological, risk model. Sometimes, for instance, the error distribution for instance may be "skewed" by the presence of a few large outliers in the empirical, georeferenced, estimator dataset (see Hosmer and Lemeshew 2002). Since seasonal empirically-oriented, vector, entomological-related, time series dependent, explanatorial, clinical, field and remote regressor, parameter estimation is based on the minimization of squared error (see Jacob et. al., 2005b), we assumed, a few extreme, seasonal, YFV-related, probablistically explanatorial, LULC observations may exert a disproportionate influence on the time series, dependent, eco-epidemiological, regressed, uncertainty-related, coefficient vlues. Calculation of confidence intervals and various significance tests for coefficients are all based on the assumptions of normally distributed errors (Fortheringham 2000). If the error distribution is significantly non-normal, confidence intervals in a seasonal, YFV-related, endemic, transmission-oriented, explanatorial, clinical, field and remote geospecified, eco-epidemiological, operationizable, geopredictive, forecasting, risk model may be too wide or too narrow. The optimal test for non-normally distributed time series-dependent, entomological-related, stochastically/deterministically, explanatorial iteratively interpolatable endmember, decomposed, biosignature,time series errors is a non-normal probability plot of the residuals in ArcGIS (see Jacob et al. 2013). We assumed that this error quantiziation would be optimally realized as a plot of the fractiles of the error distribution versus the fractiles of a seasonal, normalized distribution in an ArcGIS cyberenvironment for optimally rendering the same mean and variance for the regressed, sylvatic, YFV-related predictors. If the seasonal, geosampled, YFV-related, explanatorial, clinical, field and/or remote geospectrotemporal, ecogeographical, specified, endemic, transmission oriented, risk-dependent, eco-epidemiological, LULC decomposed, emissivity endmember, sub-meter resolution, decomposed, wavelength, transmittance, distribution pattern is normal in a regressed, empirical geosampled dataset of endemic, transmission-oriented, YFV-related, capture points, for instance, the plot would fall close to the diagonal line. A bow-shaped pattern of deviations from the diagonal would then indicate that the residuals have excessive skewness (i.e., they are not symmetrically distributed, with too many large errors in the same direction). An S-shaped pattern of deviations may indicate that the residuals have excessive kurtosis—(i.e., there are either two many or two few large YFV-related, regression time series errors in both directions).

In probability theory and statistics, kurtosis is any measure of the "peakedness" of the probability distribution of a real-valued random variable(see Hosmer and Lemeshew 2002). In a similar way to the concept of skewness, kurtosis is a descriptor of the shape of a probability distribution and, just as for skewness, there are different ways of remotely quantifying uncertainty for attaining a robust, theoretical geo-spatiotemporal, time series dependent, eco-epidemiological, normalized, explanatorial, clinical, field or remote geosampled, ecohydrologic, ecogeographic, YFV-related, probablistic, prevalance distribution, for example, from a sample of a population. There are various interpretations of kurtosis, and of how particular measures should be interpreted; these are primarily peakedness (width of peak), tail weight, and lack of shoulders (distribution primarily peak and tails, not in between) (Fotheringham 2002).

One common measure of kurtosis in entomologically dependent, empirically geo-spatiotemporally geosampled, datasets of georeferenced explanatorial, clincial, field or remote predictor variables is based on a scaled version of the fourth moment of the data or population, but Jacob et al. (2013) has argued that this non-normality really measures heavy tails, and not peakedness. In these datasets, higher kurtosis means more of the variance is the result of infrequent extreme deviations, as opposed to frequent modestly sized deviations. Jacob et al. (2012) employed an adjusted version of Pearson's kurtosis, the excess kurtosis, to provide a comparison of the shape of a given distribution to that of the normal distribution in a malaria–related, *An. arabienis s.s.,* geopredictive, eco-epidemiological, linearized, explanatorial, forecasting, endmember, decomposed, biosignature-oriented, ecohydrologic, ecogeographic, risk model. The fourth standardized moment was defined as $\beta_2 = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2} = \frac{\mu_4}{\sigma^4}$ where μ_4 was the fourth moment about the mean and σ was the standard deviation. The fourth standardized moment was bounded below by the squared skewness plus 1: $\frac{\mu_4}{\sigma^4} \geq \left(\frac{\mu_3}{\sigma^3}\right)^2 + 1$ where μ_3 was the third moment about the mean in the malarial risk model. Kurtosis in the model was then defined as the fourth cumulant divided by the square of the second cumulant which in Jacob et al. (2012) was equal to the fourth moment around the mean divided by the square of the variance of the probability distribution minus 3(i.e., $\gamma_2 = \frac{\kappa_4}{\kappa_2^2} = \frac{\mu_4}{\sigma^4} - 3$). The "minus 3" at the end of this formula made the kurtosis of the probablistically regressed, normalized, immature *An. arabiensis* s.s., habitat distribution equal to zero The fourth standardized moment must be at least 1, so the excess kurtosis must be −2 or more(Hosmer and Lemeshew 2002). This lower bound in the eco-epidemiological, time series dependent, geopredictive, explanatorial, immature, risk model was then realized by the Bernoulli distribution with $p = ½$. In probability theory and statistics, the Bernoulli distribution, is the probability distribution of a random variable which takes value 1 with success probability p and value 0 with failure probability $q = 1 - p$ (Hosmer and Lemeshew 2002).

Interestingly violations of normality in an empirically geo-spectrotemporally geosampled probabilistically regressed dataset of seasonal vector entomological–related, endemic, transmission-oriented, explanatorial, clinical, field and remote specified, georeferencable, optimizable dataset of time series, reflectance emissivity transmittance, endmember, decomposed, georeferenceable, biosignature-oriented, covariate paramterizable estimators coefficients often arises either because (a) the distributions of the dependent and/or independent variables are themselves significantly non-normal, and/or (b) the linearity assumption is violated(see Jacob et al. 2005). In such cases, a non-linear transformation of the explanatorial, endemic, transmission-oriented, geopredictive, explanatorial, ecohydrologic, ecogeographic clinical, field or remote specified variables might cure both problems. In some cases, the problem with the residual distribution is mainly due to one or two very large errors. Thus, time-series dependent,explanatorial, YFV-related, probabilistic, LULC, endemic ,transmission-oriented,geopredictive, empirically regressable reflectance, emissivity transmittance, covariate, parameter estimator coefficients should be scrutinized closely to determine if they are they genuine (i.e., not the result of data entry errors) and, if the rendered forecasts have explanability of the response (i.e., dependent) variable (e.g., similar events likely to occur again in the future for qualitatively remotely quantitating how influential the

267

regressed predictors are in the model-fitting results.If the geosampled, time series dependent, explanatorial, clinical, field and remote, sylvatic, YFV-related, geospecified, endemic, transmission oriented, risk-related, geopredictive, eco-epidemiologically regressed, endmember, decomposed, fractionalized, biosignature variables are merely errors or, if they can be explained as unique events not likely to be repeated, then an ecologist, medical entomologist, or other experimenter may have cause to remove them from an ArcGIS cyberenvironment. In some cases, however, it may be that the extreme values in the geosampled, time series-dependent, YFV-related, explanatorial regressors that provide useful information for non-violation of assumptions (e.g., homoskedascity) about the linearized explanatorial, clinical, field and remote-specified, geosampled, endemic, transmission-oriented, diagnostic values based on of some of the residual ecohydrologic, ecogeographic LULC-related, covariate, parameter estimator coefficient values and/or provide the a realistic guide to the magnitudes of forecast errors.

Importantly, one of our assumptions was that a latently autocorrelated time series dependent, explanatorial, clinical, field or remote geosampled endemic, transmission-oriented eco-epidemiological, LULC, time series, risk model geopredictor could adjust erroneous YFV estimators geospatially. By so doing, we also assumed that we would be able to construct a robust, forecasting, case distribution model. However, we realized that our calculations may be dependent on assumptions of normality and homogeneity in the YFV-related, eco-epidemiologically regressable, explanatorial, clinical, field and remote geosampled geo-spatiotemporally-dependent, non-optimized endmember, orthogonally decomposed, biosignature derivatives. If however the assumtions of normality are not violated (e.g., sufficient residual quantization of multicolinearity and other non-normal probabilities) in our seasonal geopredictive, YFV-related, ArcGIS-derived, LULC-based, eco-epidemiological, forecasting, risk model, we assumed then that the inferences drawn could be sound (e.g., stepwise regressors significance levels would not be invalidated .

Violations of homoscedasticity make it difficult to gauge the true standard deviation of regression forecast errors, usually resulting in confidence intervals that are too wide or too narrow in regressed reflectance emissivity transmittance variables (Jacob et al. 2009d, Griffith 2005). In particular, if the variance of the errors is increasing over time, confidence intervals derived from a geospatially regressively quantized, empirical geosampled datasets of operationalizable, YFV-related,clinical, field and remote-specified, endemic, transmission-oriented, endmember, decomposed, biosignature, reflectance, emissivity transmittance, covariate parameter estimator coefficients for out-of-sample predictions would tend to be unrealistically narrow. Heteroscedasticity in geo-spectrotemporally dependent, ecohydrologic, YFV-related, ecogeographically forecasted data may also have the effect of giving too much weight to small subsets of geosampled, multivariate, multitemporal, clinical, field and remote, residually, forecasted derivatives namely the subset where the error variance is largest when estimating the coefficients of statistical importance. Therefore, examining non-normal diagnostic eco-epidemiological plots of residuals versus time and residuals versus predicted value in an ArcGIS, reflectance, emissivity transmittance, seasonal, forecasting, geo-spectrotemporal LULC-related, eco-epidemioloigcal, probablistic risk model may, for instance, reveal evidence of derivatives that are getting larger (i.e., more spread-out) either as a function of time or as a function of the ecogeographically regressed value (e.g., as LULC tends to urbanize the boundaries between agro-village complex LULC and canopied LULC forest tends to increase).

Plot residuals versus some of regressed dependent, explanatorial, endemically independent, sub-mter wavelength, transmittance , frequency orthogonally decom poseable LULC variables in ArcGIS cyberenvironments can elucidate, transmission-oriented pathways (Jacob et al. 2014).

In time series-related, entomological-related, ecohydrologic, ecogeographic, endemic, transmission-oriented, forecasting, reflectance, emissivity transmittance, biosignature-oriented, non-normal, eco-epidemiological, risk models, heteroscedasticity often arises due to the effects of inflation and/or real compound growth, commonly magnified by a multiplicative seasonal pattern(See Griffith 2005, Jacob et al. 2009, Jacob et al. (2012). Some combination of logging and/or deflating in an ArcGIS cyberenvironment may stabilize the variance in this case. Seasonal regressed YFV-related, data may show periods of increased or decreased volatility over time-this is normal and is often modeled with so-called ARCH (auto-regressive conditional heteroscedasticity) models in which the error variance is fitted by an autoregressive model. Heteroscedasticity may also be a byproduct of a significant violation of the linearity and/or independence assumptions in a geo-spatiotemporal, regression-based, ecohydrologic, ecogeographic, YFV-related, endemic, transmission-oriented, explanatorial, clinical, field or remote, specified, eco-epidemiological, forecasting, reflectance, emissivity, transmittance-oriented, probabilistic, risk model in which case it may be fixed as a byproduct of fixing those problems.

Interestingly, typical seasonal, probabilsitically forecastable, ArcGIS-related, explanatorial, geopredictive, entomological-related, eco-epidemiological, reflectance, sub-meter resolution, emissivity transmittance, wavelenght, biosignature-related, risk modeling approaches have employed logistic/Poisson regression and/or discriminant analysis techniques to investigate geospatial associations between multivariate, mutlitemporal, environmental, data feature attributes. In so doing, residually forecasted, regression–related, ecohydrologic, ecogeographic, endmember, decomposed, ecogeographic, proxy biosignature, predctor variables could include seasonal, LULC, explanatory, endmember, covariate, parameter estimator, coefficient values and NDVI patterns for determining vector presence or absence for seasonal risk mapping vectors and vector-borne diseases (see Jacob et al. 2005b, Griffith 2005). For instance, in Jacob et al. (2005b) Poisson and logistic regression models were constructed in SAS (i.e., PROC REG) employing the counts of immature *Anopheles gambiae* s.l. mosquitoes for the aquatic larval habitat empirical datasets of Malindi, Kisumu,Kenya and for a combined data set of both cities.

In all models, explanatorial, time series dependent, predictor variables were added in a stepwise manner, beginning with the one that explained the most variation. For the Poisson regression this included shade for both Malindi and the combined dataset for the Kisumu and Malindi, and domestic animals for Kisumu. The final model explained 52% of the variation in Malindi, 36% in Kisumu, and 41% in the combined data sets of the eco-epidemiological Malindi and Kisumu study sites. In the logistic regression models, habitat size for Malindi, domestic animals for Kisumu, and shade for the combined data set of Kisumu and Malindi were significant. The final model explained 34% of the variation in Malindi, 5% in Kisumu and 21% for the combined data sets of Kisumu and Malindi. The R^2 of the final model for Kisumu was 0.35 ($P < 0.001$); for the validation set, it was 0.14 ($P = 0.052$). The R^2 of the final model for Malindi was 0.52 ($P < 0.001$); for the validation set, it was 0.44 ($P < 0.001$). The probability Poisson regression models with the actual counts estimated the frequencies and their variation

for all multivariable models in Kisumu and Malindi, with greater explanatory power than the parameter estimates from the logistic regression. Thus, Poisson methods was capable of ecogeographically forecasting a posteriori probability of the presence of the dependent variable (e.g., either the vector or the disease) from a set of independent variables (e.g., seasonal climate and LULC data) which was then exported into ArcGIS to generate accurate geopredictive, explanatorial *An. gambiae s.l.*, larval, habitat, vulnerability, ecohydrologic, ,time series, risk maps from sample datasets (i.e., training datasets) based on the observed similarity of environmental conditions to LULC sites.

Thus, by ecogeographically portraying a relationship in linear and non-linear regression space between two geo-spatiotemporally-geosampled explantorial, YFV-related clinical, field or remote, time series specified, georeferenced variables, a correlation coefficient may be rendered that quantitates numerically non-normalized regressors. These deviations may be cartographically delineated in an ArcGIS geodatabase employing a residually, diagnostic, error, plot estimator (e.g., an orthogonal eigenfunction decomposition algorithm). In so doing, mechanical data input error (e.g., spatially pseudoreplicated variables), non-zero, autocovariate, probabilistic, heteroskedastic parameters may be determined in a time series dependent, empirically regressed, uncertainty-oriented, eco-epidemiological dataset of YFV –related, explanatorial, diagnostic clinical, field and remote, geo-spectrotemporally geosampled, ecohydrologic, ecogeographic, covariate, parameter estimator coefficients. Thus, in an empirical, operational dataset of endemic, transmission-oriented, ARIMA time series dependent, geopredictive, YFV-related, ArcGIS constructed, heuristically optimized, risk model, time series dependnent parameter estimators, for instance, the residual explanatorial forecasts could be robustly tabulated employing an average of numerical specifications between the geo-spatiotemporally geosampled parameterizable estimators for parsimoniously defining all seasonal LULC change-oriented, probabilistic geolocations. But since these time-series, explanatorial, geopredictive variables would be unobservable, the assumption invoked would be exchangeability whereby, the set of LULC time series estimators could be permuted without affecting results in any seasonal simulated, dataset of georefernced, ArcGIS-related, YFV-related, eco-epidemiological, forecasting, risk-related, model, residualized, probablistic derivatives. We assumed that the order in which a time series mechanism generates the covariate parameter estimater coefficient values measurement indicator values across an interpolated geo-spatiotemporal, autoregressive, endemic, sylvatic, YFV-related, regression-based, eco-epidemiological, vulnearbiltiy risk map would thus be irrelevant. As such, areas with statistically higher YFV-related, endemic, transmission rates, (e.g., positive autocorrelation clusters) based on geosampled temporally dependent, ecohydrologic, ecogeographic, parasitological indicators (e.g., LULC variables, prevalence rates) in ArcGIS could then be mathematically qualitatively quantitated for precisely remotely targeting implementation of an IVM using a geospatial, eigenfunction, decomposition algorithm. In so doing, latent autocorrelation coefficients in an empirically regressed, eco-epidemiological dataset of district-level, seasonal-geosampled, explanatorial, clinical, field and remote –specified, geopredictive, YFV-related, explanatorial, time series, *Ae.aegyppti*, ovispoition,seasonal, hyperproductive, LULC predictor variables may be also sufficiently quantitated for geospatially adjusting any ArcGIS, endemic, transmission-related, optimal, vulnerability risk mapping data feature data attributes in ecogeographical space.

Importantly, the choice of techniques in software programming languages for constructing time series-related, explanatorial, geopredictive, autoregressive, entomological-related, eco-epidemiofical, YFV risk models should be able to accommodate both categorical (e.g., disease presence or absence) as well as continuous data (e.g., monthly rainfall data). C++ is one of the most popular programming languages currently and is implemented on a wide variety of hardware and operating system platforms which can accommodate continuous and categorical, explanatorial, ecohydrologic, ecogeographic time series, covariate, parameter estimator, time series dependent, probabilsitc, sub-meter resolution, wavelenght-oriented, orthogonally decomposed, fractionalized, biosignature-related, emissivity, endmember coefficients. As an efficient performance driven programming language it is employed in systems software, application software, device drivers, embedded software, high-performance server and client applications (Stroustrup, 2010). Several groups provide both free and proprietary C++ compiler software, including the GNU Project, LLVM, Microsoft and Intel. C++ has greatly influenced many other popular programming languages, most notably Java.

Currently, C++ provides more than 35 operators, covering basic arithmetic, bit manipulation, indirection, comparisons, logical operations and others. Almost all operators can be overloaded for user-defined types, with a few notable exceptions such as member access as well as the conditional operator. The rich set of overloadable operators is central to employing user created types in C++ as well and as easily as built in types so that the user employing them cannot tell the difference (Stroustrup, 2010). The overloadable operators may be also an essential part of many advanced C++ programming techniques, for geopredictive, explanatorial, forecast,eco-epidemiological, risk modeling time series dependent, YFV-related, LULC,and meteorologically-oriented, endemic, transmission-oriented, endmember decomposed, sub-meter resolution, wavelenght-oriented, emissivity transmittance-related, georeferenced, data feature, endmember attributes.

Sampling from a geospatialized, time series-related, explanatorial, YFR-related, endemic, transmission-oriented,operatiuonizable, diagnostic, explicative, clincial, field or remote geosampled, time series dependent, geo-spatiotemporal, regression model employing, eco-epidemiological-based categorical and continuous, geopredictive, variables may be efficiently conducted in C++. Forexample, , calibrating seasonal YFV-related, ecohydrologic, ecogeographic, geoclassfied LULC change-related,geo-spatiotemporal endmember variables from forest canopy to agro-village complex against instrumental climate records in C++ employing correlation and response functions may generate a robust, time series dependent, explanatorial, geopredictive, forecasting, eco-epidemiological, risk model employing bootstrapped confidence intervals to estimate the significance of both correlation and response function, sub-meter resolution, orthogonally decomposeable, wavelength, frequency transmittance, covariate coefficients. Input and output file selection, as well as analytical options may then be chosen from a user-friendly GUI. In the eco-epidemiological, C++ model the results for any time series-related, explanatorial, geopredictive, YFV-related, endemic, transmission-oriented, elucidatively geo-spectrotemporally geosampled, clinical, field and remote, explanatorial, optimally specified, probabilistic, risk-related, forecasting, *Ae aegypti* model estimators may be saved in ASCII format which may then be plotted on screen using color-coded symbols. In addition, the C++ program may statistically calibrate climate signals while

simultaneously incorporating multivariate,multitemporal, LULC-related, geo-spatiotemporal changes and yellow fever prevalence rates for parsimonious forecast, risk modeling of an empirical-geosampled dataset of operationizable, seasonal, YFV-related, explanatorial, operationizable, clinical, field and remote specified wavelength, transmittance emissivity, covariate parameter estimator coefficients for remotely forecasting case distribution. In so doing, a C++ seasonal geopredictive, YFV-related, eco-epidemiological, geopredictive, risk model may also allow present eco-dynamical representation of quantifiable, statistical relationships between continuous climatological and categorical LULC-related, explanatorial parameter estimator wavelength-oriented, risk-related, emissivity, transmittance coefficients for case distribution by means of moving intervals for implementing IVM.

Thus, different explanatorial, time series dependent, ArcGIS-related, time series LULC types and other remotely-geosampled, QuickBird synthesized meteorological observational demographical statistics, prevalence rates and known information on seasonal, YFV-related, mosquito species abundance, and occurrence were integrated in C++ for forecasting case distribution data on varying ArcGIS classified LULCs in Gulu, Uganda. The model took inputs from a table file provided the information of yearly population and yellow fever cases from 1990 to 2012.These information were employed to determine average population growth and number of yellow fever case occurrences to 2020. Our assumption was that combining ArcGIS information in C++ could robustly display all forms of ecogeographically referenced, QuickBird-derived, explanatorial, LULC information associated to yellow fever at the Gulu eco-epidemiological, study site. Since yellow fever reflects different disease ecologies we assumed that a unique landscape eco-epidemiological, explanatorial, geopredictive, ArcGIS-related, forecasting, time series dependent, risk model could elucidate endemic, transmission pathways in C++. We also assumed that specific seasonal, geopredictive, explanatorial, LULC-related, YFV-related, transmission dynamics for the Gulu eco-epidemiological study site may be elucidated for parsimoniously implementing IVM. Forexample, sylvatic YF may occur in monkeys that are infected by wild mosquitoes in Gulu, thus, identifying dense, forest canopy, LULC areas in ArcGIS geodatabase using NDVI and other remotely sensed geoparameters where wild monkeys aggregate as a hyperendemic disease transmission zone.

Therefore, we generated multiple datasets of seasonal, georefernced, YFV-related, explanatorial LULC and NDVI time series dependent (i.e., 1990 to 2012), ecohydrologic, ecogeographic covariate, parameter estimators, coefficients and then employed linearized/nonlinearized regression residuals for constructing a robust, geopredictive, eco-epidemiological, YFV-model. These model forecasted derivatives along with a series of meteorological explanatorial time series dependent variables, population density statistics, human, agricultural and census reported data were then entered into C++. We then created one hypothetical formula for geopredicting YFV case distribution at the Gulu eco-epidemiological study site. We created one covariance for the forecasting model. This formula took into consideration all the explanatorial, geo-spatiotemporal, YFV-related, forecastable variables. Our assumption was that a single covariance could facilitate the modification of multiple, explanatorial, YFV-related, geopredictive eco-epidemiological, empirically regressed, time series, geosampled variables without disturbing the model parameter estimation process. Also, we assumed that the weightage on the covariance could be employed to define the efficacy of

covariance on the overall regression-based, YFV–related, geopredictive formula. Our assumption was that C++ methods coupled with multiple,linear and non-linear, regression coefficients along with, QuickBird remotely sensed data (e.g., LULC and NDVI geoparameters) and time series dependent, explanatorial, seasonal, meteorological variables could robustly forecast, seasonal, YVF-related, case distributions at the Gulu eco-epidemiological study site. As such, our objectives in this research were to; 1) Construct multiple robust stepwise linear and non-linear regression models employing seasonal-geosampled, georeferencable, empirically-dependent, observational, *Ae aegypti* geopredictive variables 2) Create a robust classification scheme to determine the trend, nature, rate, geolocation and magnitude of ecohydrologic, ecogeographic, geoclassified, LULC change 3) Remotely synthesize NDVI-related time series dependent, explanatorial, covariate parameter estimators   and multiple geomorphological, terrain-related,DEM-derived predictors; and, 4) Generate a robust C++ model using the probabilistic, regression-based observational time series regressors   and the remote geosampled clinical and field-related data along with geo-spatiotemporal meteorological and demographic estimators for forecasting yellow fever case distribution in Gulu, Uganda.

Methods:

2.1 Study site: Uganda lies between the eastern and western sections of Africa's Great Rift Valley. The country shares borders with Sudan to the north, Kenya to the east, Lake Victoria to the southeast, Tanzania and Rwanda to the south and the Democratic Republic of Congo (DRC) to the west. Whilst the landscape is generally quite flat, most of the country is over 1,000m (3,280ft) in altitude.

Mountainous regions include the Rwenzori Mountains that run along the border with the DRC, the Virunga Mountains on the border with Rwanda and the DRC, and Kigezi in the southwest of the country. An extinct volcano, Mount Elgon, straddles the border with Kenya.

The capital city, Kampala, lies on the shores of Lake Victoria, the largest lake in Africa and second-largest freshwater inland body of water in the world. Jinja, located on the lake, is considered to be the start point of the River Nile, which traverses much of the country.

The varied scenery includes tropical forest, a semi-desert area in the northeast, the arid plains of the Karamoja, the lush, heavily populated Buganda, the rolling savannah of Acholi, Bunyoro, Tororo and Ankole, tea plantations and the fertile cotton area of Teso.

Gulu District is a district in Northern Uganda. The district is named after its chief municipal, administrative and commercial center, the town of Gulu. The District is bordered by Lamwo District to the north, Pader District to the east, Oyam District to the south, Nwoya District to the southwest and Amuru District to the west. The district headquarters at Gulu are located approximately 340 kilometers (210 mi), by road, north of Uganda's capital city, Kampala. The coordinates of the district are: 02 45N, 32 00E.

As of May 2011, Gulu District is one of the seven districts that constitute the Acholi sub-region, the historical homeland of the Acholi ethnic group. Gulu District now consists of two counties: Achwa and Omoro. Kilak County has been converted to Amuru District, and Nwoya County is now Nwoya District. In the past, Kilak and Nwoya were counties in Gulu District. The economic activity of 90 per cent of the population in the district is subsistence agriculture. The

district of Gulu, as constituted in May 2011, had a population of about 298,500. The present population of the district is unknown due to extensive urban growth

Figure 1: Base map of the Gulu study site in northern Uganda

2.2 Remote sensing data: Raster image data from the DigitalGlobe QuickBird satellite service were acquired 15 July 2013, within the study site area, covering 64 km^2. The QuickBird imagery was classified using the Iterative Self-Organizing Data Analysis Technique (ISODATA) unsupervised routine in ERDAS *Imagine* v.8.7™ (ERDAS, Inc., Atlanta, Georgia). Unsupervised classifications are commonly used for the identification of sub-meter resolution-derived, explanatorial, seasonal, ecohydrologic, ecogeographic, LULC classes associated with prolific vector insect habitats based on geo-spatiotemporal-field-geosampled count data (Wood 1991a, Wood and Washino 1994).

2.3 Land cover mapping: Base maps were generated from the QuickBird visible and NIR data and differentially corrected Global Positioning Systems (DGPS) calibrated ground coordinates of a riceland village complex at the Gulu study site with all agro-forest canopy boundaries encompassed. The differentially-based ground coordinates were acquired from a CSI max receiver which has a positional accuracy of +/- .178 (http://www.omnistar.com). Using a local DGPS broadcaster can compensate for ionospheric and ephemeris effects which can improve horizontal accuracy significantly and can bring altitude error down in a geopredictive, vector, insect, habitat, endemic, transmission-oriented, forecasting, risk model (Jacob et al. 2007).

Each georeferenced, seasonal, ecohydrologic, ecogeographic, YFV –related,

274

explanatorial, operationizable, data feature attribute was entered into the VCMS™ relational database software product (Clarke Mosquito Control Products, Roselle, IL). VCMS commonly include connectivity with hand held computers and field data collection devices including DGPS receivers, PalmOS and Windows PocketPC handhelds which have been used for malaria, Eastern Equine Encephalitis Virus(EEEV), filarsis and WNV mosquito habitat monitoring (see Jacob et al. 2009, Jacob et al. 2008).

2.4 Grid-based algorithm: A digitized grid-based algorithm was then constructed in ArcGIS employing a mathematical algorithm in order to fit the continuous and bounded sampled surfaces at the Gulu study site from an ecological dataset of clinical, field and remote geosampled, empirical, georefernced, ecohydrologic, ecogeographic, time series, data feature attributes. Multiple data layers were then created using different coded values for various QuickBird imaged field attributes which were related to the same grid cell. Programs that manipulate grids access the spatial data by setting a rectangular window defined in map coordinates (www.esri.com).

Figure 2. Gridded QuickBird map of the Gulu study site
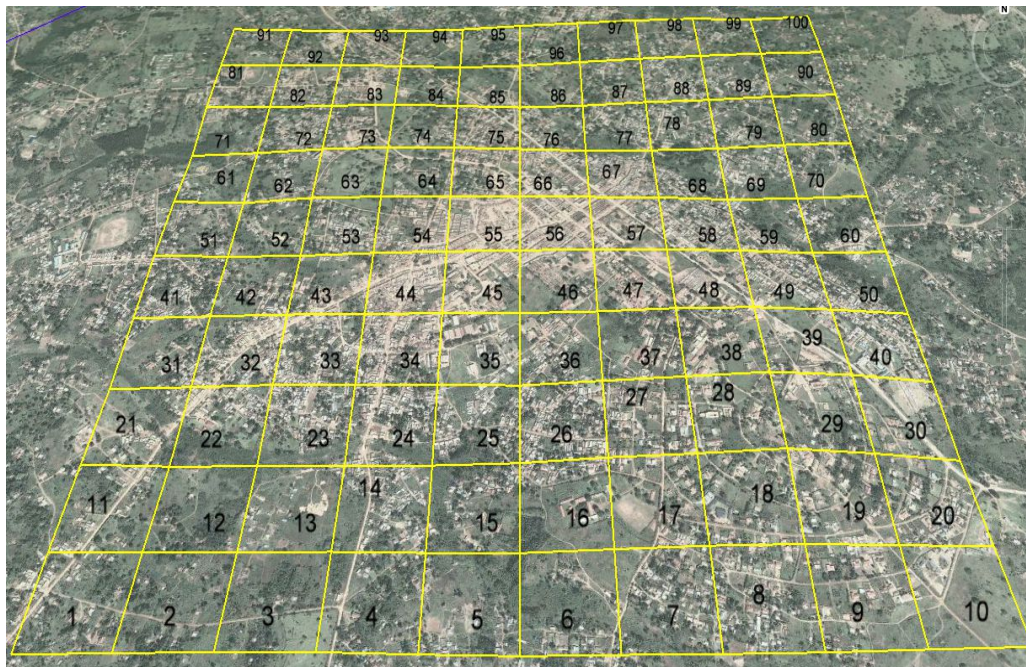
Each habitat grid cell /polygon was assigned a unique identifier. Field attribute tables were then linked to the polygons in ArcGIS. The georeferenced YFV –related polygons were used to define the sampling frame, which extended to include the external boundaries of the agro-village /forest canopied, georeferenced, boundary site. This allowed for multiple interactions enabling retrieval and transformation of the geosampled forest canopy and agro-village complex, explanatorial, LULC, data, feature attributes to be spatially separated efficiently in ArcGIS regardless of dimensionality of the geolocations within the study site.

2.5 Environmental parameters: Multiple, georeferenced, observational, informative predictors were also examined extensively using ArcGIS spatial analytical tools. The criteria involved the centrographic measures of spatial mean and distance between the time series dependent, geosampled, forest-canopy and agro-village boundaries and the georeferenced distance from the site to the nearest human habitation village complex at the eco-epidemiological study site. The data was also comprised of individual, geo-spatiotemporal-geosampled, empirical observations of the georeferenced eco-epidemiological, study site together with a battery of categorical explanatorial, field and remote specified attribute measures which were then expanded into multiple indicator geopredictive covariate parameter estimator coefficient estimates.

A LULC was performed using the QuickBird visible and NIR data in ArcGIS. Multiple LULC classes were employed including: agro-rice, forest canopy etc. The explanatorial LULC distances were then measured as Euclidean distances in the projection units of the raster and were computed within the digitized grid cell matrix. The Euclidean distance output raster contained the measured distances. The extracted Euclidean distance functions provided information according to Euclidean or, straight-line, distance between the georeferenced, forest canopy, LULC and rice-village complex LULC boundaries to the human habitation rice village complex units. All the geometric distances were quantitated in multidimensional space within the ArcGIS cyberenvironment server.

The Euclidean distances were computed as: distance (x,y) = { i (xi - yi)2 }½ in ArcGIS. Every cell in the Euclidean allocation output raster was assigned varying and constant, geopredictive, LULC values of the source to which it was closest (e.g., forest canopied LULC bordering an agro-village complex LULC). The Euclidean distance between ecological dataset of the clinical, field and remote geosampled, empirical, georefernced, YFV-related, data feature, attributes points p and q was the length of the line segment connecting them (Description: \overline{\mathbf{p}\mathbf{q}}). If p = (p1, p2,..., pn) and q = (q1, q2,..., qn) for any two geosampled, YFV-related, seasonal, endemic, transmission-oriented, georefernced points in Euclidean *n*-space, then the Cartesian coordinates the distance from p to q, or from q to p was given by: Description: \mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2 + (q_2-p_2)^2 + \cdots + (q_n-p_n)^2} = \sqrt{\sum_{i=1}^n (q_i-p_i)^2}. The position of a point in a Euclidean *n*-space is a Euclidean vector(Griffith 2003). So, p and q were Euclidean vectors, starting from the origin of the space in the empirical geosampled, explanatorial, seasonal, YFV-related data and their tips indicated two georefernced, LULC, YFR-related points. By so doing, the Euclidean norm, or Euclidean length, or magnitude of a vector measured the length of the vector using Description: \|\mathbf{p}\| = \sqrt{p_1^2+p_2^2+\cdots +p_n^2} = \sqrt{\mathbf{p}\cdot\mathbf{p}}where the last equation involved the dot product in the LULC model. A vector can be described as a directed line segment from the origin of the Euclidean space (vector tail), to a point in that space (vector tip)[Cressie 1993]. If we consider that its length is actually the distance from its tail to its tip, it becomes clear that the Euclidean norm of a vector is just a special case of Euclidean distance (i.e., the Euclidean distance between its tail and its tip). The distance between points *p* and *q* is a direction (e.g. from p to q), so it may be represented by another vector, given by Description: \mathbf{q} - \mathbf{p} = (q_1-p_1, q_2-p_2, \cdots, q_n-p_n) we assumed.

In a three-dimensional space (n=3), was an arrow from p to q, which we regarded as the position of q relative to p (i.e., a displacement vector if p and q represent two positions of the same point at two successive instants of time) in the explanatorial, clinical, field and remote geosampled, empirical, georefernced, YFV-related, eco-epidemiological, ecohydrologic, ecogeographic, forecasting, risk model.The Euclidean distance between p and q in the YFV-related risk model was just the Euclidean length of this distance (or displacement) vector: Description: Description: \|\mathbf{q} - \mathbf{p}\| = \sqrt{(\mathbf{q}-\mathbf{p})\cdot(\mathbf{q}-\mathbf{p})}.which is equivalent to equation 1, and also to: Description: Description: \|\mathbf{q} - \mathbf{p}\| = \sqrt{\|\mathbf{p}\|^2 + \|\mathbf{q}\|^2 - 2\mathbf{p}\cdot\mathbf{q}}.

The Euclidean distances were computed as: distance $(x,y) = \{ \sum_i (x_i - y_i)^2 \}^{1/2}$ in ArcGIS. Every cell in the Euclidean allocation output raster was assigned varying and constant geopredictive, explanatorial, LULC values of the source to which it was closest (e.g., forest, canopied, LULC bordering an agro-village complex LULC). The nearest source was determined by the Euclidean Distance function in ArcGIS. This function assigned ecogeographical space between the geoclassified LULCs. . The Euclidean Allocation function identified the nearest human habitation center closest to each grid cell. The Euclidean direction output raster contained the azimuth direction from each grid cell centroid to the nearest center.The distance between the geosampled and human habitation areas were categorized into specific georefernced ArcGIS Euclidean-distance, based measurement classes (e.g., 1: 0–5 km, 2: 5–10 km, and so on). For each ArcGIS mapping region, stratified sampling formulas were applied to estimate spectral, error proportions in the ecohydrologic, ecogeographic, LULC, vulnerability, eco-epidemiological, forecastimg. risk maps as in Stehman and Czaplewski (2000).

The estimates of overall and class-specific user's and producer's accuracy were calculated. The use of stratified formulas is important for validating field-sampling methods (Story and Congalton 1986) and has been used extensively for validating QuickBird-derived malaria, (Jacob et al. 2009), WNV (Jacob et al. 2010), and Eastern Equine Encephalitis Virus (EEEV) (Jacob et al. 2011), LULC-oriented risk maps. The *kappa coefficient* measures the agreement between classification and truth values. A kappa value of 1 represents perfect agreement, while a value of 0 represents no agreement. The kappa coefficient is computed as follows:

$$\kappa = \frac{N \sum_{i=1}^{n} m_{i,i} - \sum_{i=1}^{n}(G_i\, C_i)}{N^2 - \sum_{i=1}^{n}(G_i\, C_i)}$$

Where :

- **i** is the class number

- **N** is the total number of classified values compared to truth values

- **m$_{i,i}$** is the number of values belonging to the truth class **i** that have also been classified as class **i** (i.e., values found along the diagonal of the confusion matrix)

- **C$_i$** is the total number of predicted values belonging to class **i**

- **G$_i$** is the total number of truth values belonging to class **i**

In the example confusion matrix, the kappa coefficient is 0.99083

Figure 3. Land use land over map of the Gulu study site

Urban
Forrest
Riverside Vegetation
Agriculture
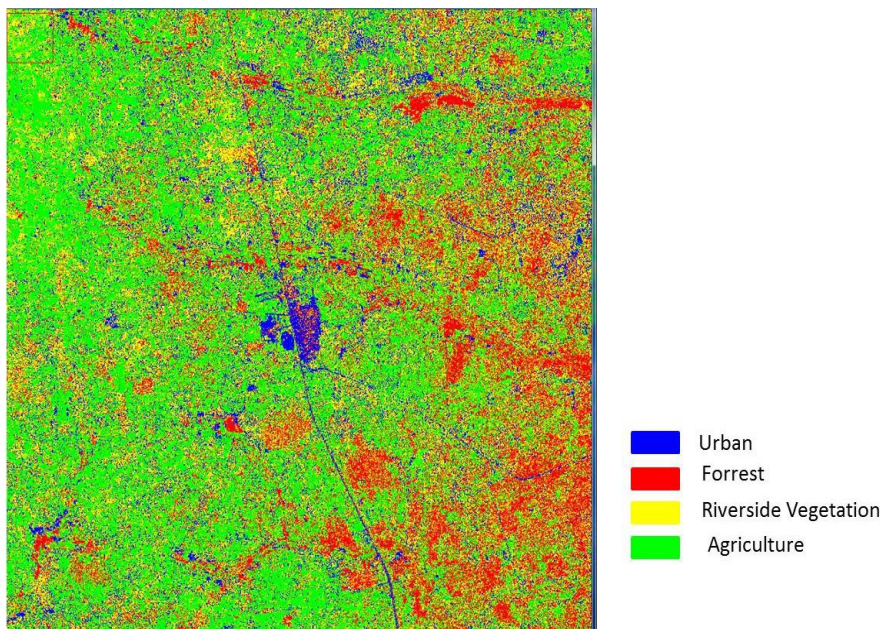
Accuracy results were thereafter computed by weighting the cell proportions by the proportion of each QuickBird classified LULC zone. Specifically, the overall accuracy ($\hat{P}$) and producer's accuracy ($\hat{P}_{Ai}$) were estimated using post-stratified formulas. We employed the post-stratified, geopredictive, time series dependent, clinical, field and remote-specified, georeferenced, explanatorial, covariate , parameter estimators as the known QuickBird pixel totals for each land-cover class (Ni+). The estimators were treated as a stratified random sample of ni+ mixels from the Ni+ mixels in that class, whereas user's accuracy of ($\hat{P}_{Ui}$) was based on the random sampling formulas:

$$(1)\ \hat{P} = \frac{1}{N} \sum_{k=1}^{q} \frac{N_{k+}}{n_{k+}} n_{kk}$$

$$(2)\ \hat{P}_{Ui} = n_{ii} / n_{i+}$$

$$(3)\ \hat{P}_{Aj} = \frac{\left(N_{j+}/n_{j+}\right)n_{jj}}{\sum_{k=1}^{q}\left(N_{k+}/n_{k+}\right)n_{kj}}$$

The producer's accuracy was then calculated to determine the probability that a reference geosampled, time series dependent, geopredictive, endemic, transmission-oriented, QuickBird-geoclassified, LULC correctly mapped and measured the errors of omission (1 - producer's accuracy). In contrast, the user's accuracy indicated the probability that a sample from the QuickBird, YFV-related, landscape ecohydrologic, ecogeographic vulnerability map actually matched the time series georeferenced data and measured the error of commission (1- use's accuracy). Kappa statistics were then calculated using SAS PROC FREQ.

Kappa statistics (κ) are appropriate for testing whether agreement exceeds chance levels for binary and nominal ratings for remotely-sensed models (Jensen 2005). The equation for κ was: $\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$, where $\Pr(a)$ was the relative observed agreement among the uncertainty based statistics and $\Pr(e)$ was the hypothetical probability of chance agreement between the empirically geosampled, clinical, field and remote, time series dependent, geopredictive, endemic, transmission-oriented, georeferenced, YFV-related, observational data feature attributes. If the explanatorial, geopredictive, residually, forecasted derivatives were in complete agreement then $\kappa = 1$ and there was no agreement among the classified YFV-related LULC data attributes other than what would be expected by chance, then $\kappa \leq 0$.

2.7 Vegetation Indices: The different modules in Spatial Analyst® extension of ArcGIS 10.2 and spatial modeller tools from ERDAS *Imagine* 9.1® were then used to perform VI calculations. NDVI was calculated using radiance, topographic, surface reflectance (*p*), or apparent reflectance (measured at the top of the atmosphere) values in the QuickBird red (0.63 to 0.69 μm) and NIR (0.76 to 0.90 μm) geospectral bands. The ratio of reflected radiance from the QuickBird red and NIR bands were then used to normalize illumination and topographic variation and to form the NDVI, which was used as an indicator of the amount and vigor of vegetation in the Gulu eco-epidemiological study site.

Initially, a sensitivity analysis was conducted prior to generating the time series-related, geopredictive, endemic, transmission-oriented, YFV-related, NDVI, observational parameters by analyzing the atmospheric and soil-perturbed responses as a continuous function of plant Leaf Area Index (LAI). LAI is a dimensionless quantity that characterizes plant canopies which defined as the one-sided green leaf area per unit ground surface area (e.g., LAI = leaf area / ground area, m^2 / m^2) in broadleaf canopies (Watson,1947). LAI is determined directly by taking a statistically significant sample of foliage from a plant canopy, measuring the leaf area per sample plot and dividing it by the plot land surface (Jensen 2005). Since canopy geometry relates directly to LAI, the indicator plays an essential role in theoretical production ecology (see Hay 2000).

Leaf Area Index was employed to generate photosynthetically active,geopredictive, endemic, transmission-oriented, explanatorily georeferenced, YFV-related, clinical, field and

remote-specified covariate, paramter estimator, coefficient estimates for determining vegetation-related regressors on specific classified LULC (e.g., rice plants on agro-village complex LULC) at the Gulu eco-epidemiological study site. We noted that there existed an inverse exponential relation between LAI and light interception which was established employing $P = P_{\max}\left(1 - e^{-c \cdot LAI}\right)$, where $P_{\max}$ designated the maximum primary production and c designated a crop specific growth coefficient (e.g., evapotranspiration).

Interestingly, physiological processes such as photosynthesis, transpiration, and evapotranspiration are related to LAI (Hay 1997). Randomly selected grid cell and geospectral, explanatorial, ecohydrologic, ecogeographic, time series-dependent, geopredictive, endemic, transmission-oriented measurements were then assessed to determine plant LAI in ArcMap®. Understanding the relation between geospectral response and LAI and gridded LULC data has allowed for the quantitation of canopy photosynthesis and evapotranspiration-related variables in a synoptic and repeatable fashion for identifying seasonally, productive, malaria, WNV and EEEV-related, mosquito, larval habitats based on geo- spatiotemporal, explanatorial, clinical, field and remote multivariate, geosampled count data (Jacob et al. 2010b, Jacob et al. 2009d).Estimations of LAI production at each time series-related, geopredictive, georeferenced, endemic transmission-oriented, YFV-related, LULC were conducted by correlation analysis with spectral reflectance ratio and measured radiance values. The best fitting waveband ratio among calculated reflectance and VI's were selected. Percent relative error and vegetation equivalent 'noise' (VEN) were calculated for soil and atmospheric influences at the Gulu epidemiological study site, separately and combined using LAI. We noted that the NDVI had a relative error of 10 percent and VEN of +/- 0.89 LAI.

We then performed Raster modeling in ArcGIS 10.2® which included performing image differencing on the NDVI layers, geoclassifying the layers into different LULC classes and calculating a wetness index using the Raster Calculator. The difference of the QuickBird visible and NIR bands was divided by their sum, which formed the functionally equivalent NDVI over the terrestrial surfaces of the study site. Sub-mixel ,time series-related, geopredictive, endemic, transmission-oriented, YFV-related, NDVI, geospectral variability (e.g. standard deviation of the QuickBird gridded agro-village and forest canopy LULC boundary radiance estimates) was then differentiated from other georefernced, land cover, explanatorial, observational, data, feature attributes. NDVI was computed directly without any bias or assumptions regarding plant physiognomy, explanatorial, YFV-related, land cover class, soil type, or climatic conditions, within a range from -1.0 to 1.0. The QuickBird visible and NIR reflectance, (*p*),we employed was as in Jacob et al (2010b)where multiple, time series dependent, Eastern Equine Encephalitis Virus (EEEV)-related, geopredictive, explanatorial, time series, dependent, endemic, transmission-oriented, covariate, paramter estimator coefficient values were robustly geospectrally quantized by the expression:

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR}\rho_{red}}$$

$$u_{cal}^2(NDVI) = \left(\frac{\partial NDVI}{\partial \rho_{NIR}}\right)^2 u_{cal}^2(\rho_{NIR}) + \left(\frac{\partial NDVI}{\partial \rho_{red}}\right) u_{cal}^2(\rho_{red}) + 2\frac{\partial NDVI}{\partial \rho_{NIR}}\frac{\partial NDVI}{\partial \rho_{red}} \bullet u_{cal}(\rho_{NIR}, \rho_{red})$$

$$\frac{\partial NDVI}{\partial \rho_{NIR}} = \frac{2\rho_{red}}{(\rho_{NIR} + \rho_{red})^2}$$

$$\frac{\partial NDVI}{\partial \rho_{red}} = \frac{-2\rho_{NIR}}{(\rho_{NIR} + \rho_{red})^2}$$

$$\frac{\partial NDVI}{\partial \rho_{NIR}}\frac{\partial NDVI}{\partial \rho_{red}} = \frac{-4\rho_{NIR}\rho_{red}}{(\rho_{NIR} + \rho_{red})^4}$$

Figure 5 Normalized Differnce Vegetation Index for the Gulu study site

To assess the accuracy of the explanatorial, ecohydrologic, ecogeographic, QuickBird-derived, YFV-related, seasonal, NDVI, thematic maps, a simple random sampling method was chosen to ensure geosampled time series selection. To provide a statistically efficient assessment of accuracy, a conservative geosampled size equation was also employed to calculate the sample sizes using a time series-dependent, geopredictive, explanatorial, clinical, field and remote specified, endemic, transmission-oriented covariate ,parameter estimator, coefficient, probabilistic, uncertainty-oriented, gridded matrix. The dataset of the DGPS points of the georeferenced YFV-related LULCs and their respective boundaries at the eco-epidemiological study site were then generated for each QuickBird thematic, YFV-related ,geopredictive, eco-epidemiological, vulnearbility seasonal, risk map. These maps were generated using the ISODATA algorithm to qualitatively regressively quantitate all the geospectrally-dependent, vegetation-oriented, NDVI and LULC temporally geosampled, geopredictive, endemic, transmission-oriented, covariate, parameter estimator, coefficient estimates associated to the

seasonal, georeferenced, YFV-related, empiricial, explanatorial, clinical, field and remote-specified data, feature attributes.

2.1.1. Conditional Expectation under Multivariate Normal (MVN) Distribution

A p-dimensional random vector $\underset{\sim}{X}$ was partitioned as $\underset{\sim}{X} = \begin{bmatrix} X_1 \ (q\times 1) \\ X_2 \ (p-q\times 1) \end{bmatrix}$, which had a multivariate normal distribution $N_p\left(\mu, \Sigma\right)$ with mean vector $\underset{\sim}{\mu} = \begin{bmatrix} \mu_1 \ (q\times 1) \\ \mu_2 \ (p-q\times 1) \end{bmatrix}$ and variance-covariance vector, $\Sigma = \begin{bmatrix} \Sigma_{11} \ (q\times q) & \Sigma_{12} \ (q\times p-q) \\ \Sigma_{21} \ (p-q\times q) & \Sigma_{22} \ (p-q\times p-q) \end{bmatrix}$, where $\underset{\sim}{X}_1$ and $\underset{\sim}{X}_2$ are two sub-vectors of dimensions q and p-q of $\underset{\sim}{X}$ respectively. We defined a transformation from $\left(X_1, X_2\right)$ to new geosampled YF discontinuous canopied and riceland *Ae.aegypti,* ovispoition, time series, explicative variables employing $\underset{\sim}{X}_1$ and $\underset{\sim}{X}'_2 = \underset{\sim}{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{X}_1$. This was achieved by linear transformation, $\begin{bmatrix} X_1 \\ X'_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{bmatrix}\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = A\underset{\sim}{X}$ (say) As any linear combination of $\underset{\sim}{X}$ is also MVN (Neter 1990) the linear transformation that that was $\left(X_1, X'_2\right)$ in the forecast, vulnerability, eco-epidemiological,,YFV model was jointly MVN distributed. Therefater we showed that $\underset{\sim}{X}_1$ and $\underset{\sim}{X}'_2$ are independent by proving that they were uncorrelated by:

$$Cov\left(\underset{\sim}{X}_1, X'_2\right) = Cov\left(\underset{\sim}{X}_1, X_2 - \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{X}_1\right)$$

$$= Cov\left(\underset{\sim}{X}_1, \underset{\sim}{X}_2\right) - Cov\left(\underset{\sim}{X}_1, \underset{\sim}{X}_1\right)\left(\Sigma_{21}\Sigma_{11}^{-1}\right)^T$$

$$= \Sigma_{12} - \Sigma_{11}\Sigma_{11}^{-1}\Sigma_{12} = 0$$

Since, $\underset{\sim}{X}_1$ and $\underset{\sim}{X}'_2$ are MVN variables and uncorrelated they are independent. Thus,

$$E\left(X'_2 | \underset{\sim}{X}_1 = \underset{\sim}{x}_1\right) = E\left(X'_2\right)$$

$$= E\left(X_2 - \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{X}_1\right)$$

$$= \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1$$

Now, as $\underset{\sim}{X}_2 = X'_2 + \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{X}_1$ the conditional distribution of $\underset{\sim}{X}_2$ given $\underset{\sim}{X}_1 = \underset{\sim}{x}_1$ is,

$$E\left(X_2 | \underset{\sim}{X}_1 = \underset{\sim}{x}_1\right) = E\left(X'_2 | \underset{\sim}{X}_1 = \underset{\sim}{x}_1\right) + \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{x}_1$$

$$= \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1 + \Sigma_{21}\Sigma_{11}^{-1}\underset{\sim}{x}_1$$

$$= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}\left(\underset{\sim}{x}_1 - \mu_1\right)$$

Similarly, $E(\bar{X}_1 | \bar{X}_2 = x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$.

We noted that the shape parameter in a Gamma distribution grews larger, when the distribution became more like a normal distribution (i.e. if X be a sampled riceland *Ae. aegypti,* oviposition, sub-meter resolution, geoclassified LULC) random variable following Gamma distribution with γ and λ as shape and scale parameters respectively and with

$$f(x|\gamma,\lambda)=\frac{\lambda^\gamma x^{\gamma-1}e^{-\lambda x}}{\lceil\gamma}; x,\lambda,\gamma>0$$

probability density function (pdf) as $\qquad$ If the shape parameter γ is large as compared to λ then Gamma distribution tends to normal distribution i.e.,

$$Gamma(\gamma,\lambda) \approx Normal\left[\left(\frac{\gamma}{\lambda}\right),\left(\frac{\gamma}{\lambda^2}\right)\right]$$

(Neter 1990).In case of lognormal distribution, if arithmetic mean m is much larger than its arithmetic standard deviations, then the distribution tends to Normal (m, s^2) (Hosmer and Lemeshew 2002). A general rule of thumb for this approximation in a entomological vector arthropod forecastd vulnerability signature paradigm is m > 6s. (see Jacob et al. 2013)We noted that X was a YFV explanatory LULC oviposition, eco-epidemioligical random variable following lognormal distribution with μ and σ as a capture point location and scale parameters respectively and with pdf as,

$$f(x|\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left[\frac{-1}{2\sigma^2}(\log x-\mu)^2\right]; x,\sigma>0 \& -\infty<\mu<\infty$$

The mean and standard deviation were defined as,

$$m=\exp\left[\mu+\frac{\sigma^2}{2}\right]$$ and, $$s=\left[e^{(2\mu+2\sigma^2)}-e^{(2\mu+\sigma^2)}\right]^{\frac{1}{2}}$$. Then, $m > 6s$,

$$Lognormal(\mu,\sigma) \approx Normal(m,s^2)$$.

Variable selection for the multiple regression models was carried out by a combination of automatic (i.e. stepwise) procedures and goodness-of-fit criteria in NLMIXED. We selected the explanatorial, time series-related, geopredictive, endemic, transmission-oriented, LULC, NDVI and other geosampled, , clinical, field and remotely eco-epidemiological covariate, parameter estimators that explained *Ae. aegypti*. prevalence in terms of vector, host, and parasite dynamics at the Gulu eco-epidemiological study site.

In the time series-related, YFV-related, NL MIXED, risk model, we let y_{ij} denote the number of covariate, time series dependent, parameter estimators employing $y_{ij}|\lambda_{ij}$ $\sim Poisson(\lambda_{ij}), \log\lambda_{ij}= \alpha_i + \beta_i(\log t_{ij} - \overline{\log t}) + e_{ij}$ and $e_{ij}|\sigma^2 \sim Normal(0,\sigma^2)$. The model specified different intercepts and slopes for each geosampled, geopredictive,explanatorial, clinical, field and remote specified, endemic, transmission-oriented, LULC data feature attribute. The procedure also estimated the random effects of the variables.

The corresponding NL MIXED statements were as follows:

```
proc nlmixed data= YFV parameters;
   parms logsig 0 beta1 1 beta2 1 alpha1 1 alpha2 1;
   if (group = 1) then beta = alpha1 + beta1*logtstd + e;
   else beta = alpha2 + beta2*logtstd + e;
   lambda = exp(beta);
   model y ~ poisson(lambda);
 random e ~ normal(0,exp(2*logsig)) subject=YFV parameters;
   estimate 'alpha1-alpha2' alpha1-alpha2;
   estimate 'beta1-beta2' beta1-beta2;
 run;
```

Probabilty Poisson regression with statistical significance was thereafter determined by a 95% confidence level. Table 1 show the regressors employed in the endemic, transmission-oriented, YFV, model in NL MIXED.

Table 1. **Environmental predictor variables sampled at the epidemiological capture point**

Variable	Description	Units
GCP	Ground Control Points	Decimal-degrees
LULC	Land Cover	meters
DEM	Catchment variable	(terrain statistics e.g.,Formazin Turbidity Unit)
Rainfall		millimeters
POP	Human Population	Discrete Integer
DISHAB	Distance between LULC	Meters
DISCAP	Distance between capture point and Epidemiological village	Meters

In the explanatorial, time series dependent, geopredictive, endemic, transmission-oriented ,YFV-related, Poissonized, eco-epidemiological, NL MIXED, explanatorial, LULC , sub-meter resolution, risk model it was assumed that the dependent variable Y, had a Poisson distribution given the independent variables X1, X2, ..., Xm, $P(Y=k| x1, x2, ..., xm) = e^{-\mu} \mu^{k} / k!$, k=0, 1, 2, ...,. In the NL MIXED model the log of the mean μ was assumed to be a linear function of the geosampled independent variables. That is, $\log(\mu) = \text{intercept} + b1*X1 + b2*X2 + ...+ b3*Xm$, implied that μ was the exponential function of the independent, geosampled, field and remote specified, YFV-related, explanatorial, geopredictive variables, where $\mu = \exp(\text{intercept} + b1*X1 + b2*X2 + ...+ b3*Xm)$. The Poisson regression model was then rewritten in the following form: $\log(\mu) = \log(N) + \text{intercept} + b1*X1 + b2*X2 + ...+ b3*Xm$, where *n* was the total number of the time series dependnet, empiricially regressed, YFV-related, endemic, transmission-oriented,

georefernced, explanatorial, georeferncable, data, feature, attribute variables at the Gulu eco-epidemiological study site.

Thereafter, the logarithm of the variable n was used as an offset, that is, a quantitative regression variable with a constant coefficient of 1 which represented each geosampled, independen, time series-related, predictive, endemic, transmission-oriented, YFV-related, explanatorial, clinic, field and remote specified, observational predictor. The log of the incidence, log (μ / n), was then modeled as a linear function of the time series-dependent independent variables. Thereafter, a maximum likelihood method was employed to estimate the covariate, parameter estimator, probability error rendered from the of regression model residuals in NL MIXED.

The parameter $\lambda_i(\mathbf{X}_i)$ was both the mean and the variance of the Poisson distribution for a specific geosampled, explanatorial, geopredictive, endemic, transmission-oriented, YFV-related LULCi in NL MIXED The data was log-transformed before the explanatorial, time series dependent, eco-epidemiolgical, probabilistically regressive quantitative data analyses to normalize the distribution and minimize residual standard error. The regression analyses assumed independent counts (i.e., n_i), taken at the georefernced, seasonally geosampled, LULC geolocations i=1, 2…n. The Poisson probability regression models assumed the response variable Y had a Poisson distribution and assumed the logarithm of its expected value was modeled by a linear combination of the time series-related, empirically regressed, geopredictive, explanatorial, endemic, transmission-oriented, YFV-related clinical, field and remote-specified, covariate, paramter estimator, coefficient values. This expression was written more compactly then as $\log(\mathrm{E}(Y|x))$ where x was an $n+1$-dimensional vector consisting of n independent variables concatenated to 1 and, thus, θ was simply a linearly linked to b.

Interestingly, in our time series-related, endemic, transmission-oriented, explanatorial, clinical, field and remote-specified, YFV-related, "Poissionzed"probabilistic, time series, eco-epidemiological, LULC risk model, θ was an input vector x and the geopredicted mean of the associated distribution rendered from the regressed covariate parameter estimator, coefficient estimates. This value was provided by $\mathrm{E}(Y|x) = e^{\theta'x}$ but, only if X ε R^n was a vector of the independent variables. Thereafter, the Poisson, time series-related, geopredictive, explanatorial, clinical, field and remote-specified, geosampled, endemic, transmission-oriented, YFV-related, eco-epidemiological, forecasting, operationizable, risk model took the form $\log(\mathrm{E}(Y|x)) = a'x + b$ where a ε R^n and b ε R. Positing salient error estimators using Poisson-derived regression estimates, the maximization of an auto-Gaussian log-likelihood function and a set of eigenvectors where lambda is the sub-space of R^n can identify and quantitate seasonal, entomological-related, observational, predictive, LULC covariate coefficients (Jacob et al. 2011b). The Gaussian distribution is a continuous probability distribution that is often employed as a first approximation to describe real-valued randomized variables that tend to cluster around a single mean value (Hosmer and Lemeshew 2000).

In our regression framework the explanatorial, forecasting, geosampled, time series-related, endemic, transmission-oriented, YFR-related, clinical, field and remote specified, covariate parameter estimator, empiricial coefficients were delineated by matrix $\mathbf{X}_i$, which was

constructed employing a $1 \times p$ vector of the geosampled measurement error indicator values. These values were based on resampled, LULC geolocations (i.e, i). The expected value of these data was given by: $\mu_i(\mathbf{X}_i) = n_i(\mathbf{X}_i) \exp(\mathbf{X}_i\boldsymbol{\beta})$ where, $\boldsymbol{\beta}$ was the vector of non-redundant parameter estimators and the Poissonized specified, time series dependent, geopredictive endemic, transmission-oriented, YFV-related predictors were then rendered by $\lambda_i(\mathbf{X}_i) = \mu_i(\mathbf{X}_i)/n_i(\mathbf{X}_i)$ (2.2).

Thereafter, the time series-dependent, geopredictive, endemic, transmission-oriented, YFV-related, Poisson probability regression LULC models were generalized by introducing an unobserved heterogeneous term for the geosampled observational, endmeic variables (i). Thus, the geo-spatiotemporal, geosampled, time series,predictive, YFV-related,LULC sub-meter resolution. data was assumed to differ randomly in a manner that was not fully accounted for by the time series, regressively quantized, explanatorial, clinical, field and remote specified covariate, paramter estimator error, uncertainty-oriented, coefficient estimates in NL MIXED. These distributions were then formulated as $E((y_i|x_i,\tau_i) = \mu_i\tau_i = e^{x_i^{\prime}\beta + \varepsilon_i}$, where the unobserved heterogeneity term $\tau_i = e^{\varepsilon_i}$ was independent of the vector of regressors x_i; thus, the distribution of y_i conditional on x_i and τ_i in the Poisson YFV-related eco-epidemiolgical, geopredictive, risk model was Poisson with a conditional variance of $\mu_i\tau_i$: $f(y_i|\mathbf{x}_i,\tau_i) = \dfrac{\exp(-\mu_i\tau_i)(\mu_i\tau_i)^{y_i}}{y_i!}$.

In our predictive, time series dependent, YFV-related, Poisson regression model construction process, we let $g(\tau_i)$ be the probability density function (pdf) of τ_i. By so doing, the distribution $f(y_i|x_i)$, was no longer conditional on τ_i in the $\mathbf{X}$ in the eco-epidemiological, forecasting, probabilsitic, risk model. Thereafter, the geo-spatiotemporal-geosampled, linearized, endemic, transmission-oriented, YFV-related model, explanatorial, clinical, field and remote specified, endemic, transmission-oriented, residually, empirically, forecasted derivatives were obtained by integrating $f(y_i|x_i,\tau_i)$ with respect to τ_i: $f(y_i|\mathbf{x}_i) = \displaystyle\int_0^{\infty} f(y_i|\mathbf{x}_i,\tau_i)g(\tau_i)d\tau_i$ .We noticed that the explanatorial, geopredictive, autoregressive, YFV-related, probabilistic, model error parameter estimators contained a constant term. As such, it was necessary to assume that $E(e^{\varepsilon_i}) = E(\tau_i) = 1$ in order to identify the mean of the distributions (see Jacob et al. 2009). We assumed that τ_i followed a gamma (θ, θ) distribution with $E(\tau_i) = 1$ and $V(\tau_i) = 1/\theta$: $g(\tau_i) = \dfrac{\theta^{\theta}}{\Gamma(\theta)}\tau_i^{\theta-1}\exp(-\theta\tau_i)$ where $\Gamma(x) = \int_0^{\infty} z^{x-1}\exp(-z)dz$ was the gamma function and θ was a positive geosampled time series-related,YFV-related, parameter estimator. Thus, the density of y_i in the time series-dependent,probabilistic, regression-based, eco-epidemiological, time series dependent, empirical, risk model was $\mathbf{X}_i$ which was further regressively quantitated employing the equation:

$$f(y_i|\mathbf{x}_i) = \int_0^{\infty} f(y_i|\mathbf{x}_i,\tau_i)g(\tau_i)d\tau_i = \frac{\theta^{\theta}\mu_i^{y_i}}{y_i!\Gamma(\theta)}\int_0^{\infty} e^{-(\mu_i+\theta)\tau_i}\tau_i^{y+y_i-1}d\tau_i = \frac{\Gamma(y_i+\theta)}{y_i!\Gamma(\theta)}\left(\frac{\theta}{\theta+\mu_i}\right)^{\theta}\left(\frac{\mu_i}{\theta+\mu_i}\right)^{y_i} = \frac{\theta^{\theta}\mu_i^{y_i}\Gamma(y_i+\theta)}{y_i!\Gamma(\theta)(\theta+\mu_i)^{\theta+y_i}}$$ (2.3)

We employed NL MIXED to perform the endemic, transmission-oriented, time series, Poisson regression analysis of the seasonal, YFV-related, eco-epidemiological, data feature

attributes with a log link function. As such, the LULC model was classified as a seasonal, log-linear, YFV-related, geopredictive, geopredictive, eco-epidemiological, risk model. The log-linear model is a mathematical model that takes the form of a function whose logarithm is a first-degree polynomial function of the parameters of the model, which makes it possible to apply multivariate linear regression (see Hosmer and Lemeshew 2000). The specific applications of log-linear seasonal, entomological-related, geopredictive, endemic, transmission-oriented, linear, eco-epidemiolgical, risk models are robust when the output quantity lies in the range 0 to ∞, for geosampled explanatorial, time series dependent, covariate, parameter estimator, coefficient values of the independent variables X, or more immediately, the transformed quantities $f_i(X)$ in the range $-\infty$ to $+\infty$(see Jacob et al. 2013, Jacob et al. 2012). This may be contrasted to explanatorial, geopredictive, logistic, eco-epidemiological, forecasting, risk models, similar to the logistic function, for which the output quantity lies in the range 0 to 1.

Interestingly, in Jacob et al. (2005b), pseudo R^2 values from a time series dependent, backward, stepwise, Poisson –related, geopredictive, explanatorial, endemic, transmission-oriented, risk model generated using multiple time series dependent, clinical, field and remote specified *An. gambiae s.l.* larval habitat explanatorial covariate, paramter estimator, coefiicnet values in GEN MOD revealed greater variation than pseudo R^2 estimates from a logistic regression model. Multivariable risk analyses were performed with the GENMOD procedure of SAS 8.01 (SAS Institute, Cary, NC) using the geosampled field and remote specified datasets. Thereafter, Poisson and logistic regression models were generated with the counts of *Anopheles* mosquitoes for the aquatic larval habitat datasets.

Interestingly, in Jacob et al. (2005b), the geosampled explanatorial datasets from the Kisumu eco-epidemiological study site did not contain many observational explanatorial, predictors with positive anopheline immature counts (i.e., zero-inflated data). The model for the Malindi study site was deemed general enough to make time series dependent, explanatorial geopredictions. Unfortunately, the authors were not able to validate the combined model. The model was over parameterized for the validation dataset (i.e., it had too many georeferenced predictors for the small sample size in the validation set). To test for collinearity, the authors employed the design matrix from a Poisson regression and ran it through a GEN MOD procedure, which indicated the absence of problematic correlation among the malarial observational predictors. Regression assumes that the predictor variables are noncollinear (Hosmer and Lemeshew 2000).

We employed methods of moment to estimate, unknown, YFV-related, endemic, transmission-oriented, predictive, LULC, sub-meter resolution parameter estimators, geo-spatiotemporally geosampled, at the Gulu eco-epidemiolgecal study site [i.e., $\theta_1, \theta_2, \ldots, \theta_k$] specifically for characterizing the distribution $f_W(w; \theta)$ of the random variable W. Two quantitate the first k moments of the true distribution (e.g., the geosampled, endemic, transmission – oriented, LULC population moments) were then expressed as functions of the θs: $\mu_1 \equiv E[W^1] = g_1(\theta_1, \theta_2, \ldots, \theta_k), \mu_2 \equiv E[W^2] = g_2(\theta_1, \theta_2, \ldots, \theta_k)$ and $\mu_k \equiv E[W^k] = g_k(\theta_1, \theta_2, \ldots, \theta_k)$.   By so doing, a sample of size n was drawn, resulting in the covariate parameter estimator clinical, field and remote-geosampled values $w_1, \ldots, w_n$. For $j - 1, \ldots, k,$ let $\hat{\mu}_j = \frac{1}{n} \sum_{i=1}^{n} w_i^j$ be the j-th sample moment, an estimate of μ_j (Hosmer and Lemeshew 2000) The method of moment's estimator for

$\theta_1, \theta_2, \ldots, \theta_k$ was then denoted by $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ which in turn was defined as the solution) to the equations: $\hat{\mu}_1 = g_1(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k), \hat{\mu}_2 = g_2(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ and $\hat{\mu}_k = g_k(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$.

The method of moments is fairly simple and yields consistent estimators (Griffith 2003). The estimating geopredictive, explanatorial, probabilsitic, YFV-related, endemic, transmission-oriented, covariate, LULC parameter estimators of a known family of probability distributions was parsimoniously quantitated. This method was superseded by Fisher's method of maximum likelihood as MLEs have higher probability of being close to unbiased estimated quantities in seasonal, clinical, field and remote specified, entomological –related, endemic, transmission-oriented, eco-epidemiological, probabilistic, risk model, residually, forecasted derivatives (Jacob et al. 2005b, Griffith 2005). As such, we employed a sample $x_1, x_2, ..., x_n$ of n independent and identically distributed (i.d.d.)geosampled, empirical, YFV-related, explanatorial, time series dependent, clinic, field and remote specified observations, quantitated from a distribution with an unknown pdf $f_0(\cdot)$.However, since it has been surmised that the function f_0 belongs to a certain family of distributions [i.e., $f(\cdot \mid \theta)$, $\theta \in \Theta$] where θ is a vector of parameters (i.e., the parametric model), $f_0 = f(\cdot \mid \theta_0)$ in the YFV risk model.The value θ_0 is unknown and is referred to as the true value of the parameter (Fotheringham 2002). It was desirable for us to qualitatively regressively probablistically quantitate the geosampled, geosampled, explanatorial, clinical, field and/or remote specified endemic, transmission-oriented, YFV-related, covariate, parameter estimator coefficient values (i.e., $\hat{\theta}$) since we assumed this would enable determining a true value θ_0 in the latent forecasted derivatives. Both the observed variables x_i and the parameter θ can be vectors in a seasonal, geopredictive, entomological-related, endemic, transmission-oriented, eco-epidemiological, robust, forecasting, risk model (Jacob et al. 2005b, Griffith 2005).

We employed the method of maximum likelihood to specify the joint density function for all the seasonal, geosampled, georeferenced, endemic, transmission-oriented, YFV-related explanatorial,, clinical, field and remote specified observations. Probability density function is most commonly associated with absolutely entomological-related, continuous univariate distributions (Jacob et al. 2005b). A random variable X has density f_X, where f_X is a non-negative Lebesgue-integrable function, if: $\Pr[a \leq X \leq b] = \int_a^b f_X(x)\, dx.$ In mathematics, the integral of a non-negative function can be regarded in the simplest case as the area between the graph of that function and the x-axis. Lebesgue integration is a mathematical construction that extends the integral to a larger class of functions; it also extends the domains on which these functions can be defined (Hosmer and Lemeshew 2000) Hence,we assumed if F_X would be the cumulative distribution function of X in a time series dependent, YFV-related, LULC endemic, transmission oriented, geopredictive, eco-epidemiological, clinical, field or remote, geosampled, risk model when $F_X(x) = \int_{-\infty}^{x} f_X(u)\, du.$ Further, if f_X was continuous at x in the model then $f_X(x) = \dfrac{d}{dx} F_X(x).$ Intuitively, a ecologist, entomologist or orther experimenter would think of $f_X(x)\, dx$ as being the probability of X falling within the infinitesimal interval $[x, x + dx]$ in the YFV model. For effectively qualitatively regressively quantitating an i.d.d. explanatorial, geo-spatiotemporally-geosampled, clincial, field and/or remote specified time series ,YFV-related geosamples, the joint density function was determined by employing $f(x_1, x_2, \ldots, x_n \mid \theta) = f(x_1 \mid \theta) \times f(x_2 \mid \theta) \times \cdots \times f(x_n \mid \theta).$

Thereafter, we examined the endemic, transmission-oriented, geopredictive, YFV-related, explanatorial, clinical, field and remote specified, observational, function from a different perspective by considering the seasonal endmeic, geosampled values $x_1, x_2, ..., x_n$ to be fixed "parameters" of this function, As such, θ was the function's variable which was allowed to vary freely; this function was the endemic, YFV-related, transmission-oriented, eco-epidemiological, probabilsitic, risk model likelihood estimate which was quantitated by $\mathcal{L}(\theta \,|\, x_1, \ldots, x_n) - f(x_1, x_2, \ldots, x_n \,|\, \theta) - \prod_{i=1}^{n} f(x_i | \theta).$ Note, that the vertical bar in $\mathcal{L}(\theta \,|\, x_1, \ldots, x_n)$ did not mean "conditional" in the risk model construction process. Instead, the vertical denoted a separation between the two input arguments: θ and the vector-valued input $x_1, \ldots, x_n$ in the eco-epidemiological, geo-spatiotermpoal, empirically probabilistcally regressed, YFV-related, forecasting, risk model. In practice it is often more convenient to work with the logarithm of the likelihood function, called the log-likelihood: $\ln \mathcal{L}(\theta \,|\, x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i | \theta),$ or the average log-likelihood: $\hat{\ell} = \dfrac{1}{n} \ln \mathcal{L}$ (see Griffith 2003). The hat over ℓ indicated that it was akin to some YFV-related parameter estimator. Indeed, $\hat{\ell}$ estimated the expected log-likelihood of any single, LULC-related, sub-meter resolution, observational predictor geosampled at the Gulu eco-epidemiological study site. The method of maximum likelihood estimates θ_0 in a robust, geopredictive, endemic, transmission-oriented, entomological-related, probabilstic, eco-epidemiological, risk model by finding a value of θ that maximizes $\hat{\ell}(\theta | x)$ (Jacob et al. 2005b, Griffith 2005). For our geopredictive, clinical, field and remote specified, explanatorial, *Ae.aegypti* endemic, transmission-oriented, YFV-related, risk model estimation we defined a maximum-likelihood estimator (MLE) of θ_0 ... $\{\hat{\theta}_{\mathrm{mle}}\} \subseteq \{\arg\max_{\theta \in \Theta} \hat{\ell}(\theta \,|\, x_1, \ldots, x_n)\}.$ An MLE is an estimated value regardless of whether it maximizes the likelihood or the log-likelihood function, since log is a monotonically increasing function(Hosmer and Lemeshew 2000).

In mathematics, a monotonic function or monotone function is a function between ordered sets that preserves the given order. This concept first arose in calculus, and was later generalized to the more abstract setting of order theory. Order theory is a branch of mathematics which investigates our intuitive notion of order using binary relations (Hosmer and Lemeshew 2000). In calculus, a function f defined on a subset of the real numbers with real values (e.g., an emprical dataset of explanatorial, clincial, field or remote geosampled YFV-related probabilistic regressors) is called monotonic if it is either entirely non-increasing or non-decreasing. It is called monotonically increasing (also increasing or non-decreasing), if for all x and y such that $x \leq y$ one has $f(x) \leq f(y)$, so f preserves the order (Davey, and Priestley 2002). Likewise, a function is called monotonically decreasing (also decreasing, or non-increasing) if, whenever $x \leq y$, then $f(x) \geq f(y)$, so it reverses the order in an entomological-related, endemic, transmission-oriented, explanatorial,forecasting, LULC risk model (Jacob et al. 2014). If the order $\leq$ in the definition of monotonicity is replaced by the strict order $<$, then one obtains a stronger requirement. A function with this property is called strictly increasing (order (Davey, and Priestley 2002). Again, by inverting the order symbol in an entomological-related, probabilistic, endemic, transmission-oriented, eco-epidemiological, risk model, a ecologist, entomologist or experimenter would find a corresponding concept strictly decreasing. Functions that are strictly increasing or decreasing are one-to-one (because for x not equal to y, either

$x < y$ or $x > y$)and so, by monotonicity, either $f(x) < f(y)$ or $f(x) > f(y)$, thus $f(x)$ would not be equal to $f(y)$ in the YFV risk model specified explanatorial clinical, field and/or remote covariate, paramter estimator coefficients. When functions between discrete sets are considered in combinatory, it is not always obvious that "increasing" and "decreasing" are taken to include the possibility of repeating the same value at successive arguments, so a ecologist, entomologist or other experimenter would find the terms weakly increasing and weakly decreasing to stress this possibility. Jacob et al. (2014) employed the definability of combinatorial functions and their linear recurrence relationships within a polylogarithmic triangularizable matrix employing subjective bilipschitz functions and other isomorphism of metric spaces for forecasting seasonal endemic onchocerciasis transmission zones in Burkina Faso.

Importantly, for regressing an empirical geosampled dataset of seasonal, explanatorial, vector, entomological-related, clinical, field and/or remote specified, geopredictive, endemic, transmission-oriented, forecasting risk modeling, covariate, parameter estimator coefficients, the terms "non-decreasing" and "non-increasing" should not be confused with the weaker negative qualifications (i.e., "not decreasing" and "not increasing"). For instance, the function of an explanatorial, YFV –related, probabilistic, time series dependent, endemic, transmission-oriented, eco-epidemiological, forecasting, risk model may first fall, then rise, then fall again. The explanatorial, clinical, field and/or remote specified parameter estimators therefore may not decrease nor increase in their statistical significance which may be neither non-decreasing nor non-increasing. The term monotonic transformation can also possibly cause some confusion as it refers to a transformation by a strictly increasing function in a robust, entomological-related, explanatorial, time series dependent, eco-epidemiological, forecasting, uncertainty-oriented, risk model framework. Notably, this is the case in some entomological risk model with respect to the ordinal properties of a utility function being preserved across a monotonic transform (see Jacob et al. 2012). A function $f(x)$ is said to be absolutely monotonic over an interval (a, b) in an entomological-related, endemic, trasnmission-oriented, probabilsitic, forecasting, eco-epidemiological, risk, model if the derivatives of all orders of f are nonnegative at all points (georefrenced time series dependnet, YFV-related clinical, field or remote geosampled explanators) on the interval (Jacob et al. 2014). By extension the vertical bar in $\hat{\ell}(\theta \mid x_1, \ldots, x_n)$ would not be then conditional in an eco-epidemiological probabilistic dataset of YFV-related, risk model, endemic, transmission-oriented, latently, ecogeographically forecasted, discontinuous forest-canopied, African riceland , LULC covariate derivatives.

Interestingly, in the exposition above, it may be assumed that the seasonal, entomological-related, explanatorial, clinical, field and/or remote specified, geopredictive, endemic, transmission-oriented, forecasting, risk-related YFV data are i.d.d. We assumed the method could be applied to a broader setting, as long as it was possible to write the joint density function $f(x_1, \ldots, x_n \mid \theta)$ in the eco-epidemiological, risk model where its parameter θ had a finite dimension that was not dependent on the sample size n. By so doing, we assumed an allowance could be made for qualiatively regressively quantizing data heterogeneity in an explanatorial, YFV-related, geopredictive, eco-epidemiological, forecasting, risk model, so that the joint density was equal to $f_1(x_1 \mid \theta) \cdot f_2(x_2 \mid \theta) \cdot \cdots \cdot f_n(x_n \mid \theta)$. Put another way, we assumed that each geosampled YFV-related, geosampled clinical, field or remote specified observation (x_i) came from a randomized probabilistic variable that had its own distribution function f_i.

The NLMIXED procedure fit multiple explanatorial, geopredictive, YFV-related, mixed, risk models where the conditional mean function was a general nonlinear function. The class of generalized linear mixed models is a special case of the nonlinear mixed models; hence some of the models can be fit with NLMIXED and the GLIMMIX procedure (www.sas.edu). Our NLMIXED procedure relied by default on approximating the marginal log likelihood through adaptive Gaussian quadrature. In the GLIMMIX procedure, MLE by adaptive Gaussian quadrature was available with the METHOD=QUAD option in the PROC GLIMMIX statement. The default estimation methods differentiated between the NLMIXED estimated, YFV-related, time series dependent, endemic, transmission-oriented, explanatorial, clinical, field or remote geosampled, covariate parameter estimator coefficient regressive quantification procedures as adaptive quadrature was available with the GLIMMIX procedure. If you choose METHOD=LAPLACE or METHOD=QUAD(QPOINTS=1) in the PROC GLIMMIX statement for a generalized linear mixed model, the GLIMMIX procedure performs MLE-based on a Laplace approximation of the marginal log likelihood (www.sas.edu). In our time series dependent, explanatorial, geopredictive, YFV-related, endemic, transmission-oriented, eco-epidemiological, probabilsitic, regression-related, risk model this approximation was equivalent to the QPOINTS=1 option in the NLMIXED procedure.

The time series dependent, geopredictive, explanatorial, YFV-related, regression model assumed independent Bernoulli outcomes denoted by Y_i as the dependent variable taken at the geosampled explanatorial, LULC sites $i = 1, 2, \ldots, n$. The estimator measurement indicator values were thereafter described by X_i, a 1-by-$(K+1)$ vector of K geo-spatiotemporal, geosampled, YFV-related, endemic, transmission-oriented, explanatorial, clinical, field and remote specified, covariate parameter estimator, coefficient values where a 1 for the intercept term represented a geosampled LULC site geolocation i. The probability of a 1 being realized for the binary outcome data was provided by: $P(Y_i = 1| X_i) = \exp(X_i\beta) / [1 + \exp(X_i\beta)]$ (2.1) where β was the $(K+1)$-by-1 vector of non-redundant parameters and $P(Y_i = 0| X_i) = 1 - P(Y_i = 1|X_i)$.

We employed the simplest form of Equation (2.1) for qualitatively assessing and geo-spatiotemporally quantizing constant probability across the multiple, randomized, time series dependent, YFV-related, endemic, transmission-oriented, covariate, paramter estimator coefficient values employing [i.e., $P(Y_i = 1| X_i) = P(Y_i = 1| \alpha)$ geosampled $= \exp(\alpha)/[1 + \exp(\alpha)]$. By so doing, the eco-epidemiological, probabilistic, risk model rendered a constant α using a bivariate regression notation. The statistical procedure was performed by denoting β_0, where $P(Y_i = 1| \alpha) \to 0$ as $\alpha \to -\infty$, $P(Y_i = 1| \alpha) \to 0.5$ as $\alpha \to 0$, and $P(Y_i = 1| \alpha) \to 1$ as $\alpha \to \infty$ in the multivariate, residually forecasted, time series dependent, geopredictive, YFRLULC-oriented, regression-related, forecasting, risk model uncertainty-oriented, matrix framework.

The logistic function or logistic curve is a common special case of the more general sigmoid function, with equation: $f(x) = \dfrac{1}{1 + e^{-x}}$ where e is Euler's number (Marsden 1985). For qualitatively regressively quantitating seasonal-geosampled values of x in the range of the empirical geosampled dataset of explanatorial, clinical, field and remote specified, YFV-related, geopredictive, covariate, parameter estimator, coefficient values from $-\infty$ to $+\infty$, the S-curve had to be obtained. A sigmoid function is a mathematical function having an "S" shape (sigmoid curve) (Hosmer and Lemeshew 2000). In our seasonal, probabilsitic, YFV-related, eco-

epidemiological, linear model, the sigmoid function was the logistic function which was defined by the formula $S(t) = \dfrac{1}{1 + e^{-t}}$. In general, a sigmoid function is real-valued and differentiable, having either a non-negative or non-positive first derivative which is bell shaped (Hosmer and Lemeshew 2000).

Interestingly, in our seasonal, explanatorial, eco-epidemiological, clincial field and remote specified, YFV-related, geopredictive, linearized, forecasting, risk model there were also a pair of horizontal asymptotes as $t \rightarrow \pm\infty$. The differential equation $\frac{d}{dt} S(t) = c_1 S(t)(c_2 - S(t))$, with the inclusion of a boundary condition provided a third degree of freedom, c_3, which in turn provided a class of functions. We noted that the number e was an important mathematical constant in the geopredictive, YFV-related, eco-epidemiological, probabilistic, risk model which was the base of the natural logarithm in the model. It was approximately equal to 2.71828, and had the limit of $(1 + 1/n)^n$ as n approached infinity. Commonly, E can a be calculated as the sum of the infinite series $e = \sum\limits_{n=0}^{\infty} \dfrac{1}{n!} = 1 + \dfrac{1}{1} + \dfrac{1}{1 \cdot 2} + \dfrac{1}{1 \cdot 2 \cdot 3} + \cdots$ (see Arflen1985).

Importantly, mathematic constants can be defined in seasonal infectious disease-related, geopredictive, explanatorial, forecasting, risk models in multiple ways. For instance, e may be a unique real geosampled, seasonal, YFV-related, covariate, parameter estimator, coefficient value, for instance such that the value of the derivative (i.e., slope of the tangent line) of the function $f(x) = e^x$ at the point $x = 0$ is equal to 1(see Jacob et al. 2012, 2009d). The function e^x so defined would then be the exponential function in an endemic, transmission-oriented, probabilistic, regression-related, eco-epidemiological, explanatorial, clinical, field or remote geosampled, risk model and its inverse would be the natural logarithm, or logarithm to base e. The natural logarithm of a positive number k can then be defined in a robust, entomological,-related, geopredictive, explanatorial, risk model constructed in ArcGIS, directly as the area under the curve $y = 1/x$ between $x = 1$ and $x = k$, in which case, e would be the number whose natural logarithm is 1 (see Jacob et al. 2012). Like the constant π, e is irrational: it is not a ratio of integers; and it is transcendental: it is not a root of any non-zero polynomial with rational coefficients (Hosmer and Lemeshew 2000).

Our geopredictive, time series dependent,YFV-related,explanatorial,endemic, transmission-oriented, regression model had the following general form $\exp\left(c + \sum\limits_{i} w_i f_i(X)\right)$, in which the $f_i(X)$ were quantities that were functions of the geosampled LULC and other explanatorial variables X,(i.e.,   a vector of YFV-related, covariate, parameter estimator coefficient values), while c and the w_i were the model parameters. We employed a log-linear plot or graph, which was a type of semi-log plot to determine all linear relationships between the YFV-related dependent variable and all the independent variables

A semi-log graph was then generated in MathLab employing the seasonal, geosampled, clincial, field and remote specified, YFV-related, explanatorial, predictor variables in order to define any probabilsitic, exponential relationships in the geosampled empirical dataset. One axis was plotted on a logarithmic scale. A logarithmic scale is a scale of measurement that displays

the value of a physical quantity using intervals corresponding to orders of magnitude, rather than a standard linear scale (Hosmer and Lemeshew 2000). Deviating from any units in a seasonal, geopredictive, entomological- related, time series dependent, eco-epidemiological, geopredictive, probabilsitic, risk model means, the logarithmic measure will change by an additive constant (see Jacob et al. 2005b).The base of the logarithm in the time series dependent, YFV-related, probabilistic, geopredictive, eco-epidemiological, clincial, field and remote-geosampled, forecasting risk model had to be specified as the scale's value which we considered to be a dimensional quantity that was expressed in the model in indefinite-base logarithmic units.

Thereafter, a semilogx graph in ArcGIS plotted the seasonal, geosampled, georefernced, YFV related, endemic, transmission-oriented, empirical, explanatorial, covariate, parameter estimator, coefficients employing logarithmic scales for the *x*-axis. In seasonal, geopredictive, entomological-related data analyses, a semi-log graph or semi-log plot is a way of visualizing data that are related according to an exponential relationship (Jacob et al. 2013).A robust time series dependent, YFV-related semilogx(Y) was then created in ArcGIS employing a plot and a base 10 logarithmic scale for the *x*-axis and a linear scale for the *y*-axis. ArcGIS then plotted the columns of Y versus their index since Y contained the geosampled, time series dependent, YFV-related, covariate parameter estimator coefficient values. Semilogx(X1,Y1,...) plotted all Yn versus Xn pairs in the time series dependent, YFV-related, seasonal dataset. If only one of Xn or Yn is a matrix, semilog x plots the vector argument versus the rows or columns of the matrix, along the dimension of the matrix whose length matches the length of the vector (www.sas.edu). Further, if the matrix is square, its columns plot against the vector if their lengths match.

Thereafter, semilogx(X1,Y1,LineSpec,...) plotted all lines defined by the Xn,Yn, *LineSpec* triples. LineSpec determined line style, marker symbol, and color of the plotted lines semilog x sets property values for all line series properties graphics objects created by semilogx. Importantly, h = semilogx(...) returned a vector of handles to regressively quantitate the YFV-related, time series dependent, geopredictive, line series graphics objects, one handle per line. A semi-log plot is useful for time series geopredictive, entomological-related, seasonal models when one of the variables being plotted covers a large range of geosampled covariate, parameter estimator, coefficient values and the other has only a restricted range – the advantage being that it can bring out features attributes in the data that would not easily be seen if both variables have been plotted linearly (see Jacob et al. 2005b).

We noted that all the time series-related, YFV-related, geopredictive, explanatorial, endemic, transmission-oriented regression-related, time series equations of the form $y = \lambda a^{\gamma x}$ formed straight lines employing the geosampled LULC and other remotely, explanatorial,time series dependent, covariate parameter estimator coefficient values. These coefficients were when plotted semi-logarithmically, since taking logs of both sides rendered $\log_a y = \gamma x + \log_a \lambda$.This was geo-visualized in our ArcGIS time series graphs as a line in slope-intercept form where γ was the slope and $\log_a \lambda$ was the vertical intercept. To facilitate use with logarithmic tables in ArcGIS for constructing robust, seasonal, vector arthropod-related, eco-epidemiological, geopredictive, risk models, logs to base 10 or e, or sometimes base 2: $\log(y) = (\gamma \log(a))x + \log(\lambda)$ may be employed (see Jacob et al. 2005b).The term log-lin was

293

used to describe a semi-log plot with a logarithmic scale on the *y*-axis, and a linear scale on the *x*-axis in our risk model. The log-lin type of a semi-log graph, defined by a logarithmic scale on the *y*-axis, and a linear scale on the *x*-axis (Anselin 1995). Likewise, a lin-log graph employed a logarithmic scale on the *x*-axis, and a linear scale on the *y*-axis. Note, that our seasonal geosampled, ArcGIS-oriented, probabilistic, time series dependent, YFV-related graph was in the form of output-input (y-x), the opposite order from (*x*, *y*). Thereafter, a regression equation for a line with an ordinate axis logarithmically scaled the seasonal, YFV-related, operational, ecoepidemiological, forecasting risk model was subsequently expressed as $\log_{10}(F(x)) = mx + b$ $F(x) = 10^{mx+b} = (10^{mx})(10^{b})$. The equation of a line on a plot in the risk model where the abscissa axis was scaled logarithmically was then $F(x) = m\log_{10}(x) + b$. Interestingly, we noted that in our ArcGIS semi-log graph, the spacing of the scale on the *y*-axis in the time series dependent, geopredictive, YFV-related eco-epidemiological, risk model was proportional to the logarithm of the number, not the geosampled, explanatorial, covariate, paramter estimator coefficient value itself. It was thus equivalent to converting the *y* values to their log, and plotting the geosampled, explanatorial, YFV-related data on lin-lin scales. A log-log graph uses the logarithmic scale for both axes, and hence is not a semi-log graph (Anselin 1995).

The Poisson process in our explanatorial, probabilistic, YFV-related, LULC, geo-spatiotemporal, geosampled, eco-epidemiological, geopredictive, risk analyses was provided by the limit of a binomial distribution based on the temporally geosampled, YFV-related, explanatorial, covariate, parameter estimator coefficient estimates using

$$P_p(n \mid N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n}.$$

(1.2). We viewed the distribution as a function of the expected number of seasonal, YFV-related, geopredictive, geoclassifed, LULC, count-related, time series dependent, regressable variables geosampled employing the sample size *N* for rigriously quantifying the fixed *p* in equation (2.1), which then was thereafter log- transformed into the linear equation:

$$P_{v/N}(n \mid N) = \frac{N!}{n!(N-n)!} \left(\frac{v}{N}\right)^n \left(1 - \frac{v}{N}\right)^{N-n},$$

Based on the sample size *N*, the distribution approached $P_v(n)$ was then equivalent to

$$\lim_{N\to\infty} P_p(n \mid N) = \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{n!} \frac{v^n}{N^n}\left(1-\frac{v}{N}\right)^N\left(1-\frac{v}{N}\right)^{-n} = \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{N^n} \frac{v^n}{n!}\left(1-\frac{v}{N}\right)^N\left(1-\frac{v}{N}\right)^{-n}$$

which was then calculated as

$$1 \cdot \frac{v^n}{n!} \cdot e^{-v} \cdot 1 = \frac{v^n e^{-v}}{n!},$$

We assumed that $Y_i \sim \text{Poi}(\mu_i, \theta)$ where we let the mean μ_i for the *i*th observation varied as a function of the LULC covariate paramter estimators for that observation. Because the mean $\mu_i > 0$, it is natural to model $\mu_i = \exp(\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i})$. Generalizing, we can write this as the vector of mean parameters $\mu = g^{-1}(X\beta)$, where g^{-1} is the exponential function, X is a design matrix of both continuous and categorical covariates, and β is a vector of parameters (regression coefficients). The *i*th row x_i' of X contains the covariates for the *i*th observation. Alternatively, we could write $g(\mu) = X\beta$ where g is the log function, and it is called the link function. This is a fairly general specification, and g can take on various forms, but here we only consider the log link. For negative binomial regression, we assume $Y_i \sim \text{NB}(\mu_i, \kappa)$, where we let the mean μ_i vary as a function of covariates. Because $\mu_i > 0$, we again let $g(\mu) = X\beta$ where g is the log link function.

A SAS-based procedure then fit a generalized linear model (GLM) to the explanatorily, geo-spatiotemporal, geosampled, YFV-related, data by MLE of the parameter vector β. In statistics,

the GLM is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution (Hosmer and Lemeshew 2000). In the explanatorial, seasonal, geopredictive, YFV-related, GLM, each outcome of the dependent variable, $\mathbf{Y}$, was assumed to be generated from a particular distribution in the exponential family which included the normal, binomial, Poisson and gamma distributions, among others. We noted that in the YFV-related model the mean, $\boldsymbol{\mu}$, of the distribution was dependent on the geosampled independent variables, $\mathbf{X}$, through: $\mathbb{E}(\mathbf{Y}) = \boldsymbol{\mu} = g^{-1}(\mathbf{X}\boldsymbol{\beta})$ where E($\mathbf{Y}$) was the expected value of $\mathbf{Y}$; $\mathbf{X}\boldsymbol{\beta}$ was the linear geopredictor, $\boldsymbol{\beta}$ was a linear combination of unknown parameters; and, g was the link function. In this framework, the variance in the risk model was a function, $\mathbf{V}$, of the mean: $\mathrm{Var}(\mathbf{Y}) = \mathrm{V}(\boldsymbol{\mu}) = \mathrm{V}(g^{-1}(\mathbf{X}\boldsymbol{\beta}))$.

Interestingly, the regression analyses assumed independent time series counts (i.e. n_i), taken at geosampled, explanatorial, geoclassified, georferenceable,LULC geolocations i=1, 2… n at the Gulu, eco-epidemiological, study site. The time series dependent, YFV-related probabilistic geoclassifed, LULC, counts were described by a set of variables denoted by matrix $\mathbf{X}_i$, where a $1 \times p$ vector of covariate paramter estimator coefficient values for a geosampled, geoclassiifed , LULC geolocation i. The expected value of these data was then given by μ_i ($\mathbf{X}_i$) =n_i ($\mathbf{X}_i$) exp ($\mathbf{X}_i\boldsymbol{\beta}$), where $\boldsymbol{\beta}$ was the vector of non-redundant parameter estimators in the time series dependnet, YFV-related, forecasting distribution model.

The Poisson rates parameter was given by λ_i ($\mathbf{X}_i$) =μ_i ($\mathbf{X}_i$)/n_i ($\mathbf{X}_i$). The rates parameter λ_i ($\mathbf{X}_i$) was both the mean and the variance of the Poisson distribution for each geosampled YFV-related geolocation i. The dependent variable was total geosampled count. The Poisson regression model assumed that the geo-spatiotemporal geosampled data was equally dispersed-that is, that the conditional variance equaled the condition mean. The procedure used MLE to find the regression coefficients. The data was then log-transformed before performing the eco-epidemiological, probablistic, risk analyses to normalize the distribution and minimize standard error.

There was considerable overdispersion in the YFV-related, eco-epidemiological, forecasting, geoclassifed, LULC, risk model; thus, we used a time series, negative binomial,l endemic, transmission-oriented, model framework to quantitate the explanatorial, clinical, field and remote specified, covariate, parameter estimators associated to the geosampled, georeferenced data. Overdispersion is often encountered when fitting very simple parametric models, such as those based on the Poisson distribution (Neter 1992). If overdispersion is a feature in a geo-spatiotemporal, entomological-related, aquatic, larval, habitat, distribution model, an alternative model with additional free parameters may provide a better fit (Jacob et al. 2005b). Thus, a time series, explanatorial, Poisson mixture model with a geopredictive, YFV-related, negative binomial distribution was employed where the mean of the Poisson distribution was itself a random variable drawn from the gamma distribution; thereby, introducing an additional free parameter in the distribution model. The family of negative binomial distributions is a two-parameter family which uses several parameterizations for treating overdispersed data (Cressie 1993). The Poisson distribution has one free parameter and does not allow for the variance to be adjusted independently of the mean (Haight 1967). A common way to deal with overdispersion for counts is to use a GLM framework, where the most common approach would

be a ''quasi-likelihood,'' with Poisson-like assumptions (i.e., quasi-Poisson) or a negative binomial model (see Hosmer and Lemeshew 2002).

For a quasi-poisson regression, the variance is assumed to be a linear function of the mean; for negative binomial regression, a quadratic function. The quasi-Poisson model and negative binomial model can account for overdispersion, and both have two parameters. Both are commonly available in software packages such as SAS, S, S-plus, or R. A natural question for the ecologist, entomologist or other experimeneter is which pakage should be employed for qualitatiuvely regressively quantitating, geosampled, georfrencable, explanatorial, interpolatable, YFV-related, probablistic, empirically regressaqble, clinical, field or remote covariate, paramter estiomator coefficients. SAS may show striking differences between quasi-Poisson regressions and negative binomial regressions for a particular time serie dependent YFV-related probablsitically geosampled data set. Unfortunately, there is surprisingly little guidance in the statistical literature, especially for the regression case for YFV-related clinical, field or remote geosampled LULC variables. Gardner et al. (1995) found little practical difference, but preferred a negative binomial model when a distributional form is required. Terceiro (2003) compared models using a Kolmogorov-Smirnov goodness-of-fit measure, and found cases where each model fit better. Potts and Elith (2006) found that a zero-inflated model was better than either quasi-Poisson or negative binomial for modeling abundance of a rare plant species, but they point out that zero-inflation is a special type of overdispersion that may be most appropriate when occurrence is rare (a specific mechanism creating excessive zeros).

For any given decomposable, empirically regressable,explanatorial, clinical, field or remote geosampled YFV-related georfrencable, geospatially interpolatable, geoclassifed, LULC, dataset, information theoretic approaches such as Akaike information criteria (AIC; Akaike 1973) or Bayesian information criteria (BIC; Schwarz 1978) might be considered to choose between a quasi-Poisson model and a negative binomial. These approaches depend on a distributional form and a likelihood; however, quasi models are only characterized by their mean and variance, and do not necessarily have a distributional form. For this reason, Burnham and Anderson (2002:67) developed quasi-AIC (QAIC), but they only used it to compare within the quasi class of models (e.g., for subset selection of covariates), and not between quasi models and models with distributional forms. Nevertheless, Sileshi (2006) compared QAIC for quasi-Poisson to AIC for negative binomial, though the validity of this approach has not been demonstrated for YFV-related probablistic , endemic , transmission-oriented, data analyses.. In theory, any model selection method that depends on full distributional likelihoods, such as Bayes factors (Raftery 1995) or minimum description length (Rissanen 1978), including the information theoretic approaches, would not help choose between a quasi-Poisson and negative binomial model.

With the lack of a demonstrated information theoretic approach, an ecologist, entomologist or other experimenter could adopt predictive or goodness-of-fit criteria as used by Gardner et al. (1995), Terceiro (2003), and Potts and Elith (2006), to choose between a quasi-Poisson and negative binomial model. However, a good understanding of the theoretical differences between them can form the basis for an a priori decision based on scientific purposes for generating robust , geoclassified, YFV-related, probablistic clinical, field or remote-specified LULC predictors with minmal forecast uncertainties. Extra-binomial (i.e., extra Poisson) variation occurs when discrete data comes in the form of counts or proportion that display

greater variability than would be predicted when fitting a model which can be resolved using a negative binomial regression (Cressie 1993).

As such, we constructed a robust negative binomial regression with a non-homogenous, gamma distributed mean in NL MIXED by making the by incorporating $\alpha = \frac{1}{\theta}$ $(\alpha > 0)$ in equation 1.2. This involved employing NLMIXED procedure to compute the conditional log-likelihood functions of the geosampled, YFV –related, clincial, field or remote geosampled data given the random effects of the explanatorial, endemic, transmission-oriented, clinical, field and remote specified, covariate parameter estimator coefficients. Note, however, that in addition to these basic equations, the NLMIXED procedure employed a number of checks for missing values and floating-point arithmetic. The NLMIXED procedure computed the conditional log-likelihood functions $l(\phi;y)$ by adding the LIST debugging option to the PROC NLMIXED statement. The following statements were used to quantitate the extra Poisson variation in our time series dependent, geopredictive, endemic, explanatorial, transmission-oriented, YFV-related, regression-based, eco-epidemiological, risk model:

$$Y \sim \text{normal}(m, v)$$

$$l(m,v;y) = -\frac{1}{2}\left(\log\{2\pi\} + \frac{(y-m)^2}{v} + \log\{v\}\right)$$

$$E[Y] = m$$
$$Var[Y] = v > 0$$

$$Y \sim \text{binary}(p)$$

$$l_1(p;y) = \begin{cases} y \log\{p\} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$l_2(p;y) = \begin{cases} (1-y) \log\{1-p\} & y < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$l(p;y) = l_1(p;y) + l_2(p;y)$$
$$E[Y] = p$$
$$Var[Y] = p(1-p)$$
$$< p < 1$$

$$Y \sim \text{binomial}(n, p)$$

$$l_c = \log\{\Gamma(n+1)\} - \log\{\Gamma(y+1)\} - \log\{\Gamma(n-y+1)\}$$

$$l_1(n, p;y) = \begin{cases} y \log\{p\} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$l_2(n, p;y) = \begin{cases} (n-y) \log\{1-p\} & n-y > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$l(n, p;y) = l_c + l_1(n, p;y) + l_2(n, p;y)$$
$$E[Y] = n p$$
$$Var[Y] = n p (1-p)$$
$$< p < 1$$
$$Y \sim \text{gamma}(a, b)$$

$$l(a,b;y) = -a\log\{b\} - \log\{\Gamma(a)\} + (a-1)\log\{y\} - y/b$$
$$E[Y] = ab$$

This parameterization of the gamma distribution differed from the parameterization employed in the GLIMMIX procedures. The following statements revealed that the equivalent reparameterization in the NLMIXED procedure fit the generalized time series, geopredictive, endemic, transmission-oriented, YFV-related, probablistic, regression-based, linear model for the gamma-distributed data in the parameterization of the GLIMMIX procedure employing

```
proc glimmix;
   YFV model y = x / dist=gamma s;
run;
proc nlmixed;
   parms b0=1 b1=0 scale=14;
   linp = b0 + b1*x;
   mu   = exp(linp);
   b    = mu/scale;
   YFV-model y ~ gamma(scale,b);
run;
```

$$Y \sim \text{negbin}(n,p) \, l(n,p;y)$$
$$= \log\{\Gamma(n+y)\} - \log\{\Gamma(n)\} - \log\{\Gamma(y+1)\}$$
$$+ n\log\{p\} + y\log\{1-p\}$$
$$E[Y] = kP = k\left(\frac{1-p}{p}\right)$$

This form of the negative binomial distribution was one of the many parameterizations in which the mass function or log-likelihood function appeared in the residually forecasted probablistic derivatives. Another parameterization we employed in this research was as follows:

$$l(n,p;y) = \log\{\Gamma(n+y)\} - \log\{\Gamma(n)\} - \log\{\Gamma(y+1)\}$$

$$+ n\log\{1 - P/(1+P)\} + y\log\{P/(1+P)\} \text{ with } P = (1-p)/p \, P > 0$$

The parameterization of the negative binomial distribution in the NLMIXED procedure differs from that in the GLIMMIX and GENMOD procedures (www.sas.edu). The following statements revealed the equivalent formulations for MLE in the GLIMMIX and NLMIXED procedures in our negative binomial time series dependent, geopredictive, explanatorial, endemic, transmission-oriented, YFV-related, regression-based model:

```
proc glimmix;
   model y = x / dist=negbin s;
run;
proc nlmixed;
   parms b0=3, b1=1, k=0.8;
   linp = b0 + b1*x;
```

```
    mu = exp(linp);
    p  = 1/(1+mu*k);
    model y ~ negbin(1/k,p);
run;
```

We then employed

$$Y \sim \text{Poisson}(m)$$

$$l(m;y) = y\log\{m\} - m - \log\{\Gamma(y+1)\}$$

$$E[Y] == m > 0$$

The negative binomial, time series dependent, geopredictive, endemic transmission-oriented, YFV-related, regression-based, probablistic, risk–related distribution was then rewritten as $f(y_i|\mathbf{x}_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i!\Gamma(\alpha^{-1})}\left(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i}\right)^{\alpha^{-1}}\left(\frac{\mu_i}{\alpha^{-1}+\mu_i}\right)^{y_i}, \quad y_i = 0,1,2,\ldots$. Thus, the negative binomial distribution was derived as a gamma mixture of the Poissonized, explanatorial, YFV-related, randomized clinical, field and remote geosampled ,time series dependent variables. The conditional mean in the models was then $E(y_i|\mathbf{x}_i) = \mu_i = e^{\mathbf{x}_i'\boldsymbol{\beta}}$ and conditional variance was $V(y_i|\mathbf{x}_i) = \mu_i[1 + \frac{1}{\theta}\mu_i] = \mu_i[1 + \alpha\mu_i] > E(y_i|\mathbf{x}_i)$.

To estimate the time series dependent, geopredictive, YFV-related, cluster-based regression models, we specified DIST=NEGBIN(p=1) in the MODEL statement in PROC REG. The negative binomial model NEGBIN1, set $p = 2$ then had the variance function $V(y_i|\mathbf{x}_i) = \mu_i + \alpha\mu_i$, which was linear in the mean. The log-likelihood function of the NEGBIN1 regression model was thereafter derived from the equation:

$$\mathscr{L} = \sum_{i=1}^{N}\left\{\sum_{j=0}^{y_i-1}\ln\left(j + \alpha^{-1}\exp(\mathbf{x}_i'\boldsymbol{\beta})\right) - \ln(y_i!) - \left(y_i + \alpha^{-1}\exp(\mathbf{x}_i'\boldsymbol{\beta})\right)\ln(1+\alpha) + y_i\ln(\alpha)\right\}$$

. The gradient for the models was then

$$\frac{\partial\mathscr{L}}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{N}\left\{\left(\sum_{j=0}^{y_i-1}\frac{\mu_i}{(j\alpha+\mu_i)}\right)\mathbf{x}_i - \alpha^{-1}\ln(1+\alpha)\mu_i\mathbf{x}_i\right\}$$

and

$$\frac{\partial\mathscr{L}}{\partial\alpha} = \sum_{i=1}^{N}\left\{-\left(\sum_{j=0}^{y_i-1}\frac{\alpha^{-1}\mu_i}{(j\alpha+\mu_i)}\right) - \alpha^{-2}\mu_i\ln(1+\alpha) - \frac{(y_i+\alpha^{-1}\mu_i)}{1+\alpha} + \frac{y_i}{\alpha}\right\}$$

.

The seasonal, YFV-related, geopredictive, negative binomial regression model variance function $V = (y_i|x_i) = \mu_i + \alpha\mu_i^2$, was referred to as the NEGBIN2 model. To estimate this model, we specified DIST=NEGBIN (*p*=2) in the MODEL statements. A test of the Poisson distribution was then performed by testing the hypothesis that $\alpha = \frac{1}{\theta_i} = 0$. A Wald test of this hypothesis was also provided which then rendered the reported *t* statistic for the estimates in the regression models. The log-likelihood function of the models (NEGBIN2) was thereafter generated by the equation

$$\mathscr{L} = \sum_{i=1}^{N}\left\{\sum_{j=0}^{y_i-1}\ln(j+\alpha^{-1}) - \ln(y_i!) - (y_i+\alpha^{-1})\ln(1+\alpha\exp(\mathbf{x}_i'\boldsymbol{\beta})) + y_i\ln(\alpha) + y_i\mathbf{x}_i'\boldsymbol{\beta}\right\}$$

, where *y* was an integer and the gradient was

$$\frac{\partial\mathscr{L}}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{N}\frac{y_i - \mu_i}{1 + \alpha\mu_i}\mathbf{x}_i$$

and

$$\frac{\partial\mathscr{L}}{\partial\alpha} = \sum_{i=1}^{N}\left\{-\alpha^{-2}\sum_{j=0}^{y_i-1}\frac{1}{(j+\alpha^{-1})} + \alpha^{-2}\ln(1+\alpha\mu_i) + \frac{y_i - \mu_i}{\alpha(1+\alpha\mu_i)}\right\}$$

.

Jacob et al. (2010c) considered a general class of negative binomial models with mean μ_i and variance function $\mu_i + \alpha\mu_i^p$ for treating overdispersion in a time series-dependent,

geopredictive West Nile Virus (WNV) mosquito-related, vector (i.e., *Culex quinquefacistus*) habitat regression, cluster-based, risk model in Birmingham, Alabama. The NEGBIN2 model with $p = 2$, was the standard formulation of the negative binomial *Cx. quinquefasciatus* probabilistic, forecasting eco-epidemiological, risk model. Although the formulation here we did derive the YFV-related geopredictive, time series dependent, risk model employing the same technique as in Jacob et al. (2010c) there were other independent values of p $(i.e., -\infty < p < \infty)$, in the district-level explanatorial models which had the same density $f(y_i|x_i)$ except that α^{-1} was replaced by $\alpha^{-1}\mu^{2-p}$.

The statisticians George Box and David Cox developed a procedure to identify an appropriate exponent (Lambda = l) to use to transform data into a "normal shape." The Lambda value indicates the power to which all data should be raised. In order to do this, the Box-Cox power transformation searches from Lambda = -5 to Lamba = +5 until the best value is found. Table 1 shows some common Box-Cox transformations, where *Y'* is the transformation of the original data *Y*. Note that for Lambda = 0, the transformation is NOT Y^0 (because this would be 1 for every value) but instead the logarithm of *Y*.

l	Y'
Table 1: Common Box-Cox Transformations	
-2	$Y^{-2} = 1/Y^2$
-1	$Y^{-1} = 1/Y^1$
-0.5	$Y^{-0.5} = 1/(Sqrt(Y))$
0	log(Y)
0.5	$Y^{0.5} = Sqrt(Y)$
1	$Y^1 = Y$
2	Y^2

Figure 2 shows non-normally distributed cycle time data. Using the Box-Cox power transformation in a statistical analysis software program provides an output that indicates the best Lambda values

Figure 2: Example of Non-normally YFV Distributed LULC Data at the Riceland Gulu study site
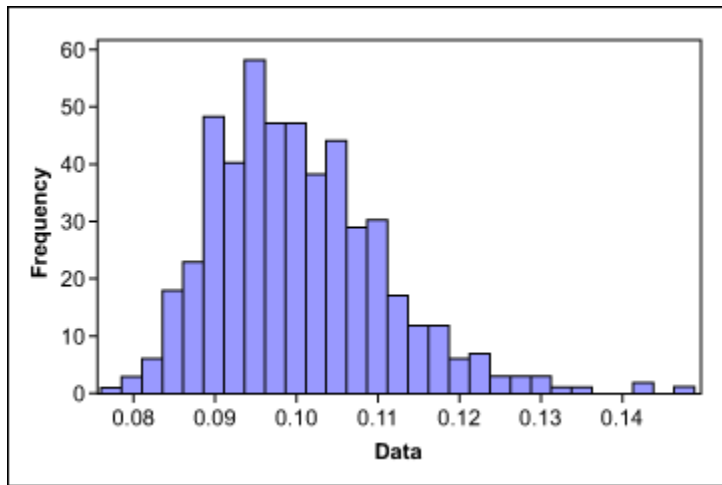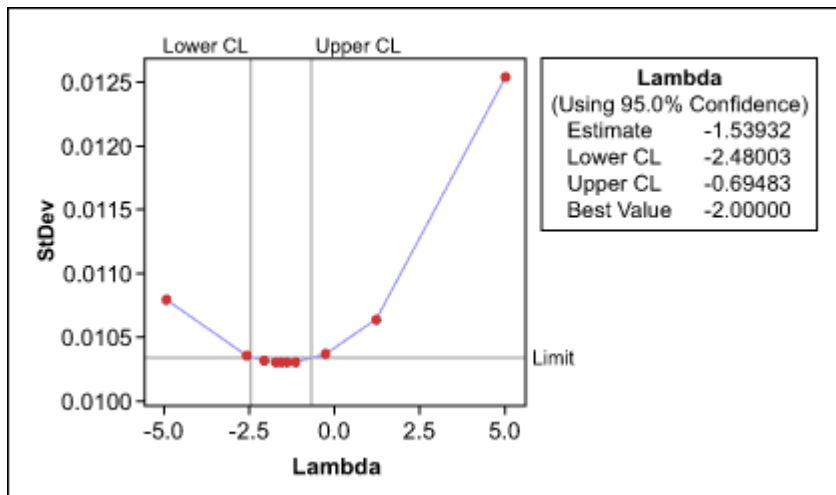

Figure 3: Example Box-Cox Plot of YFV Data at the Riceland Gulu study site

The lower and upper confidence levels (CLs) show that the best results for normality were reached with Lambda values between -2.48 and -0.69. Although the best value is -1.54 (estimate in Figure 3), the process works better if this value is rounded to a whole number; this will make it easier to transform the data back and forth. The best whole-number values here are -1 and -2 (the inverse function of Y and Y^2, respectively). The histogram in Figure 4 shows the transformed data using Lambda = -1, now more normally distributed.

Figure 4: Data Transformed Using Lambda = -1 at the Riceland Gulu study site
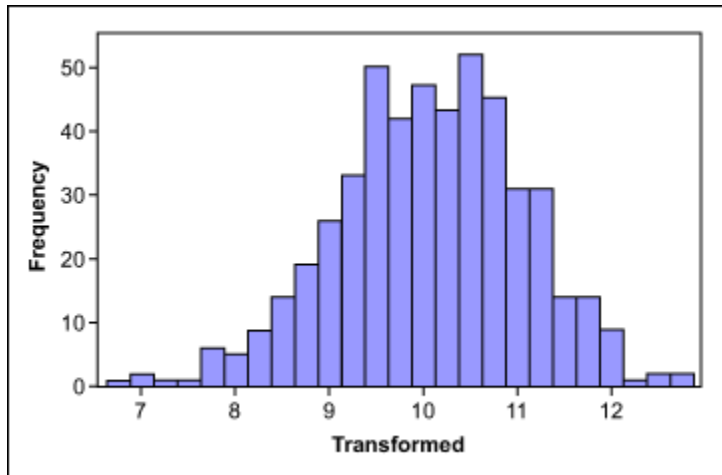
Shapiro–Wilk diagnostic test: The Shapiro–Wilk test was then employed to test the geo-spatiotemporal, geosampled, cluster-based, explanatorial, probabilsitic, YFV-related, LULC-oriented, georeferenced, geopredictive covariate coefficient estimates $x_1, ..., x_n$. In SAS/GIS, the primary test statistics for detecting the presence of non-normality is the Shapiro-Wilk (www.sas.com). Jacob et al. (2008b) used a Shapiro-Wilk test to check the normality assumption in a robust, geopredictive, autoregressive, malaria-related, mosquito, aquatic, larval habitat, distribution model of *An. gambiae s.l.* in multiple, explanatorial datasets of georeferenced, clinical, field and remote-geosampled, geopredictive, covariate, paramter estimator coefficients by constructing a W statistic. W represents the ratio of an optimal uncertainty error estimator of the residual variance based on the square of a linear combination of the ordered statistic which in turn is based on the corrected sum of squares estimator of the variance(Hosmer and Lemeshew 2002). Several diagnostics for the assessment of model misspecifications due to dependence and spatial heterogeneity were then developed using as an application of the Lagrange Multiplier principle.

In mathematical optimization, the method of Lagrange multipliers provides a strategy for finding the local maxima and minima of a function subject to equality constraints. For instance, Jacob et al. (2008) constructed a geopredictive autoregressive *An. gambiae s.l.* aquatic larval habitat, probablistic , time series dependent, regression risk model were optimized employing maximize *f(x,y)* subject to g(x,y)=C. The authors then introduced a new observational variable (λ) into the model and studied the Lagrange function which was defined as $\Lambda(x,y,\lambda) = f(x,y) + \lambda \cdot (g(x,y) - c)$. The malarial, geopredictive, eco-epidemiological, YFV-related, forecasting, risk model residual derivatives revealed that when $f(x_0,y_0)$ was a maximum of $f(x,y)$ for any constrained problem in the model. Interestingly, we noticed that there existed λ_0 such that (x_0, y_0, λ_0) was a stationary point (geosampled, georeferenced, geoclassified, YFV-related explanatorial LULC) for the Lagrange function in the model. Additionally, in the model the stationary points were those points that where the partial derivatives of Λ were zero, (i.e. $\nabla \Lambda = 0$).

We constructed a Shapiro-Wilk test statistic [i.e., $W - \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$]in ArcGIS. We noticed in our probablistic, regression-based, time series-related, YFV-related geopredictive, eco-epidemiological, risk model when $X_{(1)}$ was the ith order statistic, (i.e., the ith-smallest number in the sample dataset); $\overline{x} = (x_1 + \ldots + x_n)/n$ was the sample mean; and, the constants a_1 were rendered by$^{(a_1, \ldots, a_n)} = \frac{m^\top V^{-1}}{(m^\top V^{-1} V^{-1} m)^{1/2}}$where $m = (m_1, \ldots, m_n)^\top$ and $m1 \ldots, M_n$. Additionally, the residual expected uncertainty values from the order statistics of the i.i.d. random YFV regressors was obtained from the standard normal distribution when V was the covariance matrix of those order statistics. To perform the test, the W statistic was initially constructed by considering the regression of the ordered geosampled empirically regressed, clinical, field and remote values in ArcGIS based on the corresponding expected normalized ordered statistics, which was linearized for determining the distribution of anthropogenic populations on various time series dependent, geoclassified LULCs. After W was calculated, the hypothesis of normality was rejected in the autoregressive, explanatorial, YFV-related, risk model residual error matrix since W was less than a quintile from any geosampled value in the model.

These data were then furthered analyzed via a Q-Qin ArcGIS®. A Q-Q plot is a plot of the quintiles of two distributions against each other, or a plot based on estimates of the quintiles; the pattern of points in the plot was then used to compare the two distributions (Anselin 1995). The main step in constructing our Q-Q plot was estimating the quintiles geospatially derived from the geosampled, georeferenced, YFV-related, time series, ,geopredictive, covariate, parameter estimator coefficients.

Quantiles are values taken at regular intervals from the inverse of the CDF of a random variable (geosampled, YFV-related, LULC explanatorily, interpolatable, unmixed, biosignature endmembers). Dividing ordered data into $q$essentially equal-sized entomological-related,time series-dependent, probabilistic, eco-epidemiologist subsets is the motivation for q-quantiles; the quantiles would be the explanatorial, clinical, field or remote –specified covariate, paramter estimator coefficient values marking the boundaries between consecutive subsets. Put another way, a $k^{th}q$-quantile for a random geosampled, time series dependent,YFV –related, eco-epidemiological, geopredictive variable is a value $x$such that the probability that the random variable will be less than x at most $k/q$and the probability that the random variable will be greater than $x$is at most $(q - k)/q = 1 - (k/q)$. By so doing, we assumed that $q - 1$of the q-quantiles may be easily tabulated, one for each integer k satisfying $0 < k < q$ in the YFV-related dataset opf residually forecasted derivatives. In some cases the value of a quantile may not be uniquely determined, as can be the case for the median of a uniform probability distribution on a set of even size(Hosmer and Lemeshew 2002).

If one or both of the axes in a Q-Q plot is based on a theoretical distribution with a continuous CDF, all quintiles are uniquely defined and can be obtained by inverting the CDF in ArcGIS®(http://webhelp.esri.com/arcgisdesktop/). Interestingly if one or both of the axes in a Q-Q plot is based on a theoretical distribution with a continuous CDF generated from probablistically regressing an empirical geosampled dataset of entomological-related, geopredictive, time series variables then all the quintiles will be uniquely defined and can be

obtained by inverting the CDF (see Jacob et al. 2012). If a seasonal, YFV-related, probability distribution with a discontinuous, CDF, time series-related, LULC-oriented, and rainfall distributions are being compared, some of the quintiles may not be defined, so an explanatorily interpolated quintile may have to be plotted (see Fotheringham 2002). If the Q-Q plot is based on time-series data, there are multiple quintile estimators in use (Cressie 1993). Rules for forming Q-Q plots when determining quintiles must be estimated or interpolated are called plotting positions (Anselin 1995).

To construct the time series dependent, explanatorial, YFV-related, regression, Q-Q plot in ArcGIS, it was necessary to use an interpolated quintile estimate so that quintiles corresponded to the respected underlying probability distribution. Given the CDF functions F and $G,$ with associated quintile functions F^{-1} and G^{-1} , the inverse function of the CDF in the model represented the quintile function. The Q-Q plot then drew the qth quintile of F against the qth quintile of G for a range of the geosampled explanatorial, values of q. Thus, the Q-Q plot was a parametric curve which was then indexed over [0,1] with the geosampled, YFV-related, regression-based, covariate, parameter estimator, coefficient values in the real plane R^2. Then we employed the formula k/n for $k = 1, ..., n$, as these were the quintiles that the geosampling distribution realized in the models.

Unfortunately, the last of these, n/n, corresponded to the 100th percentile – the maximum value of the theoretical distribution, which was infinite in the time series, eco-epidemiological, probablistic, YFV, forecasting, risk model. To fix this, we shifted the geosampled, georeferenced, explanatorial, geopredictive, covariate, paramter estimator coefficient estimates over, using $(k-0.5)/n$, and spaced the geo-spatiotemporal geosampled LULC points evenly in uniform distribution, using $k/(n+1)$. By so doing, a probability plot was generated where the quartiles were the rankits, (i.e., the quintile of the expected value of the order statistic of a standard normal distribution). In ArcGIS-based statistics, rankits of a set of data are the expected values of the order statistics of a sample from the standard normal distribution which are primarily used as a graphical technique for normality testing (http://webhelp.esri.com/arcgisdesktop/). The district-level Q-Q plots then compared the shapes of the distributions while providing a graphical view of how the properties such as georefernced LULC geolocation, scale, and skewness were similar or different in the two distributions.

In terms of heuristics for the quartiles of the comparison district-level distributions, we employed the formula $k/(n+1)$ as in Jacob et al. (2011c) .Although several different formulas have been employed or proposed as symmetrical plotting positions for seasonal, entomological-related, explanatorial, geopredictive, covariate,parameter estimator, coefficient values such formulas commonly have the form $(k-a)/(n+1-2a)$ for some value of a in the range from 0 to 1/2, which commonly renders a range between $k/(n+1)$ and $(k-1/2)/n$. However, our time series-related, YFV-related, eco-epidemiological, forecasting, risk model, probabilistic residuals could not generate an accurate depiction of the regressed data as they were highly non-Gaussian. Although the georefernced points plotted in the ArcGIS-oriented, explanatorial, Q-Q plot where non-decreasing when viewed from left to right as expected (see Anselin 1995), the non-normality inherent in the geosampled, YFV-related ,geopredictive, explanatorial, covariate, paramter estimator coefficients could not be cartographically defined nor displayed. If the two distributions being compared are identical, the Q-Q plot follows the 45° line $y = x$.

(www.esri.com). Further, if geosampled distributions agree after linearly transforming the values in one distribution, then the Q-Q plot follows some line, but not necessarily the line $y = x$ (see Anselin 1995).

In our explanatorial, geopredictive, eco-epidemiological, YFV- related, forecasting, risk model, residual derivatives, the Q-Q plot was not flatter than the line $y=x$, and as such the distribution plotted on the horizontal axis was more dispersed than the distribution plotted on the vertical axis. Conversely, the generalized trend of the Q-Q plot was not steeper than the line $y=x$, and as such the distribution plotted on the horizontal axis. Generally Q-Q plots delineating time series dependent, georefernced, seasonal, entomological-related regression–based, geopredictive, eco-epidemiological, covariate, parameter estimator coefficient values are often arced or "S", indicating that one of the rendered distributions is more skewed than another, or that one distribution has heavier tails than another (see, Jacob et al. 2011c, Jacob et al. 2009d)

A "heavy tail" random variable takes extreme values, both low and high in a classic case, more frequently than does a normal random variable (Hosmer and Lemeshew 2002). Heavy tails in a explanatorial, geopredictive, clinical, field or remote, geosampled YFV-related eco-epidemiological forecasting, risk model imply high variance but a heavy tail distribution is not normal with high variance; beyond some point its extreme values are more common even than in a normal distribution with the same variance.(see Jacob et al. 2009).) Similarly a "light tail" random variable such as seen in a uniform, entomological-related, geo-spatiotemporal, distribution would take extreme low and high geosampled values less frequently than would a normal random variable. Light tails imply low variance but "light tail" extreme values are less common even than in a normal distribution with the same variance (Hosmer and Lemeshew 2002).

2.2 Autocorrelation model: Initially, a misspecification perspective was qualitatively regressively quantitated using a spatial autocorrelation estimation analysis and the seasonal YFV-related, probablistic indicators. The model was generated using the $y = X\beta + \varepsilon^*$ (i.e. regression equation) assuming the geosampled data had autocorrelation disturbances. The model also assumed that this data could be decomposed into a white-noise component, ε and a set of

$$y = XB + \underbrace{E\gamma + \varepsilon}_{= \varepsilon^*}$$

unspecified and/or misspecified sub-models that had the structure . White noise in a seasonal, autoregressive, vector, entomological –related, eco-epidemiological, risk model is a univariate or multivariate discrete-time stochastic process whose terms are independent and i.i.d. with a zero mean (Jacob et al. 2007). The misspecification term was E γ. Quantification of the topographic LULC patterns generated from the distribution of the georeferenced, explanatorial, YFV-related geopredictive, empirical, covariate coefficients was required to describe independent key dimensions of the underlying spatial processes in the geosampled data for defining a spatial pattern in the misspecification term.

A spatial autoregressive (SAR) model was then generated that used an explanatorial geopredictive, time series dependent, predictor variable **Y**, as a function of nearby geosampled, empirically regressed, YFV-related, covariate, parameter estimator, coefficient values. **Y** had an indicator value I (i.e., an autoregressive response) and/or the residuals of **Y** which were values of nearby entomological-related, regression-related, parameter estimators, the SAR model furnishes

an alternative specification that frequently is written in terms of matrix $\mathbf{W}$ (see Jacob et al. 2005b, Jacob et al. 2006). As such, its spatial covariance is a function of the matrix $(\mathbf{I} - \rho \, \mathbf{CD}^{-1})(\mathbf{I} - \rho\mathbf{D}^{-1}\mathbf{C}) = (\mathbf{I} - \rho \, \mathbf{W}^{T})(\mathbf{I} - \rho \, \mathbf{W})$, where T denotes matrix transpose (see Griffith 2003). The resulting matrix is symmetric, and may be considered a second-order specification as it includes the product of two geospatial structure matrices (i.e., $\mathbf{W}^{T}\mathbf{W}$) – adjacent geosampled districts as well as those having a single intervening unit involved in the autoregressive function. This matrix restricted positive values of the autoregressive parameter to the more intuitively interpretable range of $0 \leq \hat{\rho} \leq 1$.

Euclidean distance measurements were defined in terms of an *n*-by-*n* geographic weights matrix, $\mathbf{C}$, in ArcGIS whose C_{ij} values were; 1 if the geosampled LULC geolocations *i* and *j* were deemed nearby, and 0 otherwise. Adjusting this matrix by dividing each row entry by its row sum gave $\mathbf{C1}$, where $\mathbf{1}$ was an *n*-by-1 vector of ones which converted this matrix to matrix $\mathbf{W}$. The resulting SAR model specification, with no geosampled, georeferenced, explanatorial, YFV-related, geopredictive, autocovariate probablistic, coefficient values present (i.e., the pure spatial autoregression specification), took on the following form: $\mathbf{Y} = \mu(1-\rho)\mathbf{1} + \rho\mathbf{WY} + \boldsymbol{\varepsilon}$, where μ was the scalar conditional mean of $\mathbf{Y}$, and $\boldsymbol{\varepsilon}$ was an *n*-by-1 error vector whose parameters were statistically independently distributed normally random variates. The spatial covariance matrix for analyzing the geosampled, georeferenced, geopredictive, YFV-related, explanatorial, covariate, paremeter estimator coefficients was then calculated using $E[(\mathbf{Y} - \mu\mathbf{1})'(\mathbf{Y} - \mu\mathbf{1})] = \Sigma = [(\mathbf{I} - \rho\mathbf{W}')(\mathbf{I} - \rho\mathbf{W})]^{-1}\sigma^{2}$, where E ($\bullet$) denoted the calculus of expectations, $\mathbf{I}$ was the *n*-by-*n* identity matrix denoting the matrix transpose operation, and σ^{2} was the error variance.

Next, an autoregressive model specification was generated. The model was written as: $X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t.$ where $\varphi_1, \ldots, \varphi_p$ represented the geo-spatiotemporal, geosampled, explanatorial, clinical, field and remote specified, georeferenced, parameter estimators of the YFV-related regression model, *c* which was a constant when ε_t was the white noise. When coupled with regression and the normal probability model, an autoregressive specification results in a covariation term by characterizing spatial autocorrelation and by denoting the autoregressive parameter that with ρ, a conditional autoregressive covariance specification (Griffith 2003) which in the autoregressive,time series dependent, YFV-related eco-epidemiological, forecasting, risk model involved the matrix $(\mathbf{I} - \rho \, \mathbf{C})$, where $\mathbf{I}$ was an *n*-by-*n* identity matrix. In a robust autoregressive expression; however, the response variable is on the left-side of the equation, while the spatial lagged version of this variable is on the right side (Glantz and Slinker 2001, Anselin 1988). Therefore, one of the main objectives during the autoregressive specification process was to bring the spatially unlagged endogenous variable y exclusively on the left-hand side of the district-level, explanatorial YFV-related, regression equation in order to decorrelate the geosampled, georeferenced, geopredictive, covariate, paramter estimator, coefficient errors. This was accomplished by expanding the matrix term: $(I - \rho V)^{-1} = \sum_{k=0}^{\infty} \rho^k V^k$ as an infinite power series, which was feasible under the assumption that the underlying geospatial process in the geosampled, eco-entomological, explanatorial, clinical, field and remote-specified,

time series dependent datasets was stationary (see Bivand, 1984). The simultaneous autoregressive,probablistic, geo-spatiotemporal, error model was then rewritten as $y - \rho V y = X\beta - \rho V X\beta + \varepsilon$. Substituting this transformation rendered:

$$y = (I - \rho V)^{-1}[X\beta - \rho V(X\beta) + \varepsilon],$$

$$y = \sum_{k=0}^{\infty} \rho^k V^k (X\beta - \rho V X\beta + \varepsilon),$$

$$y = \sum_{k=0}^{\infty} \rho^k V^k X\beta - \sum_{k=0}^{\infty} \rho^{k+1} V^{k+1}(X\beta) + \sum_{k=0}^{\infty} \rho^k V^k \varepsilon,$$

$$y = X\beta + \underbrace{\sum_{k=1}^{\infty} \rho^k V^k X\beta - \sum_{k=1}^{\infty} \rho^k V^k (X\beta)}_{=0} + \sum_{k=0}^{\infty} \rho^k V^k \varepsilon,$$

$$y = X\beta + \underbrace{\sum_{k=1}^{\infty} \rho^k V^k \varepsilon}_{misspecification\cdot term} + \varepsilon$$

As a part of deciphering the geospatial surface of a geosampled, YFV-related, time series-related, geopredictive, eco-epidemiological, explanatorial, variable, it was important to consider how the geosampled, LULC, data feature attributes expanded over space and time. This was done by specifying the "case" and "order" of the connectivity.

We noticed that the misspecification term $\sum \rho^k V^k (k = 1, ..., \infty)$ in the explanatorial, geopredictive YFV, time series eco-epidemiological, model, remained uncorrelated with the exogenous variable, X, as the standard OLS assumption of the disturbances, ε, were uncorrelated with the error coefficients generated from the parameter estimators (b). The spatial lag model was expressed as: $(I - \rho V)y = X\beta + \varepsilon$. Substituting the transformation rendered: $y = \sum_{k=0}^{\infty} \rho^k V^k (X\beta + \varepsilon)$ and

$$y = X\beta + \underbrace{\sum_{k=1}^{\infty} \rho^k V^k (X\beta + \varepsilon)}_{misspecification\cdot term} + \varepsilon$$

. The misspecification term $\sum \rho^k V^k (X\beta + \varepsilon)(k = 1, ..., \infty)$ included the exogenous variables X. Consequently, the exogenous variables were correlated with the misspecification term. Under this condition, standard OLS results for the basic explanatorial, YFV-related, linear regression model $y = X\beta + \varepsilon^*$, generated from the geosampled, georeferenced, erroroneous, covariate, parameter estimator, uncertainty coefficients, provided biased estimates $\tilde{\beta}$ of the underlying regression predictors.

An autoregressive integrated moving average (ARIMA) model was then constructed in ArcGIS based on generalization of an autoregressive moving average (ARMA) model. The model was referred an ARIMA(p,d,q) model where p, d, and q were non-negative integers that refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. ARIMA models form an important part of the Box-Jenkins approach to time-series modeling (Cressie 1993). Commonly, these models are applied in such cases where data show evidence of non-stationarity or where an initial differencing step corresponding to the "integrated" part of the model can be applied to remove the non-stationarity (Griffith 2003).

An ARIMA model was fitted to the time series dependent, explanatorial, YFV-related, eco-epidemiological data to geopredict sample points (i.e., case distribution) in the series. Employing a time series of data X_t where t was an integer index and the X_T was the empirical dataset of geosampled, geopredictive, explanatorial, covariate, paramter estimator, coefficient values, then a ARMA(p,q) model was constructed by: $\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$ where L was the lag operator>In the model α_i was the explanatorial, clinical, field or remote geosampled, geo-spatiotemporal, covariate, parameter estimators of the autoregressive part of the model and the θ_i was the parameters of the moving average part (see Box and Jenkins 1985). The error terms ε_t in such a model would be generally assumed to be independently distributed variables in a robust, geopredictive, YFV-related, time series-related, cluster-based, regression model using a normal distribution with zero mean, for instance. We assume now the polynomial $\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right)$ had a unitary root of multiplicity d. This value was rewritten as: $\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right) = \left(1 + \sum_{i=1}^{p-d} \phi_i L^i\right)(1 - L)^d$. An ARIMA($p$,$d$,$q$) process then expressed this polynomial factorization property which was given by the explanatorial, geopredictive, YFV-related, time series dependent, eco-epidemiological, forecasting, risk model expression $\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$. The residuals were then classified as a particular case of an ARMA($p+d$,q) process having the auto-regressive polynomial with some roots in the unity. ARIMA model with $d>0$ is not wide sense stationary (Cressie 1993).

Because Bayesian model fit using MCMC algorithms is computationally expensive, preliminary model identification to choose the ARIMA parameters, *p*, *d*, *q*, *P*, *D*, and *Q*, was performed using standard (frequentist) tools developed for time series with Gaussian marginal errors, rather than through fitting many possible MCMC models. in a MATLAB terminal to output the figures above.

```
1  % METROPOLIS-HASTINGS BAYESIAN POSTERIOR rand('seed',12345)
2  % PRIOR OVER SCALE PARAMETERS B = 1;
3  % DEFINE LIKELIHOOD likelihood = inline('(B.^A/gamma(A)).*y.^(A-1).*exp(-
4  (B.*y))','y','A','B');
5  % CALCULATE AND VISUALIZE THE LIKELIHOOD SURFACE yy =
6  linspace(0,10,100);
7  AA = linspace(0.1,5,100);
8  likeSurf = zeros(numel(yy),numel(AA)); for iA = 1:numel(AA);
9  likeSurf(:,iA)=likelihood(yy(:),AA(iA),B); end;
10
11 figure;
12 surf(likeSurf); ylabel('p(y|A)'); xlabel('A'); colormap hot
13 % DISPLAY CONDITIONAL AT A = 2 hold on; ly =
14 plot3(ones(1,numel(AA))*40,1:100,likeSurf(:,40),'g','linewidth',3)
15 xlim([0 100]); ylim([0 100]);  axis normal
16 set(gca,'XTick',[0,100]); set(gca,'XTickLabel',[0 5]);
17 set(gca,'YTick',[0,100]); set(gca,'YTickLabel',[0 10]);
18 view(65,25)
```

```
19legend(ly,'p(y|A=2)','Location','Northeast');
20hold off;
21title('p(y|A)');
22 % DEFINE PRIOR OVER SHAPE PARAMETERS prior = inline('sin(pi*A).^2','A');
23 % DEFINE THE POSTERIOR p = inline('(B.^A/gamma(A)).*y.^(A-1).*exp(-
24(B.*y)).*sin(pi*A).^2','y','A','B');
25 % CALCULATE AND DISPLAY THE POSTERIOR SURFACE postSurf =
26zeros(size(likeSurf)); for iA = 1:numel(AA); postSurf(:,iA)=p(yy(:),AA(iA),B); end;
27
28figure
29surf(postSurf); ylabel('y'); xlabel('A'); colormap hot
30 % DISPLAY THE PRIOR hold on; pA =
31plot3(1:100,ones(1,numel(AA))*100,prior(AA),'b','linewidth',3)
32 % SAMPLE FROM p(A | y = 1.5) y = 1.5;
33target = postSurf(16,:);
34 % DISPLAY POSTERIOR psA = plot3(1:100,
35ones(1,numel(AA))*16,postSurf(16,:),'m','linewidth',3)
36xlim([0 100]); ylim([0 100]);  axis normal
37set(gca,'XTick',[0,100]); set(gca,'XTickLabel',[0 5]);
38set(gca,'YTick',[0,100]); set(gca,'YTickLabel',[0 10]);
39view(65,25)
40legend([pA,psA],{'p(A)','p(A|y = 1.5)'},'Location','Northeast');
41hold off
42title('p(A|y)');
43 % INITIALIZE THE METROPOLIS-HASTINGS SAMPLER
44% DEFINE PROPOSAL DENSITY q = inline('exppdf(x,mu)','x','mu');
45 % MEAN FOR PROPOSAL DENSITY mu = 5;
46 % DISPLAY TARGET AND PROPOSAL figure; hold on;
47th = plot(AA,target,'m','Linewidth',2);
48qh = plot(AA,q(AA,mu),'k','Linewidth',2)
49legend([th,qh],{'Target, p(A)','Proposal, q(A)'});
50xlabel('A');
51 % SOME CONSTANTS nSamples = 5000;
52burnIn = 500;
53minn = 0.1; maxx = 5;
54 % INTIIALZE SAMPLER x = zeros(1 ,nSamples);
55x(1) = mu;
56t = 1;
57 % RUN METROPOLIS-HASTINGS SAMPLER
58while t < nSamples
59    t = t+1;
60
61    % SAMPLE FROM PROPOSAL     xStar = exprnd(mu);
62
63    % CORRECTION FACTOR     c = q(x(t-1),mu)/q(xStar,mu);
64
```

```
65    % CALCULATE THE (CORRECTED) ACCEPTANCE RATIO     alpha = min([1,
66p(y,xStar,B)/p(y,x(t-1),B)*c]);
67
68    % ACCEPT OR REJECT?     u = rand;
69    if u < alpha
70        x(t) = xStar;
71    else
72        x(t) = x(t-1);
73    end end
74 % DISPLAY MARKOV CHAIN figure;
75subplot(211);
76stairs(x(1:t),1:t, 'k');
77hold on;
78hb = plot([0 maxx/2],[burnIn burnIn],'g--','Linewidth',2)
79ylabel('t'); xlabel('samples, A');
80set(gca , 'YDir', 'reverse');
81ylim([0 t])
82axis tight;
   xlim([0 maxx]);
   title('Markov Chain Path');
   legend(hb,'Burnin');
    % DISPLAY SAMPLES subplot(212);
   nBins = 100;
   sampleBins = linspace(minn,maxx,nBins);
   counts = hist(x(burnIn:end), sampleBins);
   bar(sampleBins, counts/sum(counts), 'k');
   xlabel('samples, A' ); ylabel( 'p(A | y)' );
   title('Samples');
   xlim([0 10])
    % OVERLAY TARGET DISTRIBUTION hold on;
   plot(AA, target/sum(target) , 'm-', 'LineWidth', 2);
   legend('Sampled Distribution',sprintf('Target Posterior'))
   axis tight
```

A visual analysis of the YFV-related, time series data detected the presence of a long-term (i.e., inter annual) change in the mean level, an unstable variance (which appeared to increase with the mean), and multiplicative seasonality (i.e., the size of the seasonal effect was proportional to the mean). Thus, for the preliminary Gaussian analysis, the data was transformed using a fitted Box-Cox transformation in order to stabilize the variance, to make the seasonal effect additive, and to make the data approximately normally distributed. The trend in the Box-Cox transformed series was treated as a stochastic trend, which was of first order difference stationary. The augmented Dickey – Fuller test on a lag order of 15 was used to detect the presence of a unit root, to assess whether the series needed to be integrated differenced. Gaussian SARIMA models and ARIMA models with a second-order, harmonic, seasonal component, both with $d = 1$ were computed due to the assumed presence of a unit root. Thus harmonic product was fitted with the (frequentist) R software package 'stats', and models were evaluated based on

Akaike's information criterion (AIC). The covariate matrix for the seasonal effect using second order harmonics was given by $\mathbf{x}_t^T = [\sin(2\pi t/12), \cos(2\pi t/12), \sin(2\pi t/6), \cos(2\pi t/6)]$ using two sine and cosine pairs. A time independent intercept was not included as the intercept dropped out of the equation after the first order differencing simulation exercise.

Bayesian negative binomial versions of four SARIMA models and two ARIMA models, with second-order harmonics were identified in the preliminary, YFV-related, eco-epidemiological, risk analysis .These models were implemented in SAS employing untransformed, seasonal, time series dependent, stochastically/deterministically explanatorily, interpolatable data variables using a logarithmic link function and ZQ1 transformation. Since there were only three observations with zero counts, the results was not sensitive to the choice of the transformation constant for ZQ1 which we set at c = 1. Also, versions with identity link were considered. Models were evaluated based on two criteria. The first was the deviance information criterion (DIC), which was calculated as the mean of the posterior distribution of the deviance conditional on the probablistically geosampled clinical, field or remote specified, YFV-related observations (*w*) with equal to the maximum *w* of the models compared and augmented with the number of effective estimated explanatorial parameters as penalty to prevent overfitting. Models with lower DIC were considered to have a better fit.

A second criterion was defined as the mean absolute relative error of the fitted clincial, field or remote geosampled, time series geopredictor values (MARE): MARE $= \sum_{f}^{l} |\frac{y_t - \hat{y}_t}{y_t + 1}| \Big/ (l+1-f)$, where $\hat{y}_t$ was the fitted number of YFV cases at discrete time interval *t*, and *f* and *l* were the first and last discrete time intervals, respectively, of the time period under consideration. The MARE was calculated for both the entire series except for the first ω probabilistic, YFV-related observations. Models were fitted to the entire time series dependent, forecasting eco-epidemiological, explanatorial, autoregressive, clincial, field and remote-specified, risk model (i.e., $f = \omega+1$, *l*). For the second half of the time series ($f = 205$,) the models were fitted to the first half of the time series only.

Since the posterior, explanatorial, time series dependent, geopredictive distributions estimated at each fitted, YFV-related, LULC data point was skewed, the median of the posterior distribution was taken for $\hat{y}_t$. The MARE is similar to the mean absolute percentage error (MAPE), which is applicable to series for which the variance is dependent on the mean (Hosmer and Lemeshew 2000). However, since the denominator was equal to or larger than one, this prevented problems with the larger geosampled explanatorial, clincial, field and remote, endemic, transmission-oriented, covariat, paramter estimator coefficient values caused by dividing by small numbers, and a major critique of the MAPE. The MARE statistic does not have a built-in penalty to prevent over fitting, but among models with similar value of MARE, the model with the least number of ecogeographic parameters is preferred(Griffith 2003). The MARE estimate is comparable across models with different distributional assumptions, in contrast to the DIC(Cressie 1993). Geo-predictive, explanatorial, YFV-related, eco-epidemiological, time series dependent, forecasting, risk-related, simulation models were run with three Markov chains of 11,000 iterations each including a burn-in of 1,000 iterations. Convergence was assessed by studying plots of the Gelman-Rubin convergence statistic (on estimated parameters), employing guidelines as modified by Brooks and Gelman (1992a).

We proposed a general approach to monitoring convergence of MCMC output in which $m > 1$ parallel chains were run with starting geosampled clincial, field and remote, endemic transmission-oriented, covariate, paramter estimator coefficient values that were overdispersed relative to the posterior distribution. Convergence is diagnosed when the chains have `forgotten' their initial values, and the output from all chains is indistinguishable. The gelman.diag diagnostic was applied to a single variable from the chain. It was based a comparison of within-chain and between-chain variances. We estimated the variance of the stationary distribution: the mean of the empirical variance within each chain, W, and the empirical variance from all chains combined, was expressed as $sigma.hat^2 = (n-1)W/n + B/n$ where n was the number of geo-spatiotemporally geosampled, clinical, field or remote geosampled, YFV-related iterations and B/n was the empirical between-chain variance. If the chains have converged, then both estimates are unbiased (Cressie 1993).

The convergence diagnostic was then based on the assumption that the target distribution was normal for simulating YFV-related iterations. A Bayesian credible interval was constructed using a *t*-distribution with mean *mu.hat = Sample mean of all chains were then combined* and variance was tabulated employing *V.hat=sigma.hat2 + B/(mn)*. Subsequently, the degrees of freedom was estimated by the method of moments *using d = 2*V.hat^2/Var(V.hat)*.Use of the *t*-distribution accounts for the fact that the mean and variance of the posterior distribution are estimated (Griffith 2003). The convergence diagnostic itself was *R=sqrt((d+3) V.hat /((d+1)W)*. Values substantially above 1 indicated lack of convergence. If the chains have not converged, Bayesian credible intervals based on the *t*-distribution were considered too wide, and had the potential to shrink by this factor if the MCMC run was continued.

Knowing whether the selected models and their underlying distributions fit the variation in the time series geosampled, YFV-related, temporal data adequately was of special interest to us. If these models are used to geopredict malaria cases in a discrete time interval (in this case, a month), then not only is the point estimate of the posterior, geopredictive distribution of interest, but also the entire distribution. Let F_t be the cumulative, posterior, geopredictive distribution function of $\hat{y}_t$. The lower tail residual probability $F_t(y_t)$, (*i.e.* the value of the cumulative, posterior, geopredictive YFV distribution calculated at the observed clincial, field or remote data $F_t(y_t) = P(\hat{y}_t < y_t | \mathbf{y})$) also called the probability integral transform, was calculated for each sampled month t.

Thereafter, a CDF of $F_t(y_t)$ for all months of interest allowed for the analysis of the appropriateness of the time series dependent, YFV-related, probablistic, geopredictive, explanatorial, forecasting, eco-epidemiological risk model including the assumed underlying distribution. If the model fits the data appropriately, this 'cumulative distribution function of residual probability values (C-R plot)' will follow an approximately straight diagonal line between the origin and point (1,1), similar to a Probability-Probability plot(Box and Jenkins 1985). The C-R plots were also used to assess appropriateness of models fitted to a time series with a Poisson GARIMA(1,1,0) structure.

Thus, after fitting the eco-epidemiological, forecasting, risk, time series, YFV model and obtaining posterior distributions, the $F_t(y_t)$ was calculated for each geosampled explanatorial observation. Because of the fact that the CDF for the negative binomial models is discrete, the

residual probability value was randomized by drawing a random value p_t from the uniform distribution in the interval $[F(y_t-1,\lambda_t,r),F(y_t,\lambda_t,r)]$, following a procedure by Dunn and Smyth (2005), where $F_t(y_t)$ was estimated with 30,000 samples from this distribution. This procedure is advocated by Cressie (1993) for discrete GARMA models. The appropriateness of selected models was compared using plots of their CDFs of (i.e., randomized) residual probability values, both on the entire, YFV-related, case time series and on a dataset of explanatorial georeferencable observations, where case numbers were relatively low.

It is standard practice to test time series explanatorial, entomological-related, time series dependent, eco-epidemiological, forecasting, risk model residuals for remaining autocorrelation (Jacob et al. 2005b, Griffith 2005). However, standard tools presume approximately Gaussian distributed data in these models. Therefore, the randomized residual probability values were converted into normalized randomized quantile residuals ε_t employing the quantile function (i.e., inverse CDF) of the normal distribution with zero mean and unity variance. Prior to conversion, randomized, residual probability values of zero (i.e., when all YFV-related samples from the posterior predictive distribution function were above the observed value) were set to 0.00001 and the probability values of 1(i.e., when all the YFV-related samples from the posterior predictive distribution function were below the observed value) were set to 0.99999. The normalized randomized probablistic, quantile residuals were analyzed thereafter for unquantitated latent autocorrelation with the Ljung-Box test and visual analysis of autocorrelation and partial autocorrelation functions.

We then created a geopredictive model in C++ for qualitatively quantitating, a dataset of georeferencable, clinical, field and remote geosampled, explanatorial, time series dependent, YFV-related variables based on three main covariate paramter estimators, namely rainfall, LULC change distances and human population statistics. The hypothesis was, as the rainfall increases, prevalence of *Ae. aegypti* increases, in turn increasing the chance of transfer of YFV to human population. Another hypothesis was that as the distance between human population in an agro-village compelexes and forest region decreases, the probability of *Ae aegypti* infecting human population with YFV case distribution increase. And as the population of urban LULC areas increases, the probability of occurrence of an epidemic increases from small amount of infected humans.

To generate the effective prediction model from above hypothesis, we needed to determine accurate correlation between Yellow Fever occurrences and the c estimators discussed above. This was addressed by 3-dimensional (D) regression. In our case, we had three covariate parameter estimators, rainfall, LULC and anthropogeic population. In the 3-D regression it was assumed that as population increased linearly the probability of Yellow Fever prevalence increased. It should be noted that the model was designed such that, it was dependent only on input provided.

Table 2: Regressors establishing the relation between co-variates and Yellow Fever cases, following input data table into C++

Year,	Rainfall,	LULC,	Population,	YFCases
2002,	1449,	5.3,	24067200,	0

313

2003,	1456,	5.22,	25089400,	11
2004,	1482,	5.14,	25859700,	0
2005,	1253,	5.07,	26741300,	0
2006,	1374,	4.99,	27629300,	0
2007,	1477.9,	4.92,	28581300,	0
2008,	1320.2,	4.84,	29592600,	0
2009,	1311.2,	4.77,	30661300,	0
2010,	1702.1,	4.71,	31784600,	106

RESULTS

Initially, the CLUSTER procedure in SAS hierarchically aggregated the geosampled time series, explanatorial, YFV-related observational geopredictors. PROC CLUSTER then computed all Euclidean distances in the empirical geosampled dataset based on the flexible-beta method. PROC CLUSTER thereafter generated the number of clusters in the time series-related, geopredictive, YFV-related, explanatorial, covariate parameter estimatators.

PROC CLUSTER created an output dataset which was employed by the TREE procedure in SAS to draw a diagram of the YFV-related LULC cluster hierarchy. To obtain the five-cluster solution, we first employed PROC CLUSTER with the OUTTREE= option, and then used this output dataset as the input dataset to the TREE procedure. Within PROC TREE, NCLUSTER specified the number of clusters based on the georeferenced, geosampled, time series-dependent, explanatorial, geopredictive, YFV-related, data feature attributes and the OUT= options obtained the final solution and drew a tree diagram. Since we considered all the explanatorial, georeferenced, YFV-related, geopredictor covariate paramter estimators to be equally important, we employed the STD option in PROC CLUSTER to standardize the variables to mean 0 and standard deviation 1. A data analyst can remove outliers before using PROC CLUSTER with the STD option (www.sas.edu).

The relationship between prevalence of each individual potential time series dependent, YFV-related, geopredictive, explanatorial, predictor variable geosampled in the Gulu eco-epidemiological study site was investigated by single variable regression analysis in PROC NL MIXED. We used a quadrature method to approximate a given explanatorial, geosampled, clinical, field and/or remote specified integral by a weighted sum over predefined abscissas for qualitatively regressively quantitating the seasonal, endemic, transmission-oriented, geosampled, random effects in the YFV-related, forecasting, eco-epidemiological, risk model. A good approximation can be obtained with an adequate number of quadrature points in a regressed dataset of covariate parameter estimators as well as appropriate centering and scaling of the abscissas (Hosmer and Lemeshew 2000). Adaptive Gaussian quadrature for the integral over u_i centered the integral at the empirical Bayes estimate of u_i, which we defined as the vector $\hat{u}_i$ that minimized $-\log\left[p(y_i|X_i,\phi,u_i)q(u_i|\xi)\right]$ with ϕ and ξ set equal to their current estimates. The final Hessian matrix from this optimization was then used to scale the quadrature abscissas.

We then employed $(z_j, w_j; j = 1,\ldots,p)$ to denote the standard Gauss-Hermite abscissas and weights in the seasonal, geopredictive, explanatorial, endemic, transmission-oriented, time

series dependent, clinical, field and remote geosampled, covariate paramter estimator coefficients as in (Golub and Welsch 1969 and Abramowitz and Stegun 1972). The adaptive Gaussian quadrature integral approximation was then parsimoniously quantitated using

$$\int p(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\phi},\mathbf{u}_i)q(\mathbf{u}_i|\boldsymbol{\xi})d\mathbf{u}_i \approx$$ and $$2^{r/2}|\Gamma(\mathbf{X}_i,\boldsymbol{\theta})|^{-1/2}\sum_{k_1=1}^{p}\cdots\sum_{k_r=1}^{p}\left[p(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\phi},\mathbf{a}_{j_1,\dots,j_r})q(\mathbf{a}_{j_1,\dots,j_r}|\boldsymbol{\xi})\prod_{k=1}^{r}w_{j_k}\exp z_{j_k}^2\right]$$ where r was the dimension of $\mathbf{u}_i$, $\Gamma(\mathbf{X}_i,\boldsymbol{\theta})$ was the Hessian matrix from the empirical Bayes minimization, $\mathbf{z}_{j_1,\dots,j_r}$ was a vector with elements $(z_{j_1},\dots,z_{j_r})$, and $\mathbf{a}_{j_1,\dots,j_r}=\widehat{\mathbf{u}}_i+2^{1/2}\Gamma(\mathbf{X}_i,\boldsymbol{\theta})^{-1/2}\mathbf{z}_{j_1,\dots,j_r}$.

PROC NLMIXED then selected the number of quadrature points adaptively by evaluating the log-likelihood function at the starting values of the seasonal, geopredictive, explanatorial, endemic, transmission-oriented, clinical, field and remote geosampled, covariate parameter estimator until two successive evaluations had a relative difference less than the value of the QTOL= option. The specific search sequence was described under the QFAC= option. Using the QPOINTS= option, we adjusted the number of quadrature points p to obtain different levels of accuracy in the probabilistically regressed clinical, field and remote specfiied, time series, empirical, geosampled, eco-epidemiological dataset. Setting $p=1$ resulted in the Laplacian approximation as described in Beal and Sheiner (1992), Wolfinger (1993), Vonesh (1992, 1996), Vonesh and Chinchilli (1997), and Wolfinger and Lin (1997).

The NOAD option in the PROC NLMIXED statement requested non-adaptive Gaussian quadrature. Here all $\widehat{\mathbf{u}}_i$ were set equal to zero and the Cholesky root of the estimated variance matrix of the seasonal, geopredictive, explanatorial, endemic, transmission-oriented, geo-spatiotemporal clinical, field and remote geosampled, specified, random effects was substituted for $\Gamma(\mathbf{X}_i,\boldsymbol{\theta})^{-1/2}$ employing the expression for $\mathbf{a}_{j_1,\dots,j_r}$. In this case derivatives were computed using the algorithm of Smith (1995). The NOADSCALE option requested the same scaling substitution but with the empirical Bayes $\widehat{\mathbf{u}}_i$. PROC NLMIXED then computed the derivatives of the adaptive Gaussian quadrature approximation for carrying out the default dual quasi-Newton optimization of the YFV-related explanatorial, time series dependent, clinical, field and remote variables.

Another integral approximation in PROC NLMIXED we employed was the first-order method of Beal and Sheiner (1982, 1988) and Sheiner and Beal (1985). This approximation was expressed in the geo-spatiotemporally, geosampled, YFV-related, geopredictive, linearized, endemic, transmission-oriented, explanatorial, clinical, field and remote, eco-epidemiological, specified, diagnostic,probablistic, risk model when $p(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\phi},\mathbf{u}_i)$ was normal—that is, when $$p(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\phi},\mathbf{u}_i)^{\exp}\left\{-(1/2)[\mathbf{y}_i-\mathbf{m}_i(\mathbf{X}_i,\boldsymbol{\phi},\mathbf{u}_i)]'\mathbf{R}_i(\mathbf{X}_i,\boldsymbol{\phi})^{-1}[\mathbf{y}_i-\mathbf{m}_i(\mathbf{X}_i,\boldsymbol{\phi},\mathbf{u}_i)]\right\}$$ where n_i was the dimension of $\mathbf{y}_i$, $\mathbf{R}_i$ was a diagonal variance matrix, and $\mathbf{m}_i$ was the conditional mean vector of $\mathbf{y}_i$. The first-order approximation was then derived by expanding $\mathbf{m}_i(\mathbf{X}_i,\boldsymbol{\phi},\mathbf{u}_i)$ with a one-term Taylor series expansion about $\mathbf{u}_i=\mathbf{0}$, resulting in the approximation $$p(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\phi},\mathbf{u}_i)\approx(2\pi)^{-n_i/2}|\mathbf{R}_i(\mathbf{X}_i,\boldsymbol{\phi})|^{-1/2}\exp\left(-(1/2)[\mathbf{y}_i-\mathbf{m}_i(\mathbf{X}_i,\boldsymbol{\phi},\mathbf{0})-\mathbf{Z}_i(\mathbf{X}_i,\boldsymbol{\phi})\mathbf{u}_i]'\right.$$ $$\left.\mathbf{R}_i(\mathbf{X}_i,\boldsymbol{\phi})^{-1}[\mathbf{y}_i-\mathbf{m}_i(\mathbf{X}_i,\boldsymbol{\phi},\mathbf{0})-\mathbf{Z}_i(\mathbf{X}_i,\boldsymbol{\phi})\mathbf{u}_i]\right)$$ where $\mathbf{Z}_i(\mathbf{X}_i,\boldsymbol{\phi})$ was the Jacobean matrix $\partial\mathbf{m}_i(\mathbf{X}_i,\boldsymbol{\phi},\mathbf{u}_i)/\partial\mathbf{u}_i$ evaluated at $\mathbf{u}_i=\mathbf{0}$.

Assuming that $q(\mathbf{u}_i|\boldsymbol{\xi})$ was normal with mean $\mathbf{0}$ and variance matrix $\mathbf{G}(\boldsymbol{\xi})$, the first-order integral approximation was computable in the seasonal, geopredictive, endemic, transmission-oriented, explanatorial, clinical, field and remote geosampled, YFV, geo-spatiotemporal,eco-epidemiological, probablistic, risk model in closed form after completing the square:

$$\int p(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\phi},\mathbf{u}_i)q(\mathbf{u}_i|\boldsymbol{\xi})d\mathbf{u}_i \approx (2\pi)^{-m_i/2}|\mathbf{V}_i(\mathbf{X}_i,\boldsymbol{\theta})|^{-1/2}\exp\left(-(1/2)\,[\mathbf{y}_i-\mathbf{m}_i(\mathbf{X}_i,\boldsymbol{\phi},\mathbf{0})]'\mathbf{V}_i(\mathbf{X}_i,\boldsymbol{\theta})^{-1}[\mathbf{y}_i-\mathbf{m}_i(\mathbf{X}_i,\boldsymbol{\phi},\mathbf{0})]\right)_W$$

here $\mathbf{V}_i(\mathbf{X}_i,\boldsymbol{\theta}) = \mathbf{Z}_i(\mathbf{X}_i,\boldsymbol{\phi})\mathbf{G}(\boldsymbol{\xi})\mathbf{Z}_i(\mathbf{X}_i,\boldsymbol{\phi})' + \mathbf{R}_i(\mathbf{X}_i,\boldsymbol{\phi})$. The resulting approximation for $f(\boldsymbol{\theta})$ was then minimized over $\boldsymbol{\theta} = [\boldsymbol{\phi},\boldsymbol{\xi}]$ to obtain the first-order estimates.

Interestingly, the NLMIXED procedure first displayed the "Specifications" table, listing basic information about the nonlinear, seasonally mixed, YFV-related, explanatorial, geopredictive, endemic, transmission-oriented, clinical, field and remote geosampled analyses that we specified. This included the principal variables and estimation methods. The "Dimensions" table listed the counts of important quantities in our non-linear, mixed, transmission-oriented, eco-epidemiological geosampled, probablistic, explanatorily interpolatable, endemic, risk model, including the number of observations, subjects, parameters, and quadrature points. The "Parameters" table displayed the information we provided with the PARMS statement and the value of the negative log-likelihood function evaluated at the starting seasonal, geopredictive, endemic, transmission-oriented, explanatorial, clinical, field and remote geosampled, risk model, covariate paramter estimator coefficient values. The START option in the PROC NLMIXED statement displayed the gradient of the negative log-likelihood function at the starting values of the estimator coefficient values. If you also specify the HESS option, then the starting Hessian is displayed as well[www.sas.edu]. The iteration history consists of one line of output for each iteration in the optimization process. The iteration history is displayed by default because it is important that you check for possible convergence problems. The default seasonal, explanatorial, geopredictive, endemic, transmission-oriented, geosampled, eco-epidemiological, risk model iteration included the following variables: Iter, the iteration number; Calls, the number of function calls ,NegLogLike, the value of the objective function ;Diff, the difference between adjacent function values; MaxGrad, the maximum of the absolute (projected) gradient components (except NMSIMP) ;Slope, the slope $\mathbf{g}'\mathbf{s}$ of the search direction $\mathbf{s}$ at the current parameter iterate $\boldsymbol{\theta}^{(k)}$ (QUANEW only); Rho, the ratio between the achieved and geopredicted, YFV-related, geo-spatiotemporal, endemic, clinical, field and remote geosampled values of Diff (NRRIDG only) ;Radius, the radius of the trust region (TRUREG only) ;StdDev, the standard deviation of the simplex values (NMSIMP only); Delta, the vertex length of the simplex (NMSIMP only) ;Size, the size of the simplex (NMSIMP only)

For the QUANEW method, the value of Slope was significantly negative. Interestingly, the line-search algorithm had difficulty reducing the functional, seasonal, YFV-related, geopredictive, explanatorial,endemic, transmission-oriented, geosampled, clinical, field and remote specified, forecasting, eco-epidemiological, probablistic, risk model explanatorial, clinical, field or remote-geosampled, covariate, parameter estimator, coefficient value sufficiently. When this difficulty was encountered, an asterisk (*) appeared after the iteration number. If there was a tilde $(\sim)$ after the iteration number, the BFGS update was skipped, and very high seasonal geosampled probablistic values of the Lagrange function were produced. Further, a backslash ($\backslash$) after the iteration number indicated that Powell's correction for the

BFGS update was used. For methods employing second derivatives, an asterisk (*) after the iteration number means that the computed Hessian approximation is singular and has to be ridged with a positive value (www.sas.edu). For the NMSIMP method, only one line was displayed for several of our internal iterations. This technique skipped the output for some iterations as some of the termination tests (e.g., StdDev and Size) were rather time-consuming compared to the simplex operations, and they were performed only every five simplex operations.

The ITDETAILS option in the PROC NLMIXED statement provided a detailed iteration history. Besides listing the current values of the geosampled, clinical, field or remote-specified, YFV geo-spatiotemporal parameter estimators and their gradients, the ITDETAILS option provided the following values in addition to the default output: Restart, the number of iteration restarts; Active, the number of active constraints; Lambda, the value of the Lagrange multiplier (TRUREG and DBLDOG only);Ridge, the ridge value (NRRIDG only) and Alpha, the line-search step size (QUANEW only). An apostrophe (') trailing the number of active constraints indicated that at least one of the active seasonal, YFV-related, geopredictive, explanatorial, endemic, transmission-oriented, eco-epidemiological, risk model constraints was released from the active set due to a significant Lagrange multiplier.

Interestingly, the "Convergence Status" table contained a status message describing the reason for termination of the optimization. For ODS purposes, the name of this table was "ConvergenceStatus," We then queried the nonprinting numeric variable *Status* to check for a successful optimization. This was useful in our batch processing, and for quantitating BY groups, for instance, in simulations. Successful convergence was then indicated by $Status = 0$. The "Fitting Information" table listed the final, minimized, seasonal, YFV-related, geopredictive explanatorial, endemic, transmission oriented, probabilistic, ec-epidemiologica, risk model covariate,parameter estimator, coefficient value of -2times the log likelihood as well as the information criteria of Akaike (AIC) and Schwarz (BIC), as well as a finite-sample corrected version of AIC (AICC). The criteria were computed as follows: $AIC = 2f(\hat{\theta}) + 2p, AICC = 2f(\hat{\theta}) + 2pn/(n-p-1), BIC = 2f(\hat{\theta}) + p\log(s)$where $f()$was the negative of the marginal log-likelihood function, $\hat{\theta}$was the vector of the geosampled, empirically regressable, YFV-related parameter estimators, p was the number of eco-epidemiological estimators, n was the number of observational predictors, and $s$was the number of explanatorial, clinical, field or remote-specified, YFV-related, geo-spatiotemporal subjects in the risk covariate analyses (LULC,meterological).

The "Parameter Estimates" table listed the estimates of the seasonal geosampled, YFV-related, geopredictive, explanatorial, endemic, transmission-oriented, clinical, field and remote geosampled, risk eco-epidemiologial, model, covariate, parameter estimator, coefficient values after successful convergence and optimization. Standard errors were computed from the final Hessian matrix. The ratio of the estimate with its standard error produced a t value, with approximate degrees of freedom computed as the number of subjects minus the number of random effects. A p-value and confidence limits based on this $t$distribution were also provided. Finally, the gradient of the negative log-likelihood function was displayed for each geosampled, time series, explanatorial, YFV-related, geopredictive, endemic, transmission-oriented,

geographically specified, eco-epidemiological, risk model, covariate,parameter estimator, coefficient value. We then verified that they each coefficient value were sufficiently small for the nonconstrained predictors. Following standard maximum likelihood theory as in Serfling (1980), the asymptotic variance-covariance matrix of the estimatators equaled the inverse of the Hessian matrix. We then displayed this matrix with the COV option in the PROC NLMIXED statement. The corresponding correlation form is available with the CORR option (www.sas.edu).

The "Additional Estimates" table displayed the results of all ESTIMATE statements that we specified, with the same columns as the "Parameter Estimates" table. The ECOV and ECORR options in the PROC NLMIXED statement produced tables displaying the approximate covariance and correlation matrices of the additional, seasonal,geosampled, YFV-related, geopredictive, explanatorial, endemic, transmission-oriented, probablistic, risk model, covariate, paramter estimator, coefficient value, probablistic estimates. These variables were then computed using the delta method as in Billingsley (1986) and Cox (1998).

Assuming a prediction is to be made regarding the ith subject, we supposed that $f(\boldsymbol{\theta}, \mathbf{u}_i)$ was a differentiable function for predicting the YFR-related, time series dependent, geospatotemporal forecasts. In the model $\boldsymbol{\theta}$ denoted the vector of unknown parameters and $\mathbf{u}_i$ denotee the vector of random effects. A natural point prediction was $f(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{u}}_i)$, where $\widehat{\boldsymbol{\theta}}$ was the MLE of $\boldsymbol{\theta}$ and $\widehat{\mathbf{u}}_i$ was the empirical Bayes estimate of $\mathbf{u}_i$ An approximate prediction variance

$$
\mathbf{P} = \begin{bmatrix} \widehat{\mathbf{H}}^{-1} & \widehat{\mathbf{H}}^{-1}\left(\frac{\partial \mathbf{u}_i}{\partial \boldsymbol{\theta}}\right)' \\ \left(\frac{\partial \mathbf{u}_i}{\partial \boldsymbol{\theta}}\right)\widehat{\mathbf{H}}^{-1} & \widehat{\boldsymbol{\Gamma}}^{-1} + \left(\frac{\partial \mathbf{u}_i}{\partial \boldsymbol{\theta}}\right)\widehat{\mathbf{H}}^{-1}\left(\frac{\partial \mathbf{u}_i}{\partial \boldsymbol{\theta}}\right)' \end{bmatrix}
$$

matrix for $(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{u}}_i)$ was then constructed which revealed where $\widehat{\mathbf{H}}$ was the approximate Hessian matrix from the optimization for $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\Gamma}}$ was the approximate Hessian matrix from the optimization for $\widehat{\mathbf{u}}_i$, and $(\partial \widehat{\mathbf{u}}_i / \partial \boldsymbol{\theta})$ was the derivative of $\widehat{\mathbf{u}}_i$ with respect to $\boldsymbol{\theta}$, evaluated at $(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{u}}_i)$. The approximate variance matrix for $\widehat{\boldsymbol{\theta}}$ was based on $\widehat{\mathbf{u}}_i$ which in the geo-spatiotemporally,geosampled, time series dependent, YFV-related , eco-epidemiological, forecasting, risk model an approximation to the conditional mean squared error of prediction.

The prediction variance for a general scalar function $f(\boldsymbol{\theta}, \mathbf{u}_i)$ in the model was then defined as the expected squared difference $E[f(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{u}}_i) - f(\boldsymbol{\theta}, \mathbf{u}_i)]^2$ PROC NLMIXED computed an approximation to the model derivatives employing the derivative of $f(\boldsymbol{\theta}, \mathbf{u}_i)$ which we thereafter computed with respect to each element of $(\boldsymbol{\theta}, \mathbf{u}_i)$ and evaluated it subsequently at $(\widehat{\boldsymbol{\theta}}, \widehat{\mathbf{u}}_i)$. If $\mathbf{a}_i$ was the resulting vector, then the approximate prediction variance was $\mathbf{a}_i' \mathbf{P} \mathbf{a}_i$. This approximation is known as the delta method (Billingsley 1986, Cox 1998). The EDER option in the PROC NLMIXED statement then produced a table that displayed the derivatives of the additional estimates with respect to the model covariate parameter estimators evaluated at their final estimated values.

Since nonlinear optimization is an iterative process that depends on many factors (Cressie 1993), it was difficult to estimate how much computer time was necessary to find an optimal solution satisfying one of the termination criteria. We employed the MAXTIME=, MAXITER=,

and MAXFUNC= options to restrict the amount of CPU time. The number of iterations, and the number of function was called in a single run of PROC NLMIXED. In each iteration k, the NRRIDG technique used a symmetric Householder transformation to decompose the $n \times n$ Hessian matrix $\mathbf{H}$, $\mathbf{H} = \mathbf{V}'\mathbf{TV}$, $\mathbf{V}$: orthogonal, $\mathbf{T}$: tridiagonal to compute the (Newton) search direction $\mathbf{s}$, $\mathbf{s}^{(k)} = -[\mathbf{H}^{(k)}]^{-1}\mathbf{g}^{(k)}$ $k = 1, 2, 3, \cdots$ The TRUREG and NEWRAP techniques used the Cholesky decomposition to solve the same linear system while computing the search direction. The QUANEW, DBLDOG, CONGRA, and NMSIMP techniques did not need to decompose the Hessian matrix.

The (dual) quasi-Newton method then employed the gradient to parsimoniously quantitate the geosampled, dataset of seasonal, YFV-related geopredictive, covariate, paramter estimator coefficient values. It worked especially well for medium to moderately large optimization problems where the objective function and the gradient were much faster to compute than the Hessian; but, in general, it required more iterations than the TRUREG, NEWRAP, and NRRIDG techniques, which then computed second-order derivatives. QUANEW is the default optimization algorithm because it provides an appropriate balance between the speed and stability required for most nonlinear mixed model applications (www.sas.edu).

The QUANEW technique was updated based on the value of the UPDATE= option. The original quasi-Newton algorithm updated an approximation of the inverse Hessian employing the dual quasi-Newton algorithm, which then subsequently updated the Cholesky factor of an approximate Hessian (default). We then specified four update formulas with the UPDATE= option: DBFGS performed the dual Broyden, Fletcher, Goldfarb, and Shanno (BFGS) updated the Cholesky factor of the Hessian matrix. By so doing, the default in the model was determined. DDFP performed the dual Davidon, Fletcher, and Powell (DFP) for updating the Cholesky factor of the Hessian matrix. BFGS performed the original BFGS for updating the inverse Hessian matrix. DFP performed the original for DFP updating the inverse Hessian matrix. Although the negative of the Hessian (the matrix of second derivatives of the posterior with respect to the probablistically geosampled,clinical, field and remote, YFV-related geopredictive, geo-spatiotemporal, time series dependent, estimators) was tabulated the output was positive definite thus we were able to compute the variance matrix

In each iteration, a line search was performed along the search direction to find an approximate optimum. The default line-search method employed quadratic interpolation and cubic extrapolation to obtain a step size α for satisfying the Goldstein conditions. One of the Goldstein conditions can be violated if the feasible region defines an upper limit of the step size (www.sas.edu) Violating the left-side Goldstein condition can affect the positive definiteness of the quasi-Newton update (Cressie 1993). In our case, the iterations were restarted with an identity matrix, resulting in the steepest descent or ascent search direction. We specified the line-search algorithms other than the default with the LINESEARCH= option.

The QUANEW algorithm employed its own line-search technique. No options and YFV-related covariate parameter estimators (except the INSTEP= option) controlling the line search in the other algorithms were considerable applicable. In several applications, large steps in the first iterations are troublesome (www.sas.edu). We used the INSTEP= option to impose an upper

bound for the step size α during the first five iterations. We then used the INHESSIAN=r option to specify a different starting approximation for the Hessian. Thereafter we specified only the INHESSIAN option.

The Cholesky factor of a (possibly ridged) finite difference approximation of the Hessian was used to initialize the quasi-Newton update process. Interestingly, the seasonal, geosampled, YFV-related, geopredictive, explanatorial, endemic, transmission-oriented,eco-epidemiological, clinical, field and remote geosampled, risk model covariate parameter estimator, coefficient value v of the LCSINGULAR=, LCEPSILON=, and LCDEACT= options, which controlled the processing of linear and boundary constraints which were valid only for the quadratic programming subroutine used in each iteration of the QUANEW algorithm.

The FD= and FDHESSIAN= options then specified the use of finite-difference approximations of the seasonal, geosampled, clinical, field and remote-specified, YFV-related, geopredictive, risk model derivatives. The FD= option specified that all the probablistic derivatives be approximated using function evaluations. The FDHESSIAN= option specified that second-order derivatives be then approximated using gradient evaluations. We noted that computing the robust, seasonal, YFV-related, geopredictive, explanatorial, endemic, transmission-oriented, eco-epidemiological,risk model derivatives by finite-difference approximations was very time-consuming, especially for second-order derivatives based only on values of the objective function (FD= option). Since analytical derivatives were difficult to obtain when a function was computed by an iterative process, we considered one of the optimization techniques which used first-order derivatives only (i.e., QUANEW, DBLDOG, or CONGRA). In the expressions that followed, θ denoted the parameter vector, h_i denoted the step size for the ith geosampled, explanaorial, probablistically regressed, YFV-related, geopredictive, eco-epidemiological, forecasting, risk model, covariate, parameter estimator, and e_i was a vector of zeros with a 1 in the ith position.

We also noted that the forward-difference derivative seasonal, probablistically regressed,YFV-related, geopredictive, explanatorial, endemic, transmission-oriented, specified, risk model covariate, parameter estimator, coefficient value approximations consumed less computer time, but were usually not as precise as approximations that used the central-difference formulas. For the first-order derivatives, n additional functions were required which we determined employing $g_i = \dfrac{\partial f}{\partial \theta_i} \approx \dfrac{f(\theta + h_i e_i) - f(\theta)}{h_i}$ .For second-order derivatives based on function calls only $n + n^2/2$ additional function calls were required for our dense Hessian:
$$\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \approx \frac{f(\theta + h_i e_i + h_j e_j) - f(\theta + h_i e_i) - f(\theta + h_j e_j) + f(\theta)}{h_i h_j}$$
For parsimoniously regressively qualitatively quantitating the seasonal, probablistic, YFV-related, geopredictive, explanatorial, endemic, transmission -oriented, eco-epidemiological, specified, dignaostic, risk model, second-order derivatives based on gradient calls m additional gradient were required which we determined employing: $\dfrac{\partial^2 f}{\partial \theta_i \partial \theta_j} \approx \dfrac{g_i(\theta + h_j e_j) - g_i(\theta)}{2 h_j} + \dfrac{g_j(\theta + h_i e_i) - g_j(\theta)}{2 h_i}$.

Thus, for qualitatively quantitating first-order, seasonal, YFV-related, geopredictive, time series dependent, endemic ,transmission-oriented, explanatorial, clinical, field and remote, specified, risk model derivatives, $2n$ additional function calls were required: $g_i = -\frac{\partial f}{\partial \theta_i} \approx \frac{f(\boldsymbol{\theta} + h_i \mathbf{e}_i) - f(\boldsymbol{\theta} - h_i \mathbf{e}_i)}{2h_i}$. For quantitating second-order, clinical, field and remote geosampled, specified, risk model derivatives)(i.e., $2n + 4n^2/2$)additional function calls were required. We validated these residual forecasts using the eco-epidemiological model $\frac{\partial^2 f}{\partial \theta_i^2} \approx \frac{-f(\boldsymbol{\theta} + 2h_i \mathbf{e}_i) + 16f(\boldsymbol{\theta} + h_i \mathbf{e}_i) - 30f(\boldsymbol{\theta}) + 16f(\boldsymbol{\theta} - h_i \mathbf{e}_i) - f(\boldsymbol{\theta} - 2h_i \mathbf{e}_i)}{12h_i^2}$ and subsequently therefater the model revealed $\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \approx \frac{f(\boldsymbol{\theta} + h_i \mathbf{e}_i + h_j \mathbf{e}_j) - f(\boldsymbol{\theta} + h_i \mathbf{e}_i - h_j \mathbf{e}_j) - f(\boldsymbol{\theta} - h_i \mathbf{e}_i + h_j \mathbf{e}_j) + f(\boldsymbol{\theta} - h_i \mathbf{e}_i - h_j \mathbf{e}_j)}{4h_i h_j}$ For second-order derivatives based on gradient calls, $2n$ additional gradient calls were required which we determined employing $\frac{\partial^2 f}{\partial \theta_i \partial \theta_j} \approx \frac{g_i(\boldsymbol{\theta} + h_j \mathbf{e}_j) - g_i(\boldsymbol{\theta} - h_j \mathbf{e}_j)}{4h_j} + \frac{g_j(\boldsymbol{\theta} + h_i \mathbf{e}_i) - g_j(\boldsymbol{\theta} - h_i \mathbf{e}_i)}{4h_i}$

We then used the FDIGITS= option to specify the number of accurate digits in the evaluation of the objective function in the geo-spatiotemporally geosampled, clinical, field and remote-specified, YFV-related, eco-epidemiological, time series dependent, probablistic, risk model. This specification was helpful in determining an appropriate interval size h to be used in the finite-difference, seasonal, YFV-related, geopredictive, explanatorial, endemic, transmission-oriented, geosampled, risk model formulas. The step sizes $h_j j = 1, \dots, n$ were defined as follows: For the forward-difference approximation of first-order derivatives a function call was employed in PROC NL MIXED and second-order derivatives used gradient calls as determined by $h_j = \sqrt[3]{\overline{\eta}}(1 + |\theta_j|)$. For the forward-difference approximation of second-order derivatives we used only function calls and all central-difference formulas, (i.e. $h_j = \sqrt[3]{\overline{\eta}}(1 + |\theta_j|)$). The value of η was defined by the FDIGITS= option.

We specified the number of accurate digits by using FDIGITS=r, where η was set to 10^{-r}. If the FDIGITS= option is not specified, η is set to the machine precision ε (www.sas.edu). Rows and columns of the Hessian matrix were then scaled. Newton-Raphson and double-dogleg optimization techniques were applied to the empirical, geosampled, eco-epidemiological dataset of seasonal, YFV-related, geopredictive, explanatorial, endemic, transmission oriented, risk model covariate, parameter estimators. Each element $H_{i,j} \ i, j = 1, \dots, n$ was divided by the scaling factor $d_i d_j$, where the scaling vector $d = (d_1, \dots, d_n)$ was iteratively updated in a way specified by the HESCAL=i option, as follows: $i = 0$: Importantly, no scaling was done equivalent to $d_i = 1$. $i \neq 0$: The first iteration and each restart iteration set: $d_i^{(0)} = \sqrt{\max(|H_{i,i}^{(0)}|, \varepsilon)} \ i = 1$;and then $d_i^{(k+1)} = \max\left[d_i^{(k)}, \sqrt{\max(|H_{i,i}^{(k)}|, \varepsilon)}\right] \ j = 2$: to $d_i^{(k+1)} = \max\left[0.6 d_i^{(k)}, \sqrt{\max(|H_{i,i}^{(k)}|, \varepsilon)}\right] \ j = 3 \ d_i^{(k+1)} = \sqrt{\max(|H_{i,i}^{(k)}|, \varepsilon)}$ d_i was reset in each iteration. The parameter vector $\boldsymbol{\theta} \in \mathscr{R}^n$ was subject to a set of m linear equality and inequality constraints: $\sum_{i=1}^{n} a_{i,j} \theta_j = b_i \qquad i = 1, \dots, m_e$ $\sum_{i=1}^{n} a_{i,j} \theta_j \geq b_i \qquad i = m_e + 1, \dots, m$. The seasonal, YFV-related, probablistic, geopredictive, explanatorial, endemic, transmission oriented, clinical, field and remote geosampled, risk model,eco-epidemiological, covariate, parameter estimator, coefficient values $a_{i,j}$ and right-hand sides b_i of the equality and inequality constraints were collected in the $m \times n$ matrix $\mathbf{A}$ and the m vector $\mathbf{b}$. The m linear constraints defined

a feasible region $\mathscr{G}$ in $\mathscr{F}^n$ that contained the point θ_* that minimized the geosampled data. If the feasible region $\mathscr{G}$ is empty, no solution to the optimization problem exists(www.sas.edu).

In PROC NLMIXED, all optimization techniques used the active set methods. The iteration of the probablistically regressed seasonal, YFV-related, time series dependent, geopredictive, endemic, transmission-oriented, clinical, field and remote, explanatorial, geosampled, covariate, parameter estimators started with a feasible point $\theta^{(0)}$, which we computed by the Schittkowski and Stoer (1979) algorithm implemented in PROC NLMIXED. The algorithm then moved from one feasible geosampled, eco-epidemiological, georeferncable, YFV-related point $\theta^{(k-1)}$ to a better feasible point $\theta^{(k)}$ along a feasible search directions $s^{(k)}$,

$$\theta^{(k)} = \theta^{(k-1)} + \alpha^{(k)} s^{(k)} \quad , \quad \alpha^{(k)} > 0$$

The path of the seasonal, geopredictive, YFV-related, explanatorial, endemic, transmission-oriented, eco-epridemiological, risk model, covariate, parameter estimator's points $\theta^{(k)}$ never left the feasible region $\mathscr{G}$ of the optimization problem, but it reached its boundaries. The active set $\mathscr{A}^{(k)}$ of point $\theta^{(k)}$ was then defined as the index set of all linear equality constraints and those inequality constraints were satisfied at $\theta^{(k)}$. If no constraint is active $\theta^{(k)}$, the point is located in the interior of $\mathscr{G}$, and the active set $\mathscr{A}^{(k)} = \emptyset$ is empty (Hazewinkle 2001). If our geosampled explanatorial, probablistic, clinical, field and remote specified point $\theta^{(k)}$ in iteration k hit the boundary of inequality constraint i, this constraint i became active and was added to $\mathscr{A}^{(k)}$. Each equality constraint and each active inequality constraint reduced the dimension (degrees of freedom) of the optimization problem.

The LCEPSILON=r option specified the range for active and violated linearized seasonal, YFV-related, geopredictive, explanatorial, endemic transmission-oriented, clinical, field and remote geosampled, probabilistic, eco-epidemiological, risk model, covariate, parameter estimator, empirical constraints. In practice, the active constraints can be satisfied only with finite precision (Hosmer and Lemeshew 2000). In the eco-epidemiological, forecasting, YFV-related, risk model, if the geosampled explanatorial, georefernced, clincial, field or remote specified point $\theta^{(k)}$ satisfied the condition $\left| \sum_{j=1}^{n} a_{i,j} \theta_j^{(k)} - b_i \right| \leq t$ where $t = r(|b_i| + 1)$, then the constraint i was recognized as an active constraint. Otherwise, the constraint i was classified either an inactive inequality or a violated inequality or equality constraint. Due to rounding errors in computing the projected search direction, error can be accumulated so that an iterate $\theta^{(k)}$ steps out of the feasible region (Fotheringham 2002).

Thereafter the NLMIXED procedure computed the conditional log-likelihood functions of the geosampled YFV data given the random effects. Note, however, that in addition to these basic equations, the NLMIXED procedure employed a number of checks for missing values and floating-point arithmetic. We saw the entire program employed by the NLMIXED procedure to compute the conditional log-likelihood functions $l(\phi; y)$ by adding the LIST debugging option to the PROC NLMIXED statement.

$Y \sim \text{normal}(m,v)$

$$l(m,v;y) = -\frac{1}{2}\left(\log\{2\pi\} + \frac{(y-m)^2}{v} + \log\{v\}\right)$$

$E[Y] = m$ and $\text{Var}[Y] = v$

$E[Y]_v > 0$

$Y \sim \text{binary}(p)$

$$h_1(p;y) = \begin{cases} y\,\log\{p\} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

$h_1(p;y) \quad l_2(p;y) = \begin{cases} (1-y)\,\log\{1-p\} & y < 1 \\ 0 & \text{otherwise} \end{cases}$

$l(p;y) \quad = h_1(p;y) + l_2(p;y)$

$E[Y] = p$

$\text{Var}[Y] \quad = p(1-p)$

$0 \quad < p < 1$

$Y \sim \text{binomial}(n,p) = \log\{\Gamma(n+1)\} - \log\{\Gamma(y+1)\} - \log\{\Gamma(n-y+1)\}$

$$h_1(n,p;y) = \begin{cases} y\,\log\{p\} & y > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$l_2(n,p;y) = \begin{cases} (n-y)\,\log\{1-p\} & n-y > 0 \\ 0 & \text{otherwise} \end{cases}$$

$l(n,p;y) = l_c + h_1(n,p;y) + l_2(n,p;y)$

$E[Y] = np$

$E[Y]\,\text{Var}[Y] = np(1-p)$

$0 \quad < p < 1$

$Y \sim \text{gamma}(a,b)$

$l(a,b;y) = -a\log\{b\} - \log\{\Gamma(a)\} + (a-1)\log\{y\} - y/b$

$E[Y] = ab$

$b \quad = ab^2$

$b \quad > 0$

This parameterization of the gamma distribution differed from the parameterization employed in the GLIMMIX and GENMOD procedures. The following statements revealed the equivalent re-parameterization in the NLMIXED procedure that fit a GLM for the gamma-distributed geo-spatiotemporally, geosampled, eco-epidemiological, YFV-related, explanatorial, clinical, field or remote-geosampled, covariate, parameter estimator coefficients in the parameterization of the GLIMMIX procedure:

```
proc glimmix;
   model y = x / dist=gamma s;
run;
```

```
proc nlmixed;
   parms b0=1 b1=0 scale=14;
   linp = b0 + b1*x;
   mu   = exp(linp);
   b    = mu/scale;
   model y ~ gamma(scale,b);
run;
```

$Y \sim \text{negbin}(n, p) l(n, p; y)$
$= \log\{\Gamma(n + y)\} - \log\{\Gamma(n)\} - \log\{\Gamma(y + 1)\} + n \log\{p\} + y \log\{1 - p\}$

$E[Y] = kP = k\left(\dfrac{1 - p}{p}\right)$

$Var[Y] = kP(1 - P) = k\left(\dfrac{1 - p}{p}\right)\dfrac{1}{p}$

$\geq 0 < p < 1$

This form of the negative binomial distribution is one of the many parameterizations in which the mass function or log-likelihood function appears(Hosmer and Lemeshew 2002).

Another common parameterization was also employed to quantitate the probablistically regressable , time series dependnet, clinical, field and remote YFV estimators which was:
$l(n, p; y) = \log\{\Gamma(n + y)\} - \log\{\Gamma(n)\} - \log\{\Gamma(y + 1)\} + n \log\{1 - P/(1 + P)\} + y \log\{P/(1 + P)\}$
with $P = (1 - p)/p P > 0$. Note that the parameter n in the eco-epidemiological, geo-spatiotemporal model was real-numbered so it did not have to be integer-valued. The parameterization of the negative binomial distribution in the NLMIXED procedure differed from that in the GLIMMIX and GENMOD procedures. The following statements revealed the equivalent formulations for   conducting the MLE in the GLIMMIX and NLMIXED procedures in the negative binomial regression model:

```
proc glimmix;
  model y = x / dist=negbin s;
run;
proc nlmixed;
   parms b0=3, b1=1, k=0.8;
   linp = b0 + b1*x;
   mu = exp(linp);
   p = 1/(1+mu*k);
   model y ~ negbin(1/k,p);
run;
```

$Y \sim \text{Poisson}(m)$
$l(m; y) = y \log\{m\} - m - \log\{\Gamma(y + 1)\}$

$E[Y] = m$
$E[Y] Var[Y]$
$= m > 0$

We used the regression line $(y_i - \overline{y}) = (\hat{y}_i - \overline{y}) + (y_i - \hat{y}_i)$ to generate a pseudo R^2 value where the first term was the total variation in the response y (i.e.,geosampled time series dependent, explanatorial, geopredictive, YFV-related, covariate, parameter estimator coefficients) and the second term was the variation in mean response based on the geosampled probabilistic regressors. The third term was the residual value in the eco-epidemiological, YFV-related model estimates. Squaring each of these terms and adding over all of the geosampled, observational, geopredictors generated the equation $\Sigma(y_i - \overline{y})^2 = \Sigma(\hat{y}_i - \overline{y})^2 + \Sigma(y_i - \hat{y}_i)^2$. This equation was then written as SST = SSM + SSE, where SS was notation for sum of squares and T, M, and E were the notation for total, model, and error, respectively.

We noted that the square of the sample correlation in the seasonal, YFV-related, risk model was equal to the ratio of the estimates while the sum of squares was related to the total sum of squares: r^2 = SSM/SST. This formalized the interpretation of the pseudo R^2 for explaining the fraction of variability in the geosampled, time series dependent, explanatorial, YFV-related, covariate, parameter estimatators explained by the regression model. The sample variance s_y^2 was equal to $\Sigma(y_i - \overline{y})^2/(n - 1)$, which in turn was equal to the SST/df, the total sum of squares divided by the total df. A regression equation was then constructed using the mean square model (i.e. MSM) = $\Sigma(\hat{y}_i - \overline{y})^2/(l)$, which was coincidentally equal to the SSM/df. The corresponding mean square error (i.e. MSE) was $\Sigma(y_i - \hat{y}_i)^2/(n - 2)$, which was equal to SSE/df and the estimate of the variance about the regression line (i.e., σ^2). The MSE is an estimate of σ^2 for determining whether or not the null hypothesis is true (Hosmer and Lemeshew 2000).

For the time series modeled the DFM which was equal to p, and the error degrees of freedom (dfe) which subsequently was equal to ($n - p - 1$), and the total degrees of freedom (dft). This product which was then equal to ($n - 1$), the sum of DFM and DFE. Explanatorial and response variables were numeric. The relationship between the mean of the response variable and the level of the geosampled covariate, estimator coefficients in the probablistic, time series, regression equation were assumed to be approximately linear (i.e., straight line). The corresponding table generated classified each of the time series dependent, explanatorial, YFV-related, geopredictive parameters in SAS as in Table 3.

Table 3: The YFV-related regression-based model parameters

Source	Degrees of Freedom	Sum of squares	Mean Square
Model	p MSM/MSE	$\Sigma(\hat{y}_i - \overline{y})^2$	SSM/DFM
Error	$n - p - 1$	$\Sigma(y_i - \hat{y}_i)^2$	SSE/DFE
Total	$n - 1$	$\Sigma(y_i - \overline{y})^2$	SST/DFT

In the endemic, YFV-related multiple regression analyses, the test statistic MSM/MSE had an F(p, $n - p - 1$) distribution. The null hypothesis was $\beta_1 = \beta_2 = ... = \beta_p = 0$, and the

alternative hypothesis was at least one of the geosampled, time series dependent, explanatorial, geopredictive, YFV-related, parameters $\beta_j \neq$ 0, j = 1, 2, ,,, p. The F test did not indicate which of the parameters $\beta_j \neq$ was not equal to zero, only that at least one of them was linearly related to the response variable. The ratio SSM/SST = pseudo R² (i.e., squared multiple correlation coefficient) was the proportion of the variation in the response variable that was explained by the geosampled, eco-epidemiological, geo-spatiotemporal, explanatorial, clinical, field or remote-specified, YFV-related data. The square root of pseudo R² (i.e., the multiple correlation coefficient) was the correlation then between the geosampled, time series dependent, explanatorial, geopredictive, YFV-related, parameter estimate (i.e., y_i) and the fitted values (i.e., $\hat{y}_i$.).

Additionally, from the geosampling distribution, generated from the geosampled, YFV-related, *t* time series dependent, explanatorial parameters, the probability of obtaining an F was large or larger than the one that was calculated. Since there were only two means to compare, the *t*-test and the F-test were equivalent; the relation between ANOVA and *t* was given by $F = t^2$. Significant differences by ANOVA were noted for mean numbers of the geosampled, time series dependent,explanatorial, geopredictive, YFV-related, explanatorial, clinical, field or remote-specified, parameter estimates captured throughout the sampling frame (F = 41.4, df = 1).

We then constructed a Poisson probability regression model to determine the relationship between the geosampled, time series-dependent, explanatorial, geopredictive, LULC, YFV-related, parameter estimate count data and the geosampled case distribution data. The Poisson models were built using the field and remote-sampled data. The Poisson distribution is a special case of the negative binomial distribution, where the mean approximates the standard deviation (Neter 1992). We assumed that the log of the mean,μ, was a linear function of independent variables, log(μ) = intercept + b1*X1 +b2*X2 + ...+ b3*Xm in the model which implied that μ was the exponential function of independent variables when μ = exp(intercept + b1*X1 +b2*X2 + ...+ b3*Xm) (see Jacob et al. 2005b)..

Thereafter, the NL MIXED procedure estimated the regressed probablistic, parameters estimators of the seasonal, YFV-related, geopredictive, explanatorial, eco-epidemiological, risk model numerically through an iterative fitting process. The dispersion parameter was then estimated by the residual deviance and by Pearson's chi-square divided by the degrees of freedom. Covariance, standard errors, and *p*-values were then computed for the estimated covariate, parameter estimator, coefficient values based on the asymptotic normality derived from MLE.

Note, that the sample size *N* completely dropped out of the probability function which coincidentally had the same functional form for all the geo-spatiotemporal, geosampled, eco-epidemiological, YFV-related, covariate, parameter estimator, indicator values (i.e., γ). As expected, the Poisson distribution was normalized so that the sum of probabilities equaled 1. The

ratio of probabilities was then
$$\sum_{n=0}^{\infty} P_\gamma(n) = e^{-\gamma} \sum_{n=0}^{\infty} \frac{\gamma^n}{n!} = e^{-\gamma} e^{\gamma} = 1.$$
provided by

$$\frac{P_\nu(n=i+1)}{P(n=i)} = \frac{\frac{\nu^{i+1}e^{-\nu}}{(i+1)!}}{\frac{e^{-\nu}\nu^i}{i!}} = \frac{\nu}{i+1}.$$

The Poisson distribution revealed that the geosampled, YFV-related, covariate, parameter estimator, coefficient values reached a maximum when

$$\frac{dP_\nu(n)}{dn} = \frac{e^{-\nu}n(\gamma - H_n + \ln\nu)}{n!} = 0,$$

where γ was the Euler-Mascheroni constant and H_n was a harmonic number, leading to the transcendental equation $\gamma - H_n + \ln\gamma = 0$. The model forecasts revealed that the Euler-Mascheroni constant arose in the integrals as

$$\gamma = -\int_0^\infty e^{-x}\ln x\,dx = \int_0^1 \ln\ln\left(\frac{1}{x}\right)dx = \int_0^\infty\left(\frac{1}{1-e^{-x}} - \frac{1}{x}\right)e^{-x}\,dx = \int_0^\infty \frac{1}{x}\left(\frac{1}{1+x} - e^{-x}\right)dx \quad (2.2).$$

Commonly, integrals that render γ in combination with geo-spatiotemporal constants include $\int_0^\infty e^{-x^2}\ln x\,dx = -\frac{1}{4}\sqrt{\pi}(\gamma + 2\ln 2)$ and $\int_0^\infty e^{-x}(\ln x)^2\,dx = \gamma^2 + \frac{1}{6}\pi^2$. (see Haight 1967). Thereafter, the double integrals in the time series dependent, YFV-related, regression model included $\gamma = \int_0^1\int_0^1 \frac{x-1}{(1-xy)\ln(xy)}dx\,dy$. An interesting analog of equation (2.2) in the model was then provided by $\ln\left(\frac{4}{\pi}\right) = \sum_{n=1}^\infty (-1)^{n-1}\left[\frac{1}{n} - \ln\left(\frac{n+1}{n}\right)\right] = \int_0^1\int_0^1 \frac{x-1}{(1+xy)\ln(xy)}dx\,dy = 0.241564\ldots\,\gamma.$

This solution was also provided by incorporating Marten's theorem [i.e., $e^\gamma = \lim_{n\to\infty}\frac{1}{\ln p_n}\prod_{i=1}^n \frac{1}{1-\frac{1}{\pi_i}}$,] where the product was aggregated over the seasonal, geosampled, georefernced, YFV-related, explanatorial, time series dependent, covariate, parameter estimator, coefficients values found in the ecological empirical datasets. Marten's' 3rd theorem: $\lim_{n\to\infty}\ln n\prod_{p\le n}\left(1-\frac{1}{p}\right) = e^{-\gamma},$ is related to the density of prime numbers where γ is the Euler–Mascheroni constant (see Hosmer and Lemeshew 2000). By taking the logarithm of both sides in the time series dependent, explanatorial, YFV-related, geopredictive, eco-epidemiological, forecasting, probablistic regression-related, risk model, an explicit formula for γ was then obtained using

$$\gamma = \lim_{x\to\infty}\left[\sum_{p\le x}\ln\left(\frac{1}{1-\frac{1}{p}}\right) - \ln\ln x\right].$$

This product was also given by series due to Euler, which followed from equation (2.2) by first replacing $\ln n$ by $\ln(n+1)$, in the equation $\gamma = \sum_{k=1}^\infty\left[\frac{1}{k} - \ln\left(1+\frac{1}{k}\right)\right]$ and then generating $\lim_{n\to\infty}[\ln(n+1) - \ln n] = \lim_{n\to\infty}\ln\left(1+\frac{1}{n}\right) = 0.$ We then substituted the telescoping sum $\sum_{k=1}^n\ln\left(1+\frac{1}{k}\right)$ for $\ln(n+1)$ which rendered $\ln\left(1+\frac{1}{k}\right) = \ln(k+1) - \ln k$. Thereafter, we obtained

$$\lim_{n\to\infty}\left[\sum_{k=1}^n\frac{1}{k} - \sum_{k=1}^n\ln\left(1+\frac{1}{k}\right)\right] = \lim_{n\to\infty}\sum_{k=1}^n\left[\frac{1}{k} - \ln\left(1+\frac{1}{k}\right)\right].$$

Additionally, other series in the geo-spatiotemporal, geosampled, explanatorial, geopredictive, YFV-related, time series dependent, autoregressive, probablistic regression model

327

included the equation ($\Diamond$) where $\gamma = \sum_{n=2}^{\infty} (-1)^n \frac{\zeta(n)}{n} = \ln\left(\frac{4}{\pi}\right) + \sum_{n=1}^{\infty} \frac{(-1)^{n-1} \zeta(n+1)}{2^n (n+1)}$ and where $\zeta(z)$ was

$\gamma = \sum_{n=1}^{\infty} (-1)^n \frac{\lfloor \lg n \rfloor}{n}$ (i.e., the Riemann zeta function). The Riemann zeta function $\zeta(s)$, is a function

of a complex variable s that analytically continues the sum of the infinite series $\sum_{n=1}^{\infty} \frac{1}{n^s}$, which converges when the real part of s is greater than 1 where log is the logarithm to base 2 and $\lfloor x \rfloor$ is the floor function (Titchmarsh 1986). Nielsen (1921) earlier gave a series equivalent to

$\gamma = 1 - \sum_{k=1}^{\infty} \sum_{\ell=2^{n-1}}^{2^n-1} \frac{n}{(2k+1)(2k+2)}$, and thereafter $\frac{1}{(2k+1)(2k+2)} - \frac{1}{2k+1} - \frac{1}{2k+2}$ which was then added

$0 = -\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots$ to render Vacca's formula. Additionally, we used the sums

$\gamma = \sum_{n=1}^{\infty} \sum_{k=2^n}^{\infty} \frac{(-1)^k}{k} = \sum_{k=1}^{\infty} \frac{1}{2^{k+1}} \sum_{j=0}^{k-1} \binom{2^{k-j} + j}{j}^{-1}$ as determined by Gosper et al. (1972) with $k - j$ by replacing the undefined I and then rewrote the equation as a double series for applying the Euler's series transformation to each of the geosampled, time-series dependent, georefernced, probablistic, YFV-related, explanatorial, covariate, paramter estimator, coefficients estimates.

$\binom{n}{k}$ was used as a binomial coefficient, rearranged to achieve the conditionally convergent series in the time-series dependent, georefernced, YFV-related, eco-epidemiological, forecasting, regression-related, geopredictive, risk model as the plus and minus terms were first grouped in pairs of the geosampled covariate,parameter estimator, coefficient estimates employing the resulting series of the actual geosampled measurement values. The double series was thereby

equivalent to Catalan's integral: $\gamma = \int_0^1 \frac{1}{1+x} \sum_{n=1}^{\infty} x^{2^n-1} dx.$ Catalan's integrals are a special case of

general formulas due to $J_0\left(\sqrt{z^2 - y^2}\right) = \frac{1}{\pi} \int_0^{\pi} e^{y \cos\theta} \cos(z \sin\theta) d\theta,$ where $J_0(z)$ is a Bessel function of the first kind (see Catalan, 1883).

The Bessel function is a function $Z_n(x)$ defined in a robust, probabilistic, regression model

by employing the recurrence relations $Z_{n+1} + Z_{n-1} = \frac{2n}{x} Z_n$ and $Z_{n+1} - Z_{n-1} = -2 \frac{dZ_n}{dx}$. (see Hosmer and Lemeshew 2000), which more frequently has been defined as solutions in a regression models

using the differential equation $x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (x^2 - n^2) y = 0.$ (see Watson, 1966). The Bessel

function $J_n(z)$ was defined by the contour integral $J_n(z) = \frac{1}{2\pi i} \oint e^{(z/2)(t-1/t)} t^{-n-1} dt,$ where the contour enclosing the origin and was traversed in a counterclockwise direction. This function generated: $J_n(2\sqrt{z}) = \frac{1}{\pi} \int_0^{\pi} e^{\cdots} \cos[(1-z)\sin\theta] d\theta.$ $z \equiv 1 - z'$ and $y \equiv 1 + z'$. Thereafter, to quantitate the equivalence in the geo-spatiotemporal, geosampled dataset of the YFV-related, regression-based, covariate, parameter estimators, we expanded $1/(1+x)$ in a geometric series and multiplied the sampled data by x^{2^n-1}, and integrated the term wise as in Sondow and Zudilin (2003). Other series for γ were

then employed including:
$$\gamma = \frac{3}{2} - \ln 2 - \sum_{m=2}^{\infty} (-1)^m \frac{m-1}{m} [\zeta(m) - 1]$$
and

$$\gamma = \frac{2^x}{e^{2^x}} \sum_{n=0}^{\infty} \frac{2^{nn}}{(n+1)!} \sum_{t=0}^{m} \frac{1}{t+1} - n \ln 2 + o\left(\frac{1}{2^n e^{2^n}}\right),$$
A rapidly converging limit for γ was then provided by

$$\gamma = \lim_{n \to \infty}\left[\frac{2n-1}{2n} - \ln n + \sum_{k=2}^{n} \left(\frac{1}{k} - \frac{\zeta(1-k)}{n^k}\right)\right] \lim_{n \to \infty}\left[\frac{2n-1}{2n} - \ln n + \sum_{k=2}^{n} \frac{1}{k}\left(1 + \frac{B_k}{n^k}\right)\right],$$
where B_k was a Bernoulli number. Another limit formula was then provided by the

equation $$\gamma = -\lim_{n \to \infty}\left[\frac{\Gamma\left(\frac{1}{n}\right)\Gamma(n+1)n^{1+1/n}}{\Gamma\left(2+n+\frac{1}{n}\right)} - \frac{n^2}{n+1}\right].$$ Limits to the regression model were then rendered

by $\gamma = \lim_{x \to \infty} \zeta(\zeta(z)) - 2^x + \left(\frac{4}{3}\right)^x + 1$ where $\zeta(z)$ was the Riemann zeta function.

The Riemann zeta function or Euler–Riemann zeta function, $\zeta(s)$, is a function of a complex variable s that analytically continues the sum of the infinite series $\sum_{n=1}^{\infty} \frac{1}{n^s}$ which converges when the real part of s is greater than 1(Hosmer and Lemeshew 2002). The Riemann zeta function plays a pivotal role in analytic number theory and has applications in physics, probability theory, and applied statistics (Hazewinkle 2001). In the geo-spatiotemrpoal, geosampled, probabilistic, YFV-related eco-epidemiological, forecasting risk model,the Riemann zeta function $\zeta(s)$was a function of a complex variable (i.e., $s = \sigma + it$). The notation with s, σ, and t is traditionally used in the study of the ζ-function, following Riemann(Hosmer and Lemeshew 2002).

The following infinite series converged for all empiricial, geo-spatiotemporal, explanatorial YFV-related, eco-epidemiolgical, time series dependent, clinical, field and remote-gesampled, covariate, paramter estimators coefficients with real part greater than 1, and defined $\zeta(s)$ as $$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots \qquad \sigma = \Re(s) > 1.$$ The model derivative forecast also was

defined by the integral $\zeta(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} \frac{x^{s-1}}{e^x - 1} dx$ where $\Gamma(s)$ was the gamma function.

We noted that the gamma function (i.e., Γ) in the empiricial, geo-spatiotemporal, explanatorial YFV-related, eco-epidemiological, clinical, field and remote-gesampled, geopredictive, eco-epidemiological, risk model was an extension of the factorial function, with its argument shifted down by for the regressor coefficient values. That is, if n was an empirical dataset of geo-spatiotemporal, YFV-related clinical, field or remote geosampled ,georeferenced, explanatorial, covariate paramter estimator, coefficient values then $\Gamma(n) = (n-1)!.$ The gamma function was then defined for all the geosampled,YFV-related, probablistic, coefficient values except the non-positive integers. For qualitatively regressing quantitating a positive real part, a YFV-related, eco-epidemiological, time series dependent, gamma function was defined

via a convergent improper integral: $\Gamma(t) = \int_0^{\infty} x^{t-1}e^{-x} dx$ .This integral function was extended by analytic continuation to all the time series dependent, YFV-related,geo-

spatiotemporal regressors except the non-positive integers (where the function has simple poles) for robustly yielding the meromorphic function and the gamma function. The gamma function corresponds to the Mellin transform of the negative exponential function: $\Gamma(t) = \{\mathcal{M}e^{-x}\}(t)$ (Hazewinkle 2001).

A Cantor set T_∞, was rendered given by taking the interval $[0, 1]$ (set T_0), removing the open middle third (T_1), removing the middle third of each of the two remaining pieces (T_2), and continuing this procedure ad infinitum. It was therefore the set of points in the interval $[0, 1]$. The nth iteration of the Cantor for the YF , *Ae.. aegypti* model I was implemented in the Wolfram Language as CantorMesh[n]. The sequence of binary bits produced was 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, ... (whose nth term was given by $D(n, n) = P_n(3) \pmod 3$, where $D(n, n)$ is a (central) Delannoy number and $P_n(x)$ is a Legendre polynomial . The recurrence plot for this sequence was illustrated above. This produces the set of predicticted larval habitat s immature counts $\{x\}$ such that $x = \frac{c_1}{3} + \ldots + \frac{c_n}{3^n} + \ldots$, where c_n may equal 0 or 2 for each n. This is an infinite, perfect set. The total length of the line segments in the nth iteration was $\ell_n = \left(\frac{2}{3}\right)^n$, and the number of line segments is $N_n = 2^n$, so the length of each element in the YF model was $\epsilon_n = \frac{\ell}{N} = \left(\frac{1}{3}\right)^n$ and the capacity dimension was $d_{cap} = -\lim\limits_{\epsilon \to 0^+} \frac{\ln N}{\ln \epsilon} = \log_3 2 = \frac{\ln 2}{\ln 3} = 0.6319$.

The Mellin transform in the empiricial, geo-spatiotemporal, explanatorial YFV-related eco-epidemiolgical, clinical, field and remote-gesampled, geopredictive , eco-epidemiological, risk model was an integral transform that we regarded as the multiplicative version of the two-sided Laplace transform. The two-sided Laplace transform was then defined in terms of the Mellin transform by $\{\mathcal{B}f\}(s) = \{\mathcal{M}f(-\ln x)\}(s)$ for remotely regressively quantizing the explanatorial,time series dependent, geo-spatiotemporal, YFV-related, clinical, field or remote geosampled,ecogeoreferenced, forecasting, eco-epidemiological,probabilistic, risk model, covariate, paramter estimators. We assumed that the Mellin transform may be rendered from the two-sided Laplace transform by $\{\mathcal{M}f\}(s) = \{\mathcal{B}f(e^{-x})\}(s)$. The Mellin transform may be thought of as integrating using a kernel x^s with respect to the multiplicative Haar measure, $\frac{dx}{x}$, which is invariant under dilation $x \mapsto ax$, so that $\frac{d(ax)}{ax} = \frac{dx}{x}$. In he eco-epidemiological, time series dependent, YFV-related forecasting, risk model, the two-sided Laplace transform was integrated with respect to the additive Haar measure dx, which was a translation invariant, so that $d(x + a) = dx$. We then defined the Fourier transform in the forecasting model in terms of the Mellin transform and vice versa. We then defined the two-sided Laplace transform as $\{\mathcal{F}f\}(-s) = \{\mathcal{B}f\}(-is) = \{\mathcal{M}f(-\ln x)\}(-is)$. The process was reversed by employing $\{\mathcal{M}f\}(s) = \{\mathcal{B}f(e^{-x})\}(s) = \{\mathcal{F}f(e^{-x})\}(-is)$. The Mellin transform also connected the Newton series or binomial transform together with the Poisson generating, YFV-related, time series dependent function, by means of the Poisson–Mellin–Newton cycle.

The Poisson–Mellin–Newton cycle, noted by Flajolet et al. in 1985, is the observation that the resemblance of the Nørlund–Rice integral to the Mellin transform is not accidental, but is related by means of the binomial transform and the Newton series. In the empirical, geo-spatiotemporal, explanatorial, YFV-related, probablistically geosampled, eco-epidemiolgical, explanatorial, clinical, field and remote-gesampled, geopredictive, eco-epidemiological, risk model, we let $\{f_n\}$ be a sequence, and let $g(t)$ be the corresponding Poisson generating function, that is, we let $g(t) = e^{-t} \sum_{n=0}^{\infty} f_n t^n.$ Taking its Mellin transform[i.e., $\phi(s) = \int_0^\infty g(t) t^{s-1}\, dt,$] we then regained the original sequence by means of the Nørlund–Rice integral [i.e., $f_n = \frac{(-1)^n}{2\pi i} \int_\gamma \frac{\phi(s)}{\Gamma(-s)} \frac{n!}{s(s-1)\cdots(s-n)}\, ds$] where Γ was the gamma function. The gamma function is a component in various probability-distribution functions, and, as such, it is applicable in the fields of probability and statistics, as well as combinatorics(Hosmer and Lemeshew 2002).

The Riemann zeta function for the geo-spatiotemporal, empirical, YFV-related, eco-epidemiological, forecasting, risk model was defined as the analytic continuation of the function defined for $\sigma > 1$ by the sum of the preceding series in the empirical geosampled dataset. Leonhard Euler considered the above series in 1740 for positive integer values of s, and later Chebyshev extended the definition to real $s > 1$.

Our series was a prototypical Dirichlet series that converged absolutely to an analytic function for s such that $\sigma > 1$ and diverged for all other time series dependent, YFV-related, geosampled, clinical, field or remote-specified, covariate, parameter estimator, coefficient values of s. Riemann showed that the function defined by the series on the half-plane of convergence can be continued analytically to all complex values $s \neq 1$ (Hosmer and Lemeshew 2002). For $s = 1$ the series in the model was the harmonic series which diverged to $+\infty$, and $\lim_{s \to 1}(s-1)\zeta(s) = 1$. Thus, the Riemann zeta function in our geo-spatiotemporally-geosampled YFV-related eco-epidemiological, forecasting, risk model was a meromorphic function on the whole complex s-plane, which was then subsequently holomorphic everywhere except for a simple pole at $s = 1$ with residue 1.

Another connection with the primes in the eco-epidemiological, geosampled, dataset was provided by $d(n) = \sigma_0(n)$ for the geopredictive time series dependent, explanatorial regressor numerical values from 1 to n which was found to be asymptotic to $\frac{\sum_{k=1}^n d(k)}{n} \sim \ln n + 2\gamma - 1$. De la Vallée Poussin [1898] proved that, if a large number n is divided by all primes $\leq n$, then the average amount by which the quotient is less than the next whole number is γ. An elegant identity for γ in our YFV regression model was provided by $\gamma = \frac{S_0(z) - K_0(z)}{I_0(z)} - \ln\left(\tfrac{1}{2} z\right),$ where $I_0(z)$ was a modified Bessel function of the first kind, $K_0(z)$ was a modified Bessel function of the second kind, and $S_0(z) \equiv \sum_{k=0}^{\infty} \frac{\left(\tfrac{1}{2} z\right)^{2k} H_k}{(k!)^2},$ where H_n was a harmonic number. This provided an efficient iterative algorithm for γ by

computing $B_k = \dfrac{B_{k-1}\, n^2}{k^2}$, $A_k = \dfrac{1}{k}\left(\dfrac{A_{k-1}\, n^2}{k} + B_k\right)$, $U_k V_k = U_{k-1} + A_k$ and $V_k = V_{k-1} + B_k$ with $A_0 = -\ln n B_0 = 1 U_0 = A_0$ and $V_0 = 1$. Reformulating this identity rendered the limit

$$\lim_{n\to\infty}\left[\sum_{k=0}^{\infty}\frac{\left(\frac{n^k}{k!}\right)^2 H_k}{\sum_{k=0}^{\infty}\left(\frac{n^k}{k!}\right)^2} - \ln n\right] = \gamma$$

. Infinite products involving γ also arose from the Barnes G-function since we employed n.

In mathematics, the Barnes G-function $G(z)$ is a function that is an extension of superfactorials to the complex numbers which is related to the gamma function (see Borwein and Bailey 2003). This function provides $\displaystyle\prod_{n=1}^{\infty} e^{-1+1/(2\,n)}\left(1+\frac{1}{n}\right)^n = \frac{e^{1+\gamma/2}}{\sqrt{2\pi}}$ and $\displaystyle\prod_{n=1}^{\infty} e^{-2+2/n}\left(1+\frac{2}{n}\right)^n = \frac{e^{3+2\gamma}}{2\pi}$. The Barnes G-function was linearly defined then in our time-series explanatorial, geopredictive, eco-epidemiological, YFV-related, probablistically regression–based, risk model which was then defined by $G(z+1) = (2\pi)^{z/2}\exp(-(z(z+1)+\gamma z^2)/2) \times \displaystyle\prod_{n=1}^{\infty}\left[\left(1+\frac{z}{n}\right)^n \exp(-z + z^2/(2n))\right]$, where γ was the Euler–Mascheroni constant, $\exp(x) = e^x$, and $\prod$ was capital pi notation. The Euler-Mascheroni constant was rendered by the expressions $\gamma = -\Gamma'(1) = -\psi_0(1)$, where $\psi_0(x)$ was the symmetric limit form of $\gamma = \displaystyle\lim_{s\to 1^+}\sum_{n=1}^{\infty}\left(\frac{1}{n^s} - \frac{1}{s^n}\right)$ and $\gamma = \displaystyle\lim_{x\to\infty}\left[x - \Gamma\left(\frac{1}{x}\right)\right]$.

In mathematics, the digamma function is defined as the logarithmic derivative of the gamma function: $\psi(x) = \dfrac{d}{dx}\ln\Gamma(x) = \dfrac{\Gamma'(x)}{\Gamma(x)}$. where it is the first of the polygamma functions (Borwein and Bailey 2003). The digamma function, often denoted also as $\psi_0(x)$, is related to the harmonic numbers in that $\psi(n) = H_{n-1} - \gamma$ where H_n is the nth harmonic number, and γ is the Euler-Mascheroni constant (see Sondow and Zudilin 2006). The difference between the nth convergent in equation ($\diamond$) and γ in our explanatorial, probbalistic, geosampled, YFV-related, regression-based, eco-epidemiological, forecasting risk model was then rendered by $\displaystyle\sum_{k=1}^{n}\frac{1}{k} - \ln n - \gamma = \int_n^{\infty}\frac{x - \lfloor x\rfloor}{x^2}\,dx$, where $\lfloor x\rfloor$ was the floor function which satisfied the inequality $\dfrac{1}{2(n+1)} < \displaystyle\sum_{k=1}^{n}\frac{1}{k} - \ln n - \gamma < \dfrac{1}{2n}$. The floor function $\lfloor x\rfloor$, also called the greatest integer function or integer value (Spanier and Oldham 1987), gives the largest integer less than or equal to x.

Unfortunately, in many older and current works (e.g., Honsberger 1976, Steinhaus 1999, Shanks 1993; Ribenboim 1996; Hilbert and Cohn-Vossen 1999,; Hardy 1999,), the symbol $[x]$ is used instead of $\lfloor x\rfloor$ in time series related, probablistic, regression-related models. In fact, this notation harks back to Gauss in his third proof of quadratic reciprocity in 1808. However, because of the elegant symmetry of the floor function and ceiling function symbols $\lfloor x\rfloor$ and $\lceil x\rceil$ in the geo-spatiotemporally, geosampled, YFV-related, eco-epidemiological, explanatorial,

forecasting, probablistic, risk model and because $[x]$ is such a useful symbol when interpreted as an Iverson bracket, the use of $[x]$ denoted the floor function in the model. The symbol $[x]$ was used to denote the nearest integer function since it naturally fell between $\lfloor x \rfloor$ and $\lceil x \rceil$ symbols.

In mathematics, the Iverson bracket is a notation that denotes a number that is 1 if the condition in square brackets is satisfied, and 0 otherwise(Hazewinkle 2001). More exactly,

$$[P] = \begin{cases} 1 & \text{if } P \text{ is true;} \\ 0 & \text{otherwise.} \end{cases}$$

where P is a statement that can be true or false. The Iverson bracket in our eco-epidemiological, geo-spatiotemporally,geosampled clinical, field and remote-specified, probabilistic, YFV-related ,explanatorial, geopredictive, eco-epidemiological, risk model was converted a Boolean value to an integer value through the natural map $\textbf{false} \mapsto 0; \textbf{true} \mapsto 1$, which allowed counting to be represented as summation. For instance, the Euler phi function that counted the number of positive integers in our empirical, geosampled, geo-spatiotemporal, YFV-related, clinical, field and remote-specified, covariate, parameter estimator, coefficient values in the empiricial dataset was up to n which were also coincidentally was a coprime to n

$$\phi(n) = \sum_{i=1}^{n} [\gcd(i,n) = 1], \qquad \text{for } n \in \mathbb{N}^+.$$

which then was subsequently expressed by . More generally the notation allowed moving boundary conditions of summations (or integrals) as a separate factor into the summand, freeing up space around the summation operator while simutaneously allowing the geosampled, YFV-related, covariate, paramter estimators to be manipulated algebraically. For example, a YFV-related rendered explanatorial, time series dependent, eco-epidemiological, forecast threshold of

$$\sum_{1 \leq i \leq 10} i^2 = \sum_i i^2 [1 \leq i \leq 10].$$

maintained the first sum of the index i which was limited to be in the range 1 to 10. The second sum was allowed to range over all the explanatorial, geopredictive, time series dependent, geosampled integer values but only where i was strictly less than 1 or strictly greater than 10 and the summand was 0 which of corse did not contribute nothing to the sum in the regression model estimation process. Use of the Iverson bracket can permit easier manipulation of these expressions(Hosmer and Lemeshew 2000).

Another use of the Iverson bracket in the geo-spatiotemporally, geosampled YFV-related, explanatorial, eco-epidemiological, clinical, field and remote specified, forecasting, risk model

$$\sum_{\substack{1 \leq k \leq n \\ \gcd(k,n)=1}} k = \frac{1}{2} n \varphi(n)$$

was to simplify equations with special cases. For example, the formula was valid for $n > 1$ in our geopredictive, time series dependent, YFV-related, eco-epidemiological, regression model but was off by $1/2$ for $n = 1$. To get an identity valid for all the positive explanatorial, geopredictive, time series dependent, YFV-related, endemic transmission-oriented, eco-epidemiological, risk model$n$ all geosampled probablistic values for which $\phi(n)$ had to be defined and a correction term involving the Iverson bracket had to be added:

$$\sum_{\substack{1 \leq k \leq n \\ \gcd(k,n)=1}} k = \frac{1}{2} n (\varphi(n) + [n = 1])$$

. By so doing,the Kronecker delta notation was then a specific case of Iverson notation when the condition expressed equality (e.g., $\delta_{ij} = [i = j]$). The indicator function has a set membership as its condition: $\mathbf{I}_A(x) = [x \in A]$ (Hazewinkle 2001).

The simplest interpretation of the Kronecker delta in the time series dependent, eco-epidemiological, YFV-related forecasting, risk model, residual derivatives was then defined as the discrete version of the delta function which was subsequently defined by $\delta_{ij} \equiv \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j. \end{cases}$

The Kronecker delta was implemented in *Mathematica* as KroneckerDelta[i, j], as well as in a generalized form KroneckerDelta[$i, j, ...$] that returned 1 if all arguments was equal and 0 otherwise was rendered by the residual derivatives from the time series dependent, probablistic, YFV-related, eco-epidemiological, explanatorial, clinical, field and remote-specified, geosampled risk model. It had the contour integral $\delta_{mn} = \frac{1}{2\pi i} \oint_\gamma z^{m-n-1} \, dz,$ representation where γ was a contour corresponding to the unit circle and m and n where the geosampled integer covariate, paramter estimator, coefficient values. In three-space, the Kronecker delta satisfied the identities $\delta_{ii} = 3$, $\delta_{ij} \epsilon_{ijk} = 0$, $\epsilon_{ipq} \epsilon_{jpq} = 2\delta_{ij}$, $\epsilon_{ijk} \epsilon_{pqk} = \delta_{ip} \delta_{jq} - \delta_{iq} \delta_{jp}$, where Einstein summation was implicitly assumed, $i, j = 1$, 2, 3, and ϵ_{ijk} was the permutation symbol.

In mathematics, especially in applications of linear algebra to physics, the Einstein notation or Einstein summation convention is a notational convention that implies summation over a set of indexed terms in a formula, thus achieving notational brevity. As part of mathematics it is a notational subset of Ricci calculus; however, it is often used in applications in physics that do not distinguish between tangent and cotangent spaces. It was introduced to physics by Albert Einstein in 1916 mathematics, Ricci calculus constitutes the rules of index notation and manipulation for tensors and tensor fields(Hosmer and Lemeshew 2002). It is also the modern name for what used to be called the absolute differential calculus (the foundation of tensor calculus), developed by Gregorio Ricci-Curbastro in 1887–96, and subsequently popularized in a paper written with his pupil Tullio Levi-Civita in 1900.

A component of a tensor in the eco-epidemiological, explanatorial, clinical, field or remote geosampled, YFV-related, geo-spatiotemporally, georeferenced, eco-epidemiological, forecasting risk model was used as a coefficient of a basis element for the tensor space. The tensor is the sum of its components multiplied by their basis elements (Hazewinkle 2001).For the eco-epidemiological, YFV-related geopredictive, risk model tensors and tensor fields were expressed in terms of their components, and operations on tensors and tensor fields were expressed in terms of operations on their components. The description of tensor fields and operations on them in terms of their components was focus of our Ricci calculus calculations. Where needed, the notation extended to components of non-tensors, particularly multidimensional arrays(Hazewinle 2001).Henceforth, we defined an eco-epidemiological, probablistic, YFV-related, explanatorial, geo-spatiotemporally, geosampled, eco-epidemiological tensor as a linear sum of the tensor product of vector and covector basis elements. The resulting tensor components were labelled by indices of the basis. Each index had one possible value per dimension of the underlying vector space in the model residual derivatives. The number of indices equaled the order of the tensor.

For compactness and convenience, the notational convention implied certain things, in our geo-spatiotemporal, forecasting risk model notably that the summation over indices may be

repeated within a term employing a universal quantification over the free indices (i..e, those not so summed). Expressions in the notation of the Ricci calculus were interpreted as a set of simultaneous YFV-related geopredictive equations relating and their eco-epidemiological, components were classified as time series functions over a manifold. More specifically these functions of the coordinates on the manifold allowed intuitive manipulation of the model expressions with familiarity of only a limited set of rules.

Technically, the Kronecker delta was a tensor defined by the relationship $\delta_l^k \frac{\partial x_i'}{\partial x_k} \frac{\partial x_l}{\partial x_j'} = \frac{\partial x_i'}{\partial x_k} \frac{\partial x_k}{\partial x_j'} = \frac{\partial x_i'}{\partial x_j'}$, in the eco-epidemiological, YFV-related, forecasting, risk model since, by definition, the seasonal, geosampled explanatorial, geopredictive time series dependent, geocoordinates x_i and x_j were independent for $i \neq j$, $\frac{\partial x_i'}{\partial x_j'} = \delta_j'^i$, so $\delta'^i_j = \frac{\partial x_i'}{\partial x_k} \frac{\partial x_l}{\partial x_j'} \delta_l^k$, and δ_j^i was really a mixed second-rank tensor. This model output satisfied $\delta_{ab}{}^{jk} = \epsilon_{abi} \epsilon^{jki} = \delta_a^j \delta_b^k - \delta_a^k \delta_b^j, \delta_{abjk} = g_{aj} g_{bk} - g_{ak} g_{bj}, \epsilon_{aij} \epsilon^{bij} = \delta_{ai}{}^{bi} = 2\delta_a^b$. The sign function and Heaviside step function was also easily expressed in this notation as $\text{sgn}(x) - [x > 0] - [x < 0] H(x) = [x > 0]$. The floor and ceiling functions in the empirical geosampled, explanatorial, geopredictive, time series –related, probablistic, YFV-related, regression model was then expressed: $\lfloor x \rfloor = \sum_{n=-\infty}^{\infty} n[n \leq x < n+1], \lceil x \rceil = \sum_{n=-\infty}^{\infty} n[n-1 < x \leq n]$.

In mathematics, the Kronecker delta or Kronecker's delta, is a function of two variables, usually integers (Haight 1967). The function is 1 if the variables (e.g., empirical geosampled, explanatorial, geopredictive, time series dependent, YFV-related regressors) are equal, and 0 otherwise: $\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j, \end{cases}$ where the Kronecker delta δ_{ij} is a piecewise function of variables i and j. For example, $\delta_{1\,2} = 0$, whereas $\delta_{3\,3} = 1$. Further, in linear algebra, the identity matrix can be written as $(\delta_{ij})_{i,j=1}^n$ and the inner product of vectors can be written as $a \cdot b = \sum_{ij} a_i \delta_{ij} b_j$. The Kronecker delta is used in many areas of mathematics, physics and engineering, primarily as an expedient to convey in a single equation what might otherwise take several lines of text and the tracheotomy of the reals which may then be expressed as $[a < b] + [a = b] + [a > b] = 1$ (Hosmer and Lemeshew 2000). The Macaulay brackets in the YFV-related, eco-epidemioloigical, forecasting, risk model were then be expressed $\{x\} = x[x \geq 0]$.

The floor function satisfied the identity $\lfloor x + n \rfloor = \lfloor x \rfloor + n$ for all the geosampled, YFV-related, explanatorial, covariate, parameter estimator coefficients n. A number of explanatorial, geometric-like sequences with a floor function in the numerator of our eco-epidemiological, probablistic, risk model were done analytically. For instance, the sums of the form $\sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{t^n}$ were qualitatively quantiated analytically for the rational x in the model derivatives. For $x = 1/m$ a unit

fraction $\sum_{n=1}^{\infty} k^{-n} \left\lfloor \frac{n}{m} \right\rfloor = \frac{k}{(k-1)(k^m-1)}$ was generated. Sums of this form lead to the Devil's staircase-like behavior.

A plot of the Devil's staircase-like behavior is a map winding number W resulting from mode locking as a function of Ω for the circle map $\theta_{n+1} = \theta_n + \Omega - \frac{K}{2\pi} \sin(2\pi\theta_n)$ with $K=1$ (Bailey and Crandall 2001).Interestingly, since the circle explanatorial, geopredictive, time series dependent, YFV-related, regressor map became a mode-locked where the vulnerability risk map winding number was independent of the initial starting argument θ_0. At each seasonal, geosampled, YFV-related observation Ω, the map winding number was some geosampled, covariate, paramter estimator, coefficient value. The result was a monotonic increasing "staircase" for which the rational, geosampled, explanatorial, clinical, field and remote covariate coefficient had the largest steps. The Devil's staircase continuously mapped the interval $[0,1]$ onto $[0,1]$, but it was constant almost everywhere. For $K=1$, the measure of quasiperiodic states (Ω irrational) on the Ω-axis in the YFV-related, risk model output was 0, and the measure of mode-locked state was 1. The dimension of the Devil's staircase in the model $\approx 0.8700 \pm 3.7 \times 10^{-4}$. For irrational $\alpha > 0$, continued fraction convergent in the forecasts p_n/q_n, and $\epsilon_n \equiv q_n\alpha - p_n$,

$$\lfloor n\alpha + \epsilon_N \rfloor = \begin{cases} \lfloor n\alpha \rfloor & \text{for } n < q_{N+1} \\ \lfloor n\alpha \rfloor + (-1)^N & \text{for } n = q_{N+1} \end{cases}$$

This lead to the rather amazing relating sums of the floor function of multiples of α based on the continued fraction of α by

$$\sum_{n=1}^{\infty} \lfloor n\alpha \rfloor z^n = \frac{p_0 z}{(1-z)^2} + \sum_{n=0}^{\infty} (-1)^n \frac{z^{q_n} z^{q_{n+1}}}{(1-z^{q_n})(1-z^{q_{n+1}})}$$

The symbol γ was then $\gamma \equiv e^{\gamma} \approx 1.781072$. This led to the radical representation of the geosampled, georefernced, YFV-related explanatorial, geopredictive, regressed covariate, parameter estimator, coefficient values as $e^{\gamma} = \left(\frac{2}{1}\right)^{1/2} \left(\frac{2^2}{1\cdot3}\right)^{1/3} \left(\frac{2^3\cdot4}{1\cdot3^2}\right)^{1/4} \left(\frac{2^4\cdot4^4}{1\cdot3^6\cdot5}\right)^{1/5}$ which was related to the double series $\gamma = \sum_{n=1}^{\infty} \frac{1}{n} \sum_{k=0}^{n-1} (-1)^{k+1} \binom{n-1}{k} \ln(k+1)$ and $\binom{n}{k}$ a binomial coefficient. Thereafter, another proof of product in the geo-spatiotemporal, geo-sampled, YFV-related, regression-related, forecasting, eco-epidemiological, risk model was provided by the equation $\frac{\pi}{2} = \left(\frac{2}{1}\right)^{1/2} \left(\frac{2^2}{1\cdot3}\right)^{1/4} \left(\frac{2^3\cdot4}{1\cdot3^3}\right)^{1/8} \left(\frac{2^4\cdot4^4}{1\cdot3^6\cdot5}\right)^{1/16}$. The solution was then made even clearer by changing $n \to n+1$. Both these regression-based formulas were also analogous to the product for e which was then rendered by the computation: $e = \left(\frac{2}{1}\right)^{1/1} \left(\frac{2^2}{1\cdot3}\right)^{1/2} \left(\frac{2^3\cdot4}{1\cdot3^3}\right)^{1/3} \left(\frac{2^4\cdot4^4}{1\cdot3^6\cdot5}\right)^{1/4}$.

Importantly, the distribution of a empirically probablistic, YFV-related, random, explanatorial, endemic, transmission-oriented, LULC, NDVI and other remotely ecogeographical covariate paramter estimator geosampled at the Gulu eco-epidemiological study site was discrete. Further, X was a discrete random variable, if $\sum_u \Pr(X = u) = 1$. It then followed that such a random variable could assume only a finite or countably infinite number of the empirically regressed, explanatorial, time series dependent, geopredictive endemic, transmission-oriented, geosampled coefficient values. For the number of potential values to be countably infinite even though their probabilities sum to 1 requires that the probabilities decline to zero fast enough: for

example, if $\Pr(X = n) = \frac{1}{2^n}$ for n = 1, 2, ..., the sum of probabilities 1/2 + 1/4 + 1/8 + ... = 1 would be efficiently quantified (Hosmer and Lemeshew 2000).

The Poisson probability model allowed the mean of the geo-sampled, geopredictive YFV-related, explanatorial, clinical, field and remote specified, eco-epidemiological, observational indicators to depend on a linearized time series dependent, geopredictive, endemic, transmission-oriented ,covariate, paramter estimator, coefficient through a nonlinear link function. This Poisson regression feature allowed the response probability distribution to be any member of an exponential family of YFV-related, seasonal, endemic, transmission-oriented, explanatorial distribution. The response variable (i.e., Y) then represented prevalence rates which was numeric and had nonnegative integer values.

We constructed a Poisson regression model employing the geo-spatiotemporal, seasonal-geosampled, YFV-related, covariate, paramter estimator coefficient, measurement values. Our model was generalized by introducing an unobserved heterogeneity term for each geosampled, time series observation i. The weights were then assumed to differ randomly in a manner that was not fully accounted for by the other seasonal-geosampled covariate parameter estimators. The process was formulated as $E(y_i \mid x_i, \tau_i) = \mu_i \tau_i = e^{x_i'\beta + \varepsilon_i}$ where the unobserved heterogeneity term $\tau_i = e^{\varepsilon_i}$ was independent of the vector of regressors x_i. Then the distribution of y_i was conditional on x_i and had a Poisson specification with conditional mean and conditional variance $\mu_i \tau_i : f(y_i \mid x_i, \tau_i) = \frac{\exp(-\mu_i \tau_i)(\mu_i \tau_i)^{y_i}}{y_i!}$. We then let $g(\tau_i)$ be the pdf of τ_i. Then, the distribution $f(y_i \mid x_i)$ was no longer conditional on τ_i. Instead it was obtained by integrating $f(y_i \mid x_i, \tau_i)$ with respect to $g(\tau_i)$: $f(y_i \mid x_i) = \int_0^\infty f(y_i \mid x_i, \tau_i) g(\tau_i) d\tau_i$.

```
lines(C,lty=2)
C=trans3d(x,y,z,mat)
lines(C,col="blue")}
```

We found that an analytical solution to this integral existed in our probablistic, YFV-related, model when τ_i was assumed to follow a gamma distribution. The model also revealed that y_i, was the vector of the geosampled predictor covariate, paramter estimator coefficients while x_i, was independently Poisson distributed with $P(Y_i = y_i \mid x_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, y_i = 0,1,2,...$ and the mean parameter — that is, the mean number of geosampling events per geo-spatiotemporal period — was given by $\mu_i = \exp(x_i'\beta)$ where β was a $(k+1) \times 1$ parameter vector. The intercept in the model was β_0 and the coefficients for the k regressors were $\beta_1, ..., \beta_k$. Taking the exponential of $x_i'\beta$ ensured that the mean parameter μ_i was nonnegative. Thereafter, the conditional mean was provided by $E(y_i \mid x_i) = \mu_i = \exp(x_i'\beta)$.

The clinical, field and remote-geosampled, covarite parameter estimators were then evaluated using $\ln[E(y_i \mid x_i)] = \ln(\mu_i) = x_i'\beta$. Note, that the conditional variance of the count random

variable was equal to the conditional mean (i.e., equidispersion) in the eco-epidemiological, YFV model [i.e., , $V(y_i|x_i)=E(y_i|x_i)=\mu_i$]. In a log-linear model the logarithm of the conditional mean is linear(Hosmer and Lemeshew 2002). The marginal effect of any YFV-related, probbalistic, geosampled regressor in the geo-spatiotemporal eco-epidemiological, time series dependent, risk model was then provided by $\frac{\delta E(y_i|x_i)}{\delta x_{ji}}=\exp(x_i'\beta)\beta_j=E(y_i|x_i)\beta_j$. Thus, a one-unit change in the jth regressor in the model led to a proportional change in the conditional mean $E(y_i|x_i)$ of β_j.

The standard estimator for our Poisson probability risk model was the MLE. Since the observations were independent, the log-likelihood function in the model was then: $\sum_{i=1}^{N}(-\mu_i+y_i\ln\mu_i-\ln y_i!)=\sum_{i=1}^{N}(-e^{x_i'\beta}+y_i x_i'\beta-\ln y_i!)$. Given the geosampled dataset of the clinical, field and remote YFV-related, covariate parameter estimators (i.e., θ) and an input vector x, the mean of the predicted Poisson distribution was then provided by $E(Y|x)=e^{\theta'x}$. By so doing, the Poisson distribution's pmf was then rendered by $p(y|x;\theta)=\frac{e^{y(\theta'x)}e^{-e^{\theta'x}}}{y!}$ The pmf in a targeted geo-spatiotemporal, predictive, seasonal, eco-epidemiological, probablistic, risk model can be the primary means for defining a discrete probability distribution, and, as such, functions could exist for either scalar or multivariate, multitemproal, field geosampled, random variables, given that the distribution is discrete(Jacob et al. 2013). Gu and Novak (2005) found that a targeted geo-spatiotemporal, predictive, seasonal, risk model is vital for IVM.

Since the geosampled data consisted of m vectors $x_i \in \mathbb{R}^{n+1}, i=1,\ldots,m,$ along with a set of m values $y_1,\ldots,y_2 \in \mathbb{R}$ then, for the covariate parameter estimator's θ, the probability of attaining a particular set of the geosampled, explanatorial, clinical, field or remote observations was provided by the equation $p(y_1,\ldots,y_m|x_1,\ldots,x_m;\theta)=\prod_{i=1}^{m}\frac{e^{y_i(\theta'x_i)}e^{-e^{\theta'}x_i}}{y_i!}$ .Consequently, we found the set of θ that made this probability as large as possible in the model probilistic regression estimates. To do this, the explanatorial equation was first rewritten as a likelihood function in terms of θ[i.e., $p(y_1,\ldots,y_m|x_1,\ldots,x_m;\theta)=\prod_{i=1}^{m}\frac{e^{y_i(\theta'x_i)}e^{-e^{\theta'}x_i}}{y_i!}$].Note the expression on the right hand side in our model had not actually changed. Next, we used a log-likelihood [i.e., $\ell(\theta|X,Y)=\log L(\theta|X,Y)=\sum_{i=1}^{m}(y_i(\theta'x_i)-e^{\theta'x_i}-\log(y_i!))$. Because the logarithm is a monotonically increasing function, the logarithm of a function achieves its maximum value at the same points as the function itself, and, hence, the log-likelihood can be used in place of the likelihood in MLE and related techniques (Hosmer and Lemeshew 2002). Finding the maximum of a function in a time series dependent, geopredictive, eco-epidemiological, risk model often involves taking the derivative of a function and solving for the parameter estimator being maximized, and this is often easier when the function being maximized is a log-likelihood rather than the original likelihood function (Jacob et al. 2005).

We noticed that the that the geo-spatiotemporal, geosampled,YFV-related, eco-epidemiological, explanatorial, clinical, field and remote, probablistically regressed, parameters θ only appeared in the first two terms of each derivative in the summation. Therefore, given that

we were only interested in finding the optimal value for θ in the geopredictive, YFV-related, regression, time series dependent, probabilistic, eco-epidemiological, clinical, field or remote-specified, risk model we dropped the $y_i!$ and simply wrote $\ell(\theta \mid X, Y) = \sum_{i=1}^{m} \left(y_i (\theta' x_i) - e^{\theta' x_i} \right)$. Thereafter, to find a maximum, we solved an equation $\dfrac{\partial \ell(\theta \mid X, Y)}{\delta \theta} = 0$ which had no closed-form solution. However, the negative log-likelihood (LL)[i.e., $-\ell(\theta \mid X, Y)$] was a convex function, and so standard convex optimization was applied to find the optimal value of θ.

We found that given the probbality Poisson process in our YFV-related, regression, eco-epidemiological, probabilistic, risk model had the limit of a binomial distribution which was $p_p(n \mid N) = \dfrac{N!}{n!(N-n)!} p^n (1-p)^{N-n}$. Viewing the distribution as a function of the expected number of successes[i.e., $v \equiv Np$] in the model derivatives, instead of the sample size N for fixed P, then rendered the equation (2.1) which then became $p_{v/N}(n \mid N) = \dfrac{N!}{n!(N-n)!} \left(\dfrac{v}{N} \right)^n \left(1 - \dfrac{v}{N} \right)^{N-n}$ Our model revealed that as the sample size n became larger, the distribution approached P when the following equations aligned $\lim_{N \to \infty} p_p(n \mid N)$ $\lim_{N \to \infty} \dfrac{N(N-1)\dots(N-n+1)}{n!} \dfrac{v^n}{N^n} \left(1 - \dfrac{v}{N} \right)^N \left(1 - \dfrac{v}{N} \right)^{-n}$, $\lim_{N \to \infty} \dfrac{N(N-1)\dots(N-n+1)}{N^n} \dfrac{v^n}{n!} \left(1 - \dfrac{v}{N} \right)^N \left(1 - \dfrac{v}{N} \right)^{-n}$

$1 \cdot \dfrac{v^n}{n!} \cdot e^{-v} \cdot 1$ and $\dfrac{v^n e^{-v}}{n!}$. Note, that the sample size n had completely dropped out of the probability function, which had the same functional form for all values of v in the YFVmodel.

Thereafter, as expected, the probability Poisson regression distribution was normalized so that the sum of probabilities was equal to 1, since $\sum_{n=0}^{\infty} P_v(n) = e^{-v} \sum_{n=o}^{\infty} \dfrac{v^n}{n!} = e^{-v} e^v = 1$. The ratio of probabilities was then provided by the equation $\dfrac{P_v(n = i+1)}{P(n = i)} = \dfrac{\dfrac{v^{i+1} e^{-v}}{(i+1)!}}{\dfrac{e^{-v} v^i}{i!}} = \dfrac{v}{i+1}$. The YFV model revealed that the Poisson distribution reached a maximum when $\dfrac{dP_v(n)}{dn} = \dfrac{e^{-v} n (\gamma - H_n + \ln v)}{n!} = 0$ where g was the Euler-Mascheroni constant and H_n was a harmonic number, leading to the equation $\gamma - H_n + \ln v = 0$ which could not be solved exactly for n.

Next, the moment-generating function of the geospatiotemporal, geosampled, probablistic, YFV-related, eco-epidemiological, explanatorial, time series dependent, Poisson probability distribution was given by $M = M = e^{-v} e^{v e^t} = e^{v(e^t - 1)}$, $M = M = v e^t e^{v(e^t - 1)}$ and $M = M = \left(v e^t \right)^2 e^{v(e^t - 1)} + v e^t e^{v(e^t - 1)}$, when the equation

$R = v\left(e^t - 1\right)$, $R' = v e^t$ so $R = R'(0) = v$. The raw moments were also computed directly by summation, which yielded an unexpected connection with the exponential polynomial $\varphi_n(x)$ and Stirling numbers of the second kind[i.e. $\phi_n(x) = \sum_{k=0}^{\infty} \frac{e^{-x} x^k}{k!} k^n = \sum_{k=1}^{n} x^k S(n,k)$] which in this research was the Dobiński's formula.

In combinatorial mathematics, Dobinski's formula states that the number of partitions of a set of n members is $\frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}$. This number has come to be called the nth Bell number B_n, where the proof is rendered as an adaptation to probabilistic language(see Hosmer and Lemeshew 2002). In our YFV-related, probbalistic, time series dependnet, explanatorial, clinical, field or remote geosampled, eco-epidemiological, forecasting regression model the formula $\phi_n(x) = \sum_{k=0}^{\infty} \frac{e^{-x} x^k}{k!} k^n = \sum_{k=1}^{n} x^k S(n,k)$ was viewed as a particular case, for $x=0$, employing the relation $\frac{1}{e} \sum_{k=x}^{\infty} \frac{k^n}{(k-x)!} = \sum_{k=0}^{n} \binom{n}{k} B_k x^{n-k}$. The expression given by the model's Dobinski's formula was then revealed as the n th moment of the Poisson distribution with expected value 1. This value was derived by dividing the generating function formula for a Stirling number of the second kind $S(n,k)$ by $m!$, yielding $\frac{m^n}{m!} = \sum_{k=1}^{n} \frac{S(n,k)}{(m-k)!}$. Then $\sum_{m=1}^{\infty} \frac{m^n}{m!} \lambda^m = \left(\sum_{k=1}^{n} S(n,k) \lambda^k \right)\left(\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \right)$, and $\sum_{k=1}^{n} S(n,k) \lambda^k = e^{-\lambda} \sum_{m=1}^{\infty} \frac{m^n}{m!} \lambda^m$. Dobinski's formula was the number of partitions of a set of the geosampled, probablistic, YFV-related, covariate parameter estimator size (i.e. n) which equaled the nth moment of that distribution.

We partitioned a set of n elements into M non-empty sets (i..e, m set blocks). For example, the empirical, YFV-related, explanatorial, clinical, field and remote geosampled datset was transformed into n elements into m nonempty sets (i.e., m set blocks), (i.e., a Stirling set number) . For example, the empirical geosampled, dataset of YFV-related explanatorial, time series-related regressors $\{1, 2, 3\}$ was partitioned into three subsets in one way: $\{\{1\}, \{2\}, \{3\}\}$; into two subsets in three ways: $\{\{1, 2\}, \{3\}\}$, $\{\{1, 3\}, \{2\}\}$, and $\{\{1\}, \{2, 3\}\}$; and into one subset in one way: $\{\{1, 2, 3\}\}$. The Stirling numbers of the second kind have been variously denoted $S(n, m)$ (by Jacob et al. (2012) $S_n^{(m)}$, $\mathfrak{S}_n^m$ (Jordan 1965), $S_2(n, m)$, or Knuth's notation $\left\{ \begin{matrix} n \\ m \end{matrix} \right\}$ (Graham et al. 1994; Knuth 1997). The Stirling numbers for qualitatively regressively quantiating the explanatorial, clinical, field or remote geosampled covariate, paramter estimators of the second kind can be computed from the sum $S(n,k) = \frac{1}{k!} \sum_{l=0}^{k} (-1)^l \binom{k}{l} (k-l)^n$, with $\binom{n}{k}$ a binomial coefficient, or the generating functions $x^n = \sum_{m=0}^{n} S(n, m)(x)_m$ $\sum_{m=0}^{n} S(n, m) x(x-1) \cdots (x-m+1)$, where $(x)_m$ is the falling factorial (Roman 1984, pp. 60 and 101). For qualitatively quantiating the Stirling numbers of the second

kind we used the Wolfram Language as StirlingS2[n, m], and denoted $S_n^{(m)}$. Designed for the new generation of programmers, the Wolfram Language has a vast depth of built-in algorithms and knowledge, all automatically accessible through its unified symbolic language (http://www.wolfram.com/language/).

First, according to Stirling's formula, we can replace the factorial of a large number n with the approximation:

$$n! \simeq n^n e^{-n}\sqrt{2\pi n} \qquad \text{as } n \to \infty.$$

Thus,

$$\binom{n}{k}p^k q^{n-k} = \frac{n!}{k!(n-k)!}p^k q^{n-k}$$

$$\simeq \frac{n^n e^{-n}\sqrt{2\pi n}}{k^k e^{-k}\sqrt{2\pi k}(n-k)^{n-k}e^{-(n-k)}\sqrt{2\pi(n-k)}}p^k q^{n-k}$$

$$= \sqrt{\frac{n}{2\pi k(n-k)}}\frac{n^n}{k^k(n-k)^{n-k}}p^k q^{n-k}$$

$$= \sqrt{\frac{n}{2\pi k(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}$$

Next, use the approximation $\frac{k}{n} \to P$ to match the root above to the desired root on the right-hand side.

$$\binom{n}{k}p^k q^{n-k} \simeq \sqrt{\frac{1}{2\pi n\frac{k}{n}\left(1-\frac{k}{n}\right)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}$$

$$\simeq \frac{1}{\sqrt{2\pi npq}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k} \qquad p+q=1$$

Finally, we rewrote the expression in the yellow fever model as an exponential and use s the Taylor Series approximation for ln(1+x): $\ln(1+x) \simeq x - \frac{x^2}{2} + \frac{x^3}{3} - ...$ n Note that k cannot be fixed or it would quickly fall outside the range of interest as $n \to \infty$. What is needed is to let k vary but always be a fixed number of standard deviations from the mean, so that it is always associated with the same point on the <u>standard normal distribution</u> (see Hosmer and Lemeshew 2002) . For the yellow fever model we defined $k = np + x\sqrt{npq}$ for some fixed x. Then when $x = 1$, k we had 1 standard deviation from the mean. From this definition we have the approximations $k \to np$ and $\frac{k}{n} \to P$ as $n \to \infty$.However, the left-hand side requires that k be an integer. Keeping the notation but assuming that k is the nearest integer given by the definition, this is seen to be inconsequential in the limit by noting that as $n \to \infty$ the change in x required to

make k an integer becomes small and successive integer values of k produce converging values on the right-hand side:

$$\frac{\frac{1}{\sqrt{2\pi npq}}e^{\frac{-(k+1-np)^2}{2npq}}}{\frac{1}{\sqrt{2\pi npq}}e^{\frac{-(k-np)^2}{2npq}}} = e^{\frac{2np-2k-1}{2npq}} \simeq e^{\frac{-x}{\sqrt{npq}}} \simeq e^0 = 1 \qquad \textbf{as } n \to \infty.$$

The model construction consisted of transforming the left-hand side to the right-hand side by three approximations such that:

$$\binom{n}{k}p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}}\exp\left\{\ln\left(\frac{np}{k}\right)^k + \ln\left(\frac{nq}{n-k}\right)^{n-k}\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-k\ln\left(\frac{k}{np}\right) + (k-n)\ln\left(\frac{n-k}{nq}\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-k\ln\left(\frac{np+x\sqrt{npq}}{np}\right) + (k-n)\ln\left(\frac{n-np-x\sqrt{npq}}{nq}\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-k\ln\left(1+x\sqrt{\frac{q}{np}}\right) + (k-n)\ln\left(1-x\sqrt{\frac{p}{nq}}\right)\right\} \qquad p+q=1$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-k\left(x\sqrt{\frac{q}{np}} - \frac{x^2 q}{2np} + \ldots\right) + (k-n)\left(-x\sqrt{\frac{p}{nq}} - \frac{x^2 p}{2nq} - \ldots\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{(-np-x\sqrt{npq})\left(x\sqrt{\frac{q}{np}} - \frac{x^2 q}{2np} + \ldots\right) + (np+x\sqrt{npq}-n)\left(-x\sqrt{\frac{p}{nq}} - \frac{x^2 p}{2nq} - \ldots\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{(-np-x\sqrt{npq})\left(x\sqrt{\frac{q}{np}} - \frac{x^2 q}{2np} + \ldots\right) - (nq-x\sqrt{npq})\left(-x\sqrt{\frac{p}{nq}} - \frac{x^2 p}{2nq} - \ldots\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{\left(-x\sqrt{npq} + \frac{1}{2}x^2 q - x^2 q + \ldots\right) + \left(x\sqrt{npq} + \frac{1}{2}x^2 p - x^2 p - \ldots\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-\frac{1}{2}x^2 q - \frac{1}{2}x^2 p - \ldots\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}\exp\left\{-\frac{1}{2}x^2(p+q) - \ldots\right\}$$

$$\simeq \frac{1}{\sqrt{2\pi npq}}\exp\left\{-\frac{1}{2}x^2\right\}$$

$$= \frac{1}{\sqrt{2\pi npq}}e^{\frac{-(k-np)^2}{2npq}}$$

The Stirling numbers in the geospatiotemporally, geosampled, YFV-related, explanatorial, clinical, field or remote-geosampled eco-epidemiological, risk model of the second kind for three elements were $S(3,1)=1$, $S(3,2)=3$, $S(3,3)=1$. Since a set of n elements can only be partitioned in a single way into 1 or n subsets (Griffith 2003), we employed $S(n,1)=S(n,n)=1$. The triangle of Stirling numbers of the second kind was

1
1 1
1 3 1
1 7 6 1
1 15 25 10 1
1 31 90 65 15 1 where the nth row in the eco-peimdeiological, risk model corresponded to the coefficients of the Bell polynomial $\Phi_n(x)$ (see Figure 5). In the model $\sum_{n=k}^{\infty} S(n,k) \frac{x^n}{n!} = \frac{1}{k!}(e^x - 1)^k$ and

$\sum_{n=1}^{\infty} S(n,k) x^n = \sum_{n=k}^{\infty} S(n,k) x^n = \frac{x^k}{(1-x)(1-2x)\cdots(1-kx)} = \frac{(-1)^k}{\binom{x-1}{x}_k}$ for $z < 1/m$ (Abramowitz and Stegun 1972, Stanley 1997,), where $(x)_m$ was a Pochhammer symbol. Another generating function was given by $\sum_{k=1}^{n} S(n,k)(k-1)! z^k = (-1)^n \mathrm{Li}_{1-n}(1+1/z)$ for $n \geq 2$, where $\mathrm{Li}_n(z)$ was the polylogarithm.

Figure 5 Stirling numbers of the second kind intimately connected with the Poisson distribution using the Dobiński's formula $B_n(x) = \sum_{k=0}^{n} x^k S(n,k)$ **where** $B_n(x)$**is a geo-spatiotemporal YFV- Bell polynomial**
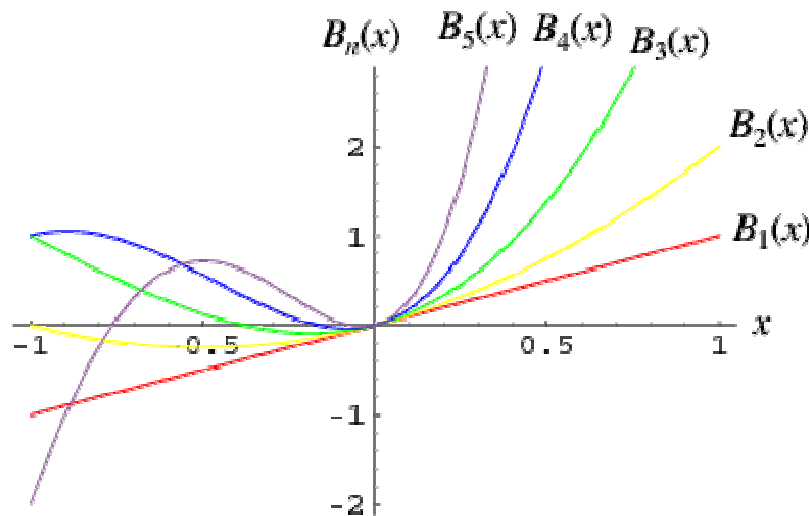
The polylogarithm $\mathrm{Li}_n(z)$, was the function $\mathrm{Li}_n(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^n}$ in the geo-spatiotemporal, geosampled , YFV-related eco-epidemiological, explanatorial, clinical, field and remote-specified eco-epidemiological, risk model defined in the complex plane. Its definition on the whole complex plane then followed uniquely via analytic continuation. Note that the similar notation $\mathrm{Li}(z)$ was emplouyued for computing the logarithmic integral. The polylogarithm was also denoted $F(z,n)$ which was equal to $\mathrm{Li}_n(z) = z\,\Phi(z,n,1)$, where $\Phi(z,n,a)$ was the Lerch transcendent .

The Lerch transcendent is generalization of the Hurwitz zeta function and polylogarithm function. Many sums of reciprocal powers can be expressed in terms of it. We defined this transcendent in the geo-spatiotemporal, geosampled, eco-epidemiological, explanatorial, clinical, field and remote, interpolatable, YFV-related eco-epidemiological risk model by

$$\Phi(z, s, a) = \sum_{k=0}^{\infty} \frac{z^k}{(a+k)^s},$$

for $|z| < 1$ and $a \neq 0, -1$. The polylogarithm was implemented in the Wolfram Language as PolyLog[n, z]. The polylogarithm arose in the closed form of the integrals of the Fermi-Dirac distribution

$$\int_0^{\infty} \frac{k^s \, dk}{e^{k-\mu}+1} = -\Gamma(s-1)\mathrm{Li}_{1+s}(-e^{\mu}),$$

in the model where $\Gamma(z)$ waSs the gamma function, and the Bose-Einstein distribution

$$\int_0^{\infty} \frac{k^s \, dk}{e^{k-\mu}-1} = \Gamma(s+1)\,\mathrm{Li}_{1+s}(e^{\mu}).$$

The Fermi-Dirac distribution arose in the study of half-integer as $P(k) = \dfrac{k^s}{e^{k-\mu}+1}$ whose integral was given by

$$\int_0^{\infty} \frac{k^s \, dk}{e^{k-\mu}+1} = e^{\mu}\,\Gamma(s+1)\,\Phi(-e^{\mu}, s+1, 1) = -\Gamma(s+1)\,\mathrm{Li}_{1+s}(-e^{\mu}),$$

where $\Phi(z, s, a)$ was the Lerch transcendent and $\mathrm{Li}_n(z)$ was a polylogarithm. The $z = 1$ in the geo-spatiotemporal, geosampled, YFV-related, eco-epidemiological, explanatorial, clinical, field or remote-specified, geosampled, probabilistic, risk model reduced to $\mathrm{Li}_s(1) = \zeta(s)$, where $\zeta(s)$ was the Riemann zeta function. Note, that the meaning of $\mathrm{Li}_s(z)$ for fixed complex s was not completely well-defined in the model resdiual forecasts since it was dependent on how s was quantitated in four-dimensional (s, z)-space.

The polylogarithm of negative integer order arose in sums of the form

$$\sum_{k=1}^{\infty} k^n r^k = \mathrm{Li}_{-n}(r) = \frac{1}{(1-r)^{n+1}} \sum_{i=0}^{n} \binom{n}{i} r^{n-i},$$

where $\binom{n}{i}$ was an Eulerian number in the YFV model. Polylogarithms also arose in sum of generalized harmonic numbers $H_{n,r}$ as

$$\sum_{n=1}^{\infty} H_{n,r} z^n = \frac{\mathrm{Li}_r(z)}{1-z}$$

in the YFV-related eco-epidemiological risk model derivatives for $|z| < 1$. Special forms of low-order polylogarithms in the model also included $\mathrm{Li}_{-2}(x) = \dfrac{x(x+1)}{(1-x)^3}$, $\mathrm{Li}_{-1}(x) = \dfrac{x}{(1-x)^2}$, $\mathrm{Li}_0(x) = \dfrac{x}{1-x}$, $\mathrm{Li}_1(x) = -\ln(1-x)$. At arguments -1 and 1, the general geo-spatiotemporal, regressively quantitated polylogarithms became $\mathrm{Li}_n(-1) = -\eta(n)$ and $\mathrm{Li}_n(1) = \zeta(n)$, where $\eta(x)$ was the Dirichlet function and $\zeta(x)$ was the Riemann zeta function. The polylogarithm for argument $1/2$ was also evaluated analytically for n when $\mathrm{Li}_1\left(\tfrac{1}{2}\right) = \ln 2$, $\mathrm{Li}_2\left(\tfrac{1}{2}\right) = \tfrac{1}{12}\left[\pi^2 - 6(\ln 2)^2\right]$ and $\mathrm{Li}_3\left(\tfrac{1}{2}\right) = \tfrac{1}{24}\left[4(\ln 2)^3 - 2\pi^2 \ln 2 + 21\zeta(3)\right]$.

The derivative of the geo-spatiotemporal, eco-epidemiological, georeferncable, explanatorial, clinical, field and remote geosampled, time series dependent, YFV model polylogarithm was itself a polylogarithm, $\dfrac{d}{dx}\mathrm{Li}_n(x) = \dfrac{1}{x}\mathrm{Li}_{n-1}(x)$. Bailey *et al.* showed that

$$\frac{\mathrm{Li}_m\left(\frac{1}{64}\right)}{6^{m-1}} - \frac{\mathrm{Li}_m\left(\frac{1}{8}\right)}{3^{m-1}} - \frac{2\,\mathrm{Li}_m\left(\frac{1}{4}\right)}{2^{m-1}} + \frac{4\,\mathrm{Li}_m\left(\frac{1}{2}\right)}{9} - \frac{5\,(-\ln 2)^m}{9\,m!}$$

$$+ \frac{\pi^2\,(-\ln 2)^{m-2}}{54\,(m-2)!} - \frac{\pi^4\,(-\ln 2)^{m-4}}{486\,(m-4)!} - \frac{403\,\zeta(5)\,(-\ln 2)^{m-5}}{1296\,(m-5)!} = 0.$$

A number of remarkable identities exist for polylogarithms, including the amazing identity satisfied by $\mathrm{Li}_{17}\left(\alpha_1^{-17}\right)$, where $\alpha_1 = \left(x^{10} + x^9 - x^8 - x^6 - x^5 - x^4 - x^3 + x + 1\right)_2 \approx 1.17628$ where the smallest Salem constant, i.e., the largest positive root of the polynomial in Lehmer's Mahler measure problem may be efficiently quantitated (Cohen *et al.* 1992; Bailey and Broadhurst 1999; Borwein and Bailey 2003).

We used the Pochhammer symbol $(x)_n$ to denote the falling factorial $(x)_n = x(x-1)(x-2)...(x-n+1)$ in the YFV-related, eco-epidemiological, time series dependent, forecasting, probablistic, risk model. If x and n are nonnegative integers, [i.e., $0 \le n \le x$], then $(x)_n$ is the number of one-to-one functions that map a size-n set into a size-x set(Cressie 1993). At this junction we let f be any function from a size-n set A into a size-x set B. Thus, in the eco-epidemiological, YFV-related, forecasting, risk model residual forecasts $u \in B$ .We then let $f^{-1}(u) = \{v \in A : f(v) = u\}$. Then $\{f^{-1}(u) : u \in B\}$ was a partition of A. This equivalence relation was the "kernel" of the function f.

Any function from A into B factors into one function that maps a member of A to that part of the kernel to which it belongs, and another function, which is necessarily one-to-one, that maps the kernel into B(Jensen 2005). The first of these two factors was completely determined by the partition π, in the YFV model that is the kernel. The number of one-to-one functions from π into B was then $(x)_{|\pi|}$, in the operationizable, time series dependent YFV-related regression, model when $|\pi|$ was the number of parts in the partition π. Therefore, the total number of functions from a size-n set A into a size-x set B was $\sum_\pi (x)_{|\pi|}$ in the model when the index π ran through the set of all partitions of A. On the other hand, the number of functions from A into B was clearly x^n. Thus, we had $x^n = \sum_\pi (x)_{|\pi|}$ Since X was a geo-spatiotemporal, Poisson-distributed, seasonal, clinical, field and remote geosampled, YFV-related, explanatorial, random variable with expected value 1, then the nth moment of this probability distribution was $E(X^n) = \sum_\pi E\left((X)_{|\pi|}\right)$ but all of the factorial moments $E((X)_k)$ of this probability distribution was equal to 1 in the model derivatives also. Thereafter, we had, $E(X^n) = \sum_\pi 1$, which was the number of optimal; partitions of the set A in the model.

Next, the central moments in the geo-spatiotemporally,geosampled , operationizable, risk model was computed as $v(1+3v)$ so the mean, variance, skewness, and kurtosis were rendered as $\frac{\mu_3}{\sigma^3} = \frac{v}{v^{3/2}} = v^{-1/2}$, $\frac{\mu_4}{\sigma^4} - 3 = \frac{v(1+3v)}{v^2} - 3$ and $\frac{v + 3v^2 - 3v^2}{v^2} = v^{-1}$, respectively. The characteristic function for the Poisson distribution in the geopredictive, autoregressive, eco-epidemiological, risk model was then revealed as $\phi(t) = e^{v(e^{it}-1)}$ and the cdf was $K(h) = v(e^h - 1) = v\left(h + \frac{1}{2!}h^2 + \frac{1}{3!}h^3 + ...\right)$ so $\kappa_r = v$.The

345

mean deviation of the Poisson distribution mode was then subsequently rendered by $MD = \frac{2e^{-\nu}\nu^{[\nu]+1}}{[\nu]!}$.
The CDFs of the Poisson and chi-squared distributions were then related in the eco-epidemiological, YFV-related, explanatorial, probablistic, clinical, field and remote, geosampled, forecasting risk model as $F_{Poisson}(k;\lambda) = 1 - F_{\chi^2}(2\lambda; 2(k+1))$ integer k and $P_r(X=k) = F_{\chi^2}(2\lambda;2k) - F_{\chi^2}(2\lambda;2(k+1))$. The Poisson distribution was then expressed in terms of $\lambda \equiv \frac{\nu}{x}$ whereby, the rate of changes were equal to the equation $P_\nu(n) = \frac{(\lambda x)^n e^{-\lambda x}}{n!}$. The moment-generating function of the Poisson distribution generated from the geosampled, time series dependent, explanatorial, YFV-related, predictor variables was also rendered by $M(t) = e^{(\nu_1+\nu_2)(e^t-1)}$ Given a random variable x and a pdf[e.g., $P(x)$], if there exists an $h > 0$ such that $M(t) \equiv \langle e^{tx} \rangle$ for $|t| < h$, where $\langle y \rangle$ denotes the expectation value of y, then $M(t)$ is called the moment-generating function(Hazewinkle 2001). Commonly, for a continuous distribution in a seasonal, linear, regression-based, time-series dependent, explanatorial, probabilistic, regression model[e.g., $\int_{-\infty}^{\infty} e^{tx} P(x)dx \int_{-\infty}^{\infty}\left(1 + tx + \frac{1}{2!}t^2 x^2 + ...\right)P(x)dx$] the equation $1 + tm_1' + \frac{1}{2!}t^2 m_2' + ...$ is used where m_r' is the r (i.e., the raw moment)(Hosmer and Lemeshew (2002). For quantifying independent X and Y, the moment-generating function in a robust model must satisfy the equation $M_{x+y}(t) = \langle e^{t(x+y)} \rangle, \langle e^{tx} e^{ty} \rangle, \langle e^{tx} \rangle \langle e^{ty} \rangle$ and $M_x(t)M_y(t)$ if, the independent variables $x_1, x_2, ..., x_N$ have Poisson distributions with probability parameters $\mu_1, \mu_2, ..., \mu_N$ and $X = \sum_{j=1}^{N} x_j$ (Haight 1967). This was evident since the cumulant-generating function was $K \equiv \sum_j K_j(h) = (e^h - 1)\sum_j \mu_j = \mu((e^h - 1))$.

In the endemic transmission-oriented, YFV-related, forecasting clinical, field and remote-geosampled, eco-epidemiological, risk model the directed Kullback-Leibler (K-L) divergence between Pois(λ) and Pois(λ_0) was then provided by $D_{KL}(\lambda \| \lambda_0) = \lambda_0 - \lambda + \lambda \log \frac{\lambda}{\lambda_0}$. In probability theory and information theory, the K-L divergence along with information divergence, information gain, relative entropy are a non-symmetric measures of the difference between two probability distributions P and Q in a model(Cressie 1993). For remotely quantifying the probability distributions P and Q of a geosampled discrete random variable the K–L divergence was defined by $D_{KL}(P \| Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)}$. The model revealed that the average of the logarithmic difference between the probabilities P and Q was the average quantified using the probabilities P. The K-L divergence is only defined if P and Q both sum to 1 and if $Q(i) > 0$ for any i such that $P(i) > 0$(Hosmer and Lemeshew 2002).

In our geo-spatiotemporal, YFV-based, probabilistic, regression-related, forecasting, explanatorial, clinical, field or remote geosampled, *Ae.aegypti*, eco-epidemiological, risk model, if the quantity ln 0 appeared in the formula it was interpreted as zero. For distributions P and Q

of the continuous random variable in the geosampled eco-epidemiological YFV-related, empirical datasets, K-L divergence was defined to be the integral[i.e.,

$$D_{KL}(P \| Q) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)} dx]$$

where p and q denoted the densities of P and Q. More generally, since P and Q were probability measures over the eco-epidemiological, regressed, geosampled dataset X, and Q which was absolutely continuous with respect to P, then the K-L divergence from P to Q was optimally defined as

$$D_{KL}(P \| Q) = -\int_x \ln \frac{dP}{dQ} dP$$

in the risk model when $\frac{dQ}{dP}$ was the Radon–Nikodym derivative of Q with respect to P provided the expression on the right-hand side existed. In mathematics, the Radon–Nikodym theorem is a result in measure theory that states that given a measurable space (i.e., X, Σ), if a σ-finite is measured on (i.e., X, Σ) then the expression is absolutely continuous with respect to a σ-finite measure μon (X, Σ)(Hosmer and Lemeshew 2002). By employing the derivative, a measurable function f was rendered on X $(0, \infty)$, such that

$$\nu(A) = \int_A f d\mu$$

for any, operationizable, time series dependent, measured eco-epidemiological, clinical, field or remote geosampled value in thye risk model which then subsequently revealed the statistical significance of the geosampled, YFV-related, explanatorial, covariate, parameter estimator, coefficient values.

Likewise, since P was absolutely continuous with respect to Q in the time series dependent YFV-related, eco-epidemiological, probablistic, regression model. The explanatorial, predictor, covariate, paramter estimator, coefficient values were then defined employing:

$$D_{KL}(P \| Q) = \int_x \ln \frac{dP}{dQ} dP = \int_x \frac{dP}{dQ} \ln \frac{dP}{dQ} dQ$$

which was recognized as the entropy of P relative to Q. We found that if μ was any measure on X in the model, then $p = \frac{dP}{d\mu}$ and $q = \frac{dQ}{d\mu}$ existed, and the K-L divergence from P to Q was given as

$$D_{KL}(P \| Q) = \int_x p \ln \frac{p}{q} d\mu$$

. The bounds for the tail probabilities of the Poisson random variable were then optimally derived in the, eco-epidemiological, regression model using a Chernoff bound argument where $X \sim Pois(\lambda)$ $P(X \geq x) \leq \frac{e^{-\lambda}(e\lambda)^x}{x^x}$, for $x < \lambda$ and for $P(X \geq x) \leq \frac{e^{-\lambda}(e\lambda)^x}{x^x}$ for $x < \lambda$.

In probability theory, the Chernoff bound, provides exponentially decreasing bounds on tail distributions of sums of independent random variables. It is a sharper bound than the known first or second moment based tail bounds such as Markov's inequality or Chebyshev inequality, which only yield power-law bounds on tail decay. However, in our geo-spatiotemporal, YFV-related, eco-epidemiological, explanatorial, clinical, field and remote, geosampled forecasting, risk model, the Chernoff bound required that the variates be independent - a condition that neither the Markov nor the Chebyshev inequalities commonly require. In probability theory, Markov's inequality renders an upper bound for the probability that a non-negative function of a random variable is greater than or equal to some positive constant (Cressie 1993).

We let $X_1, ..., X_n$ be independent Bernoulli random variables, each having probability $p >$ 1/2. Then the probability of simultaneous occurrence of more than $n/2$ of the geosampling, time

series, YFV-related, probablistic sampling events that had an exact value S in the model when $S = \sum_{i=\lceil \mu \rceil+1}^{n} \binom{n}{i} \mu^i (1-\mu)^{n-i}$. the Chernoff bound revealed that S had the following lower bound: $S \geq 1 - e^{-2n\left(\mu - \frac{1}{2}\right)^2}$. We noticed that if X was any geosampled, eco-epidemiological, random variable and $a > 0$, then $\Pr(|X| \geq a) \leq \dfrac{\mathrm{E}(|X|)}{a}$. In the language of measure theory, Markov's inequality states that if (X, Σ, μ) is a measure space, f is a measurable extended real-valued function, and $\epsilon \geq 0$, then $\mu(\{x \in X : |f(x)| \geq \epsilon\}) \leq \dfrac{1}{\epsilon} \int_X |f| \, d\mu$. (Cressie 1993). We then used Chebyshev's inequality to determine the variance bound to the probability that the geo-spatiotemporal-seasonal, geosampled, explanatorial, probabilistic random variable deviated far from the mean in the model. Specifically we used $\Pr(|X - \mathrm{E}(X)| \geq a) \leq \dfrac{\mathrm{Var}(X)}{a^2}$ for any $a>0$. $\mathrm{Var}(X)$ was the variance of X, defined as: $\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}(X))^2]$. Chebyshev's inequality follows from Markov's inequality by considering the random variable $(X - \mathrm{E}(X))^2$ for which Markov's inequality also reads $\Pr((X - \mathrm{E}(X))^2 \geq a^2) \leq \dfrac{\mathrm{Var}(X)}{a^2}$ (see Cressie 1993). Further, in Markov's inequality if x takes only nonnegative, geosampled, eco-epidemiological, YFV-related clinical, field or remote geosampled values, then $P(x \geq a) \leq \dfrac{(x)}{a}$. can be re-written $(x) = \int_0^\infty x P(x) \, dx = \int_0^a x P(x) \, dx + \int_a^\infty x P(x) \, dx.$ However, since $P(x)$ in the eco-epidemiological, YFV-related, risk model was a prevalence rate value, the rendered regression residuals were ≥ 0. It should be stipulated that $x \geq 0$ was renderd from the resdiual YFV regressors so $(x) = \int_0^a x P(x) \, dx + \int_a^\infty x P(x) \, dx \geq \int_a^\infty x P(x) \, dx \geq \int_a^\infty a P(x) \, dx = a \int_a^\infty P(x) \, dx = a P(x \geq a).$ In order to determine the statistically significant time series dependent, explanatorial, time series dependnet, clinical, field or remote explanatorial, covariate, paramter estimator, coefficient values.

$$\zeta(s) = \prod_{k=1}^{\infty} \frac{1}{1 - \dfrac{1}{p_k^s}}$$

We then considered the Euler product where $\zeta(s)$ was the Riemann zeta function and p_k was k the prime. $\zeta(1) = \infty$. Thereafter, by taking the finite product up to k=n in the eco-epidemiological, clinical, field or remote explanatorial, time series dependent, YFV-related, probablistic, risk model and pre-multiplying by a factor $1/\ln p_n$, we were able to employ

$$\lim_{n \to \infty} \frac{1}{\ln p_n} \prod_{k=1}^{n} \frac{1}{1 - \dfrac{1}{p_{k=e^\gamma}}} = e^\gamma$$

$n \to \infty$ to render which was equivalent to 1.781072. By so doing, g became the Euler-Mascheroni constant which also represented the limit of the sequence g= $\gamma = \lim_{n \to \infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln n \right) = \lim_{n \to \infty} (H_n - \ln n) =$ in the regressed residuals where H_n was the harmonic number which had the form $H_n = \sum_{k=1}^{n} \frac{1}{k}$ in the forecasted residual derivatives. A harmonic number can be expressed analytically as $H_n = \gamma + \psi_0(n+1)$ where γ is the Euler-Mascheroni constant

(Hosmer and Lemeshew 2002). Our model revealed that the Euler product attached to the Riemann zeta function $\zeta(s)$ represented the sum of the geometric series rendered from the geo-spatiotemporal, geosampled, empirical, probalistic dataset of explanatorial, observational predictor, YFV-related, time series dependent, covariate, paramter estimator, coefficients as

$$\prod_{p}(1-p^{-s})^{-1} = \prod_{p}\left(\sum_{n=0}^{\infty} p^{-ns}\right) = \sum_{n=1}^{\infty}\frac{1}{n^s} = \zeta(s)$$

. A closely related result was also obtained by

$$1 + \frac{1}{p_k} = \frac{1 - \frac{1}{p_k^2}}{1 - \frac{1}{p_k}}.$$

noting that We also considered the variation of when with the $+$sign changed to a $-$sign and the $\ln p_n$ in the model which moved from the denominator to the numerator rendering

$$\lim_{n\to\infty}\ln p_n \prod_{k=1}^{n}\frac{1}{1+\frac{1}{p_k}} = \lim_{n\to\infty}\ln p_n \prod_{k=1}^{n}\frac{1}{\frac{1-\frac{1}{p_k^2}}{1-\frac{1}{p_k}}} = \frac{\prod_{k=1}^{\infty}\frac{1}{1-\frac{1}{p_k^2}}}{\lim_{n\to\infty}\frac{1}{\ln p_n}\prod_{k=1}^{t}\frac{1}{1-\frac{1}{p_k}}} = \frac{\zeta(2)}{e^{\gamma}} = \frac{\pi^2}{6e^{\gamma}}$$

$$= .0915.$$

We then tested the model for overdispersion with a likelihood ratio test. This test quantified the equality of the mean and the variance imposed by the Poisson distribution against the alternative that the variance exceeded the mean. For the negative binomial distribution, the variance = mean + k mean^2 (k>= 0, the negative binomial distribution reduced to Poisson when k=0)(Haight 1967). The null hypothesis was H_0: k=0 and the alternative hypothesis was H_a : k>0 . To carry out the test, we employed the following steps initially and then ran the model using negative binomial distribution and a record log-likelihood (LL) value. We then recorded LL for the Poisson model. We used the likelihood ratio (LR) test, that is, we computed LR statistic, -2(LL (Poisson) – LL (negative binomial). The asymptotic distribution of the LR statistic had probability mass of one half at zero and one half – chi-sq distribution with 1 d.f. To test the null hypothesis further at the significance level , we then used the critical value of chi-sq distribution corresponding to significance level 2, that is we rejected H_0 if LR statistic $>^2$ (1-2 , 1 df).

Next, we assumed that our geo-spatiotemporal,geosampled eco-epidemiological, forecasting, risk model explanatorial, predictor YFV-related, covariate, paramter estimator, coefficient estimates were based on the log of the mean, which was a linear function of independent variables, log() = intercept + b1*X1 +b2*X2 + ...+ b3*Xm. This log-transformation implied that was the exponential function of the independent variables, = exp(intercept + b1*X1 +b2*X2 + ...+ b3*Xm). Instead of assuming as before that the distribution of the seasonal geosampled coefficients[i.e., Y], was Poisson, we assumed a negative binomial distribution. That meant, relaxing the generalized linear Poisson regression specification assumption about the equality of the mean and variance since in our model we found that the variance of negative binomial was equal to + k2 , where k>= 0 was a dispersion parameter. The maximum likelihood method was then used to estimate k as well as the covariate, parameter estimators of the YFV-related, time series dependent, eco-epidemiological, risk-related, probablistic, model for log(). Fortunately, the SAS syntax for running negative binomial regression was almost the same as for Poisson regression. The only change was the dist option in the MODEL statement was used

instead of dist = poisson, dist = nb. The pmf of the negative binomial distribution with a gamma distributed mean in the geopredictive, explanatorial, interpolatble, YFV-related, eco-epidemiological, risk model was then expressed as $f(k) \equiv \Pr(X = k) = \binom{k+r-1}{k}(1-p)^r p^k$ using the geosampled, explanatorial, covariate, parameter estimator, coefficient estimates (i.e., $k = 0, 1, 2, \ldots$). In this equation, the quantity in parentheses was the binomial coefficient, which was equal to $\binom{k+r-1}{k} = \frac{(k+r-1)!}{k!\,(r-1)!} = \frac{(k+r-1)(k+r-2)\cdots(r)}{k!}$. This quantity was also alternatively written as $\frac{(k+r-1)\ldots(r)}{k!} = (-1)^k \frac{(-r)(-r-1)(-r-2)\ldots(-r-k+1)}{k!} = (-1)^k \binom{-r}{k}$ for explaining "negative binomialness' in our regression model. Results from both a Poisson and a negative binomial model residual forecasts revealed that the geo-spatiotemporal, geosampled, explanatorial, covariate coefficient estimates were highly significant, but virtually furnished no predictive power.

Thus, instead of assuming that the distribution of the geosampled time series explanatorial, geopredictive, YFV-related, geo-spatiotemporal estimates (i.e., Y) was Poisson, we were able to assume that Y had a negative binomial distribution. We relaxed the assumption about equality of mean and variance (i.e., Poisson distribution property), since the variance of negative binomial was equal to $\mu + k\mu^2$, where $k >= 0$ which represented a dispersion parameter. The maximum likelihood method was then used to estimate k, as well as the geosampled time series, YFV-related, parameter estimators of the regression model for $\log(\mu)$. Thereafter, we noted that for the negative binomial distribution, the variance was equal to the mean + k mean^2 (i.e. $k >= 0$) as the negative binomial distribution reduced to Poisson when k was 0.

In the risk based, explanatorial, time series, exploratory dependent, YFV-related, forecasting, eco-epidemiological, regression analyses, the null hypothesis was: H_0: k=0 and the alternative hypothesis was: H_a : k>0. We recorded the log-likelihood (i.e. LL) for the seasonal, YFV-related, forecasting, risk model. We employed the likelihood ratio (LR) test to compute the LR statistic using -2(LL) (Poisson) and the LL (i.e., negative binomial). The asymptotic distribution of the LR statistic had probability mass of one half at zero and one half – chi-square distribution with 1 df. To test the null hypothesis at the significance level α in our endemic, YFV-related, transmission-oriented, geopredictive, eco-epidemiological, risk model, we employed the critical value of chi-square distribution corresponding to significance level 2α, that was rejection of H_0, if LR statistic $> \chi^2_{(1-2\alpha,\ 1\ df)}$. We generated the log of the mean, μ, which was a linear function of independent variables, $\log(\mu) = $ intercept + b1*X1 +b2*X2 + ...+ b3*Xm, in the time series explanatorial, clinical, field and remote geosampled, YFV-related, risk model which implied that μ was the exponential function of the independent variables when $\mu = \exp(\text{intercept} + b1*X1 + b2*X2 + ...+ b3*Xm)$. By so doing, LULC change was found to be significantly associated to the YFV-related, explanatorial, covariate, parameter estimatators empirically geosampled at the Gulu epidemiological study site.

The continuous random geosampled, seasonal, YFV-related, variable x had a chi-square probability distribution, with df, and its mathematical function was given by the equation   where $\nu > 0$. Mean and variance of a random variable   this

$$f(x) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2} \\ 0 \end{cases}$$

distribution was then quantitated as $E(x) = \mu = \nu$ and $Var(x) = \sigma^2 = 2\nu$. Note ν was a geosampled, YFV-related, times series parameter estimator. Different shapes of the probability distribution resulted from different values of ν.

Figure 6. Chi-square probability distribution plot of the regressed time series YFV-related endemic, variables geosampled at the Gulu eco-epidemiological study site
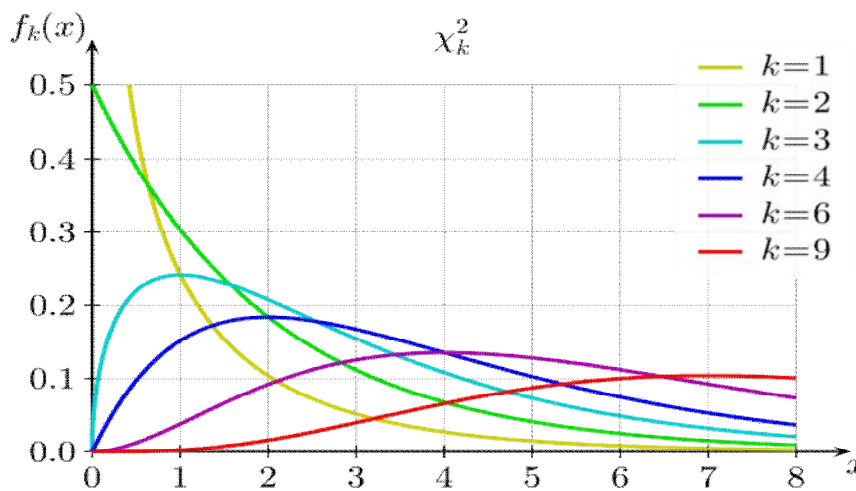
For finding the probabilities and quintiles from the time series, YFV-related, chi-square distribution we employed $P(x < 3.84)$ with $\nu = 1$. For   integration we then employed

$$\int_0^{3.84} \frac{1}{2^{1/2}\Gamma(1/2)} x^{1/2-1} e^{-x/2} dx$$

We also used the *pchisq()* function: where > pchisq(q = 3.84, df = 1)[1] 0.94996 To find the $1 - \alpha/2$ quantile from a chi-square distribution, we then used   where we solved for c. We also used dthe *qchisq()* function:>

$$1 - \alpha/2 = \int_0^c \frac{1}{2^{\nu/2}\Gamma(1/2)} x^{1/2-1} e^{-x/2} dx$$

alpha<-0.1> qchisq(p = 1 - alpha/2, df = 1)[1] 3.8415. By so doing, the *dchisq()* function then allowed us to evaluate *f(x)* so that we could parsimoniously plot the seasonal YFV-related distribution as:

```
> curve(expr = dchisq(x = x, df = 1), from = 0, to = 5, col
    = "red", lwd = 2, main = "Chi-square distribution with
    1 DF", ylab = "f(x)", xlab = "x", n = 1000)
> abline(h = 0)
```

For our population of geo-spatiotemporal, YFV-related, probablistic, discrete values, the kth q-quantile was the data value where the CDF crossed k/q. That is, x was a kth q-quantile for a random, geosampled, explanatorial, clinical, field or remote –related, geopredictive, eco-epidemiological, time series dependent, geopredictive variable X if $\Pr[X < x] \leq k/q$ (or equivalently, $\Pr[X \geq x] \geq 1 - k/q$) and $\Pr[X \leq x] \geq k/q$ (or equivalently, $\Pr[X > x] \leq 1 - k/q$). For a finite population of N values indexed $1,...,N$ from lowest to highest, the kth q-quantile of the eco-epidemiological, YFV-related, parameter estimatator population was computed via the value rendered from $I_p = N\dfrac{k}{q}$. If I_p is not an integer, then round up to the next integer to get the appropriate index; the corresponding data value is the kth q-quantile (Hazewinkle 2001). On the other hand, we assumed that if I_p is an integer values in a geo-spatiotemporal, geosampled YFV-related eco-epidemiological, forecasting, risk model, then any number from the data value can be taken as the quantile, and it is conventional (though arbitrary) to determine the average of any two geosampled values. We instead employed the empirically geosampled, explanatorial, clinical, field and remote variable integers where k and q, were the "p-quantile" based on a response variable (prevalance of YFV) p with $0 < p < 1$. By so doing, p replaced k/q in the YFV, eco-epidemiological, geopredictive, risk model.

Although the Q-Q plots rendered from the time series dependent, YFV-related ,geopredictive covariate,parameter estimators coefficients was based on accurately tabulated quartiles, the Q-Q plot was not able to quantize which georeferenced LULC explanatorial, summarized, diagnostic point in the Q-Q plot determined a given quantile. For instance, it was not possible to determine the median of the time series plotted distributions. Commonly Q-Q plots indicate deciles to make determinations such as this possible. The slope and position of the geo-spatiotemporal, YFV-related, probabilistic, regressors between the quartiles did not render a measure of the relative district-level geolocation and relative scale of the samples. If the median of the distribution plotted on the horizontal axis is 0, the intercept of a regression line is a measure of the geolocation, and the slope is a measure of scale (Cressie 1993). The distance between medians of relative LULC geolocation was not reflected in the Q-Q plots. The probability plot of error correlation coefficients (i.e., the correlation coefficient between the paired sample quartiles) was thus not quantifiable.

A major problem is that the "r-squared" statistic for both of the prediction equations is fairly small ($R^2= 26\%$ and 4.7% respectively) which suggests that the prediction lines do not fit the data very well. One way to improve the predictive model might be to combine the information in both of the images. The "Normalized Density Vegetation Index (NDVI)" does just that by calculating a new value that indicates plant vigor— NDVI= ((NIR – Red) / (NIR + Red). Figure &shows the process for calculating NDVI for the sample grid location— ((121-14.7) / (121 + 14.7))= 106.3 / 135.7= .783 . The scatter plot on the right shows the yield versus NDVI plot and regression line for all of the field locations.

the R^^2 value is a higher at 30% indicating that the combined index is a better predictor of yield.
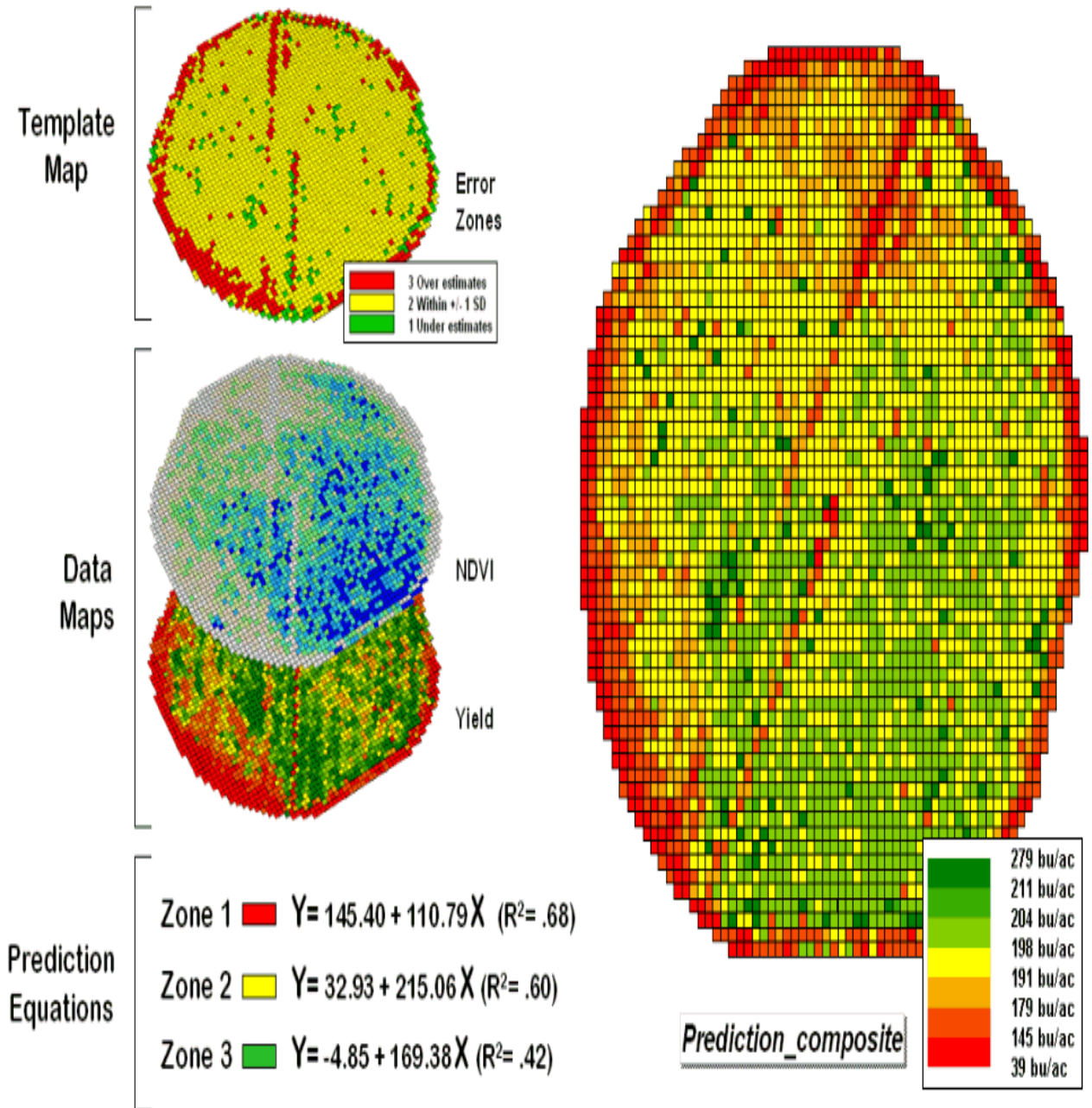
Commonly in a seasonal vector entomological-related cluster-based forecasting, regression model, the closer the correlation coefficient is to one, the closer the distributions are to being shifted, scaled versions of each other(see Jacob et al. 2009d, Jacob et al. 2010a,b). Unfortunately for quantitating the seasonal vector distributions with a single shape parameter, the probability plot correlation coefficient plot in SAS could not provide a method for estimating , clinical, field or remote specified YFV-related shape parameter by computing the correlation coefficient for different geosampled, eco-epidemiological, geo-spatiotemporal values of the shape parameter. The use of Q-Q plots was not able to quantitatively assess nor compare the geo-spatiotemporal, explanatorial, eco-epidemiolotgical, clinical, field and/or remote geosampled, distribution of the YFR-related samples to the standard normal distribution [i.e., $N(0,1)$].

Figure 7. a plot of the Chi-square distribution, when $\nu = 1$ degree of freedom in the seasonal geopredictive YFV-related endemic transmission-oriented risk model
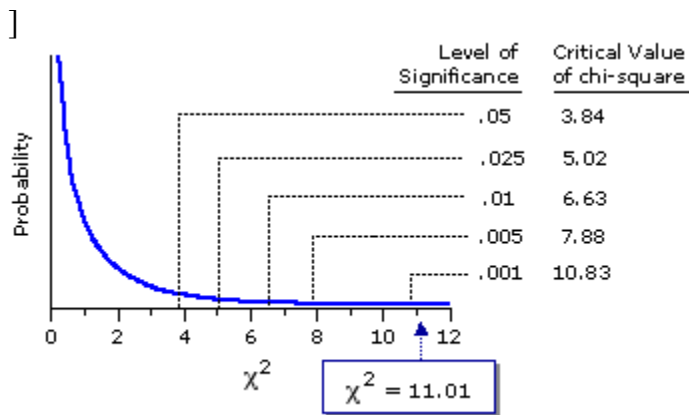
Bootstrapping allows for estimation of statistics through the repeated resampling of data. In this page, we will demonstrate several methods of bootstrapping a confidence interval about an R-squared statistic in SAS. We will be using the **hsb2** dataset that can be found here. We will begin by running an OLS regression, predicting **read** with **female**, **math**, **write**, and **ses**, and saving the R-squared value in a dataset called **t0**. The R-squared value in this regression is 0.5189.

```
ods output FitStatistics = t0;
proc reg data = hsb2;
  model read = female math write ses;
run;
quit;
```

The REG Procedure
Model: MODEL1
Dependent Variable: read reading score

Number of Observations Read   200
Number of Observations Used   200

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	10855	2713.73294	52.58	<.0001
Error	195	10064	51.61276		
Corrected Total	199	20919			

Root MSE	7.18420	R-Square	0.5189	
Dependent Mean	52.23000	Adj R-Sq	0.5090	
Coeff Var	13.75493			

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	6.83342	3.27937	2.08	0.0385
female		1	-2.45017	1.10152	-2.22	0.0273
math	math score	1	0.45656	0.07211	6.33	<.0001
write	writing score	1	0.37936	0.07327	5.18	<.0001
ses		1	1.30198	0.74007	1.76	0.0801

*store the estimated r-square;
data _null_;
 set t0;
 if label2 = "R-Square" then
 call symput('r2bar', cvalue2);
run;

To bootstrap a confidence interval about this R-squared value, we will first need to resample. This step involves sampling with replacement from our original dataset to generate a new dataset the same size as our original dataset. For each of these samples, we will be running the same regression as above and saving the R-squared value. **proc surveyselect** allows us to do this resampling in one step.

Before carrying out this step, let's outline the assumptions we are making about our data when we use this method. We are assuming that the observations in our dataset are independent. We are also assuming that the statistic we are estimating is asymptotically normally distributed.

We indicate an output dataset, a seed, a sampling method, and the number of replicates. The sampling method indicated, urs, is unrestricted random sampling, or sampling with replacement. The samprate indicates how large each sample should be relative to the input dataset. A samprate of 1 means that the sampled datasets should be of the same size as the input dataset. So in this example, we will generate 500 datasets of 200, so our output dataset bootsample will have 100,000 observations.

```
%let rep = 500;
proc surveyselect data= hsb2 out=bootsample
    seed = 1347 method = urs
            samprate = 1 outhits rep = &rep;
run;
ods listing close;
```

The SURVEYSELECT Procedure

Selection Method   Unrestricted Random Sampling

```
Input Data Set              HSB2
Random Number Seed              1347
Sampling Rate                1
Sample Size                200
Expected Number of Hits          1
Sampling Weight              1
Number of Replicates          500
Total Sample Size          100000
Output Data Set          BOOTSAMPLE
```

With this dataset, we ran our regression model, specifying by replicate so that the model will be run separately for each of the 500 sample datasets. After that, we use a data step to convert the R-squared values to numeric.

```
ods output  FitStatistics = t (where = (label2 =  "R-Square"));
proc reg data = bootsample;
  by replicate;
  model read = female math write ses;
run;
quit;
* converting character type to numeric type;
data t1;
  set t;
  r2 = cvalue2 + 0;
run;
```

The explanatorial, geosampled, time series-related, eco-epidemioligical, YFV-related, geopredictive, clinical, field and remote geosampled covariate, parameter estimator, coefficient values were then input into an eigenfunction decomposition algorithm to quantitate latent autocorrelation error coefficients in the linear geo-spatiotemporally interpolatable, residual variance estimates. Results indicated that positive spatial autocorrelation (PSA) (i.e., aggregation of like terms in geo-space) and negative spatial mautocorrelation (NSA) were detected for the geosampled data. Eigenvectors were then extracted from the matrix $(\mathbf{I}-\mathbf{11}'/n)$ $\mathbf{C}$ $(\mathbf{I}-\mathbf{11}'/n)$, employing the eco-entomological geosampled, explanatorial, time series-related, geopredictive variables. We noted that denoting the autoregressive parameter captured the latent

autocorrelation in the time series –related, YFV, eco-epidemiological, risk model. This quantification involved ρ, a conditional autoregressive covariance specification, which involved the matrix $(\mathbf{I} - \rho\ \mathbf{C})$, where $\mathbf{I}$ was an *n*-by-*n* identity matrix. The residual autocorrelation error components were then calculated as the matrix $\mathbf{C}$ raised to the power 1 Interestingly, only adjacent geosampled, time series dependent, explanatorial, geopredictive, clinical, field and remote-specified, YFV-related, data were involved in the autoregressive function, which in actuality was a first-order specification, with the autoregressive term being $\mathbf{CY}$.

An important matrix was then generated from **C1** in the explanatorial, time series dependent, YFV-related, endemic, transmission-oriented, forecasting, eco-epidemiological, risk model which was the vector of the number of geoampled, covariate, parameter estimators neighbors in the Gulu eco-epidemiological study site. Next, the inverse of the elements of **C1** were inserted into the diagonal of a diagonal matrix, (i.e., $\mathbf{D}^{-1}$) rendering matrix $\mathbf{W} = \mathbf{D}^{-1}\mathbf{C}$ which became a stochastic matrix (i.e., each of its row sums equaled 1). One appealing feature of this matrix was that the autoregressive term in the autoregressive model became $\mathbf{WY}$, which generated averages, rather than sums, of the neighboring geosampled estimated LULC values. Because a covariance matrix for a robust, seasonal, vector, insect –related, probabilistic, distribution model must be symmetric (Jacob et al. 2006, Jacob et al. 2005b), we employed a matrix $\mathbf{W}$ specification with a conditional autoregressive model by making the individual geosampled LULC and other explanatorial variance no constants using $(\mathbf{I} - \rho\ \mathbf{D}^{-1}\mathbf{C})\mathbf{D}^{-1} = (\mathbf{D}^{-1} - \rho\ \mathbf{D}^{-1}\mathbf{CD}^{-1})$.

An appealing feature of this version for our geopredictive, eco-epidemiological, YFV-related, distribution model was that it restricted values of the autoregressive parameter to the more intuitively interpretable range of $0 \leq \hat{\rho} \leq 1$. The model then furnished an alternative specification which was also written in terms of matrix $\mathbf{W}$. The spatial covariance was then a function of the matrix $(\mathbf{I} - \rho\ \mathbf{CD}^{-1})(\mathbf{I} - \rho\mathbf{D}^{-1}\mathbf{C}) = (\mathbf{I} - \rho\ \mathbf{W}^{T})(\mathbf{I} - \rho\ \mathbf{W})$, where $\mathbf{T}$ denoted the matrix transpose. The resulting matrix was symmetric, and was considered a second-order specification, as it included the product of two spatial structure matrices (i.e., $\mathbf{W}^{T}\mathbf{W}$), which also restricted values of the autoregressive parameter to the more intuitively interpretable range of $0 \leq \hat{\rho} \leq 1$ in the forecasting risk model.

In statistics, Moran's *I* is a measure of spatial autocorrelation (Moran 1950). We defined Moran's *I* as $I = \dfrac{N}{\sum_i \sum_j w_{ij}} \dfrac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$ where N was the number of LULC case units indexed by i and j; X was the other geosampled, time series-related, explanatorial, YFV-related, clinical, field and remote specified variables of ; $\bar{X}$ was the mean of X; and w_{ij} was an element of a matrix of spatial weights. The expected value of Moran's *I* under the null hypothesis of no spatial autocorrelation was then $E(I) = \dfrac{-1}{N-1}$ (see Griffith 2003). We noted that variance equaled $\mathrm{Var}(I) = \dfrac{NS_4 - S_3 S_5}{(N-1)(N-2)(N-3)(\sum_i \sum_j w_{ij})^2}$ where

$$S_1 = \frac{1}{2}\sum_i \sum_j (w_{ij} + w_{ji})^2$$

$$S_2 = \frac{\sum_i (\sum_j w_{ij} + \sum_j w_{ji})^2}{1}$$

$$S_3 = \frac{N^{-1}\sum_i(x_i - \bar{x})^4}{(N^{-1}\sum_i(x_i - \bar{x})^2)^2}$$

$$S_4 = \frac{(N^2 - 3N + 3)S_1 - NS_2 + 3(\sum_i\sum_j w_{ij})^2}{1}$$

$$S_5 = S_1 - 2NS_1 + \frac{6(\sum_i\sum_j w_{ij})^2}{1}$$

Negative (positive) LULC values indicated negative (positive) spatial autocorrelation in the eco-epidemiological, YFV-related time series dependent, eco-epidemiological, clinical, field and remote specified model. Values ranged from −1 (indicating perfect dispersion) to +1 (perfect correlation). A zero value indicated a random time series-related, YFV-oriented, LULC geospatial patterns. For statistical hypothesis testing, Moran's I values can be transformed to Z-scores in which values greater than 1.96 or smaller than −1.96 indicate spatial autocorrelation that is significant at the 5% level (see Griffith 2003).

Because the Moran's scatterplot utilizes a standardized variable y (see Anselin, 1995), plotting georeferenced, time series-related, YFV, explanatorial, geopredictor variables geosampled in varying explanatorial, LULC transitional zones allowed the data to be geolocated in one of the four planar graph quadrants centered at (0,0). The slope of the line in the scatterplot is equivalent to the MC (Griffith, 2003). Thus, since the geopredictive, YFV-related, time series distribution slope-related coefficinets in the scatterplot was negative, we assumed that this would generate a checkerboard pattern, or a sort of spatial competition in which high geopredictor covariate coefficient indicator values geosampled from a georeferenced explanatorial, LULC site geolocation would tend to be co-located with lower covariate coefficient values in neighboring LULC geolocations [i.e., NSA]. We generated a graph in ArcGIS to depict a standardized variable in the x-axis versus the spatial lag of the standardized variables employing a weighted geographical matrix for seasonally quantitating the YFV-related explanatorial, clinical, field and remote –specified regressors. The spatial lag in their model was a summary of the effects of the neighboring prolific geosampled LULC –related feature data attributes which was obtained by means of the weights matrix.

Next, an eigenfunction decomposition algorithm was generated based on the weights matrix, for uncovering relationships between the determinant of the matrix and its extreme time series dependent, eco-epidemioloigical, YFV-derived eigenvalues. We also quantitated regression values and the mininium eigenvalue of the weights matrix and the eigenvectors of a row-standardized asymmetric version of the matrix and its symteric similarity matrix counterpart components. In addition , a conjecture was posited pertaining to the estimation and subsequent quantitation of the largest eigenvalue rendered from the binary geographic weights matrix especially when the specified estimate obtained with the matrix powering began to oscillate between two tracjectories in its convergence. The decomposition algorithm then outlined the extreme, YFV-related, time series, dependent eigenvalues. By so doing, ArcGIS planar graphs with the georeferenced, clinical, field or remote geospatiotemporally-geosampled, endemic , transmission-oriented, YFV-related variables were cartographically described.

We also qunatiated the Geary'C ratio. Geary's C is a measure of spatial autorrelation (Griifth 2003). Geary C is invcrsely related to the Morans I but it is not identical. Morans I is a measure of global autocorrelation while Geary's C is more sensitive to local spatial autocorrelation . Like temporal correlation , autocorrelation means that adjacent observations of the same phenomena are correlated in geographic space. However , temporal autocorrelation is about actually only a proximity in time measured while autocorrelation is more in regard to proximity to spatialized processes. We defined Gearcy C for geospatially quantizing the geosampled, YFV-related, explanatorial, time series, clinical, field and remote specified variables using

$$C = \frac{(N-1)\sum_i \sum_j w_{ij}(X_i - X_j)^2}{2W \sum_i (X_i - \bar{X})^2}$$

where N was the number of cases indexed by i and j; X was the variables of interest; $\bar{X}$ was the mean of X; w_{ij} was a matrix of spatial weights; and W was the sum of all w_{ij} = The value of Geary's C lies between 0 and 2. 1 means no spatial autocorrelation. Values lower than 1 demonstrate increasing positive spatial autocorrelation, whilst values higher than 1 illustrate increasing negative spatial autocorrelation (Griffith 2003).

Estimation results, for these models, appear in Table 4. Positive and NSA spatial filter component pseudo-R^2 values are reported. These values did not exactly sum for the complete spatial filter; however, they are very close to their corresponding totals, suggesting that any induced multicollinearity was quite small.

Table 4. Global Spatial analyses of the geosampled explanatorial geopredictive YFV-related explanatory covariate coefficient count data in the Gulu study site

Study Site	n	Transformation	MC	GR
Gulu	152	LN(count + 1)	0.068	0.791

A generalized linear model (GLM) was extended to account for latent, non-spatial ,YFV-related,time series, explanatorial covariate correlation effects, which allowed inferences to be drawn for a much wider range of geographic sampling configurations generated from the geosampled, georefernced, time series, geopredictive, estimates than those utilized by employing a GLMM. The GLMM included a random effect, which was specified as a random intercept that was assumed to be normally distributed with a mean of zero, a constant variance, and non-zero spatial autocorrelation. This varying intercept term compensated for the non-constant mean associated with the seasonal, YFV-related, negative binomial model, GLM specification. The spatial structuring of random effects was implemented with a conditional autoregressive model and was achieved in this research with a spatial filter. The autocorrelation components revealed 16% redundant information in the ecologically geosampled datasets. The GLMM estimation results appear in Table 5.

Table 5. Poisson spatial filtering model results for the sampled YFV-related count data at the Gulu study site

Spatial Statistics	Gulu
SF: # of eigenvectors	3
SF: MC	0.594
SF: GR	0.386
SF pseudo-R^2	0.177
Positive SA SF: # of eigenvectors	4
Positive SA SF: MC	0.603

*SF denotes spatial filter

*SA denotes spatial autocorrelation

Table 6 lists the improvements of fit in the adjusted and unadjusted models for all the seasonal, geosampled, YFV-related, model specifications and random error in the geospatial analyses. The unadjusted model compared the univariate model to a model containing only the intercept term. Interactions were examined, and significant interactions were included. Improvement of fit was also calculated for the first-order interaction models to determine whether including significant interactions improved fit compared to the full main effects model. Convergence problems prevented obtaining results of a saturated model to determine whether the presented model fit as well as the saturated model.

Table 6. Comparison of improvement of fit measured by likelihood ratio between unadjusted and adjusted effects models

Unadjusted effects				**Adjusted effects**			
		Improvement				**Improvement**	
Variable	**Deviance**	χ^2	**df**	**Deviance**	χ^2		**df**
Intercept	983.1241						
LULC	983.4208	14.4751	1	901.432	20.748		1
RFALL	983.5936	12.1098	1	885.169	3.7497		1
NDVI	982.6438	14.1147	1	896.257	19.397		1
HPOP	986.3168	8.00961	1	890.007	8.6375		1
ELEV	986.5872	9.96053	1	901.328	20.632		1
Full Main Effects							

Generalized autoregressive moving average (GARMA) models were then extended to generalized, seasonal, autoregressive, integrated moving average (GSARIMA) models for robustly quantiating observation-driven, YFV-related, non-Gaussian, non-stationary and/or

explanatorial, georeferenced, time series count data. The models were applied to monthly YFR case time series in the Gulu eco-epidemiological, study site.

Bayesian negative binomial versions of SARIMA models and two ARIMA models, with second order harmonics were identified in the preliminary analysis, in SAS employing untransformed YFV-related data, using a logarithmic link function and ZQ1 transformation. Since there were only three observations with counts in the geopredictive, YFV-related, endemic, transmission-oriented, eco-epidemiological, risk model (rainfall, LULC and human population), the results would not have been sensitive to the choice of the transformation constant for ZQ1 which was set at c = 1. Also, versions with identity link were considered. Models were evaluated based on two criteria. The first was the deviance information criterion (DIC), which was calculated as the mean of the posterior distribution of the deviance conditional on the first, time series geosampled, YFV-related ω observations where ω was equal to the maximum w of the models compared which were then augmented with the number of effective estimated parameters as penalty to prevent over fitting. Models with lower DIC were considered to have a better fit. A second criterion was then defined as the mean absolute relative error of fitted values (MARE): MARE = $\sum_{f}^{l} |\frac{y_t - \hat{y}_t}{y_t + 1}| \Big/ (l + 1 - f)$, where $\hat{y}_t$ was the fitted number of previous yellow fever cases at discrete time interval t, and f and l were the first and last discrete time intervals, respectively, of the time period under consideration. The MARE was calculated for both the entire series (except for the first YFV-related ω observations), where models were fitted to the entire time series using $f = \omega + 1$, $l = n$.

We noticed that the posterior, geopredictive, seasonal-geosampled, explanatorial, YFV-related distributions estimated at each fitted data point were skewed and the median of the posterior distribution was taken for $\hat{y}_t$. The MARE is similar to the mean absolute percentage error (MAPE), which is applicable to series for which the variance is dependent on the mean (Anselin 1995). However, since the denominator was equal to or larger than one in our seasonal, explanatorial, geopredictive, time series-related, YFV-related, endemic, transmission-oriented, risk model, this prevented problems with large geosampled, LULC, explanatorial covariate coefficients. The MARE statistic does not have a built-in penalty to prevent over fitting, but among models with similar value of MARE, the model with the least number of parameters is preferred (Cressie 1993). The MARE estimates were comparable across models with different distributional assumptions, in contrast to the DIC. Models were run with three Markov chains of 11,000 iterations each including a burn-in of 1,000 iterations. Convergence was assessed by studying the time series YFV-related plots of the Gelman-Rubin convergence statistic on the estimated endemic, transmission-oriented, parameters geosampled at the Gulu eco-epidemiolgical, study site.

Gelman and Rubin convergence statistics proposed a general approach to monitoring convergence of MCMC output in which $m > 1$ parallel chains were updated with initial geosampled YFV-related covariate coefficient values that were overdispersed relative to each target distribution, which had to be normally distributed. Convergence was diagnosed when the chains had 'forgotten' their initial, seasonal, explanatorial, geopredictive, time series dependent, YFV-related, endemic, transmission-oriented, eco-epidemiological, risk model covariate coefficient values, and the output from all chains is indistinguishable. The Gelman Diagnostic

function made a comparison of within-chain and between-chain variances, and is similar to a classical analysis of variance. A large deviation between these two variances indicates non-convergence

Since the Gelman Diagnostic function relies heavily on overdispersion with respect to the target distribution (Gilks 1996) we considered using MCMC twice, first to estimate the target distributions, and secondly to overdisperse initial seasonal, explanatorial, geopredictive, time series dependent, YFV-related, endemic, transmission-oriented, risk model, covariate, coefficient values with respect to them. This helped identify multimodal YFV-related target distributions. If multiple modes are found, it remains possible that more modes exist(Cressie 1993). When multiple modes are found, and if chains are combined with the Combine functionin PROC MCMC, each mode may be probably represented in a proportion correct to the distribution (ww.sas.edu).

PROC MCMC used a random walk Metropolis algorithm to obtain posterior samples from the empirical geosampled dataset of clinical, field and remote, geo-spatiotemporal, covariate coefficients .We wanted want to obtain samples (i.e., T) from a univariate distribution with pdf $f(\theta|y)$. In our YFV-related eco-epidemiological, forecasting, risk model θ^t was the tth sample from f. To use the Metropolis algorithm, we needed to have an initial value θ^0 and a symmetric proposal density $q(\theta^{t+1}|\theta^t)$. For the $(t+1)$th iteration, the algorithm generated a sample from $q(\cdot|\cdot)$ based on the current sample θ^t, and it made a decision to either accept or reject the new sample. If the new sampled accepted, the algorithm repeated itself by starting at the new sample. If the new sample is rejected, the algorithm starts at the current point and repeats(http://support.sas.com/)

We noticed that $q(\theta_{new}|\theta^t)$ in the time series dependent, YFV-related, eco-epidemiological, forecasting, risk model was a symmetric distribution. The proposal distribution was then an easily quantiated distribution from which to sample where $q(\theta_{new}|\theta^t)=q(\theta^t|\theta_{new})$, meaning that the likelihood of θ_{new} from θ^t was the same as the likelihood of jumping back to θ^t from θ_{new}. The most common choice of the proposal distribution is the normal distribution $N(\theta^t,\sigma)$ with a fixed σ(Cressie 1993). The Metropolis algorithm for qualitatively quntitating the empirical geosampled dataset of geo-spatiotemporally, explanatorial, clinical, field or remote specified, YFV-related geopredictive variables was summarized as follows. We set $t=0$ .We chose a starting point θ^0. This was an arbitrary point as long as $f(\theta^0|y)>0$ .We generated a new sample, θ_{new}, by using the proposal distribution $q(\cdot|\theta^t)$ .We calculated the following quantity $r=\min\left\{\dfrac{f(\theta_{new}|y)}{f(\theta^t|y)},1\right\}$ .We then geosampled u from the uniform distribution $U(0,1)$. We set $\theta^{t+1}=\theta_{new}$ if $u<r$ or $\theta^{t+1}=\theta^t$ otherwise. Then we set $t=t+1$. Note that the number of iterations kept increasing regardless of whether a proposed explanatorial, time series sample was acceptedor not. This algorithm defined a chain of random variates whose distribution converged to the desired distribution[i.e., $f(\theta|y)$}, and so from some point forward, the chain of samples was a sample from the distribution of the georfernced YFV-related regressors. In Markov chain terminology, this distribution is called the stationary distribution of the chain, and in Bayesian statistics, it is the posterior distribution of the model parameter estimators (Cressie 1993). The

reason that the Metropolis algorithm works is beyond the scope of this documentation. More detailed descriptions and proofs are present in many standard textbooks, including Roberts (1996) and Liu (2001).

We employed the random-walk Metropolis algorithm in PROC MCMC.By default , PROC MCMC assumed that all the geo-spatiotemporally-geosampled YFV-related clinical, field and remote –geosampled,time series covariate coefficients in the empirical datset were independent. The logarithm of the posterior density was calculated as follows $\log(p(\theta \, y_i) - \log(\pi(\theta)_i - \sum_{i=1}^{n} \log(f(y_i|\theta))$ where θ was a vector of parameters. The term $\log(\pi(\theta))$ was the sum of the prior densities in the PRIOR and the HYPERPRIOR statements. The term $\log(f(y_i|\theta))$ was the log-likelihood specified in the MODEL statement. The statement specified the log likelihood for a single geo-spatiotemporally geosampled, YFV-related, explanatorial, clinical, field or remote-specified observation in the parameter estimator dataset.

PROC MCMC evaluated every statement for each explanatorial, time series dependent, geosampled YFV-related explanatorial, operationizable, observation. The procedure cumulatively added the log-likelihood for each geopredictor. Stements between the BEGINNODATA and ENDODATA were then invasively evaluated. We noted that at the last YFV- related LULC explanatorial, covariate, the log of the prior and hyperprior distributions were added to the sum of the log-likelihood as to render a robustified log of the posterior distribution.

PROC MCMC updated each block of geosampled, georefernced, YFV-related paarmter while holding all the other covariates constant during iteration. The procedure used mutliple steps to calculate the logs of the posterior distribution. By so doing, all outputs converged parsimoniously. All parameter estimators were then updated. In other words, the procedure did not calculate the conditional probability distribution explicitly for each empirically divided block of geosampled, explanatorial, clinical, field or remote specified explanatorial covariate coefficient value, but instead employed the full joint distribution in the Metropolis algorithm for every block update. Since in our explanatorial, time series dependent forecasting eco-epidemiological risk model we determined dependency based on a step-wise analyses, [i.e., $\log(f(y|\theta)) \neq \sum_i \log(f(y_i|\theta))$], we used the PROC option JOINTMODEL.

In our YFV-related geo-spatiotemrpoal forecasting risk model, we assumed that the inputs were independent and that the joint log likelihood was the sum of individual log-likelihood functions. We specified the log-likelihood of one geosampled, explanatorial, time series dependent, clinical, field and remote-specified variable of one observation in the MODEL statement. PROC MCMC evaluated that function for each, YFV-related, observational predictor in the paramter estimator dataset and cumulatively summed them up. If observations are not independent of each other, this summation produces the incorrect log-likelihood (Hosmer and Lemeshew 2002).

The following statementrs fit a simple, YFV-related model with an unknown mean (mu) in PROC MCMC where the variance in the likelihood was deemed to be known.

363

```
proc mcmc data=exi yfv=7 outpost=p1;
    parm mu;
    prior mu ~ normal(0, sd=10);
    model y ~ normal(mu, sd=1);
  run
```

The following statements then specified the log-likelihood for the entire geosampled emepirical datawseet of YFV-related regressors

```
  proc mcmc data=a outpost=p2 jointmodel;
    array data[1] / nosymbols;
    begincnst;
      rc = read_array("exi", data, "y");
      n = dim(data, 1);
    endcnst;

    parm mu;
    prior mu ~ normal(0, sd=10);
    ll = 0;
```

The MODEL statement indicated a normal likihood for each explanatorial, time series dependent, clinical, field or remote, YFV-related explanatorial varaible y.The 'potential scale reduction factor' (PSRF) was then estimated employing a factor by which the scale of the current distribution for the target distribution was reduced if the simulations were continued for an infinite number of iterations. Each PSRF declined to 1 as the number of iterations approached infinity in the regressed dataset of YFV-related endemic transmission-oriented risk model covariate coefficient values. PSRF is also often represented as R-hat(Cressie 1993). PSRF was calculated for each marginal posterior distribution in x, together with upper and lower confidence limits. Approximate convergence was then diagnosed when the upper limit was close to 1. The recommended proximity of each PSRF to 1 varies(Cressei 1993). Our goal was to achieve PSRF < 1.1. PSRF is an estimate of how much narrower the posterior might become with an infinite number of iterations(Griffith 2003). By so doing, when PSRF = 1.1, for example, was revealed in the regressed seasonal, explanatorial, geopredictive, time series-related, YFV-related, endemic transmission-oriented, risk-related, model, covariatem coefficient values, it was then interpreted as a potential reduction of 10% in posterior interval width, given infinite iterations. The multivariate form bounds above the potential scale reduction factor for any linear combination of the sampled variables were then quantitated.

The confidence limits in the regressed, seasonal ,explanatorial, geopredictive ,time series –related,YFV-oriented, endemic, transmission-oriented, risk model, covariate coefficient indicator values were based on the assumption that the target distribution is stationary and normally distributed. The transform argument was used to improve the normal approximation. A large PSRF indicates that the between-chain variance is substantially greater than the within-chain variance, so that longer simulation is normally needed for proper quantitation (Hosmer and Lemeshew 2000). If a PSRF is close to 1, then the associated chains are likely to have converged to one target distribution. A large PSRF (perhaps generally when a PSRF > 1.2) indicates convergence failure, and can indicate the presence of a multimodal marginal posterior

distribution in which different chains may have converged to different local modes or the need to update the associated chains longer, because burn-in has yet to be completed (Gilks 1996).

We then used Gelman Diagnostics to quantitate the program as to work with the geospatial objects in the eco-epidemiological, forecasting risk model. We employed the gelman.diag function in the coda package . We used two methodologies to estimate the variance of the stationary distribution and cartographically described the time series YFV-related regressors. This estimation was conducted by determining the variance of the stationary distribution. By so doing, the mean of the empirical variance within each chain, *W*, and the empirical variance from all chains combined was robustly quantized .The resdiual forecasts were expressed as *sigma.hat^2 = (n-1)W/n + B/n* where *n* is the number of iterations and *B/n* is the empirical between-chain variance

The Gelman Diagnostic was programmed to work with objects of class demonoid. Since the chains converged,the seasonal explanatorial, geopredictive, time series-related, YFV, endemic, transmission-oriented, eco-epidemiological, risk model, covariate coefficient,estimates were deemed unbiased. We noted that the first method underestimated the variance, since the individual chains had no time to range all over the stationary distribution, and the second method overestimated the variance, since the initial values were chosen to be overdispersed. Regardless, our convergence diagnostic was based on the assumption that each target distribution is normal. A Bayesian estimation probability interval was then constructed using a t-distribution with mean *mu.hat* = Sample mean of all the chains were combined with the variance using *V.hat = sigma.hat2 + B/(mn),*and degrees of freedom estimated by the method of moments *d = 2*V.hat^2 / Var(V.hat).*Use of the t-distribution accounted for the fact that the mean and variance of the posterior distribution estimated. The convergence diagnostic itself was *R=sqrt((d+3) V.hat /((d+1)W).*Values substantially above 1 indicate lack of convergence (Gilks 1996). If the chains have not converged, then Bayesian probability intervals based on the t-distribution were deemed too wide, and thus had the potential to shrink by this factor if the MCMC run is continued.

The multivariate version of Gelman and Rubin's diagnostic was also performed. Unlike the univariate proportional scale reduction factor, the multivariate version did not include an adjustment for the estimated number of degrees of freedom. A list is returned with the following components: PSRF (a list containing the YFV-related explanatorial, geo-spatiotemporal point-estimates of the potential scale reduction factor (labeled Point Est.) and the associated upper confidence limits (e.g., labeled Upper C.I.) and MPSRF (e.g.,point-estimate of the multivariate potential scale reduction factor).

Figure 8. Posterior predictive distributions for the 12 months in 2006 of the yellow fever case count series in Gulu Uganda.
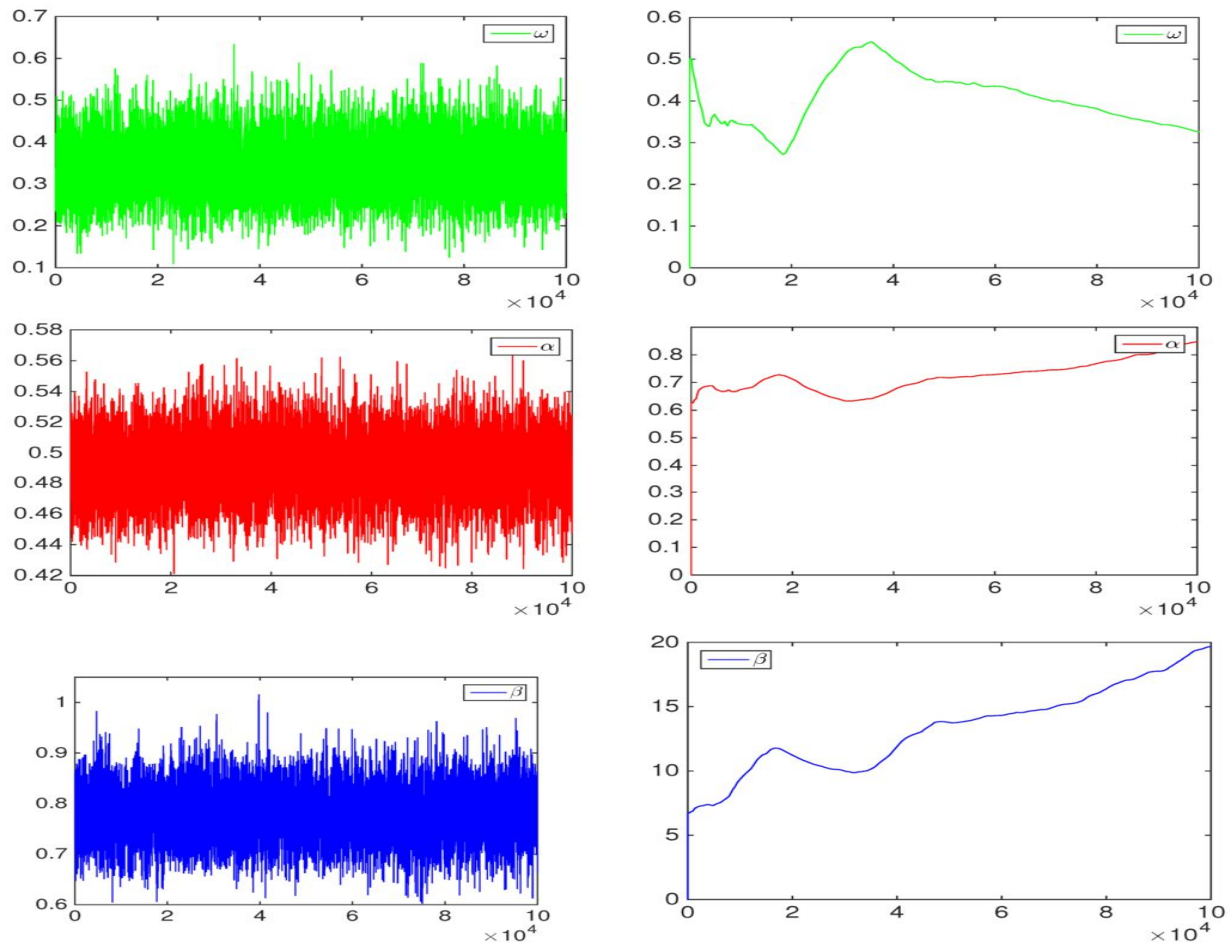
The seasonal YFV-related time series revealed long-term changes in the mean, unstable variance and seasonality of the sampled covariates. A negative binomial Bayesian was then modeled in SAS. When estimating time serie dependent, YFV-related parameter estimators of a negative binomial distribution for describing seasonal vector entomological -related count data, the MCMC chain can become extremely autocorrelated because the parameters are highly correlated (Jacob et al. 2009d, Jacob et al. 2012)

We employed a lognormal and gamma mixed negative binomial (NB) regression model for the geosampled, time series-related, YFV-related counts, and presented efficient closed-form Bayesian inferences. Unlike conventional Poisson models, the proposed approach had two free parameters to include two different kinds of random effects which allowed the incorporation of prior information, such as sparsity in the YFV-related clinical, field and remote-specified time series, regression coefficients. By placing a gamma distribution prior on the negative binomial dispersion parameter r, and connecting a lognormal distribution prior with the logic of the negative binomial probability parameter p, efficient Gibbs sampling and vibrational Bayes

inference were both developed. The closed-form updates were obtained by exploiting conditional conjugacy via both a compound Poisson representation and a Poly-Gamma distribution based data augmentation approach.

After fitting the seasonal YFV-related negative-binomial Bayesian models, both a GSARIMA and a GARIMA deterministic seasonality models were selected based on different criteria. Posterior geopredictive distributions indicated that our YFV-related negative-binomial models provided better geopredictions than Gaussian models, especially when counts were low. The GSARIMA models were able to capture the autocorrelation in the series. We then let $\mathbf{y}^T = (y_t, y_{t+1}, \ldots, y_{t+n})$ be a TFV-related time series of count data of length n arising from a negative binomial distribution $y_t \sim NegBin(\lambda_t, \psi)$ with $E(y_t) = \lambda_t$ and $V(y_t) = \lambda_t + \dfrac{\lambda_t^2}{\psi}$. The limiting form of the negative binomial distribution, that is $\psi \to \infty$, was the Poisson distribution. The YFV-related seasonal explanatory geopredictive $GARMA(p,q)$ model was then written: $g(\lambda_t) = \Phi_p(B)[\mathbf{x}_t^T\boldsymbol{\beta} - g(y_t)] + g(y_t) - \Theta_q(B)[g(y_t) - g(\lambda_t)] + g(y_t) - g(\lambda_t)$ where $g(.)$ was a link function, $\Phi_p(B) = 1 - \phi_1 B^1 - \ldots - \phi_p B^p$, and $\Theta_q(B) = 1 - \theta_1 B^1 - \ldots - \theta_q B^q$. B was a backshift operator with $B^d y_t = y_{t-d}$. Note that in this research $B^d B^D y_t = y_{t-(d+D)}$). $\boldsymbol{\beta}^T = (\beta_0, \beta_1, \beta_2, \ldots, \beta_v)$ was a vector of coefficients for $\mathbf{x}_t^T = (x_0, x_{1,t}, x_{2,t}, \ldots, x_{v,t})$ which included an intercept multiplier taken as $x_0 = 1$ when V was the time dependent covariates. In the GARMA framework, seasonal-sampled YFV-related count data was then modeled via a logarithmic and an identity link function. To avoid the problem of taking the logarithm of the YFV-related observations with value zero under the logarithmic link, we employed a transformation y_t' of y_t where $y_t' = \max(y_t, c), 0 < c \le 1$. This time series explanatory geopredictive YFV-related autoregressive model then was translated into

$$\log(\lambda_t) = \Phi_p(B)\{\log[\exp(\mathbf{x}_t^T\boldsymbol{\beta}) + c] - \log(y_t + c)\} + \log(y_t + c) - \Theta_q(B)\log[(y_t + c)/(\lambda_t + c)] + \log[(y_t + c)/(\lambda_t + c)]$$

The time series YFV-related explanatory geopredictive models were then extended to

$GSARIMA(p,d,q) \times (P,D,Q)_s$ analogues by including seasonality (S) and differencing (I) components as follows: $-\Theta_q(B)\Theta_Q^*(B^s)[g(y_t) - g(\lambda_t)] + g(y_t) - g(\lambda_t)$. where S was the length of the period (i.e., $s = 12$ for monthly rainfall data with an annual cycle), $\Phi_P^*(B^s) = 1 - \phi_1^* B^s - \ldots - \phi_P^* B^{sP}$, $\Theta_Q^*(B^s) = 1 - \theta_1^* B^s - \ldots - \theta_Q^* B^{sQ}$, $\Phi_p(B)$, $\Theta_q(B)$, and which was then quantiated employing . binomial $GARMA(1,1)$ and $GSARIMA(0,0,0) \times (1,1,0)_s$ models with log link function and ZQ1 transformation .

Thereafter , a seasonal YFV explanatory geopredictive endemic transmission-oriented Bayesian model risk model was In Bayesian inference, prior distributions need to be assigned to all model parameters(Cressie 1993). A weakly stationary YFV-related time series model was assumed and, therefore, the auto correlation and moving average parameters were constrained using an algorithm provided by Jones (1987]. For this purpose, the autoregressive and moving average parameters in the likelihood in the explanatory geopredictive YFV-related risk model

were re-parameterized and prior distributions were adopted based on the new parameterization. For instance , the non -seasonal YFV-related autoregressive parameters $\phi_1,...,\phi_p$ were re-parameterized in terms of **r**, $\mathbf{r}^T = (r_1,...,r_p)$, where $\phi_p = 2r_p - 1$ and

$$\phi_{p-i} = 2r_{p-i} - 1 - \sum_{k=1}^{i} (2r_{p-k} - 1)(2r_{p-k+1} - 1), i = 1,...,p-1$$

. The following prior distributions were then assumed: $r_i \sim Beta\left(\left[\frac{1}{2}(i+1)\right], \left[\frac{1}{2}i+1\right]\right), i = 1,...,p$, where $[x]$ denoted the integer part of x. Further priors chosen were $\beta_0,...,\beta_v \sim N(0,1000)$ and $\psi \sim Ga(0.01,0.01)$.

For the first W seasonal YFV –related observations, we employed the residuals on the geopredictor scale (e.g., $\log(y'_t) - \log(\lambda_t)$ were a case of a logarithmic link function was set to zero in a seasonal YFV-related explanatory geopredictive endemic transmission-oriented risk model. A restriction was thereafter put on the mean λ_t itself, that is $\lambda_t \geq 0$ when the identity link was used for parameter estimator quantitation in the risk model . The GSARIMA YFV-related time series models were then estimated using SAS which employed MCMC simulation methods.

For the purpose of Gaussian SARIMA model identification, a Box-Cox transformation was additionally identified by fitting to the seasonal explanatory geopredictive endemic transmission oriented YFV –related case count time series. The fitted Box-Cox parameters revealed a power of 0.271 and, given that the series contained observations with zero counts, a constant of 0.0219 was added to each seasonal YFV-related observation prior for transformation. As observed for the original series, the presence of long-term change in the mean level was apparent in the transformed time series data .

The augmented Dickey–Fuller test supported the presence of a unit-root (p = 0.17) in the Box-Cox transformed series and the series was then differenced. In this research the autoregressive YFV-related risk model was $y_t = \rho y_{t-1} + u_t$ in SAS where y_t was the seasonal sampled variable of interest (e.g., time series LULC change from forested canopy to rice-village complex) , t was the time index, ρ was a coefficient, and u_t was the error term. A unit root is present if $\rho = 1$ (Box and Jenkins 1985). The time series explanatory geopredictive regression model was then written as $\nabla y_t = (\rho - 1)y_{t-1} + u_t = \delta y_{t-1} + u_t$ where ∇ was the first difference operator.

A finite difference is a mathematical expression of the form f(x + b) − f(x + a). If a finite difference is divided by b − a, one gets a difference quotient (Marsden 1985) . The approximation of derivatives by finite differences plays a central role in finite difference methods for the numerical solution of differential equations, especially boundary value problems(Ross 2007). The function difference divided by the point difference is known as the difference quotient (also known as Newton's quotient): $\frac{\Delta F(P)}{\Delta P} = \frac{F(P+\Delta P) - F(P)}{\Delta P} = \frac{\nabla F(P+\Delta P)}{\Delta P}$. If ΔP is infinitesimal, then the difference quotient is a derivative, otherwise it is a difference: thus, If $|\Delta P| = \iota$: $\frac{\Delta F(P)}{\Delta P} = \frac{dF(P)}{dP} = F'(P) = G(P)$; If $|\Delta P| > \iota$: $\frac{\Delta F(P)}{\Delta P} = \frac{DF(P)}{DP} = F[P, P+\Delta P]$.

We then defined the point range in the seasaonl YFV-related explanatory geopredictive endemic trsnmission-oriented risk model Since ΔP was finite in our model, there is (at least—in the case of the derivative—theoretically) a point range, where the boundaries are $P \pm (0.5) \Delta P$ (depending on the orientation—$\Delta F(P)$, $\delta F(P)$ or $\nabla F(P)$): LB = Lower Boundary;   UB = Upper Boundary(Marsden 1985)Then LB $= P_0 = P_0 + 0\Delta_1 P = P_ń - (Ń\text{-}0)\Delta_1 P$;

$P_1 = P_0 + 1\Delta_1 P = P_ń - (Ń\text{-}1)\Delta_1 P$;
$P_2 = P_0 + 2\Delta_1 P = P_ń - (Ń\text{-}2)\Delta_1 P$;
$P_3 = P_0 + 3\Delta_1 P = P_ń - (Ń\text{-}3)\Delta_1 P$;
$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$
$P_{ń\text{-}3} = P_0 + (Ń\text{-}3)\Delta_1 P = P_ń - 3\Delta_1 P$;
$P_{ń\text{-}2} = P_0 + (Ń\text{-}2)\Delta_1 P = P_ń - 2\Delta_1 P$;
$P_{ń\text{-}1} = P_0 + (Ń\text{-}1)\Delta_1 P = P_ń - 1\Delta_1 P$;
UB $= P_{ń\text{-}0} = P_0 + (Ń\text{-}0)\Delta_1 P = P_ń - 0\Delta_1 P = P_ń$;
$\Delta P = \Delta_1 P = P_1 - P_0 = P_2 - P_1 = P_3 - P_2 = ... = P_ń - P_{ń\text{-}1}$;
$\Delta B = UB - LB = P_ń - P_0 = \Delta_ń P = Ń\Delta_1 P$.

In this research derivatives in the seasonal YFV-related risk model were regarded as functions themselves, harboring their own derivatives. Thus each function is home to sequential degrees ("higher orders") of derivation, or differentiation. This property can be generalized to all difference quotients. As this sequencing requires a corresponding boundary splintering, it is practical to break up the point range into smaller, equi-sized sections, with each section being marked by an intermediary point (P_i), where LB $= P_0$ and UB $= P_ń$, the nth point, equaling the degree/order. Recurrence relations can be written as difference equations by replacing iteration notation with finite difference(Jacob et al. 2013).

In this research the primary difference quotient ($Ń = 1$) was $\dfrac{\Delta F(P_0)}{\Delta P} = \dfrac{F(P_ń) - F(P_0)}{\Delta_ń P} = \dfrac{F(P_1) - F(P_0)}{\Delta_1 P} = \dfrac{F(P_1) - F(P_0)}{P_1 - P_0}$. A derivative was then generated using the seasonal YFV-related explanatory covariate coefficients. The difference quotient as a derivative needs no explanation, other than to point out that, since P_0 essentially equals $P_1 = P_2 = ... = P_ń$ (as the differences are infinitesimal), the Leibniz notation and derivative expressions do not distinguish P to P_0 or $P_ń$: $\dfrac{dF(P)}{dP} = \dfrac{F(P_1) - F(P_0)}{dP} = F'(P) = G(P)$. There are other derivative notations, but these are the most recognized, standard designations.

In calculus, Leibniz's notation, uses the symbols dx and dy to represent "infinitely small" (or infinitesimal) increments of x and y, just as Δx and Δy represent finite increments of x and y[Marsden 1985]. For y as a function of x, or $y = f(x)$, the derivative of y with respect to x, which later came to be viewed as $\lim\limits_{\Delta x \to 0} \dfrac{\Delta y}{\Delta x} = \lim\limits_{\Delta x \to 0} \dfrac{f(x + \Delta x) - f(x)}{(x + \Delta x) - x}$, was, according to Leibniz, the quotient of an infinitesimal increment of y by an infinitesimal increment of x, or $\dfrac{dy}{dx} = f'(x)$, where the right hand side is Lagrange's notation for the derivative of f at x. From the point of view of modern infinitesimal theory, Δx is an infinitesimal x-increment, Δy is the corresponding y-increment, and the derivative is the standard part of the infinitesimal ratio: $f'(x) = \text{st}\left(\dfrac{\Delta y}{\Delta x}\right)$ Then

one sets $dx = \Delta x$, $dy = f'(x)dx$, so that by definition, $f'(x)$ is the ratio of *dy* by *dx*. Similarly, although mathematicians sometimes now view an integral $\int f(x)\,dx$ as a limit $\lim_{\Delta x \to 0} \sum_i f(x_i)\,\Delta x$, where Δx is an interval containing x_i, Leibniz viewed it as the sum (the integral sign denoting summation) of infinitely many infinitesimal quantities *f(x) dx*. From the modern viewpoint, it is more correct to view the integral as the standard part of an infinite sum of such quantities

In order to quantitate a divided difference for our , however, did require further elucidation, as it equals the average derivative between and including LB and UB:

$$P_{(tn)} = LB + \frac{TN-1}{UT-1}\Delta B = UB - \frac{UT-TN}{UT-1}\Delta B \quad (P_{(1)} = LB; \; P_{(ut)} = UB)$$

$$F'(P_{\tilde{a}}) = F'(LB < P < UB) = \sum_{TN=1}^{UT=\infty} \frac{F'(P_{(tn)})}{UT}$$

In this interpretation, $P_{\tilde{a}}$ represents a function extracted, average value of P (midrange, but usually not exactly midpoint), the particular valuation depending on the function averaging it is extracted from. More formally, $P_{\tilde{a}}$ is found in the <u>mean value theorem</u> of calculus, which says: For any function that is continuous on [LB,UB] and differentiable on (LB,UB) there exists some $P_{\tilde{a}}$ in the interval (LB,UB) such that the secant joining the endpoints of the interval [LB,UB] is parallel to the tangent at $P_{\tilde{a}}$. Essentially, $P_{\tilde{a}}$ denotes some value of P between LB and UB—hence

$P_a := LB < P < UB = P_0 < P < P_{\tilde{n}}$ which links the mean value result with the divided difference?

$$\frac{DF(P_0)}{DP} = F[P_0, P_1] = \frac{F(P_1) - F(P_0)}{P_1 - P_0} = F'(P_0 < P < P_1) = \sum_{TN=1}^{UT=\infty} \frac{F'(P_{(tn)})}{UT},$$

$$= \frac{DF(LB)}{DB} = \frac{\Delta F(LB)}{\Delta B} = \frac{\nabla F(UB)}{\Delta B},$$

$$= F[LB, UB] = \frac{F(UB) - F(LB)}{UB - LB},$$

$$= F'(LB < P < UB) = G(LB < P < UB).$$

As there is, by its very definition, a tangible difference between LB/P_0 and $UB/P_{\tilde{n}}$, the Leibniz and derivative expressions *do* require divarication of the function argument. Higher order quotients were then derived for the seasonal YFV-related risk model using :

$$\frac{\Delta^2 F(P_0)}{\Delta_1 P^2} = \frac{\Delta F'(P_0)}{\Delta_1 P} = \frac{\frac{\Delta F(P_1)}{\Delta_1 P} - \frac{\Delta F(P_0)}{\Delta_1 P}}{\Delta_1 P},$$

$$= \frac{\frac{F(P_2) - F(P_1)}{\Delta_1 P} - \frac{F(P_1) - F(P_0)}{\Delta_1 P}}{\Delta_1 P},$$

$$= \frac{F(P_2) - 2F(P_1) + F(P_0)}{\Delta_1 P^2};$$

$$\frac{d^2 F(P)}{dP^2} = \frac{dF'(P)}{dP} = \frac{F'(P_1) - F'(P_0)}{dP},$$

$$= \frac{dG(P)}{dP} = \frac{G(P_1) - G(P_0)}{dP},$$

$$= \frac{F(P_2) - 2F(P_1) + F(P_0)}{dP^2},$$

$$= F''(P) = G'(P) = H(P)$$

$$\frac{D^2 F(P_0)}{DP^2} = \frac{DF'(P_0)}{DP} = \frac{F'(P_1 < P < P_2) - F'(P_0 < P < P_1)}{P_1 - P_0},$$

$$\neq \frac{F'(P_1) - F'(P_0)}{P_1 - P_0},$$

$$= F[P_0, P_1, P_2] = \frac{F(P_2) - 2F(P_1) + F(P_0)}{(P_1 - P_0)^2},$$

$$= F''(P_0 < P < P_2) = \sum_{TN=1}^{\infty} \frac{F''(P_{(in)})}{UT},$$

$$= G'(P_0 < P < P_2) = H(P_0 < P < P_2).$$

Thereafter, a third order YFV-related quotient was derived using:

$$\frac{\Delta^3 F(P_0)}{\Delta_1 P^3} = \frac{\Delta^2 F'(P_0)}{\Delta_1 P^2} = \frac{\Delta F''(P_0)}{\Delta_1 P} = \frac{\frac{\Delta F'(P_1)}{\Delta_1 P} - \frac{\Delta F'(P_0)}{\Delta_1 P}}{\Delta_1 P},$$

$$= \frac{\frac{\frac{\Delta F(P_2)}{\Delta_1 P} - \frac{\Delta F'(P_1)}{\Delta_1 P}}{\Delta_1 P} - \frac{\frac{\Delta F'(P_1)}{\Delta_1 P} - \frac{\Delta F'(P_0)}{\Delta_1 P}}{\Delta_1 P}}{\Delta_2 P},$$

$$= \frac{\frac{F(P_3) - 2F(P_2) + F(P_1)}{\Delta_1 P^2} - \frac{F(P_2) - 2F(P_1) + F(P_0)}{\Delta_1 P^2}}{\Delta_1 P},$$

$$= \frac{F(P_3) - 3F(P_2) + 3F(P_1) - F(P_0)}{\Delta_1 P^3};$$

$$\frac{d^3 F(P)}{dP^3} = \frac{d^2 F'(P)}{dP^2} = \frac{dF''(P)}{dP} = \frac{F''(P_1) - F''(P_0)}{dP},$$

$$= \frac{d^2 G(P)}{dP^2} = \frac{dG'(P)}{dP} = \frac{G'(P_1) - G'(P_0)}{dP},$$

$$= \frac{dH(P)}{dP} = \frac{H(P_1) - H(P_0)}{dP},$$

$$= \frac{G(P_2) - 2G(P_1) + G(P_0)}{dP^2},$$

$$= \frac{F(P_3) - 3F(P_2) + 3F(P_1) - F(P_0)}{dP^3},$$

$$= F'''(P) = G''(P) = H'(P) = I(P);$$

$$\frac{D^3 F(P_0)}{DP^3} = \frac{D^2 F'(P_0)}{DP^2} = \frac{D F''(P_0)}{DP} = \frac{F''(P_1 < P < P_3) - F''(P_0 < P < P_2)}{P_1 - P_0},$$

$$\neq \frac{F''(P_1) - F''(P_0)}{P_1 - P_0},$$

$$= \frac{\frac{F'(P_2 < P < P_3) - F'(P_1 < P < P_2)}{P_1 - P_0} - \frac{F'(P_1 < P < P_2) - F'(P_0 < P < P_1)}{P_1 - P_0}}{P_1 - P_0},$$

$$= \frac{F'(P_2 < P < P_3) - 2F'(P_1 < P < P_2) + F'(P_0 < P < P_1)}{(P_1 - P_0)^2},$$

$$= F[P_0, P_1, P_2, P_3] = \frac{F(P_3) - 3F(P_2) + 3F(P_1) - F(P_0)}{(P_1 - P_0)^3},$$

$$= F'''(P_0 < P < P_3) = \sum_{TN=1}^{UT=\infty} \frac{F'''(P_{(tn)})}{UT},$$

$$= G''(P_0 < P < P_3) = H'(P_0 < P < P_3) = I(P_0 < P < P_3).$$

Finally an $\acute{N}$th order for the seasonal YFV-related risk model was generated:

$$\Delta^{\acute{n}} F(P_0) = F^{(\acute{n}-1)}(P_1) - F^{(\acute{n}-1)}(P_0),$$

$$- \frac{F^{(\acute{n}-2)}(P_2) - F^{(\acute{n}-2)}(P_1)}{\Delta_1 P} - \frac{F^{(\acute{n}-2)}(P_1) - F^{(\acute{n}-2)}(P_0)}{\Delta_1 P},$$

$$- \frac{\frac{F^{(\acute{n}-3)}(P_3) - F^{(\acute{n}-3)}(P_2)}{\Delta_1 P} - \frac{F^{(\acute{n}-3)}(P_2) - F^{(\acute{n}-3)}(P_1)}{\Delta_1 P}}{\Delta_1 P},$$

$$- \frac{\frac{F^{(\acute{n}-3)}(P_2) - F^{(\acute{n}-3)}(P_1)}{\Delta_1 P} - \frac{F^{(\acute{n}-3)}(P_1) - F^{(\acute{n}-3)}(P_0)}{\Delta_1 P}}{\Delta_1 P},$$

$$\frac{\Delta^{\acute{n}} F(P_0)}{\Delta_1 P^{\acute{n}}} = \frac{\sum_{I=0}^{\acute{N}} \binom{-1}{\acute{N}-I} \binom{\acute{N}}{I} F(P_0 + I \Delta_1 P)}{\Delta_1 P^{\acute{n}}};$$

$$\frac{\nabla^{\acute{n}} F(P_{\acute{n}})}{\Delta_1 P^{\acute{n}}}$$

$$= \frac{\sum_{I=0}^{\acute{N}} \binom{-1}{I} \binom{\acute{N}}{I} F(P_{\acute{n}} - I \Delta_1 P)}{\Delta_1 P^{\acute{n}}};$$

$$\frac{d^{\hat{n}}F(P_0)}{dP^{\hat{n}}} = \frac{d^{\hat{n}-1}F'(P_0)}{dP^{\hat{n}-1}} = \frac{d^{\hat{n}-2}F''(P_0)}{dP^{\hat{n}-2}} = \frac{d^{\hat{n}-3}F'''(P_0)}{dP^{\hat{n}-3}} = \cdots = \frac{d^{\hat{n}-r}F^{(r)}(P_0)}{dP^{\hat{n}-r}},$$

$$= \frac{d^{\hat{n}-1}G(P_0)}{dP^{\hat{n}-1}}$$

$$= \frac{d^{\hat{n}-2}G'(P_0)}{dP^{\hat{n}-2}} = \frac{d^{\hat{n}-3}G''(P_0)}{dP^{\hat{n}-3}} = \cdots = \frac{d^{\hat{n}-r}G^{(r-1)}(P_0)}{dP^{\hat{n}-r}},$$

$$= \frac{d^{\hat{n}-2}H(P_0)}{dP^{\hat{n}-2}} = \frac{d^{\hat{n}-3}H'(P_0)}{dP^{\hat{n}-3}} = \cdots = \frac{d^{\hat{n}-r}H^{(r-2)}(P_0)}{dP^{\hat{n}-r}},$$

$$= \frac{d^{\hat{n}-3}I(P_0)}{dP^{\hat{n}-3}} = \cdots = \frac{d^{\hat{n}-r}I^{(r-3)}(P_0)}{dP^{\hat{n}-r}},$$

$$= F^{(\hat{n})}(P) = G^{(\hat{n}-1)}(P) = H^{(\hat{n}-2)}(P) = I^{(\hat{n}-3)}(P) = \cdots$$

$$\frac{D^{\hat{n}}F(P_0)}{DP^{\hat{n}}} = F[P_0, P_1, P_2, P_3, \ldots, P_{\hat{n}-3}, P_{\hat{n}-2}, P_{\hat{n}-1}, P_{\hat{n}}],$$

$$= F^{(\hat{n})}(P_0 < P < P_{\hat{n}}) = \sum_{TN=1}^{UT=\infty} \frac{F^{(\hat{n})}(P_{(tn)})}{UT}$$

$$= F^{(\hat{n})}(LB < P < UB) = G^{(\hat{n}-1)}(LB < P < UB) = \cdots$$

The quintessential application of the divided difference was then employed as presentation of a definite integral, which in this research was nothing more than generating a finite difference for the risk model using::

$$\int_{LB}^{UB} G(p)\,dp = \int_{LB}^{UB} F'(p)\,dp = F(UB) - F(LB),$$

$$= F[LB, UB]\Delta B,$$

$$= F'(LB < P < UB)\Delta B,$$

$$= G(LB < P < UB)\Delta B.$$

Given that the mean value, derivative expression form provides all of the same seasonal YFV-related information as the classical integral notation, the mean value may was then the expression, that only supported standard ASCII text, or in cases that only required the average derivative (such as when finding the average radius in an elliptic integral). This is especially true for definite integrals that technically have (e.g.) 0 and either π or 2π as boundaries, with the same divided difference found as that with boundaries of 0 and $\frac{\pi}{2}$ (thus requiring less averaging effort):

$$\int_0^{2\pi} F'(p)\,dp = 4\int_0^{\frac{\pi}{2}} F'(p)\,dp = F(2\pi) - F(0) = 4(F(\tfrac{\pi}{2}) - F(0)),$$

$$= 2\pi F[0, 2\pi] = 2\pi F'(0 < P < 2\pi),$$

$$= 2\pi F[0, \tfrac{\pi}{2}] = 2\pi F'(0 < P < \tfrac{\pi}{2}).$$

This also become when dealing with iterated and multiple integrals ($\Delta A = AU - AL$, $\Delta B = BU - BL$, $\Delta C = CU - CL$):

$$\int_{CL}^{CU} \int_{BL}^{BU} \int_{AL}^{AU} F'(r, q, p)\,dp\,dq\,dr$$

$$= \sum_{TC=1}^{UC=\infty} \left(\sum_{TB=1}^{UB=\infty} \left(\sum_{TA=1}^{UA=\infty} F'(R_{(tc)} : Q_{(tb)} : P_{(ta)}) \frac{\Delta A}{UA} \right) \frac{\Delta B}{UB} \right) \frac{\Delta C}{UC},$$

$$= F'(CL < R < CU : BL < Q < BU : AL < P < AU)\Delta A\,\Delta B\,\Delta C.$$

Hence, $F'(R, Q : AL < P < AU) = \sum_{TA=1}^{UA=\infty} \frac{F'(R, Q : P_{(ta)})}{UA}$; and

$$F'(R : BL < Q < BU : AL < P < AU) = \sum_{TB=1}^{UB=\infty} \left(\sum_{TA=1}^{UA=\infty} \frac{F'(R : Q_{(tb)} : P_{(ta)})}{UA} \right) \frac{1}{UB}.$$

This model was estimated and testing for a unit root is equivalent to testing $\delta = 0$ (where $\delta = \rho - 1$). Since the test is done over the residual term rather than raw seasonal YFV data, it was not possible to use standard t-distribution to provide critical values. Therefore this statistic τ has a specific distribution simply known as the table. There are three main versions of the test:1. test for a unit root[$\nabla y_t = \delta y_{t-1} + u_t$];2. test for a unit root with drift:$\nabla y_t = a_0 + \delta y_{t-1} + u_t$;and,3. Test for a unit root with drift and deterministic time trend:$\nabla y_t = a_0 + a_1 t + \delta y_{t-1} + u_t$Each version of the test has its own critical value which depends on the size of the sample. For our explanatory geopredictive YFV-related endemic transmission-oriented risk model , the null hypothesis was that there was a linear relationship with case distribution and changes in LULC from forested canopy to rice-village complex [i.e., unit root, $\delta = 0$] . To test the statistical power of our seasonal YFV-related model we attempted to distinguish between true unit-root processes ($\delta = 0$)and near unit-root processes [i.e., δ is close to zero ("near observation equivalence" problem)].

Our intuition behind the test for the seasonal, YFV-related, explanatory, predictive, endemic, transmission-oriented, risk model   was as follows. If the series y was stationary or trend stationary in the YFV-related model, then we assumed the residual derivatives would have a tendency to return to a constant or deterministically trending mean. Therefore, large geosampled values(e.g., forested canopy to agro-village complex LULC change over 10 years) will tend to be followed by smaller values(monthly mean rainfall during the dry season) (negative changes), and small values by larger values (positive changes). Accordingly, the level of the series in the seasonal, YFV-related, endemic transmission-oriented, eco-epidemiological,

geopredictive, risk model would then be a significant observational predictor of next period's change, and will have a negative coefficient. If, on the other hand, the series is integrated in the YFV model then positive changes and negative changes will occur in the residual risk-related derivatives with probabilities that do not depend on any current level of the series; (e.g., in a random walk).It is notable to mention that the regression-based YFV-related, probablistic, explanatorial, equation: $\nabla y_t = a_0 + u_t$ was rewritten as $y_t = y_0 + \sum_{i=1}^{t} u_i + a_0 t$ as to quantify any time series dependent, explanatorial deterministic trends coming from $a_0 t$ for quantitating a stochastic intercept term coming from $y_0 + \sum_{i=1}^{t} u_i$.

Thereafter, P=plots of the LULC autocorrelation function (ACF) and the partial auto correlation function (PACF) of the differenced series showed significant auto correlation at lags of three and twelve months. Based on the preliminary analysis of the Box-Cox transformed series, four Gaussian SARIMA models and two Gaussian ARIMA models with second order harmonics (SOH) were initially selected, based on AIC (see Table 1).

Table 7. Akaike's information criterion (AIC) for selected (Gaussian) models on Box-Cox transformed seasonal YFV-related data sampled at the Gulu study site

Model	Excluding Rainfall	Including Rainfall
SARIMA(3',1,0)X(1,0,0)$_{12}$	1725.52	1729.58
SARIMA(3',1,0)X(0,0,1)$_{12}$	1726.86	1729.42
SARIMA(0,1,3')X(1,0,0)$_{12}$	1725.71	1729.28
SARIMA(0,1,3')X(0,0,1)$_{12}$	1726.91	1729.41
ARIMA(3',1,0)-SOH	1722.75	1719.11
ARIMA(3',1,0)-SOH	1721.63	1719.62

Legend: SOH: second order harmonics. For all these models, where applicable, the autoregressive (ϕ_1 and ϕ_2) or moving average parameters (θ_1 and θ_2) corresponding to the first two lags were omitted.

SARIMA parameters, p, d, q, P, D, and Q, was constructed employing standard (frequentist) tools developed for time series with Gaussian marginal errors, rather than through fitting many possible MCMC models in SAS. A visual analysis of the explanatorial, YFV-related time series moderl ouput revealed the presence of a long-term (inter annual) change in the mean level, an unstable variance which appeared to increase with the mean and multiplicative seasonality. The size of the seasonal effect in the geosampled eco-epidemiological dataset of clinical, field or remote-specified, YFV-related, time series dependnet, explanatorial endemic, transmission-oriented, covariate paramter estimators was

proportional to the mean. Thus, for the preliminary Gaussian analysis, the geosampled YFR-related, reference data was transformed using a fitted Box-Cox transformation in order to stabilize the variance and to make the seasonal effect additive for approximating a normal distribution. The trend in the Box-Cox transformed series was treated as a stochastic trend, which was (first order) difference stationary. The augmented Dickey – Fuller test on a lag order of 15 was used to detect the presence of a unit root, to assess whether the series needed to be integrated (i.e., differenced). Gaussian SARIMA models and ARIMA models with a second order harmonic seasonal component, both with $d = 1$ because of the presence of a unit root, were fitted with a frequentist framework. The SAS 'models were then evaluated based on Akaike's information criterion (AIC). The covariate matrix for the seasonal effect using second order harmonics (i.e., two sine and cosine pairs) was given by $x_t^T = [\sin(2\pi t/12), \cos(2\pi t/12), \sin(2\pi t/6), \cos(2\pi t/6)]$. A time independent intercept was not included because the intercept dropped out of the equation after first order differencing.

We found that the ARIMA-SOH, sub-meter resolution, *Ae. egypti* , oviposition, eco-georferenecable, eco-epidemiological , forecast, vulnerability, geospectrotemporal,YFV-related models had the lower (better) AIC compared to SARIMA models. A SARIMA model is usually

Time series analysis and forecasting is an efficient versatile tool in diverse applications such as in economics and finance, hydrology and environmental management fields just to mention a few. Among the most effective approaches for analyzing time series data, the method propounded by Box and Jenkins, the Autoregressive Integrated Moving Average (ARIMA) was employed in this study. In this paper, we used Box-Jenkins methodology to build ARIMA model for Nigeria's monthly inflation rates for the period November 2003 to October 2013 with a total of 120 data points. In this research, ARIMA (1, 1, 1) (0, 0, 1)12 model was developed, and

obtained as $\hat{y}_{t+1} =$ $0.3587 y_t + 0.6413 y_{t-1} - 0.8840 e_{t-11} - 0.7308912 e_{t-12} + 0.8268 e_t$

ARIMA-SOH models including rainfall as a covariate parameter estimator had a slightly lower AIC than ARIMA-SOH models without rainfall. However, for the SARIMA models, the inverse was true. Bayesian negative binomial variants of these selected, probablistically regressed, YFV-related, time series models were built. In order to establish ω, the model with the largest lag required, *w,* needed to be identified for comparison of the DIC of these Bayesian models. This was the model $GSARIMA(3,1,0) \times (1,1,0)_{12}$ with $w = 18$ Models with logarithmic link function performed better than models with identity link.

Based on the DIC, the best negative binomial model was the negative binomial $GARIMA(3,1,0)$ model with parameters for the first two lags (ϕ_1 and ϕ_2) omitted (fixed to zero), with deterministic harmonic seasonality and with rainfall preceding malaria with two months (Table 2). However, based on the MARE on the out of sample predictions for the second half of the time series, when the model was fitted to the first half, the negative binomial $GSARIMA(3',1,0) \times (1,1,0)_{12}$ model (the 'prime' in the "3" indicating that also here the parameters for the first two lags were fixed to zero) without rainfall as covariate, was preferred.

This model also had the best overall MARE. The parameter and deviance estimates for this model, henceforth "$\text{GARIMA}(3',1,0)\text{-SOH-RF}$", are detailed in Table 3.

Table 8. Parameter estimates (mean and 95% credible interval) of selected seasonal YFV- related negative binomial models fitted to the entire time series

Parameter	GARIMA(3',1,0)-SOH-RF	GSARIMA(3',1,0) X(1,0,0)12
βrain	-0.23(-0.54, -0.03)	
$\beta\sin(2\pi t/12)$	-0.09(-0.45,-0.03)	
$\beta\cos(2\pi t/12)$	-0.14 (-0.48,-0.3)	
$\beta\sin(2\pi t/6)$	0.17(0.03,0.22)	
$\beta\cos(2\pi t/6)$	0.19(0.04,0.28)	
$\phi 3$	-0.10(-0.26, 0.01)	-0.11(-0.21,-0.05)
$\phi 1^*$		0.14(0.04.0.26)
ψ		5.33(4.71.6.04)
Amplitude AH$	13(3.17, 6.33)	
Amplitude SAH$	0.17(0.11,0.39)	
Phase shift AH$	5.83(2.37,7.11)	
Phase shift SAH$	-0.57(-1.16,-0.31)	

GARIMA(3',1,0)-SOH-RF = GARIMA(3,1,0) model with parameters for the first two lags (ϕ_1 and ϕ_2) omitted, second order harmonics and rainfall lagged at 2 months (in m); GSARIMA(3',1,0)x(1,0,0)$_{12}$ = GSARIMA(3,1,0)x(1,0,0)$_{12}$ model with parameters for the first two lags (ϕ_1 and ϕ_2) fixed to zero; AH = annual harmonic, SAH = semi-annual harmonic; $^$ = derived parameter, phase shift = phase shift of the cosine function expressed in months.

Despite the $\text{GSARIMA}(3',1,0) \times (1,1,0)_{12}$model having a higher (worse) DIC than the $\text{GARIMA}(3',1,0)\text{-SOH-RF}$ model, the out of sample MARE of the $\text{GSARIMA}(3',1,0) \times (1,1,0)_{12}$model was 5.7 per cent better than the out of sample MARE of the $\text{GARIMA}(3',1,0)\text{-SOH-RF}$ model, and required less than half the number of fitted parameters. This indicates that the $\text{GARIMA}(3',1,0)\text{-SOH-RF}$ model was probably over-fitting the data, describing the random error rather than the underlying process. The $\text{GSARIMA}(3',1,0) \times (1,1,0)_{12}$ model was selected for further analysis. posterior predictive distributions for the last 12 months of the series by the $\text{GSARIMA}(3',1,0) \times (1,1,0)_{12}$model and those by a (Bayesian) Gaussian $\text{SARIMA}(3',1,0) \times (1,1,0)_{12}$ model on Box-Cox transformed data, when fitted to the entire data set. Differences in the posterior predictive distributions between the two models are apparent with the Gaussian model predictive distributions having longer right tails.

The C-R plot of the negative binomial $\text{GSARIMA}(3',1,0) \times (1,1,0)_{12}$ model fit was compared to that of a (Bayesian) Gaussian $\text{SARIMA}(3',1,0) \times (1,1,0)_{12}$ on Box-Cox transformed data in Figure 3. The C-R plot on the entire series is not entirely satisfactory for either model. For the Gaussian $\text{SARIMA}(3',1,0) \times (1,1,0)_{12}$, the posterior predictive distribution appears to be platykurtic (for values of the residual probability below 0.5, there are too few observations, and for values above 0.5, there are too many). For the negative binomial $\text{GSARIMA}(3',1,0) \times (1,1,0)_{12}$ model, for randomized residual probability values below about 0.5, cumulatively fewer observations had these values than the posterior density distributions had indicated. Therefore, on average, the part of the posterior density distributions below the median was spread out too much to the left. The lower boundaries of credibility intervals of the distributions were thus on average too low. For the values above 0.5, the cumulative distribution function followed the diagonal. We then compared both models for the last 50 months of the series only, where numbers of monthly cases were smaller than 35. For these low numbers, the negative binomial $\text{GSARIMA}(3',1,0) \times (1,1,0)_{12}$ model was much more appropriate

We then generated the normal Q-Q plot for the normalized, randomized, probablistic, quantile residuals of the $\text{SARIMA}(3',1,0) \times (1,1,0)_{12}$ model, for which the distribution was slightly leptokurtic. A plot of these quantile residuals against time appeared as a random scatter at first sight, but upon closer inspection, extreme residuals occurred more often during periods with stronger relative changes. This was because the residuals, ε_t, were positively correlated with a relative change in YFV cases, with linear regression line $\varepsilon_t = 1.85 \log\left(\frac{y_t}{y_{t-1}}\right) + 0.22$, $R^2 = 0.93$

Initially rainfall and an LULC column represent rainfall for specific year and interface distance between human population and forest regions respectively were entered into C++ model. TheYFV-related cases represented the number of yellow fever cases reported for a particular year. From these input table, we created a formula of form as below,Z = a0 + a1*X + a2*Y and YFV cases = a0 + a1*Rainfall + a2*LULC. We then determine the values of constants a0 a1 and a2. We predicted the value of YFV-related cases without considering effect of human population in initial stage and then later we add the effect of population over yellow fever prevalence.

To determine a relationship between our YFV-related explanatorial, endemic, transmission-oriented, seasonal data, we need tr 3-D regression. We also used Least square regression plane of the form Z = a0 + a1*X + a2*Y that provided the best fit for the input data points sampled at the Gulu study site. It should be noted here, there can be multiple solutions of these equations but values of interests are those, which provides the best fit to the input data. For this purpose, only those values were used which minimized the following least square value function: $F(a0, a1, a2) = \sum(Zi - a0 - a1*X - a2*Y)^2$. The solution of this formula was found with the matrices as the system $\partial F/\partial a0 = 0$, $\partial F/\partial a1 = 0$, $\partial F/\partial a2 = 0$ is a linear system of equation. The matrix can be used to solve the constants is as follows:

$$\begin{bmatrix} \Sigma\,Xi^2 & \Sigma\,XiYi & \Sigma\,Xi \\ \Sigma\,XiYi & \Sigma\,Yi^2 & \Sigma\,Yi \\ \Sigma\,Xi & \Sigma\,Yi & N \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \Sigma\,XiZi \\ \Sigma\,YiZi \\ \Sigma\,Zi \end{bmatrix}$$

Solving these equation will provide with the unique values of constants. In our case these values were as follows: a0 = 60.9239, a1 = 0.1957 a2= -65.4218. So the formula can be written as follows:
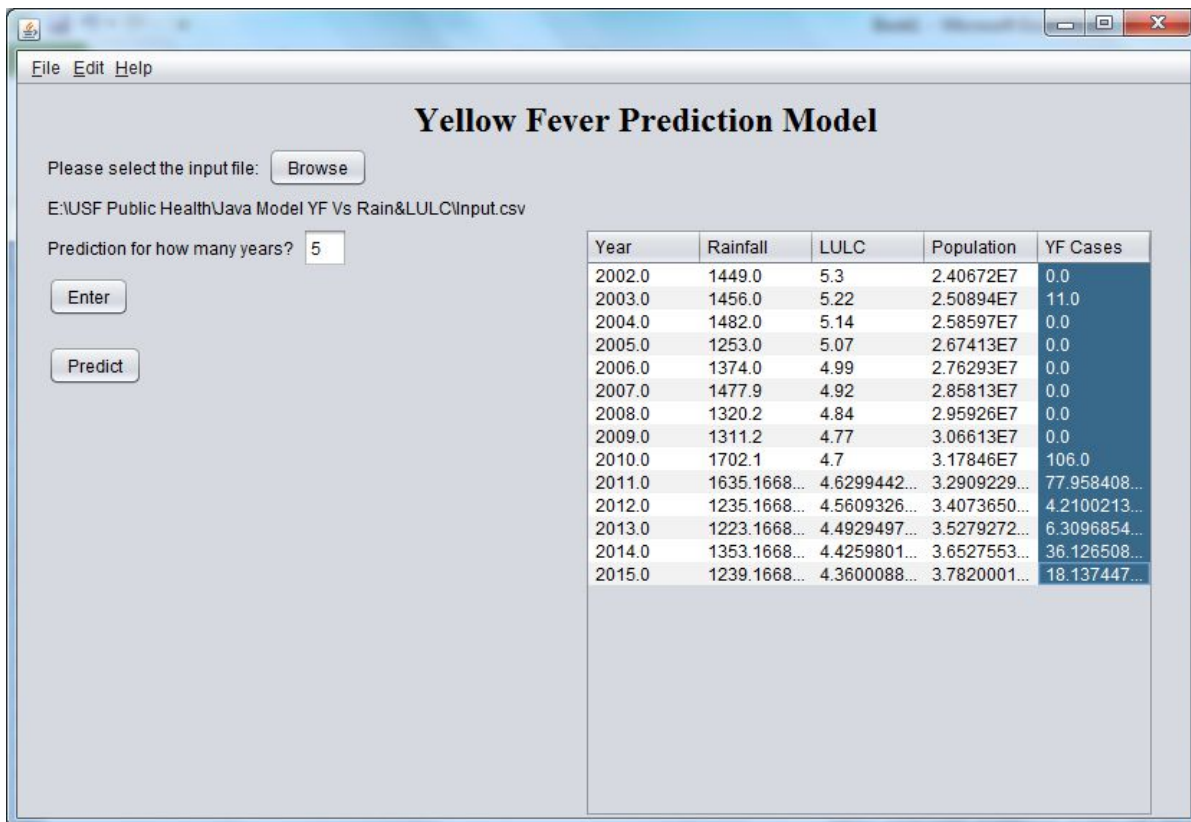
YFCases = 60.9239 + 0.1957 * Rainfall - 65.4218 * LULCThese YFCases values predicted does not consider effect of population on it. So to consider population data on this prediction, we can re-write the formula as below:

YFCases = (1+PI_Factor) * [60.9239 + 0.1957 * Rainfall - 65.4218 * LULC]

where, (Population of Prediction Year – Population of Last Available Year)
PI_Factor =Population of Last Available Year.

Table 9 Values generated by a formula for quantitating the effect of rainfall, LULC and population yellow
fever cases.

**Table 10. The table of geopredictor YFR-related cases at the Gulu study site**

Year	Rainfall	LULC	Population	YF Cases
2002.0	1449.0	5.3	2.40672E7	0.0
2003.0	1456.0	5.22	2.50894E7	11.0
2004.0	1482.0	5.14	2.58597E7	0.0
2005.0	1253.0	5.07	2.67413E7	0.0
2006.0	1374.0	4.99	2.76293E7	0.0
2007.0	1477.9	4.92	2.85813E7	0.0
2008.0	1320.2	4.84	2.95926E7	0.0
2009.0	1311.2	4.77	3.06613E7	0.0
2010.0	1702.1	4.7	3.17846E7	106.0
2011.0	1635.1668...	4.6299442...	3.2909229...	77.958408...
2012.0	1235.1668...	4.5609326...	3.4073650...	4.2100213...
2013.0	1223.1668...	4.4929497...	3.5279272...	6.3096854...
2014.0	1353.1668...	4.4259801...	3.6527553...	36.126508...
2015.0	1239.1668...	4.3600088...	3.7820001...	18.137447...

Year	Rainfall	LULC	Population	Yellow fever Cases
2002	1449.00	5.30	2406.7200	0
2003	1456.00	5.22	2508.9400	11
2004	1482.00	5.14	2585.9700	0
2005	1253.00	5.07	2674.1300	0
2006	1374.00	4.99	2762.9300	0
2007	1477.90	4.92	2858.1300	0
2008	1320.20	4.84	2959,2600	0
2009	1311.20	4.77	30661300	0
2010	1702.10	4.70	31784600	106
2011	1635.17	4.63	32909229	78
2012	1235.17	4.56	34073651	4
2013	1223.17	4.49	35279273	6
2014	1353.17	4.43	36527553	36
2015	1239.17	4.36	37820002	18

It should be noted that rainfall data is predicted randomly using the current mean and variance of the rainfall data.

Figure 9: 3-D representation of our C++ YFV-related geopredictive risk model
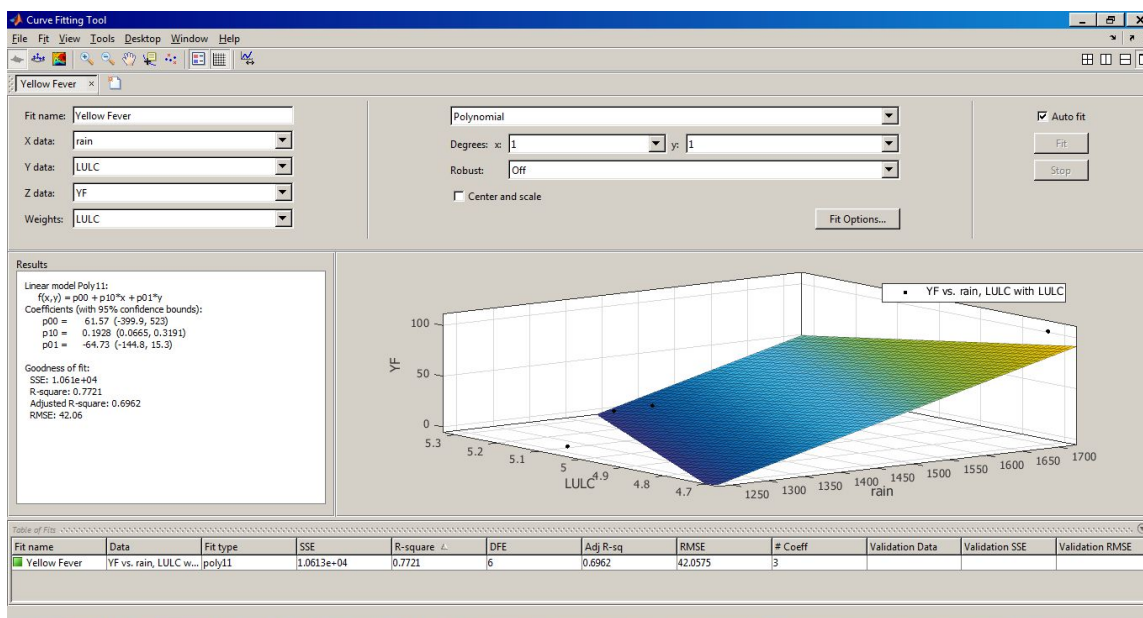
Discussion

A landscape approach using remote sensing SAS and ArcGIS technologies was developed to discriminate between the agro-villages at the Gulu eco-epidemiological study site for determining high and low risk areas of endemic, ecohydrologic, ecogeographic, YFV-related transmission variables. QuickBird visible and NIR data were digitally processed to generate a cartographic model of risk-related landscape elements in Geospatial Analyst^{TM}. The landscape elements surrounding the agro-village complexes and the geoclassified canopied forest LULC where defined using a robust NDVI map. The relationships between specific landscape morphological elements were probabilistically regressively quantitated using stepwise discriminant analysis and previous prevalence rates as the respone (dependent) variable in PROC STEPDISC. The BSSCP and TSSCP options displayed the LULC between-class SSCP matrix and the total-sample corrected SSCP matrix. By default, the significance level of an Ftest from the analysis covariance was used as the selection criterion. The variable under consideration was the dependent variable, and the geo-spectrotemporal, ecogeorfeerenceable YFV , explicative diagnostic clinical, field and remote geosampled *Ae. egypti* , oviposition LULC variables already chosen act as covariates The analyses indicated that the most important elements in terms of explaining LULC abundance at the Gulu eco-epidemiological study site was the proportion of seasonal change between quantitated, forest, canopied LULC and agro-village complexes at the Gulu study site.

In the explanatorial, time series, YFV-related, endemic, transmission-oriented,3-D, DEM , the Flow direction operation determined if natural water flow influenced a forest, canopied, LULC mixel and/or a neighboring rice complex, agro-village, LULC mixel. Flow direction was calculated for every central mixel employing input blocks of various level of mixels (e.g., 3x3). During each geosample frame the value of the central mixel with the value of its neighbors was calculated. The output map contained flow directions as N (to the North), NE (to the North East), etc. For determining the steepest slope, a method for each input block of LULC mixels was performed in an ArcGIS operation. In so doing, the procedure calculated the height difference between the forest canopied LULC mixels and each of the neighboring, riceland, LULC mixels in the scene. If, for a neighboring seasonal, endemic, transmission-oriented, sub-mter resolution, *Ae. egypti* , oviposition, YFV-related, LULC mixel, the difference was positive (i.e. central mixel had larger spectral value than the specified neighbor), then for the corner neighbors, the spectral covariate parameterizable estimator, differences were divided by Euclidean distance measurements from forest canopied, LULC and agro-village complex boundaries to the georefernced village-complex centroids. Meanwhile for geo-spatiotemporally quantitating horizontal neighbor mixel ovisposition values in the seasonal, ecohydrologic, ecogeographic, YFV-related, explanatorial LULC, covariate, parameter estimator, mixel differences were divided by Euclidean distance measurements. By so doing, the steepness between the forest canopy LULC mixel and its neighboring agro-village complex LULC were determined.

Conversely, to geospectrally validate the lowest reflectance wavelenght emissivities in the multitemporal, QuickBird, seasonal, ecohydrologic, YFV-related, LULCs, we employed each input block of mixels. Thereafter, the ArcGIS operation calculated the difference between the central mixel and each of the neighbor mixels in an explanatorial, geopredictive, time series, LULC, eco-epidemiological, forecasting, risk map format. We noted that from all neighbors with

a positive height difference, (i.e. when central LULC mixel had larger value than a specific neighbor), the position of the neighbor with largest positive height difference was the output flow direction for the central mixel. Further, LULC, sub-resolution *Ae. egypti* , oviposition pixels along the edges of the input forecast canopy inhomogebeosu map (i.e., margins and corners) were defined in the output map.

Interestingly, diagnostically if all neighboring LULC mixels of a central rice-village complex had a large explanatorial, clinical, field and remote, YFV-related, covariate, parameter estimator, coefficient mesurement value than the central mixel itself (i.e. a sink or pit), the undefined value was displayed in the output, geopredictive, endemic, transmission-oriented, ecohydrologic, ecogeographic, vulnerability, eco-georeferenceable, time series, LULC map. When a central LULC mixel had an undefined value, it was displayed in the output map. Neighbor mixels that had the undefined LULC values were ignored during the calculation. If then fdata from neighboring agro-village complex LULC pixels were considered, three adjacent neighbor mixels from a canopied, forested, explanatorial LULC in a single row or column were identified in SAS/GIS and then parsimoniously robustly quantitated to determine slope and height values. Conversely, if neighboring canopied, agro-forested, LULC pixels were considered, three neighboring agro-village complex LULC mixels were remotely   identified and their slope and height value thereafter qualitatively quantitated.

We then constructed a Poisson regression model in PROC NL MIXED from the multiple explanatorial, empirically geosampled, geopredictive, time series dependent, YFV-related, covariate, parameter estimator coefficient values. A Poisson process is a commonly employed starting point for analyzing seasonal, explanatorial, geopredictive, vector, arthropod-related, endemic, transmission-oriented, ecohydrologic, ecogeographic, eco-epidemiological, risk models of stochastically interpreted, entomological count, data variation when quantitating regressively qualitatively a theoretical expectation (see Jacob et al. 2005b).The statistical analysis of continuous data that is non-negative is a common task in quantitative ecology (McDonald 2008). In our seasonal, Poissonized probabilty paradigm both the number of gamma variates, and their average size, was regressed separately. The model had a composite link that varied within the probabilistically geosampled, geospectrotemporal or geo-spatiotemporal, wavelength, frequency, signature, YFV-related, clinical, field and remote-specified, time series dependent, geoclassified, LULC, covariate, parameterizable, regression estimators. Unlike a normal distribution, the Poisson is a natural distribution for count data (Neter et al., 1990), but overdispersion in our regression coefficients suggested that the eco-epidemiologcal, YFV-related, probabilistic, *Ae. aegypti* risk model was inappropriate for differentiating the geo-spatiotemporal, explanatorial, geopredictive variables, seasonally geosampled at the Gulu eco-epidemiological study site.

Overdispersion in explanatorial, seasonal, geopredictive, entomological-related, eco-epidemiological, probabilistic, risk-related, data analyses is often accounted for by using models with different assumptions about how the variance changes with the expectation (Jacob et al. 2014, Jacob et al. 2009). The choice of these assumptions can naturally have apparent consequences for statistical inference generated from probabilistic, sylvatic YFV-related, geoclassified LULC-related, geo-spatiotemporal or geo-spectrotemporal, regression based models. For instance, we proposed a parameterization of the negative binomial distribution, where two overdispersed-related, seasonal, endmember, LULC, sub-meter resolution YFR-

related, probabilsitic geo-spectrotemporally geosampled, *syla Ae. egypti* , oviposition explanatorial, covariate, parameter estimators (i.e., LULC changes from forested canopy and rainfall) were introduced into an endemic, transmission-oriented, eco-epidemiological, forecasting, regression-related, probablistic, ecohydrologic, ecogeographic, geo-spatiotemporal, risk model. The model then regressively qualitatively quantitated the various quadratic mean–variance relationships. Taking the Poisson mean as a gamma distributed random variable lead to regression model we subsequently employed to obtain various forms of mean-variance relationship, in particular both linear and quadratic.This chose was dependent on assumptions about the gamma mixing distribution parameters (Hosmer and Lemeshew 2000). A eco-epidemiological, linear mean-variance, sylvatic YFV, forecasting, vulnerability model was obtained by allowing the gamma shape parameter to vary across the explanatorial, diagnostic, clinical, field or remote geosampled, time series observations and keeping the scale parameter constant, whereas the quadratic form arose from taking the shape parameter as a constant by letting the scale vary. However, the regression residuals indicated that an inappropriate model fit existed due to overdispersion caused by spatial outliers.

Outliers in seasonal, geopredictive, explanatorial, mosquito, risk-based, endemic, transmission-oriented, forecasting, sylvatic, risk models can occur by chance in any distribution, but they are often indicative either of measurement error or that the geosampled population has a heavy-tailed distribution (i.e., probability distributions whose tails are not exponentially bounded) (see Jacob et al., 2005b Jacob et al. 2009d). Assuming that the distribution of the geopredictive, explanatorial, disganostic, clinical, field and remote-specified, time series dependent, geo-spectrotemporal, YFV-related, ecogeographic endemic, transmission-oriented, *Ae. egypti*, oviposition data geosampled at the Gulu eco-epidemiological study site was in fact Poisson, the square root transformation may have tended to homogenize the variance of the eco-epidemiological, probabilistically regressed, geoclassified, LULC ecohydrologic, variables and/or the logarithmic transformation may have not stabilized the positive variance-mean relationships in the residually forecasted derivatives. Failure of the Poisson assumption of equidispersion in the endemic, YFV-related, forecasting, eco-epidemiological, risk model may then have been then due to the failure of the assumption of homoscedasticity associated with a normalized distribution for "Poissionizing" the empirical geosampled, time series-related, probablistic, eco-epidemiological dataset of explanatorial, clinical, field and remote-geosampled, geopredictive, ecogeographical, YFV-related geocalssified LULC variables.

Fortunately, we found that the negative binomial regression with a non-homogenous gamma-distributed mean in PROC NL MIXED could handle the overdispersion in the time series-related, geo-spectrotemporally geosampled explanatorial, geopredictive, YFV-related, endemic, transmission-oriented, probabilistic, dataset of operationizable, endmember, sylvatic, *Ae. aegypti* LULC, sub-meter resolution ,vulnerability model, explicatively diagnostic clinical, field and remote, covariate paramter estimator coefficients. The parameterized, time series-related, explanatorial, probablistic, negative binomial distribution converged with the Poisson distribution and controlled the deviation from the Poisson in the risk model. This made the negative binomial distribution suitable as a robust alternative to the Poissonian for regressively qualitatively quantitating the seasonal, geo-spectrotemporal or geo-spatiotemporal, geosampled, YFV-related, empirically predictive, *Ae. egypt*, oviposition endemic, transmission-oriented, parameterized, eco-epidemiological, clinical, field or remote covariate, parameter estimator

coefficient values. Further, since the negative binomial distribution has one more parameter than the Poisson (Neter et al., 1990), the second parameter adjusted the variance independently of the mean in the, time series, YFV, linear, probabilistically regressed, geoclassified, LULC risk model.

The negative binomial distribution in PROC NL MIXED especially in its alternative parameterization, was efficiently employed as an alternative to the Poissonian distribution when regressively qualitatively quantitating an empirical geosampled dataset of explanatorial, clinical, field and remote–specified, YFV-related, geosampled, geoclassified LULC, georeferenceable variables. This statistical distribution is useful for evaluating whether discrete ,explicative time series dependent, explanatorial, geopredictive, YFV-related, uncoalesced LULC, *Ae. egypti*, oviposition, endember, data feature attributes over an unbounded positive range sample variance exceeds the sample mean. In such cases, the geo-spectrotemporal or geo-spatiotemporal, geosampled observational, explanatorial prognosticators with respect to a time series dependent Poissonian, seasonal, YFV-related, explanatorial distribution, for which the mean is equal to the variance (see Haight 1967), will be qualitatively, probabilistically, regressively quantitated even if the geosampled, LULC data are overdispersed. Hence, although a Poisson distribution is not an appropriate. sylvatic ,YFV African riceland, capture point, geo-spectrotemrpoal or geo-spatiotemporal, forecast-oriented, sub-meter resolution, vulnerability, model for regressively quantitating an geo-spatiotemporal, empirically, probabilistically, regressable dataset of eco-epidemiological, seasonal, ecogeorefencable, geopredictive, diagnostic, time series, clinical, field and remote specified variables and explanatorial, LULC covariate, parameterizable estimator coefficients, a negative binomial with a non-homogenous gamma distributed mean is.

Further, since the negative binomial distribution has one more parameter than the Poisson, the second parameter can be employed to adjust the variance independently of the mean in any robust explanatorial, seasonal, geopredictive, sylvatic YFV-related, linear, endemic, transmission-oriented, eco-epidemiological, probablistic, riceland African agro-irrgation, prognosticative, risk model. Forexample, if an arbovirologist medical entomologist or other YF experimenter considers a sequence of negative binomial distributions from a probabilstically regressed dataset of seasonal geo-spatiotemporally or geo-spectrotemporally geosampled, time series dependent, clinical, field and remote-geospecified, YFV-related, geopredictive, LULC-related, time series, covariate, *Ae. egypti* parameterizable, oviposition,specified, estimators where the stopping parameter r goes to infinity, the probability of success in each trial, p, would go to zero in such a way as to keep the mean of the distribution constant. In the YF , predictive, eco-epidemiological, sub-mter resolution, grid-stratified, orthogonal vulnerability, oviposition model Denoting this mean λ in the explanatorial, seasonal, empirically regressed, YFV-related, eco-epidemiological, probabilty, geoclassified, endmember, sub-meter resolution, LULC, risk model, observational, explanatorial predictors thereafter, the parameter p was validated using

...p=r/(lambda+r) $\lambda = r\dfrac{p}{1-p} \quad \Rightarrow \quad p = \dfrac{\lambda}{r+\lambda}.$ Under this parameterization of the mass function the geopredictive, explanatorial, parameter estimators were efficiently quantitated by

$$f(k;r,p) = \frac{\Gamma(k+r)}{k! \cdot \Gamma(r)}(1-p)^k p^r = \frac{\lambda^k}{k!} \cdot \frac{\Gamma(r+k)}{\Gamma(r)\,(r+\lambda)^k} \cdot \frac{1}{\left(1+\frac{\lambda}{r}\right)^r}.$$

We successfully employed PROC NL MIXED to fit all of the time series-related, explanatorial, predictive, YFV-related, endemic, transmission-oriented, clinical, field and remote geospecified, empirical, Poissonized and negative binomial regression model, LULC-related, covariate, *Ae. egypti*, oviposition, parameterized estimators. The syntax structure and progression across the models constructed allowed us to clearly demonstrate the differences, and similarities, between these models. PROC NL MIXED is a highly flexible procedure that can be used to run and qualitatively quantitate a large variety of seasonally geopredictive, explanatorial, geosampled, endemic, transmission-oriented, time series, clinical, field and remote geosampled, forecasting, eco-epidemiological, geoclassified, sub-meter resolution, sylvatic LULC, risk models. PROC NLMIXED fits specified nonlinear, mixed, time series dependent, explanatorial, probabilistic, YFV-related, endemic, transmission-oriented, endmember, clinical, field and remote geo-specified, eco-epidemiological, eco-georferenceable, estimators by maximizing an approximation to the likelihood integrated over the random effects.

Fortunately, different approximations to the integral are available in PROC NLMIXED, .For our endmember, LULC, sub-meter resolution YFV-related explanatorial, time series dependent, explanatorial, eco-epidemiological--related risk model we employed the adaptive Gaussian quadrature and a first-order Taylor series approximation. The default method in PROC NLMIXED computed an integral in the adaptive Gaussian quadrature as described in Pinheiro and Bates (1995). Another approximation method we employed was the first-order method of Beal and Sheiner (1982, 1988). We quantitated the approximation $p(y_i|\mathbf{X}_i, \boldsymbol{\phi}, \mathbf{u}_i)$ as normal—that is, $p(y_i|\mathbf{X}_i, \boldsymbol{\phi}, \mathbf{u}_i) = (2\pi)^{-n_i/2} |\mathbf{R}_i(\mathbf{X}_i, \boldsymbol{\phi})|^{-1/2}$ and $\exp\left\{-(1/2)\left[y_i - \mathbf{m}_i(\mathbf{X}_i, \boldsymbol{\phi}, \mathbf{u}_i)\right]' \mathbf{R}_i(\mathbf{X}_i, \boldsymbol{\phi})^{-1}\left[y_i - \mathbf{m}_i(\mathbf{X}_i, \boldsymbol{\phi}, \mathbf{u}_i)\right]\right\}$ where n_i was the dimension of y_i, $\mathbf{R}_i$ was a diagonal variance matrix, and $\mathbf{m}_i$ is the conditional mean vector of y_i. Here, the first-order approximation for our explanatorial, eco-epidemiological, geopredictive, YFV-related forecasting, LULC-related, *Ae. egypti*, oviposition, risk model was obtained by $\mathbf{m}_i(\mathbf{X}_i, \boldsymbol{\phi}, \mathbf{u}_i)$ with a one-term Taylor series expansion about $\mathbf{u}_i = \mathbf{0}$, resulting in the approximation $p(y_i|\mathbf{X}_i, \boldsymbol{\phi}, \mathbf{u}_i) = (2\pi)^{-n_i/2} |\mathbf{R}_i(\mathbf{X}_i, \boldsymbol{\phi})|^{-1/2} \exp\left(-(1/2)\left[y_i - \mathbf{m}_i(\mathbf{X}_i, \boldsymbol{\phi}, 0) - \mathbf{Z}_i(\mathbf{X}_i, \boldsymbol{\phi})\mathbf{u}_i\right]'\right)$ and $\mathbf{R}_i(\mathbf{X}_i, \boldsymbol{\phi})^{-1}\left[y_i - \mathbf{m}_i(\mathbf{X}_i, \boldsymbol{\phi}, 0) - \mathbf{Z}_i(\mathbf{X}_i, \boldsymbol{\phi})\mathbf{u}_i\right]$ where $\mathbf{Z}_i(\mathbf{X}_i, \boldsymbol{\phi})$ was the Jacobian matrix [i.e., $\partial \mathbf{m}_i(\mathbf{X}_i, \boldsymbol{\phi}, \mathbf{u}_i)/\partial \mathbf{u}_i$] evaluated at $\mathbf{u}_i = \mathbf{0}$.

The Jacobian matrix is the matrix of all first-order partial derivatives of a vector-valued function for any paridym (e.g., a explanatorial, geopredictive, clinical, field and remote-geosampled, sylvatic, YFV-related time series dependent, eco-epidemiological, forecasting, risk model) (Griffith 2003). Specifically, in the model $f : \mathbb{R}^n \to \mathbb{R}^m$ was a function which took as input the vector $x \in \mathbb{R}^n$ and rendered an output t vector $f(x) \in \mathbb{R}^m$. Then the YFV Jacobian matrix J of f was an $m \times n$ matrix, which was defined and arranged as follows:
$$\mathbf{J} = \frac{d\mathbf{f}}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$
or, component-wise: $\mathbf{J}_{i,j} = \frac{\partial f_i}{\partial x_j}$ in AUTOREG. This matrix, whose entries were functions of x, was also denoted by Df, J_f, and $\partial(f_1,...,f_m)/\partial(x_1,...,x_n)$. The Jacobian matrix was important because if the function f was differentiable at a geo-spatiotemrpoally geosampled, eco-geoefernced, YFV-related, explanatorial, sylvatic,LULC point x (e.g., georeferenced, forest canopy, LULC, eco-epidemiological, explanatorial, clinical, field or remote forecastable regressor) then there would exist a slightly stronger condition in the matrix than that which could be quantiated merely by all existing partial derivatives. The Jacobian matrix

defined a robust , geospatial optimal, geopredictive, explanatorial, clinical, field and remote-geosampled, *Ae. egypti* , oviposition, YFV-related, time series dependent, eco-epidemiological, forecasting, vulnearbility, risk map $\mathbb{R}^n \to \mathbb{R}^m$, which was the best linear approximation of the function *f* near an LULC agroecoystem and forest canopied, georefernced, geopredictive, variable point *x* in the matrix. This YFV linear map was thus the generalization of the usual notion of derivative, and was the derivative or the differential of *f* at *x*.

If *m = n*, the Jacobian matrix is a square matrix, and its determinant, a function of x_1, …, x_n, is the Jacobian determinant of f (Griffth 2003).Thus, optimal, geospatially geopredictive, explanatorial, diagnostic, geo-spectrotemporal, endmember, LULC, sub-meter resolution clinical, field and remote-geosampled, YFV-related, time series dependent, eco-epidemiological, forecasting, LULC-related, vulnearbility specified, probabilistically weighted, Jacobian matrices could carry important information about the local behavior of *f*. In particular, the function *f* can be regressively qualitatively quantitated locally in the neighborhood of a point x where an inverse function is differentiable if and only if the Jacobian determinant is nonzero at x (i.e., Jacobian conjecture).

The Jacobian conjecture in the plane, first stated by Keller (1939), states that given a ring map F of $\mathbb{C}[x,y]$(e.g., the polynomial ring in two georferenceable, clinical, field or remote specified, geo-spatiotemporal, YFV-related, probablistic, geopredictive variables over the complex geosampled discrete, explanatorial, geoclassified, ovispoition, sylvatic, LULC,integer values$\mathbb{C}$) to itself that fixes $\mathbb{C}$and sends x, $y$to f, $g$respectively, where $F$is an automorphism if the Jacobian $f_x\,g_y - f_y\,g_x$is a nonzero element of $\mathbb{C}$. The condition can easily be shown to be necessary, but proving sufficiency has been an open problem since Keller (1939). The Jacobian conjecture is one of Smale's problems. Smale's problems are a list of 18 challenging problems for the twenty-first century proposed by Field medalist Steven Smale who was inspired in part by Hilbert's famous list of problems presented in 1900 (Hilbert's problems) in part in response to a suggestion by V. I. Arnold on behalf of the International Mathematical Union (Hochster 2004).

Regardless, suppose *N > 1* is a fixed, geosampled seasonal, ecogeoreferncable, explanatorial, probablistically regressable, clincial, field or remote specified, time series dependent integer where the polynomials f_1, ..., f_N with forecasting, eco-epidemiological, risk model variables X_1, ..., X_N and their measurement predictor indicator values in an algebraically closed field *k*. It then suffices to assume *k = C* in any probabilsitically empirically, regressed dataset of an probabilistic, eco-epidemiological, sylvatic, YFV-related, geosampled, forecasting, georeferncable,LULC risk model derivatives. A ecologist, entomologist or other experimenter could optimally define a vector-valued function *F*: $k^N \to k^N$ by setting:$F(c_1, ..., c_N) = (f_1(c_1, ...,c_N),...,f_N(c_1,...,c_N))$ in AUTOREG. By so doing, the Jacobian determinant would occur when changing the, geo-spatiotemporally, geosampled, explanatorial, endmember, LULC, sub-meter resolution, YFV-related, time series dependent, *Ae. egypti* , ovipositionm forecast, vulnerability, variables in the operationizable datset of multi-variable integrals (see substitution rule for multiple variables).If *m = 1* in the probablistic, explanatorial, YFV-related, time series dependent, eco-epidemiological,clinical, field or remote geosampled, georeferenced, forecasting, risk model, *f* , for instance, would be a scalar field and the Jacobian matrix would be reduced to a row vector of partial derivatives of *f*—(i.e. the gradient of *f*).

Assuming that $q(u_i|\xi)$ is normal with mean 0 and the variance matrix $G(\xi)$, is the first-order integral approximation, a robust dataset of empirically probablistically regressed, time series dependnet, endemic, eco-epidemiological, explanatorial, YFV-related, forecasting, clinical, field or remote-geosampled, risk model, covariate, parameter estimator, coefficient values may be probabilistically qualiatively derived in closed form after completing the square:

$$\int p(y_i|X_i,\phi,u_i)q(u_i|\xi)du_i \approx (2\pi)^{-n_i/2}|V_i(X_i,\theta)|^{-1/2}\exp\left(-(1/2)[y_i-m_i(X_i,\phi,0)]'V_i(X_i,\theta)^{-1}[y_i-m_i(X_i,\phi,0)]\right)$$ where

$V_i(X_i,\theta)=Z_i(X_i,\phi)G(\xi)Z_i(X_i,\phi)'+R_i(X_i,\phi)$. The resulting approximation for $f(\theta)$ would then be optimally heuristically minimized over $\theta=[\phi,\xi]$ to obtain the first-order estimates. PROC NLMIXED uses finite-difference derivatives of the first-order integral approximation when carrying out the default dual quasi-Newton optimization (www. http://support.sas.com/documentation/).

NLPQN sub-routine uses dual quasi-Newton optimization techniques, and it is one of the two sub-routines available that can solve problems with nonlinear constraints in SAS. These techniques would work well for medium to moderately large explanatorial, optimization problems in an empiricially regressable, operationizable, endemic, datset of geo-spatiotemporally geosampled, geopredictive, *Ae. egypti* , oviposition, YFV-related time series dependent, eco-epidemiological, forecasting, geoclassified LULC, eco-epidemiological risk model covariate, paramter estimators where the objective function and the gradient would be much faster to clinical, field and remote-geosampled compute than the Hessian matrix. In mathematics, the Hessian matrix or Hessian is a square matrix of second-order partial derivatives of a scalar-valued function, or scalar field which describes the local curvature of a function of many geopredictive variables (Cressie 1993). The NLPQN sub-routine does not need to compute second-order derivatives, but it generally requires more iterations than the techniques that compute second-order derivatives. The two categories of problems solved by the NLPQN subroutine are unconstrained or linearly constrained problems and explanatorily nonlinearly constrained problems(www.sas.edu). Traditionally unconstrained or linearly constrained problems do not use the *"nlc"* or *"jacnlc"* module arguments, but the nonlinearly constrained problems in the e *Ae. egypti*, oviposition, xplanatorial, eco-epidemiological, eco-georeferenceable, clinical, field and remote-geosampled, YFV-related, time series dependent, eco-epidemiological, endmember, LULC, sub-meter resolution model may employ the arguments to specify the explanatorial, nonlinear constraints and the probabilsitically geopredictive, heuristically optimaizable, Jacobian matrix, for instance, of their first-order derivatives, respectively.

The type of optimization problem specified in SAS determines the algorithm that the subroutine invokes(www.sas.edu). The algorithms are very different, and they use different sets of termination criteria. Here the PROC NL MIXED was employed to fit the time series dependent, explicative, diagnostic, clinical, field or remote, geo-spectrotemrpoally geosampled, YFV-related, geopredictive, eco-epidemiological, forecasting, probabilistic, eco-epidemiological, risk model, covariate, parameter estimators. Of all the SAS/STAT procedures that perform at least one type of regression analysis (e.g., CATMOD, GENMOD, GLM, LOGISTIC, PROC REG NLIN, ORTHOREG, PROBIT, RSREG, and TRANSREG procedure) PROC NL MIXED models fit are viewed as generalizations of the geosampled random coefficient models fit.

SAS/ETS procedures have been updated for specialized applications in time-series or simultaneous systems by the NL MIXED procedure. (www.sas.edu). This generalization allowed the geo-spectrotemporal, geosampled, YFV-related, explanatorial, clinical, field or remote specified random coefficients to enter the model nonlinearly. Within PROC MIXED a maximum likelihood and restricted maximum likelihood (REML) estimation was performed on the explanatorial, clinical, field and remote –specified, regressors. In the risk model, the analog to the REML method in PROC NLMIXED involved a high-dimensional integral over all of the fixed-effects, YFV-related explanatorial, endmember, LULC, sub-meter resolution, *Ae. egypti*, oviposition geo-spatiotemrpoal, geosampled, probabilistically geoclassifed, covariate parameter estimators, but this integral was not available in closed form. PROC NLMIXED enables analyze data that are normal, binomial, or Poisson or that have any likelihood programmable with SAS statements(www.sas.edu).

PROC NLMIXED did not implement the same probabilsitic estimation techniques available with the NLINMIX macro or the default estimation method of the GLIMMIX procedure. Instead these techniques were based on the estimation methods of Lindstrom and Bates (1990), Breslow and Clayton (1993), and Wolfinger and O'Connell (1993). The time series dependent, eco-epidmeiological, probabilistically regressed, endmember, LULC, sub-meter resolution, YFV residually forecasted derivatives iteratively fit a set of generalized estimating equations. PROC NLMIXED directly maximized an approximate integrated likelihood We proposed the following logistic nonlinear mixed model for the YFV-related geoclassified LULC data: $y_{ij} = \dfrac{b_1 + u_1}{1 + \exp[-(d_{ij} - b_2)/b_3]} + e_{ij}$ Here, y_{ij} was represented by the jth measurement on the ith tree ($i = 1, \ldots, 5$; $j = 1, \ldots, 7$), d_{ij} was the correspondingseasonal covariate, paramter estimators, b_1, b_2, b_3 where the fixed-effects parameters, $u_1$were the random-effect parameters assumed to be independnetly distributed [i.e., $N(0, \sigma_u^2)$], and e_{ij}were the residual errors assumed to be $N(0, \sigma_e^2)$and independent of the u_{i1}. This model had a logistic form, and the random-effect clinical, field and remote geosampled YFV-related LUC covariate parameter estimators [e.g.,] u_{i1}enter the model linearly. The statements to fit this nonlinear mixed model was follows:

```
proc nlmixed data=tree;
   parms b1=160 b2=70 b3=35 s2u=100 s2e=6;
   num = b1+u1;
   ex  = exp(-(seasonal-b2)/b3);
   den = 1 + ex;
   model y ~ normal(num/den,s2e);
   random u1 ~ normal(0,s2u) subject=tree;
run;
```

The PROC NLMIXED statement then invoked the procedure and inputs the *tree* data set. The PARMS statement identifies the unknown parameters and their starting values. Here there were three fixed-effects YFV-related LULC geoclassified *Ae. egypti*, oviposition, parameters (*b1*, *b2*, *b3*) and two variance components (*s2u*, *s2e*).

The next SAS programming statements for specifying an explicative diagnostic clinical, field or remote geosampled, time series, logistic mixed model was a new geopredictive variable *u1* which included identification of the random effect. These statements were evaluated for every probabilsitically regressed, time series dependent, empiricial YFV-related, geoclassified, explanatorial, LULC observation in the dataset. The NLMIXED procedure computed the log likelihood function and its derivatives. The MODEL statement then defined the dependent variable (e.g., YFV-related prevalance) and its conditional distribution given the random effects at the ecoepidemioloigcal, riceland agro-ecosystem study site. Here a normal (Gaussian) conditional distribution was specified with mean *num/den* and variance *s2e*. Meanwhile,the RANDOM statement defined the single random effect to be *u1*, and then specified that it followed a normal distribution with mean 0 and variance *s2u*. The SUBJECT= argument in the RANDOM statement defined a single YFV-related explanatorial, clinical, field or remote geo-spatiotemporally, geosampled, geopredictive, risk-related endmeic, transmission-oriented, eco-epidemioloigcal variable for indicating when the random effect obtained new realizations; in this case, it changes according to the values of the *tree* variable. PROC NLMIXED assumed that the input data set was clustered according to the levels of the *tree* variable; that is, all explanatorial, clinical, field or remote-related *Ae. egypti* , oviposition,YFV-oriented geoclassified LULC, eco-georferenceable, observations from the same tree occured sequentially in the input dataset.

The GLIMMIX procedure also fit a mixed model for quantiating the non-normal data with nonlinearity in the conditional mean function in our YFV-related LULC, endmeic , transmission-oriented risk-oriented, eco-epidemiological, analyses. The GLIMMIX procedure by default probabilistically estimated the geosampled, LULC, sub-meter resolution YFV-risk-related, endmember, predicor variables in a generalized linear mixed model framework by pseudo-likelihood techniques. Given the probabilsitically regressable, random geosampled, YFV-related, geopredictive, LULC, eco-epidemiological, risk variables $X = X_1, X_2, ..., X_n$ and a set E of dependencies between these random variables, where $\{X_i, X_j\} \notin E$ implies X_i was conditionally independent of X_j given X_i's neighbors, the pseudolikelihood of $X = x = (x_1, x_2, ...x_n)$ was $\Pr(X = x) = \prod_i \Pr(X_i = x_i | X_j = x_j \text{ for all } j \text{ for which } \{X_i, X_j\} \in E).$ where X was a vector of the geosampled variables, x was a vector of the covariate paramter estimator coefficient values. $X = x$ above meant that each geo-spatiotemrpoally, geosampled variable X_i in the vector X had a corresponding value x_i in the vector x. The expression $\Pr(X = x)$ was the probability that the vector of variables X had coefficient values equal to the vector x. Because situations can often be described using state variables ranging over a set of possible empiricallly regressable,vector arthropod-related clinical, field or remote geo-spatiotemrpoally geosampled forecastable, covariate paramter estimator coefficient values(Jacob et al 2011), the expression $\Pr(X = x)$ may robustly represent the probability of a certain state among all possible YFV-related, *Ae. egypti* , oviposition, seasonal LULC change states allowed by the state variables.

The GLIMMIX procedure by default estimated the probabilsitically geosampled YFV-related, explanatorial, clincial, field or remote geo-spatiotemrpoal, geosampled LULC values in a generalized linear mixed model framework by pseudo-likelihood techniques. Alternatively

PROC NLMIXED by default performs maximum likelihood estimation by adaptive Gauss-Hermite quadrature. Hermite-Gauss quadrature, also called Hermite quadrature, is a Gaussian quadrature over the interval $(-\infty, \infty)$ with weighting function $W(x) = e^{-x^2}$ (Abramowitz and Stegun 1972,). The abscissas for quadrature order n may be then given by the roots x_i of the Hermite polynomials $H_n(x)$, which occur symmetrically about 0. The YFV-related, geoclassified, time series dependent, LULC weights would be

$$-\frac{A_{n+1}\,\gamma_n}{A_n\,H_n'(x_i)\,H_{n+1}(x_i)} \quad w_i = \frac{A_n}{A_{n-1}} \quad \frac{\gamma_{n-1}}{H_{n-1}(x_i)\,H_n'(x_i)},$$

where A_n is the coefficient of x^n in $H_n(x)$. For Hermite polynomials, $A_n = 2^n$, so $\frac{A_{n+1}}{A_n} = 2$. Additionally, $\gamma_n = \sqrt{\pi}\,2^n\,n!$, so

$$w_i = -\frac{2^{n+1}\,n!\,\sqrt{\pi}}{H_{n+1}(x_i)\,H_n'(x_i)} = \frac{2^n\,(n-1)!\,\sqrt{\pi}}{H_{n-1}(x_i)\,H_n'(x_i)} = \frac{2^{n+1}\,n!\,\sqrt{\pi}}{[H_n'(x_i)]^2} = \frac{2^{n+1}\,n!\,\sqrt{\pi}}{[H_{n+1}(x_i)]^2} = \frac{2^{n-1}\,n!\,\sqrt{\pi}}{n^2\,[H_{n-1}(x_i)]^2},$$

where it would follow the recurrence relation [e.g., $H_n'(x) = 2\,n\,H_{n-1}(x) = 2\,x\,H_n(x) - H_{n+1}(x)$] to obtain $H_n'(x_i) = 2\,n\,H_{n-1}(x_i) = -H_{n+1}(x_i)$, (Abramowitz and Stegun 1972 p. 890). In a The error term woulod be

$$E = \frac{n!\,\sqrt{\pi}}{2^n\,(2\,n)!}\,f^{(2n)}(\xi).$$

Thus, if h is a function in a seasonal, geo-spatiotemporally geosampled, clinical, field or remote, endmember, LULC, sub-meter resolution, *Ae. egypti* , oviposition, YFV-related, observational explanatorial, geopredictive, ecogeoreferenceable, observational variable y is normally distributed $\mathcal{N}(\mu, \sigma^2)$, the expectation of h would correspond to the following integral:

$$E[h(y)] = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) h(y)\,dy$$

However as this does not exactly correspond to the Hermite polynomial, an ecologist, entomologist, or experimenter may need a change of variable to

$$x = \frac{y-\mu}{\sqrt{2}\sigma} \Leftrightarrow y = \sqrt{2}\sigma x + \mu$$

Coupled with an integration by substitution method then the geo-spatiotemporal, YFV-related, eco-epidemiological, forecatsing, *Ae. egypti* , oviposition, vulnerability model would be expressed as

$$E[h(y)] = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} \exp(-x^2) h(\sqrt{2}\sigma x + \mu)\,dx \qquad E[h(y)] \approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^{n} w_i h(\sqrt{2}\sigma x_i + \mu)$$

leading to:

Let $I \subseteq \mathbb{R}$ be an interval and $\varphi : [a,b] \to I$ be a continuously differentiable function in an explanatorial, probabilistically, clinical, field or remote geosampled, YFV-related, eco-epidemiological, forecasting, risk model. Suppose that $f : I \to \mathbb{R}$ is then a continuous function in the model. Then $\int_{\varphi(a)}^{\varphi(b)} f(x)\,dx = \int_a^b f(\varphi(t))\varphi'(t)\,dt$. In other notation: the substitution $x = \varphi(t)$ would yield $dx/dt = \varphi'(t)$ in the eco-epidemiological, endmember, LULC, sub-meter resolution sylvatic, YFV-related, forecasting, geoclassified, ovispoition, eco-epidemiological, risk model and thus, formally, $dx = \varphi'(t)\,dt$, which would be the required substitution for dx in the forecasted derivatives. A ecologist, entomologist or experimenter could view the method of integration by substitution as a major justification of Leibniz's notation for integrals and derivatives for efficiently qualitatively quantitating an empirically probabilistically regressable. For example, consider y as a function of an georeferencable, clinical, field or remote geosampled,YFV-related, geoclassified LULC variable x, or $y = f(x)$. If this is the case, then the derivative of y with

respect to x, may be viewed as the limit $\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$ which according to Leibniz would be the quotient of an infinitesimal increment of y by an infinitesmal increment of x where $\frac{dy}{dx} = f'(x),$ where the right hand side is Langrange's notation for the derivative of f at x in the eco-epidemioloigcal, risk model. From the point of view of modern infinitesimal theory the quotient of an infinitesimal increment of y by an infinitesimal increment of x may be empolyed for constructing a robust, forecasting, probabilistic, time series dependent, risk model geopredictors, from a probabilistically regressed resdiualized dataset (e.g., explanatorial, clinical, field or remote specified endemic, transmission-oriented, YFV-related, eco-epidemiological LULC-related, covariate paramter estimator datasets), on the right hand side using Lagrange's notation for the derivative of f at x (see Cressie 1993). From the point of view of modern infinitesimal theory, Δx then would be an infinitesimal x-increment, Δy the corresponding y-increment, and the derivative would be the standard part of the infinitesimal ratio: $f'(x) = \pi t \left(\frac{\Delta y}{\Delta x} \right)$ in the YFV, endmember, LULC, sub-meter resolution, *Ae. egypti* , oviposition, risk model.

Lagrange multiplier, or score, time series, probabilistic, YFV-related, explanatorial statistics may be computed in eco-epidemiological, forecasting, risk modeling cases using the NOINT or NOSCALE option where restrictions are placed on the intercept or scale parameters. These statistics may assess the validity of the restrictions, which may be computed as $\chi^2 = \frac{s^2}{V}$ where s is the component of the score vector evaluated at the restricted maximum corresponding to the restricted parameter and $V = \mathbf{I}_{11} - \mathbf{I}_{12} \mathbf{I}_{22}^{-1} \mathbf{I}_{21}$. The matrix $\mathbf{I}$ is the information matrix (1 refers to the restricted parameter, and 2 refers to the rest of the parameters) in SAS(www.sas.edu). Under regularity conditions, this statistic may have an aympototic, chi-square distribution with one degree of freedom when for example, YFV-related LULC p-values are computed based on a limiting distribution. If a ecologist, entomologist or experimenter sets $k = 0$ in a negative binomial, sylvatic YFV-related, geoclassified, LULC, eco-epidmeioloigcal, risk model, s then the score statistic of Cameron and Trivedi (1998) for testing for overdispersion in a Poissonian probability model against alternatives of the form would be $V(\mu) = \mu + k\mu^2$.

PROC NLMIXED used a subset of the optimization code underlying PROC NLP to qualiatively regressively quantaite the empirically probabilistically, time series regressed geo-spatiotemrpoal geosampled, eco-epidemiological, endmember, sub-meter resolution clinical, field and remote-specified LULC covariate parameter estimator dataset. The programming statement functionality employed by PROC NL MIXED had a subset of the optimization code underlying PROC NLP which had many optimization-based options for robustly regressing the geosampled coefficients values. PROC NLMIXED has close ties with the NLP procedure in SAS/OR software (www.sas.edu). Also, the programming statement functionality used by PROC NLMIXED is the same as that used by PROC NLP and the MODEL procedure in SAS/ETS software.

Interestingly, the GLIMMIX procedure generalized. The GLIMMIX procedure can fit time series dependent, entomological-related, regressive, prognosticative, explanatorial, endemic, transmission-oriented, YFV-related, probablistic, eco-epidemiological, *Ae. egypti* , oviposition vulnerability models to other vital data (seasonally meterological, geomorphological etc.) in other modules (e.g., the NL MIXED and GENMOD) when the derivatives render quantitated correlations and no constant variability. As such, the model response will not be necessarily normally distributed [i.e., generalized linear mixed models (GLMM)] (See Jacob et al. 2009d. Jacob et al. 2008b).

GLMMs, like linear mixed models, assume normality (Gaussianism), probablistic, randomized effects. Conditional on these probabilsitic, random effects, seasonal explanatorial, operationizable, geosampled, geopredictive, probabilistically regressable, time series dependent, sylvatic, YFV-related endemic, transmission-oriented, geoclassified sub-meter resolution, LULC, risk-related covariate, parameter estimators can have any distribution in the exponential family. Importantly, the exponential family comprises many of the elementary discrete and continuous distributions. The binary, binomial, Poisson, and negative binomial distributions, for example, are discrete members of this family. The normal, beta, gamma, and chi-square distributions are representatives of the continuous distributions in this family. In the absence of random effects, the GLIMMIX procedure can fit empiricially regressable, seasonal, geosampled, explanatorial, generalizable, clinical, field and remote, *Ae. egypti*, oviposition endemic, endmember, transmission-oriented, geopredictive, LULC risk-related, explanatorial, probabilistic, regressonable, linearized model, covariate paramter estimator coefficients by the NL MIXED GENMOD procedure. GLMMs may be useful for probablistically estimating trends in endemic, YFV-related, count data models and for designing randomly selected treatments or randomly selected blocks for geopredicting the uncertainty probabilities of explanatorily regressed, geosampled, interpolatable, covariate parameter estimator, coefficient, skewed forecasts.

For continuous geo-spectrotemporal, risk-related empirically probabilistically regressed, clincial, field or remote specified, outcomes in geo-spatiotemporal, geosampled, YFV-related, forecasting,LULC-related, eco-epidemiological, time series, risk models, the most commonly used statistical technique in literature is multiple regression analysis. The least squares procedure gives minimum-variance unbiased estimators of the regression coefficients when the variance of the error term is assumed to be the same regardless of the values of the covariate paramter estimator (i.e., homoscedasticity) (Hosmer and Lemeshew 2002). Further, the usual testing and confidence interval procedures for YFV-related model cuurently make the additional assumption that the errors are normally distributed. Because these assumptions are rarely satisfied in time series, YFV-model LULC construction practice, certain adjustments must be made to improve the validity of the results. An arbovirologist, medical entomologist, or YF experimenter may describe some of these adjustments, including retransformation of outcome variables employing the smear factor, accounting for missing cases using multiple imputation and attrition weights and improving statistical inference with bootstrap methods.For robust, optimal, explanatorial, clinical, field or remote geosampled, count outcomes, common methods of analysis for an explanatorial, dataset of clinical, field or remote geosampled, YFV-related, geoclassified LULC-related are the Poisson and negative-binomial regressions. However, in many situations the probability of a zero occurrence (e.g., dry deason sampling of *Ae. egypti*

immature habitats on forest canopied and/or ricleand agroecosystem LULCs) may exceed that assumed by the model. An arbovirologist, medical entomologist, or YF experimenter may choose to describe some methods that aim to overcome this problem, namely zero-inflated and two-part models. Further, analyzing joint modeling of multivariate, multitemporal, explanatorial, clincial, field and remote-related, sylvatic YFV-related, operationizable, regressed outcomes may display correlations among some or all hyperendemic, seasonal, ovispoition, observational geoclassified, eco-georferenceable, LULC predictors as well as their non-normal values. The correlations can arise from repeated observations of the same geosampling units and/or shared probabilistically tabulated random effects in an experimental design, involving geo-spatiotemporal, geosampled proximity-oriented, multivariate, multitemporal, YFV-related,, *Ae. egypti* , oviposition eco-epidemiological, forecast, vulnerability related,explanatorial, sub-mter resolution, LULC observations.

Unfortunately, the GLIMMIX procedure does not fit hierarchical, seasonal, entomological-related, eco-epidemiological, LULC risk models such as an YFV-related, eco-epidemiological, risk-related,explanatorial model with non-normal, probabilistically regressed, random effects.As such, a simple structural model may not then be considered for parsimoniously qualitatively quantitating random geosampled, explanatorial, clincial, field and remote geo-specified, endemic, transmission-oriented, sub-meter resolution, seasonal, risk-related probablistic predictors represented between and within-group, YFV-related, explanatorial, LULC variations. Methods for eco-epidemiologically estimating the cumulates of the two explanatorial, time series, dependent YFV-related model components of variation may be then proposed, based on homogeneous polynomials in the geosampled seasonal, eco-epidemiological data. Emphasis may be placed on situations in which the number of seasonal, hierarchical, vector, probablistic, risk model, explanatorial, interpolatable, clinical, field and remote geosampled, sylvatic, YFV-related ricland or forest canopied, inhomogeneous, observations per geoclassified LULC group are quite small. In some cases an essentially unique probablistic covariate, parameter estimator may be available, whereas in others there is a family of possible consistent estimators(see Fotheringham 2002). By so doing, the choice of the polynomial may then be considered for qualitatively regressively probabilsitically quantitating the geo-spatiotemporal, geosampled, *Ae. egypti* , oviposition YFV-related, geoclassified LULC explanatorial, operationizable, covariate, parameter estimator coefficients.

Importantly, the GLIMMIX procedure in NL MIXED employing a distribution of the response variable conditional on normally distributed probablistic, randomized effects in the probablistically empirically regressable dataset of explanatorial, time series dependent, clinical, field and remote-geosampled, YFV-related, endemic, transmission-oriented, explanatorial, geoclassified, LULC covariate, parameter estimator, coefficient values may render robust residually forecastable, derivatives. Frequently in seasonal, vector, arthropod-related, observational, eco-epidemiologcal, risk models, the responses (e.g., forecasted, optimal, unbiased, asymptotical, stochastic/deterministic, interpolatable, residual, covariate parameter estimators) can have a non-normal distribution. Normally distributed entomological, data is needed to use a number of statistical tools, such as individually seasonally geosampled habitat control charts, C_p/C_{pk} analysis, t-tests and the analysis of variance (ANOVA)(see Jacob et al. 2005b). The MIXED procedure will then assume then that the entomological responses are

normally (i.e., Gaussian) distributed. By so doing, the GLIMMIX procedure in NL MIXED will incorporate probablistic, random effects in a frequentist, endemic transmission-oriented, risk-related paridym which can allow for subject-specific (i.e., conditional) and population-averaged (i.e., marginal) inferences. Further, the procedure would allow only for marginal inferences in an explanatorial, geo-spatiotemporal or geo-spatiotemporal, eco-epidemiological, endmember, LULC, sub-meter resolution, clinical, field or remote-geosampled, YFV-related, forecast, time series eco-epidemiological, vulnerability model to be probabilistically regressively, qualitatively quantitated for heuristically optimally,optimizing and geospatially targeting important endemic, transmission-oriented, clinical, field and remote –specified geopredictive variables.

Importantly, the GENMOD procedure would fit certain empirically, geosampled, probablistically regressable, explanatorial, diagnostic, clinical, field and remote-specified, seasonal, endemic, transmission-oriented, YFV-related, GLMs for independent probabilistic, risk-related, data analyses by maximum likelihood. The procedure could also handle probablistically correlated, explanatorial, time series, YFV-related regressors through the marginal GEE approach of Liang and Zeger (1986) and Zeger and Liang (1986). The GEE implementation, however, in the GENMOD procedure is a marginal method that does not incorporate random effects. The GEE estimation in the GENMOD procedure relies on R-side covariance only, and, as such, the unknown seasonal, geosampled, YFV-related, explanatorial, geopredictive, parameterized, explanatorial, clinical, field or remote geosampled, eco-georeferencable, *Ae. egypti*, oviposition, parameterizable, covariate estimator coefficient values would not be estimated by the method of moments. For optimal quantization of explanatorily georeferenced, time series dependent, empirically regressable, endmember, LULC, sub-meter resolution, YFV-related, parasitological indicators, the GLIMMIX procedure in NL MIXED would allow G-side random effects and R-side covariances to be robustly probablistically estimated when constructing the forecasting, eco-epidemiological, explanatorial, georefernceable, clinical, field or remote specified, risk model *Ae. egypti* , oviposition LULC sub-mter resolution, predictors. Thereafter, PROC GLIMMIX can fit, marginal, time series dependent, probablistic, GEE-type, Y sylvatic, FV-related, explanatorial, geopredictive, endemic, transmission-oriented, risk-related, sub- models, but the covariance parameters would have to be estimated by the method of moments.

In an empirically explanative, regressable dataset of robustifiable, seasonal, YFV-related geopredictive, forecastable, eco-epidemiological, risk model, georefernceable, geo-spectrotemporally geosampled, clinical, field and remote specified covariate, parameter estimators, the method of moments could be optimally defined as a method of probability estimation. The experimenter would start by deriving optimizable regression equations that are related to the population moments (i.e., the expected, explicatvely geosampled, ArcGIS-derived, sylvatic, YFV-related, LULC, geoclassified, endemic transmission-oriented, *Ae. egypti* , oviposition, time series, covariate, parameterizable estimator, coefficient values of powers of the random variable under consideration) to the geolocations of interest (e.g., productive, geosampled, georeferenced, *Ae.egypti,* aquatic, larval habitat geosampled in a , rice-village complex tillering habitat LULC polygon). Then a sample may be drawn and the population moments may be probablistically regressively estimated from the sample. The equations may then be solved for the explanatorial, clinical, field or remote, geosampled, regressable parameter

estimators of interest in a seasonal, empirical, YFV-related,, endmember, LULC, sub-meter resolution. probablistically, geopredictive, eco-epidemiological, SAS risk model using the sample moments in place of the unknown population moments. In so doing, system reliability moments may be robustly generated by expanding the system function in a multivariable, multitemporal, explanatorial, clinical, field and remote-specified, time series dependent, endemic, YFV-related, transmission-oriented, eco-epidemiological, probablistic, risk model employing Taylor series expansion about the geo-statistically expected values of each of the component reliabilities.

Interestingly, the method of moments is an approximate, rather than an exact, method, because of the omission of higher order terms in the Taylor series expansion. Further, traditional mathematical approaches to fitting nonlinear, orthogonizable, eco-epidemiological, explanatorial, clinical, field or remote geosampled, probabilistically regressable, geo-spatiotemporally, geosampled, YFV-related, mixed model, covariate, parameter estimator coefficients would involve the Taylor series expansions, around either zero or the empirical, best, linear, unbiased predictions of the random effects which may be defined in PROC NL MIXED. In so doing, the first-order method of Beal and Sheiner (1982) and Sheiner and Beal (1985), may be implemented in PROC NLMIXED for probabilistically regressively quantitating time series specified, explanatorial datasets of operationizable, endemic, transmission-oriented, emprically regressable, clinical, field and remote geosampled covariate, parameter estimator, coefficient values. Although the basis for the estimation method of Lindstrom and Bates (1990), is presently not available in PROC NLMIXED, the closely related Laplacian approximation is available; it is equivalent to adaptive Gaussian quadrature with only one quadrature point. The Laplacian approximation and its relationship to the Lindstrom-Bates method are discussed by Beal and Sheiner (1992), Wolfinger (1993), Vonesh (1992,1996), and Wolfinger and Lin (1997).

Further, parallel literature exists in the area of GLMMs, in which probabilistically tabulated random effects appears a part of the explanatorial, linear probabilsic predictors inside of a link function. Taylor-series methods that are similar to those just described are discussed in articles such as Harville and Mee (1984), Stiratelli, Laird, and Ware (1984), Gilmour, Anderson, and Rae (1985), Goldstein (1991), Schall (1991), Engel and Keen (1992), Breslow and Clayton (1993), Wolfinger and O'Connell (1993), and McGilchrist (1994), but such methods have not been implemented in PROC NLMIXED for regressively qualitatively quantitating YFV-related, explanatorial, clinical, field and remote-specified, geo-spatiotemporally, geosampled, endemic, transmission –oriented, endmember, ovipsoition, sub-meter resolution, LULC covariate, parameter estimator, explanatorial coefficients because they can produce biased results in certain binary data situations (see Rodriguez and Goldman 1995, Lin and Breslow 1996).

Instead, a numerical quadrature approach is available in PROC NLMIXED, for robustly regressively quantitating an emprically probabilistically regressable dataset of explanatorial, clinical, field or remote geosampled, time series dependent, eco-epidemiological geopredictors as discussed in Pierce and Sands (1975), Anderson and Aitkin (1985), Crouch and Spiegelman (1990), Hedeker and Gibbons (1994), Longford (1994), McCulloch (1994), Liu and Pierce (1994),and Diggle, Liang, and Zeger (1994). The models fit by PROC NLMIXED may qualitatively quantitate an empirical, geosampled, georeferncable dataset of explanatorial, clinical, field and remote-specified, YFV-related endemic,transmission-oriented, covariate,

parameter estimators which can be viewed as generalizations of the random coefficient models fit by the MIXED procedure. This generalization can allow the random coefficients to enter a geopredictive, explanatorial, time series dependent, eco-epidemiological, YFV-related, probablistic, risk model framework to qualitatively regressively quantitate the geosampled data linearly.

Importantly, because of generalized nonlinear formulations in an empirically probablistically, time series dependent, explanatorily regressable dataset of YFV-related, clinical, field and remote-geosampled, eco-epidemiological, covariate, parameter estimator coefficients, no direct analog to the REML method is available in PROC NLMIXED; only standard maximum likelihood may be used. Also, PROC MIXED will assume the data to be normally distributed, whereas PROC NLMIXED enables data analyses that are normal, binomial, or Poisson or that has any likelihood programmable with SAS statements. PROC NLMIXED does not implement the same estimation techniques that are available with the NLINMIX and GLIMMIX macros. These macros are based on the estimation methods of Lindstrom and Bates (1990), Breslow and Clayton (1993), and Wolfinger and O'Connell (1993), and they iteratively can fit a set of generalized estimating equations (refer to Chapters 11 and 12 of Littell et al. 1996 and to Wolfinger 1997). In contrast, PROC NLMIXED directly maximizes an approximate integrated likelihood. PROC NLMIXED has close ties with the NLP procedure in SAS/OR$^®$ software. PROC NLMIXED employs a subset of the optimization code underlying PROC NLP and has many of the same optimization based options. Also, the programming statement functionality that is used by NLMIXED is the same as that used by PROC NLP and the MODEL procedure in SAS/ETS$^®$ software.

Interestingly, if an experimenter consider the limit as $r \to \infty$, the second factor will thereafter converge to one in NLMIXED and the third to the exponent function: $\lim_{r \to \infty} f(k; r, p) = \frac{\lambda^k}{k!} \cdot \mathbf{1} \cdot \frac{1}{e^\lambda}$ in a robust, explanatorial, geopredictive, eco-epidemiological, time series dependent, YFV-related, regression-related, probablistic, risk model. This tabulation could be performed employing the mass function of any explanatorial, geopredictive Poisson-distributed, time series-related, sylvatic, YFV-related, randomized variable(e.g., LULC change from forest canopy to agro-village complex over 10 years) with expected value λ. In other words, the alternatively parameterized negative binomial distribution would converge to the Poissonian distribution in the endemic, transmission-oriented, eco-epidemiological, probabilistic, risk model. By doing so, r would control the deviation from the Poissonian in the residually forecasted probabilistic derivatives. Thus, this processing makes the negative binomial distribution suitable as a robust alternative to the Poissonian for seasonal, eco-epidemiological, probablistic, eco-epidemiological, risk modeling time series dependent, geopredictive, explanatorial, YFV-related, endemic, transmission-oriented, geo-spatiotemporally, geosampled, clinical, field and remote variables which approach the Poisson for large r, but which has larger variance than the Poisson for small r[e.g., $\text{Poisson}(\lambda) = \lim_{r \to \infty} \text{NB}\left(r, \frac{\lambda}{\lambda + r}\right)$]in GEN MOD.

Interestingly, the negative binomial distribution with size = n and prob = p in an explanatorial, geopredictive, robust, time series-related, YFV-related, eco-epidemiological, endemic, transmission-oriented, risk model may have the density $\Gamma(x+n)/(\Gamma(n) x!) p^n (1-p)^x$

for *x = 0, 1, 2, ..., n > 0* and *0 < p ≤ 1* in R. R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and Marcos. The R language is widely used among statisticians and data miners for developing statistical software and risk-related, eco-epidemiological, forecasting, data analysis. R is an implementation of the S programming language combined with lexical scoping semantics inspired by Scheme (Ihaka, 1998).

S was created by John Chambers while at Bell Labs. R was created by Ross Ihaka and Robert Gentlemanat the University of Auckland, New Zealand, and is currently developed by the *R* Development Core Team, of which Chambers is a member. R is a GNU project. The source code for the R software environment is written primarily in C, Fortran, and R. R is freely available under the GNU General Public License, and pre-compiled binary versions are provided for various operating systems. R uses a command line interface; however, several graphical user interfaces are available for use with R. To download R, CRAN mirror is commonly employed (http://www.r-project.org/). Importantly, in R the explanatorial, seasonal, clinical, field, and remote specified, probabilistic, geo-spatiotemporally, geosampled, YFV-related, endemic, transmission-oriented, eco-epidemiological, covariate, parameter estimator, coefficient values would represent the number of failures in the geoprediction of a seasonal, probabilistic, georeferencable, stochastically/deterministically explanatorily iteratively interpolatable, YFV-related variables (e.g., case distribution, geoclassified ArcGIS-derived riceland LULCs). This representation would occur in a sequence of Bernoulli trials before a target number of successes is reached. The mean in the time series dependent, explanatively probabilistic, YFV-related, endemic, transmission-oriented, risk model would then be *n(1-p)/p* and variance *n(1-p)/p^2*.

Importantly, a negative binomial distribution can also arise as a mixture of Poisson distributions with mean distributed as a gamma distribution (i.e., pgamma explanatorial, YFV-related, diagnostic, clinical, field or remote-specified, geo-spatiotemporally geosampled probabilistic, LULC regression variable) with scale parameter (1 - prob)/prob and shape parameter size) (seeHosmer and Lemeshew 2002).This definition would allow all the seasonal, geosampled,predictive, explanatorial, sylvatic YFV-related, eco-epidemiological, endemic, transmission-oriented, non-integer values from an empirical frican expanding agro-irrigated village dataset size to be employed in the model construction process where dgamma(x, shape, rate = 1, scale = 1/rate, log = FALSE) pgamma(q, shape, rate = 1, scale = 1/rate, lower.tail = TRUE, log.p = FALSE)qgamma(p, shape, rate = 1, scale = 1/rate, lower.tail = TRUE, log.p = FALSE) may be parsimoniously, robustly estimated. Possible arguments then could be resolved in a robust, probabilsitic, empirical dataset of seasonally empirically regressable, time series dependent, geopredictive, YFV-related, endemic, transmission-oriented, explanatorial, risk model, residually forecasted derivatives in R which would be x, q (i.e., vector of YFV-related quintiles) p (vector of probabilities), *n* (number of geosampled, clinical, field and remote-specified, time series-related, YFV-related observations), rate for devising an alternative way to specify the scale of an eco-georefernceable, capture point, seasonal, *Ae egypti,* aquatic, larval habitat on an agro-village complex LULC geolocation. In R commonly shape for positive, scale strictly log, log.p[logical; occurs if TRUE, probabilities/densities p are returned as log(p)] log or log.p logical; if TRUE, probabilities/densities p are returned as log(p) (http://www.r-project.org/). Further, if scale is omitted in an operationizable, empirical dataset of heurstically optimizable seasonal, explanatorial, probabilistically geo-spatiotempoally, geosampled, YFV-

related, eco-epidemiological, endemic, transmission-oriented, grid-startified, LULC, rforecast, vulnerability mordel, derivatives would assume the default value of 1.

Interestingly, the Gamma distribution with probabilistically regressed seasonal geosampled, explanatorial, geopredictive, YFV-related parameters shape = *a* and scale = *s* would have the density $f(x) = 1/(s^a\ Gamma(a))\ x^{(a-1)}\ e^{-(x/s)}$ for $x \geq 0$, $a > 0$ and $s > 0$. Here *Gamma(a)* in the time series dependent, explanatorial, geopredictive, YFV-related, probabilistic, endemic, transmission-oriented, eco-epidemiological, sub-mter resolution, LULC, risk model, residually forecasted derivatives would be the function implemented by R's gamma(Cressie 1993).Thereafter, *a=0* would correspond to the trivial distribution with all mass (e.g., explanatorial, parameterizable , sub-mtrer resolution LULC, covariate, *Ae. egypti* , oviposition parameterizable estimator coefficients) at every geo-spatiotemporally geosampled time series dependent, endemic points in the YFV-related, eco-epidemiological, probabilistic, risk model. The mean and variance in the seasonal, endemic, transmission-oriented, YFV-related, probabilistic, eco-epidemiological, risk model could then be regressively qualitatively quantitated $E(X) = a*s$ and $Var(X) = a*s^2$. By so doing, the cumulative hazard $H(t) = -log(1 - F(t))$ in the risk model, residual, forecasted *Ae. egypti* , oviposition LUCsub-mter resolution, derivatives would then be -pgamma(t, ..., lower = FALSE, log = TRUE.

Note that for a small explanatorial dataset of geopredictive, time series-related, eco-epidemiological, endmember YFV-related, probabilistic, *Ae. egypti*, oviposition Poissonized, covariate coefficient, regressed LULC value of shape and moderate scale and large parts of the mass, the Gamma distribution could be represented as seasonal, geosampled values of *x* in computer arithmetic. As such, rgamma may return a probabilistically regressable, empirically, geosampled dataset of eco-georefernced, time series-related, explanatorial, predictive, diagnostic, clinical, field or remote geosampled, sylvatic,YFV-related, endemic, transmission-oriented, covariate, parameter estimator, coefficient values (e.g., seasonal change in LULC from forested canopy to agro-village complex) which may be then be eco-epidemiologically represented in a time series, risk model, SAS framework efficiently. This could also happen during the regressive covariate, parameter estimation process for a very large geosampled, empirically eco-epidemiological dataset of geopredictive, explanatorial, probablistic, YFV-related, endemic, transmission-oriented, diagnostic,clinical, field or remote, covariate, parameterizable,covariate estimator, coefficient values of scale since the actual generation would be done for scale = 1.

In so doing, dgamma would t render the density, pgamma which could then provide the distribution function, qgamma which in turn would render the quantile function while rgamma would generate random deviates in the probablistic forecasted, clinical, field or remote-specified, sylvatic YFV-related, risk model derivatives. Interestingly, probabilistic, invalid arguments in a seasonal explanatorial, geopredictive, YFV-related, endemic, transmission-oriented, probabilsitic, eco-epidemiological,sub-mter resolution , grid-stratieid, oviposition, risk model, parameter estimation process would result in a return value NaN, with a warning in SAS. The length of the result in the covariate, parameter estimation process may be then determined by *n* for rgamma, which would be the maximum of the lengths of the seasonal geosampled, explanatorial, geopredictive, sylvatic YFV-related, numerical *Ae. egypti*, oviposition parameterizable copvariate estimators for the other functions. The numerical estimators other

than *n* would then be recycled to the length of the result. Only the first elements of the logical parameter estimators could be employed in a dataset of empirically regressable, probabilsitic, explanatorial, YFV-related, endemic transmission-oriented, risk model covariate, parameter estimator coefficients.

Interestingly, the pgamma in a seasonal, YFV-related, explanatorial, geo-spatiotemporally geosampled, probabilistically regressable, eco-epidemiological, forecasting, eco-georferenceable, LULC, risk model may be closely related to the incomplete gamma function. As defined by Abramowitz and Stegun (1972) this is $P(a,x) = 1/Gamma(a) \text{ integral}\_0^x t^{(a-1)} exp(-t) dt P(a, x)$ may be regressively probabilistically qualitatively quantitated as pgamma(x, a) in an empirical dataset of unbiased residually forecastable, seasonal, YFV-related, explanatorial, stochastically/deterministically iteratively interpolatable, *Ae. egypti*, oviposition derivatives. Other authors for example Karl Pearson in his 1922 tables omitted the normalizing factor for defining the incomplete gamma function $\gamma(a,x)$ as $gamma(a,x) = \text{integral}\_0^x t^{(a-1)} exp(-t) dt$, [i.e., pgamma(x, a) * gamma(a)]. Alternatively, the 'upper' incomplete gamma function, $Gamma(a,x) = \text{integral}\_x^{Inf} t^{(a-1)} exp(-t) dt$ may be employed for computing pgamma employing x, a, lower = FALSE * gamma(a) in an eco-epidemiological, YFV-related, probabilsitic, diagnostic, risk model. Note, however that pgamma(x, a, ..) for generating a robust, explanatorial, geopredictive, seasonal, sylvatic YFV-related, endemic, transmission-oriented, probabilistic, eco-epidemiological, ovispoition, capture point, risk model would require $a > 0$, whereas the incomplete gamma function may be defined for negative *a*. In that case, an experimenter may use gamma_inc(a,x) instead (for *$\Gamma(a,x)$*) from package gsl. (http://dlmf.nist.gov/8.2#i).Importantly, dgamma may be computed via the Poisson density, using code dbinom for generating a robust, linearized, seasonal, explanatorial, YFV-related ,endemic, transmission-oriented, probabilistic, eco-epidemiological, risk model.

However, in seasonal, sylvatic, YFV-related, endemic, transmission-oriented, explanatorial, geopredictive, probabilistic, risk model, the normalized binomial distribution with size = *n* and prob = *p* would have a heurstically optimizable,explanatively probabilsically regressable, geosampled, eco-epidemiological, operationizable, interpolatable, covariate, parameterizable, geo-spectrotemporal, estimator, explanatorial, *Ae. egypti* , oviposition, sub-mter resolution, geoclassifiable, LULC coefficients may still be computed in R or any other sophisticated statistical program and then exported into a ArcGIS cyberenvironment for generateaing a geo-spatially robustifible, vulnerabity-oriented, eco-epidemiological, probabilistic, risk map. If an element of x is not an integer in a seasonal explanatorial, geopredictive, YFV-related, endemic, transmission-oriented, probabilsitic, geo-spectrotemporal, eco-epiudemiological, risk model, the result of dbinom would be zero, with a warning. In so doing, *p(x)* could then be ideally computated using Loader's algorithm in R (http://www.r-project.org/), if so desired. The quantile for a robustifiable, explanative, time series-related, explanatorial, geopredictive, seasonal, robustifiable, YFV-related, forecasting, eco-epidemiological, probabilistic,geospectrotemrpoal, geospatial, risk model may then be also defined as the smallest geosampled explanatorial, endemic, transmission-oriented, covariate, parameterizable, elucidative, estimator, coefficient value x such that $F(x) \geq p$, where F is the distribution function in GEN MOD. By so doing, dbinom would render the density, pbinom which would provide the distribution function, qbinom which in turn would render the quantile function in an explanatorial, clinical, field or remote-specified, probabilistically regressed dataset of, diagnostic

georferenecable, clinical, field or remote geosampled, YFV-related, eco-epidemiological, risk model estimators. In R this could occur employing rbinom which then would generate random deviates from the empirically geosampled dataset of explanatorial, time series-related, geo-spectrotemporally geosampled, endemic, transmission-oriented, risk-related geopredictors. If size is not an integer, NaN would also be returned in the residually forecasted explanative derivatives in GEN MOD. The length of the result could then be determined by n for rbinom in the R derived, eco-epidmeiological, forecastable, risk model, and the maximum of the lengths of the numerical parameters for the other functions.

Importantly, for generating a robust, seasonal, YFV-related dbinom in SAS a saddle-point expansion may be used(see Catherine Loader (2000)]. In so doing, pbinom can use pbeta while qbinom which would employ the Cornish–Fisher Expansion to include a skewness correction for normal approximation, followed by a robust search. For a continuous, YFV-related, time series-related, geopredictive, explanatorial, regressed when robustly quantitating a georferenceable datatste of eco-epidemiological, diagnostic, geo-spectrotemporally geosampled, space-time, distributions (like the normal) in as endemic, transmission-oriented, geoclassifiable LULC, eco-epidemiological, risk model, the most useful functions for doing problems may involve solving probability calculations such as the "p" and "q" functions (c. d. f. and inverse c. d. f.), because the probabilty densities calculated by the "d" function can only be used to calculate probabilities via integrals (see Jacob et al. 2012). Unfortunately, R does not compute integrals. For a discrete, explanative, seasonal, probabilistic, YFV-related, regressive distribution like the binomial, the "d" function would however, calculate the density (e.g., pdf), which in this case would be a probability $f(x) = P(X = x)$ and, thus would be useful in calculating probabilities in the forecasted eco-epidemi9ological YFV-related derivatives. R has functions to handle many probability distributions (https://www.r-project.org/).

The Cornish–Fisher expansion is a mathematical expression used to approximate the quintiles of a random explanative variable based only on its first few cumulates .For instance, suppose an experimenter lets x be a random geo-spectrotemporally geosampled, time series dependent, probabilistic, explanatorial, diagnostic, clinical, field and/or remote-derived, YFV-related, geopredictive, linearized variable with a density function $f(x)$ with a mean of zero and a variance of 1. By letting β_1 be the skewness of this distribution and then letting β_2 be its kurtosis, all the geosampled explanatorial, time series dependent, clinical, field or remote-specified, YFV-related, geopredictive variables may be regressively qualitatively quantitated. Further, if a experimenter lets z be a normally distributed random variable and lets z_α be the possible, probablistic, geosampled, explanatorial, diagnostic, clinical, field or remote, YFV-related, covariate, parameterizable, *Ae. egypti* , prolific, oviposition , sub-mter resolution decomposable, estimator coefficient values of z in the risk model at the αth percentile, an illustration of this last definition when $\alpha = 0.95$, $z_\alpha = 1.96$ may be optimally heuristically derived In so doing,

$$\omega_\alpha = z_\alpha + \frac{1}{6}(z_\alpha^2 - 1)\beta_1 + \frac{1}{24}(z_\alpha^3 - 3z_\alpha)(\beta_2 - 3) - \frac{1}{36}(2z_\alpha^3 - 5z_\alpha)\beta_1^2 - \frac{1}{24}(z_\alpha^4 - 5z_\alpha^2 + 2)\beta_1(\beta_2 - 3)$$

where ω_α is the corresponding value for $f(x)$ in the dataset of endemic, probabilistic, transmission-oriented, YFV-related, eco-epidemiological, geo-predictive, risk model, optimally residually forecasted derivatives. Kachitvichyanukul, and Schmeiser (1988) commutated distributions for standard distributions, including dnbinom for the negative binomial, and dpois for the Poisson distribution.

Herewith, a hypothetical, seasonal, sylvatic YFV-related,explanatorial, ovispoition, capture point, forecast, vulnerability,, eco-epidemiolgical, probabilsitic, time series dependent, robustifiable, explicative, *Ae. egypti* , oviposition,sub-meter resolution,LULC model graphics are provided for constructing a robust, endemic, transmission-oriented, risk model in R:

```
# Compute P() for X Binomial(100,0.5)
sum(dbinom(100, 0.5))
## Using "log = TRUE" for an extended lulc, rainfall, and human population range :
n <- 100
k <- seq(0, n, by = 20/per season)
plot (k, dbinom(k, n, pi/10, log = TRUE), type = "l", ylab = "log lulc change forest canopy to agro-village density",
    main = "dbinom(*, log=TRUE) is better than   log(dbinom(*))")lines(k, log(dbinom(k, n, pi/10)), col = "rainfall", lwd = 2)
mtext("dbinom(k, log=TRUE)", adj = 0)
mtext("extended anthropogenic population range"), adj = 0, line = -1, font = 4)
mtext("log(dbinom(k))", col = " adj = 1)
```

In the R-derived seasonal, explanatorial, YFV-related, operationizable, eco-epidemiological, clinical, field or remote-specified, geopredictive, forecasting, risk model uncertainty-oriented, extreme points (i.e.,geospatial, outliers)would be omitted by making dbinom render 0.

An alternative optimaizable parameterization for constructing a robustiifable explanative, seasonal, sylvatic YFV-related, endemic, transmission-oriented, probabilistic, eco-epidemiological, forecasting, regression-related, time series, risk model would be by employing the mane (mu), and size, (i.e., the dispersion parameter) where prob = size/(size+mu) in R or SAS. The variance in the eco-epidemiological, ovipsotion, hypeproductive, explanatorial, diagnostic, clinical, field or remote geosampled, time series dependent, risk model then would be mu + mu^2/size which would be regressively, qualitatively quantitatable during the endmember parameterization process. If an element of x is not a time series dependent, explicatively represented YFV-related, diagnostic, endemic, transmission-oriented, parameterizable, covariate, coefficient, geosampled, discrete integer value, the result of dnbinom would be zero, with a warning. The quantile would then could be defined as the smallest explanatorial, geosampled, time series-related, YFV-related, eco-epidemiological, forecasting value *x* such that $F(x) \geq p$, where *F* is the distribution function. In so doing, Dnbinom could then render the density, pnbinom which would in turn render the distribution function in the risk model, eco-epidemiological, residual, forecasted *Ae. egypti* , oviposition, sub-mter resolution decomposed derivatives.

Thereafter, qnbinom could render the quantile function, in SAS for optimally regressively qualitatively quantitating endemic, transmission-oriented, parameterizable, covariate, coefficient, estimator levels of significance in the probabilistic, *Ae. egypti* , oviposition, YFV-related, eco-epidemiological, geopredictive, risk model outputs or rnbinom would generate random deviates to conduct the same task in R. Invalid size or prob would result in return value NaN, with a

warning regardless of which program is employed for the estimator quantitation process. Importantly, dnbinom would compute elucidative, binomial probabilities in R using codes contributed by Catherine Loader. Alternatively, Pnbinom would use pbeta whilest qnbinom would employ the Cornish–Fisher Expansion in SAS to include a skewness correction to a normalized approximation, followed by a search. Interestingly, rnbinom could be intrinsically ideally employable for qualitatively regressively quantitating the explanative, georferenceable, YFV-related derivation as a gamma mixture of Poissons. The normalized negative binomial distribution, in its alternative parameterization form, however in SAS and/or R can be used as an alternative to the Poisson distribution for a quantifying geo-spatiotemporal, geosampled explanatorial, geopredictive, YFV-related, overdispersed parameterizable, covariate, estimators over an unbounded positive range whose sample variance exceeds the sample mean.

An overdispersed exponential family of generalizable, seasonal explanatorial, geopredictive, YFV-related, explanative, vulnerability, linear model distributions was employed to quantitate exponential family in PROC GEN MOD and dispersion model of distributions which also included those probability distributions, parameterized by θ and τ, whose density functions was expressed in the form $f_Y(\mathbf{y}|\boldsymbol{\theta},\tau) = h(\mathbf{y},\tau)\exp\left(\dfrac{\mathbf{b}(\boldsymbol{\theta})^{\mathrm{T}}\mathbf{T}(y) - A(\boldsymbol{\theta})}{d(\tau)}\right) \cdot \tau$. Importantly, the dispersion parameter, typically is known and is usually related to the variance of the distribution(seeCressie 1993). For qualitatively probabilistically regressively quantitating a pmf for the seasonal geo-spectrotemrpoally geosampled, YFV-related, seasonal, eco-epidmeiolgical, normalized, *Ae. egypti* , oviposition-related, seasonal, discrete distribution the functions $h(\mathbf{y},\tau)$, $\mathbf{b}(\boldsymbol{\theta})$, $\mathrm{T}(y)$, $A(\boldsymbol{\theta})$, and $d(\tau)$ were employed . For determining the scalar Y and θ in the probabilsitic, negative binomial, YFV-related, endemic, transmission-oriented, eco-epidemiological, optimized, risk model, these functions reduced to $f_Y(y|\theta,\tau) = h(y,\tau)\exp\left(\dfrac{b(\theta)T(y) - A(\theta)}{d(\tau)}\right)$ and $\boldsymbol{\theta}$ was related to the mean of the distribution . If $\mathbf{h}(\boldsymbol{\theta})$ is the identity function, then the distribution is said to be in canonical form (or natural form) (Hosmer and Lemeshew 2000).

Note that any distribution in our seasonal, predictive, YFV-related, explanatorial, endemic, transmission-oriented, probabilistic, eco-epidemiological, risk model was converted to canonical form by rewriting θ as θ' and then applying the transformation $\theta = \mathbf{b}(\theta')$. Interestingly, we noted that it was possible to convert $A(\boldsymbol{\theta})$ in terms of the new parameterization, even if $\mathbf{b}(\theta')$ was not a one-to-one function in the eco-epidemiological, explanatorial, YFV-related, forecasting, probabilisitic, risk model. Further, $\mathbf{T}(y)$ was an additional identity since in our model τ was known, and as such the canonical parameter or natural parameter in the model was efficiently regressively quantitated. We then optimally determined whether the canonical parameter could be or not to be related to the mean through $\mu = \mathrm{E}(\mathbf{Y}) = \nabla A(\boldsymbol{\theta})$. For scalar Y and θ in a robustifiable empirical dataset of endmember, sub-meter resolution, explanatorial, predictive seasonal, YFV-related, endemic, transmission-oriented, risk model, georferenceable, parameterizable covariate estimators may reduce to $\mu = \mathrm{E}(Y) = A'(\theta)$. (see Jacob et al. 2013). In so doing, the variance of our seasonal geosampled, explanatorial, geopredictive explanatorial, YFV-related, noramlized distribution were shown to be $\mathrm{Var}(\mathbf{Y}) = \nabla\nabla^{\mathrm{T}}A(\boldsymbol{\theta})d(\tau)$. For optimally regressively qualitatively quantitating scalar Y and θ, in the probabilistic, forecasting, YFV-

related, risk model, residually forecasted derivatives, this was further reduced to $\mathrm{Var}(Y) = A''(\theta)d(\tau)$ .Thereafter, all the seasonal geosampled, sylvatic ,YFV-related, endemic, transmission-oriented, probabilistic, eco-epidemiological, risk variables were specialized in a ecogeographically, spatially, weighted matrix.

The probabilistically estimation of the spatial filter equation employing the ecogeographically weighted, SAS/GIS-derived, transitional matrix accounted for substantial variation in our geosampled, explanatorial, geopredictive, YVR-related, time series dependent, explanatorial, diagnostic, geospatialized, clinical, field and remote geo-spectrotemporally, geosampled, parameterized covariate, estimator, with the regressed data exhibiting both a PSA and NSA spatial filters that were roughly of equal importance. Approximately 50 percent of the variance accounted for by the relative pseudo-R^2 values was probabilistically attributable to each spatial filter component. This finding made sense because, on the one hand, common factors may dictate which portions of geoclassified, riceland agro-cosystem, and forecast sparsely canopied LULCs are suitable for seasonal, sylvatic YFV-related mosquitoes in Gulu. Based on PSA configurations in the geospatial, agro-village riceland LULC explanatorial distributions,the field-geosampled, time series dependent, count data aggregated based on similar traits (sparsely shaded, discontinuous canopied geolocations). On the other hand, expansion of agro-urban territories into forested canopy LULC resulted in NSA being generated in the normalized distribution of the empirically geo-spectrotemrpoally geosampled, diagnostic, clinical, field or remote, YFV-related, geo-spatiotemporal, data feature, ecogeorferenceable hyperproductive *Ae. egypti*, oviposition attributes at the eco-epidemiological, study site. Interestingly, these two components tended to be of equal importance in the overall spatial autocorrelation mixture in the explanatorial, autoregressive seasonal, YFV-related, geopredictive, endemic, transmission-oriented, forecasting,probabilistic,eco-epidemiological, risk model.

Importantly, in the Gulu, eco-epidemiological study site, geo-specific,time series dependent, geoclassifiable, LULC changes of forest canopy LULC to agro-village complex LULC may have contained inconspicious NSA that was masked by PSA. Generally, positive autocorrelation occurs in explanatorial, geo-spectrotemporal, eco-epidemiolgical, seasonal, geopredictive, entomological-related, geoclassifiable, LULC, endemic transmission-oriented, eco-epidemiological, forecasting, risk model where similar seasonally explanative, immature count data values from geo-spatiotemporal geosampled clinical, field or remote georferenceable clusters occur in geospace on a specific LULC, while negative autocorrelation occurs where unlike data geosampled attribute values cluster in geospace on an LULC (see Jacob et. al. 2008, Jacob et al. 2005). Negative spatial autocorrelation, however, was encountered in the empirical data analyses of the geo-spectrotemrpoal, geo-spatiotemporal, eco-epidemiological, geosampled georeferenced, explanatorial, sylvatic YFV-related, endmember uncoalesced, ovispoition, capture point, LULC, time serioes, predictor variables less frequently than PSA.

Because the MC is asymptotically normally distributed (Griffith, 2003), MC may sometimes fail to detect hidden NSA in highly heterogeneous seasonal, explanatorily probabilistically regressed, diagnostic, geo-spectrotemrpoal, geospatialized, diagnostic, normalized, YFV-related clinical, field or remote geosampled, parameterizable, covariate coefficient, estimator datasets, although all of the geo-visual and conventional numerical evidence suggests the presence of PSA. The reason for this omission in a geo-spatiotemporal,

probabilistically regressedable, explicative, YFV-related eco-epidemiological, risk model can include the continuity in geospace and time of the sampleding of the georeferenceable, forested, sparsely canopied, agro-village complex LULCs and their explanatorily quantative, LULC geopredictive, time series, dependent, change variables for instance. Nevertheless, a series of circumstances in empirical situations can have deleterious affects on the seasonal geospatial distribution of an explanatorial dataset of explicative diagnostic, YFV-related, seasonal, endemic transmission-oriented, geopredictive, regressed variables displaying NSA. The models we generated commonly by capturing the negatively geospatially autocorrelated phenomena in ta probabilistically geo-spectrotemrpoally/geospatiotemporally geosampled, diagnostic datatste of georferenceable, clinical, field and remote-specified, YFV-related, explanatorial parameterizable, covariate, *Ae. egypti*, oviposition estimator coefficients, which may be attributable to the geospatial competition for land surface by the immature *Aedes* at a riceland agro-ecosystem, eco-epidemiological, village, study site. No mosquito species exhibit globally dispersed offspring and the density of propagules and progeny usually decrease with distance in the urban environments (Walker and Lynch, 2007); hence, the observed geospatialization and explanatorial, distributions of a regressed dataset of geosampled georeferencable, immature, geopredictive diagnsotci YFV-related, clinical, field or remote geopredictive variables may be the result of dispersal and competition of adult *Aedes*. Indeed, ecogeographically competitive phenomena can provide sound conceptual examples of NSA in a geopredictive, seasonal, explanatorial, medical entomological-related immature habitat, probablistic explicative distribution as the manifestation of the geosampeld immature georferenceable habitats with high productivity, based on field geosampled larval/pupal count, may be due to the influence of neighboring immature habitats. Negative geospatial autocorrelation naturally materializes with competitive geolocational processes in an empirical dataset of a georeferenced, geospectrotemrpoal/geospatiotemporal, explanative dataset of vector, arthropod-related, larval habitat, explanatorial, probablistically, endemic, transmission-oriented, orthogonally decomposeable, parameterizable, covariate estimator, coefficient values especially those geosampled in urbanizing environments (see Jacob et al. 2012, Griffith, 2006).

Apart from qualitatively probabilistically regressively quantitating latent autocorrelation coefficients, understanding the presence of NSA in geo-spectrotemporally, geo-spatialized, robustifiiable, geosampled, georeferenced, forested, sparsely shaded- canopied discontinuous LULC to agro-village complex LULC change-related explanatorial, geopredictive riceland variables is of interest for many other reasons for implementing IVM for yellow fever control. Foremost, is that the amount of quanatizable residual autocorrelation being dealt with in an ArcGIS-related, LULC change quantitation procedure while more than a trace, is rather small. This means that any elucidative, ArcGIS-related, endemic, transmission-oriented, forecasting, risk model seasonal, diagnostic, YFV-related, ovipsoition, hypeproductive, explanative, time series dependent, clinical, field and/ or remote-geosampled explanatorial, endemic transmission oriented, geopredictive, regression-based, normalized distribution generated in an ArcGIS cyberenvironment would be robust. As such, subsets of the elements of the transitional weights matrix constructed from forested canopy to agro-village complex LULC uncoalesced explanatorial parameter estimators, for example, might be optimally heuristically determined to be associated with different seasonal environmental, geosampled, geopredictive, risk-related, probabilistically geo-classified, parameterizable covariate estimator coefficients. This then

would enable determining optimal time frames for implementing IVM in an eco-epidemiological, riseland agro-ecosystem, study site.

Importantly, a seasonal, robust, explanatorial, geopredictive, YFV–related, probablistic, SAR, eco-epidemiological, risk-related, risk model may also account for a sizeable amount of the detected PSA in a specific, SAS/GIS geo-classified LULC, endemic transmission-oriented, thus allowing field teams to detect important geo-spectrotemrpoally, geospatiotemrpoally geosampled geolocations (e.g., a geospatial aggregation of prolific, georeferenceable,*Aedes* larval habitats in a riceland agro-village ecosystem based on immature field geosampled count data). Loss of information from geospatial aggregation in the presence of PSA in some environments may tend to not be severe in seasonal, YFV-related, data analyses because, perhaps on average, locally geospatially, probabilistically aggregated, regressed coefficient, time series values are similar. In contrast, loss of seasonal *Aedes* mosquito's habitat geospatial aggregation suitability in an agro-riceland eco-epidemiological, village complex, geoclassifiable LULC in the presence of NSA may result in a dramatic loss of information reflecting a sharp decline in variance for any seasonal SAS/GIS constructed, endemic, transmission-oriented, risk model estimator. For example, as candidate eigenvectors increase, the percentage of variance accounted for in a seasonal, forecast, vulnerability, explanatorial, YFV, endemic, transmission-oriented, probabilistic, diagnostic, eco-epidemiological, geo-predictiev, risk model with their respective, regression coefficient values may decrease while latent autocorrelation coefficient estimates increase. This model feature is particularly relevant for identifying NSA for optimally geospatially adjusting georeferenced, explanatorial, diagnostic, georferenceable, clinical, field or remote-geo-spectrotemrpoally, geo-spatiotemrpoally geosampled, YFV-related, time series-related, eco-epidemiological, *Ae. egypti* , oviposition geopredictive,LULC, sub-meter resolution, variables for implementing IVM as geolocations of prolific endemic, transmission-oriented activities may be regressively qualitatively quantitated geo-spatiotemporally. Intensively quantifying geo-spatiotemporal, geosampled geolocations in SAS/GIS cyberenvironments can distinguish geopredictive variables that influence vector mosquito larval/pupal productivity (Jacob al. 2009, Jacob et al., 2008, Jacob et al. 2007,).

Additionally, determining seasonal NSA in a seasonal explanatorial, forecast, vulnerability, sylvatic YFV-related, explicative, diagnostic, eco-epidemiological, georeferenceable, endemic, transmission-oriented, probablistic, heuristically optimizable, African expnding, ricland agro-irrigated , village, risk model can index the nature and degree to which a fundamental statistical assumption is violated which, in turn, can indicate the extent to which conventional inferences are compromised when non-zero spatial autocorrelation is overlooked. Autocorrelation may complicate statistical analysis of geo-spatiotemporal, geosampled, *Aedes* aquatic larval habitats, eco-georeferenced, explanatorial, geopredictive variables on specific georeferenced LULC forest canopied change sites to agro-village complex sites for instance, by altering the variance of the geosampled geopredictive variables, thereby changing the uncertainty probabilities commonly associated with making incorrect statistical decisions (Jacob et al., 2005).Also, regressively quantifying time series dependent, latent autocorrelation, probabilistic, uncertainty components can signify the presence extent of redundant information in georeferenced seasonal-geosampled, eco-epidemiological, YFV-related, diagnostic, time series depemndnet, geospectrotemrpoal, geospatialized, clinical, field or remote,time series dependent data.

Because it is redundant information, spatial autocorrelation represents duplicate information contained within georeferenced data, linking it to missing values estimation as well as to notions of effective sample size and degrees of freedom (Griffith, 2003). For probabilistically, normally distributed seasonal, *Ae. egypti* , oviposition, sub-mter resolution, sylvatic YFV-related, explanatorial, transmission-oriented, geopredictive, eco-epidemiological, endemic *Ae. egypti*, oviposition , sub-meter esolution LULC ariables these latter two quantities could establish a correspondence between geo-spectrotemporally geospatially autocorrelated and zero spatial autocorrelation, independent-geosampled, quantizable, explicative, seasonal, explanatorial, forecast, vulnerability, YFV-related observations. Richardson and Hémon (1981) promote this view for correlation coefficients computed for pairs of geospatially explanatorily distributed variables. Haining (1991) demonstrates an equivalency between their findings and the results obtained by removing spatial dependency effects with filters analogous to those employed in constructing robust time series impulse-response functions.

Accordingly, more spatially autocorrelated than independent, YFV-related, probabilistically, explanatorial, endemic observations are needed in calculations involving geo-spatiotemporal, geosampled, LULC variables to attain a valid informative statistic. Conventional regression model protocols offer a guide for regressive qualitatively quantitating local spatial autocorrelation as part of a battery of seasonal, time series dependent, YFV-related, endemic, transmission-oriented, eco-epidemiological, georeferenced risk model residual diagnostics, because both the Moran's scatterplot and the MC link directly to the analyses of the georeferenced geo-spectrotemporal, geosampled, explanatorial, geopredictor variables. The measurement of latent autocorrelation explanatorial, LULC components in a seasonally, autoregressive, sylvatic, YFV-related, endemic, transmission-oriented, eco-epidemiological,forecast, vulnerability model therefore, can describe the overall topographic risk pattern across a agro-village landscape, supporting geo-spatiotemporal geopredictions for any empirical geosampled, probablistically regressed dataset while regressive qualitatively quantitating striking deviations in the covariate, parameter estimator coefficients for compensating for unknown variables missing from any model specification.

Further, through qualitatively time series, regressive quantitation of higher-order geospatial interdependences, autoregressive disturbances, a seasonal, ArcGIS-derived, YFV-related, explanatorial, LULC-oriented, geopredictive, risk model could generate more seasonal robust inferences. We assumed a higher-order, geospatial statistical, probabilistic framework may derive generalized moments (GM) estimators for the autoregressive parameters of the disturbance process and the variances of the error components in a eco-epidemiological, clinical, field or remote geosampled, seasonal, YFV-related, endemic, transmission-oriented, probabilistic, forecast, vulnerabity model, while defining a feasible generalized two-stage least squares estimator for the regression parameters. Geo-spatiotemporal, autoregressive, YFV-related, explanatorial, clinical, field or remote-specified, geopredictive, endemic, transmission-oriented, forecasting, regression-related, probabilistic, risk-based analyses of georeferenced variables frequently employ risk model–based statistical inferences, the dependability of which is commonly based on the correctness of posited assumptions about the residual, time series dependent, LULC variance error estimates (see Jacob et al., 2007a; Jacob et al., 2006b, 2008c). Thereafter, the joint asymptotic distribution of the GM and the explanatorial, time series dependent, LULC, probabilistic geopredictive estimators can easily be attained, enabling

specification tests and proper interpolation of multi-dimensional, geospatial, probabilsitic, error component models with dependent observations in a time series, robust explanatorial, endemic, transmission-oriented, YFV-related, LULC, risk model format. Generation of an improved GM estimator for regressively qualitatively quantitating a autoregressive parameter from a spatial error model constructed in ArcGIS from a datset of empirically probabilistically regressable, eco-epidemiological clinical, field and remote georeferencable, YFV-related, meteorological and NDVI-related, geopredictor, time series dependent, explanatorial, variables, may also reveal differentiating unobservable regression disturbances and observable regression residuals.

Unfortunately, we were not able to quantitate a geo-spatiotemporally, geosampled, YFV-related clinical, field or remote,SASGIS-derived derived variable within a Q-Q plot. The standard error of a geosampled, quantile, sylvatic YFV-related, explanatorial estimate may be estimated via the bootstrap. In statistics, bootstrapping can refer to any test or metric that relies on random sampling with replacement (Hosmer and Lemeshew 2002).Bootstrapping allows assigning measures of accuracy (defined in terms of bias, variance, confidence intervals, prediction error or some other such measure) to sample estimates (Hazewinkle 2001).This technique may allow estimation of the geosampling distribution of almost any statistic using random sampling methods for constructing a robust, time series dependent, LULC, forecasting, YFV-related, eco-epidemiological, risk model. In practice this model would fall into a broader class of resampling methods. Bootstrapping is the pratice of estimating properties of an estimator (e.g., tabulated LULC explanatorial, geo-spatiotemporally georeferenced, YFV-related, geopredictive, parameter variance) by measuring those properties when geosampling from an approimating distribution.

One standardized choice for parsimoniously regressively qualitatively quantitating an empirical geosampled dataset of observed, YFV-related, clinical, field or remote–specified, geo-spatiotemporal *Ae. egypti* , oviposition data would be empirical distribution function. In statistics, the empirical distribution function, or empirical cdf, is associated with the empirical measure of the sample(Hosmer and Lemeshew 2000). This cdf is a step function that jumps up by $1/n$ at each of the n data points (Hazewinkle 2001). The empirical distribution function may estimate the true underlying cdf of the geosample points in a empirically probabilistically regressed dataset of clinical, field or remote geosampled, time series dependent observational, explanatorial predictors while converging with probability 1 according to the Glivenko–Cantelli theorem

The theory of probability, the Glivenko–Cantelli theorem determines the asymptotic behaviour of the empirical distribution function as the number of independent and identically distributed observations grows (see Hazewinkle 2001). The uniform convergence of more generalizable empirical measures becomes an important property of the Glivenko–Cantelli classes of functions or sets. The Glivenko–Cantelli classes arise in Vapnik–Chervonenkis theory, with applications to machine learning. related to statistical learning theory and to empirical processes (Fotheringham 2002). Applications for robustly, qualitatively, regressively quantiating an empiricial geosampled dataset of explanatorial, time series dependent, clinical, field or remote specified, LULC, covariate parameter estimator coefficients values may be be found by making use of M-estimators.

Let $X_1, \ldots, X_n$ be randomized, YFV-related, probabilistic, regressable, clinical, field or remote specified, LULC, time series dependent elements defined on a measurable space $(\mathcal{X}, \mathcal{A})$. Then for a measure Q set: $Qf = \int f dQ$ in the residual derivaties rendered. Measurability issues will be ignored here. Therafter, if a ecologist, entomologist, or experimenter lets $\mathcal{F}$ be a class of measurable functions $f : \mathcal{X} \to \mathbf{R}$ and defines $\|Q\|_{\mathcal{F}} = \sup\{|Qf| : f \in \mathcal{F}\}$ an empirical measure $\mathbb{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}$ may be robustly defined where δ here stands for the Dirac measure. A Dirac measure is a measure δ_x on a set X (with any σ-algebra of subsets of X) defined for a given

$$\delta_x(A) = 1_A(x) = \begin{cases} 0, & x \notin A; \\ 1, & x \in A \end{cases}$$

$x \in X$ and any (measurable) set $A \subseteq X$ by $\qquad$ where $\mathbf{1}_A$ is the indicator function of A (Hazewinkle 2001).The Dirac measure is a probability measure, and in terms of probability it represents the almost sure outcome x in the sample space X (Griffith 2003).

We may also quantitate a single measured value at x; however, treating the Dirac Mesure in a geo-spatiotemporally, geosampled, *Ae. egypti*, oviposition eco-epidemiological, YFV-related clinical, field or remote, geoclassified, LULC, forecasting, risk model would not render robust optimally mappable residual derivatives when attempting to quantiate a sequential Diarac Delta predictor as a limit of a delta sequence. The Dirac measures are the extreme points (e.g., regressed outliers) of a convex set of probability measures plotted against x. The output would then be a back-formation from the Delta-function (Schwartz distribution (please see Griffith 2003). Treating the Dirac measure as an atomic measure is not correct when we consider the sequential definition of Dirac delta, as the limit of a delta sequence in an eco-epidemiological, YFV-related, probabilistic, risk model. The Dirac measures are the extreme points of the convex set of probability measures on X. The name is a back-formation from the Dirac delta function, considered as a Schwartz distribution, for example on the real line; measures can be taken to be a special kind of distribution. The identity $\int_X f(y) \, d\delta_x(y) = f(x)$, which, in the form $\int_X f(y)\delta_x(y) \, dy = f(x)$, is often taken to be part of the definition of the "delta function", holds as a theorem of Lebesgue integration.

The empirical measure may induce constructing a robust, geopredictive, time series-related, sylvatic, YFV-related, eco-epidemiological, geo-spatiotemporally geosampled, risk map $\mathcal{F} \to \mathbf{R}$ given by: $f \mapsto \mathbb{P}_n f$. Now suppose P in the underlying true distribution based on the regression exercise. Then the probabilsically residualized, forecasted, explanatorial, clinical, field or remote-specified, geo-spatiotemrpoally geosampled oviposition, LULC variables would be unknown. However, if we utilize Empirical Process Theory in the model construction process, then identifying classes not for F would hold only if: uniform large integers values are employed in the forecasting, risk model [e.g., $\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0$,] and/or a uniform central limit theory is frameworked [e.g., $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) \rightsquigarrow \mathbb{G}, \quad \text{in } \ell^{\infty}(\mathcal{F})$]. In probability theory, an empirical process is a stochastic process that describes the proportion of objects in a system in a given state(Cressie 1993). In mean field theory, limit theorems (as the number of objects becomes large) are considered and generalise the central limit theorem for empirical measures

(Hazewinkle 2001). Applications of the theory of empirical processes arise in non-parametric statistics (Griffith 2003).Thus, for X_1, X_2, ... X_n independent and identically-distributed randomly geo-spatiotemporally geosampled, endemic, transmission-oriented, geopredictive variables in R with common cumulative distribution function F(x), the empirical distribution

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i),$$

function my be defined by where I_C is the indicator function of the set C.For every (fixed) x, $F_n(x)$ is a sequence of random variables which converge to F(x) almost surely by the strong law of large numbers. That is, F_n converges to F pointwise. Glivenko and Cantelli strengthened this result by proving uniform convergence of F_n to F by the Glivenko–Cantelli theoreumA class $\mathcal{C}$ is called a Glivenko–Cantelli class (or GC class) with respect to a probability measure P if any of the following equivalent statements is true.1. $\|P_n - P\|_{\mathcal{C}} \to 0$ almost surely as $n \to \infty$.2. $\|P_n - P\|_{\mathcal{C}} \to 0$ in probability as $n \to \infty$.3. $\mathbb{E}\|P_n - P\|_{\mathcal{C}} \to 0$, as $n \to \infty$ (convergence in mean).A class is called uniformly Glivenko–Cantelli if the convergence occurs uniformly over all probability measures P on (S,A): $\sup_{P \in \mathcal{P}(S,A)} \mathbb{E}\|P_n - P\|_{\mathcal{C}} \to 0$; and $\sup_{P \in \mathcal{P}(S,A)} \mathbb{E}\|P_n - P\|_{\mathcal{F}} \to 0$. (Cressie 1993). In the former case F would be considered a Glinvenko-Cantelli class variable and in the latter case the regressors would be conmsidered Donsker variable [i.e., $\forall x, \sup_{f \subset \mathcal{F}} |f(x) - Pf| < \infty$]. In the former case $\mathcal{F}$ is called Glivenko-Cantelli class, and in the latter case (under the assumption) the class $\mathcal{F}$ is called Donsker or P-Donsker. A Donsker class is Glivenko-Cantelli in probability by an application of Slutsky's theorem (see Hosmer and Lemeshew 2002).

Let $\{X_n\}$, $\{Y_n\}$ be sequences of scalar/vector/matrix, empirically, probablistically geosampled, YFV-related, geo-spatiotemporally, geosampled, explanatorial, LULC-related, random habitat, capture point, seasoal hyperproductive elements.If X_n converges in distribution to a random element X;and Y_n converges in probability to a constant c, then $X_n + Y_n \xrightarrow{d} X - c; X_n Y_n \xrightarrow{d} cX; X_n/Y_n \xrightarrow{d} X/c$, provided that c is invertible,where $\xrightarrow{d}$ denotes convergence in distribution.This theorem follows from the fact that if X_n converges in distribution to X and Y_n converges in probability to a constant c, then the joint vector (X_n, Y_n) converges in distribution to (X, c). Next, an ecologist, entomologist or other experimenter can apply the continuous mapping theorem for recognizing the explanatorial, regressionable, YFV-related, functions $g(x,y)=x+y$, $g(x,y)=xy$, and $g(x,y)=x^{-1}y$ as continuous (for the last function to be continuous, x has to be invertible).These statements would be true for a single f, by standard LLN, CLT arguments under regularity conditions, and the difficulty in the Empirical Processes comes in because joint statements are being made for all $f \in \mathcal{F}$. Intuitively then, the set $\mathcal{F}$ cannot be too large, and as it turns out that the geometry of $\mathcal{F}$ plays a very important role.

One way of measuring how big the function set for probabilistic, eco-epidemiological, risk modeling, time series dependent, *Ae. egypti*, oviposition YFV-related, clinical, field and remote-specified data $\mathcal{F}$ is to use the so-called covering numbers. The covering number $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ is the minimal number of balls $\{g : \|g - f\| < \varepsilon\}$ needed to cover the set $\mathcal{F}$ assuming that there is an underlying norm on $\mathcal{F}$. The entropy is the logarithm of the covering number (Cressie 1993).Two sufficient conditions may be provided under which it can be proved that the set $\mathcal{F}$ is Glivenko-Cantelli or Donsker. A class $\mathcal{F}$ is P-Glivenko-Cantelli if it is P-

measurable with envelope F such that $P^*F < \infty$ and satisfies: $\forall \varepsilon > 0 \quad \sup_Q N(\varepsilon\|F\|_Q, \mathcal{F}, L_1(Q)) < \infty.$ The next condition is a version of the celebrated Dudley's theorem. If $\mathcal{F}$ is a class of functions such that $\int_0^\infty \sup_Q \sqrt{\log N(\varepsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q))} d\varepsilon < \infty$ then $\mathcal{F}$ is P-Donsker for every probability measure P such that $P^*F^2 < \infty$. In the last integral, the notation means $\|f\|_{Q,2} = \left(\int |f|^2 dQ\right)^{\frac{1}{2}}$.

Let $(x_1, \ldots, x_n)$ be the independently explanatorial, geosampled, dataset of eco-epidemiological, time series dependent, clinical, field or remote geosampled, geo-spatiotemporal, YFV-related, randomized geopredictive variables with the common cdf $F(t)$. Then the empirical distribution function maye be defined as $$\hat{F}_n(t) = \frac{\text{number of elements in the sample} \leq t}{n} = \frac{1}{n}\sum_{i=1}^n 1\{x_i \leq t\},$$ where $1\{A\}$ would be the indicator of event A (e.g., climatic regressor). For a fixed t, the indicator $1\{x_i \leq t\}$ would then be a Bernoulli random variable with parameter $p = F(t)$, hence $n\hat{F}_n(t)$ would be a binomial random variable with mean $nF(t)$ and variance $nF(t)(1 - F(t))$ in the YFV, forecasting, eco-epidemiological, time series dependent risk model. The Bernoulli distribution is a discrete distribution having two possible outcomes labelled by $n = 0$ and $n = 1$ in which $n = 1$ ("success") occurs with probability p and $n = 0$ ("failure") occurs with probability $q = 1 - p$, where $0 < p < 1$. It therefore has probability density function $$P(n) = \begin{cases} 1 - p & \text{for } n = 0 \\ p & \text{for } n = 1, \end{cases}$$ which can also be written $$D(n) = \begin{cases} 1 - p & \text{for } n = 0 \\ 1 & \text{for } n = 1. \end{cases}$$ $P(n) = p^n(1-p)^{1-n}$. The corresponding distribution function is This implies that $\hat{F}_n(t)$ would be an unbiased forecast, *Ae. egypti*, oviposition, YFV-related, LULC, clinical, field or remote-specified, sub-meter resolution, LULC estimator for $F(t)$.

By the strong law of large numbers, the probabilsitic, clincial, field or remote, YFV – related, LULC parameter estimators $\hat{F}_n(t)$ converges to $F(t)$ as $n \to \infty$ almost surely, for every value of t: $\hat{F}_n(t) \xrightarrow{a.s.} F(t)$, thus the estimator $\hat{F}_n(t)$ is consistent. This expression asserts the point-wise convergence of the empirical distribution function to the true cdf. There is a stronger result, called the Glivenko–Cantelli theorem, which states that the convergence in fact happens uniformly over t: [3] $\|\hat{F}_n - F\|_\infty \equiv \sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F(t)| \xrightarrow{a.s.} 0.$ The sup-norm in this expression is called the Kolmogorov–Smirnov statistic for testing the goodness-of-fit between the empirical distribution $\hat{F}_n(t)$ and the assumed true cdf F. Other norm functions may be reasonably employed here instead of the sup-norm for qualitatively probalistically regressively quantitating an empirical dataset of explanatorial, clinical, field or remote-specified, geo-spatiotemporal, sylvatic YFV-related, sub-meter resolution, ovispoition, LULC, geosampled variables. For example, the L^2-norm gives rise to the Cramér–von Mises statistic.

The Cramér–von Mises statsitic is based on a non-parametric test for testing a hypothesis H_0 which states that independent and identically-distributed, random variables $X_1, \ldots, X_n$ (e.g., an empirical dataset of geo-spatiotemporal, LULC, explanatorial, clinical, field

or remote geosampled, YFV-related,predictors variables have a given continuous distribution function $F(x)$. The Cramér–von Mises test is

$$\omega_n^2\,[\,\Psi\,(F(x))] = \int_{-\infty}^{+\infty} \left[\sqrt{n}(F_n(x) - F(x))\right]^2 \Psi\,(F(x))\,dF(x),$$

where $F_n(x)$ is the empirical distribution function constructed from the sample $X_1,\ldots,X_n$ and $\Psi\,(t)$ is a certain non-negative function defined on the interval $[0,\,1]$ such that $\Psi\,(t)$, $t\,\Psi\,(t)$ and $t^2\,\Psi\,(t)$ are integrable on $[0,\,1]$. Tests of this type, based on the "square metric" , were first considered by H. Cramér [C] and R. von Mises [M]. N.V. Smirnov proposed putting $\Psi\,(t) \equiv 1$, and showed that in that case, if the hypothesis H_0 is valid and $n \to \infty$, the statistic $\omega^2 = \omega_n^2$ would then have the limit of an "omega-squared" distribution, independent of the hypothetical distribution function $F(x)$. A statistical test for testing H_0 in a geo-spatiotemporally, geosampled, probabilsitic, YFV-related , forecasting, LULC, clinical, field or remote-specified, ,ecoepidemiological, endemic, transmission-oriented,risk model based on the statistic ω_n^2, would be an ω^2(i.e., Cramér–von Mises–Smirnov) test, where the explanatorial, interpolatable, geosampled numerical value of ω_n^2 may be found using the following representation:

$$\omega_n^2 = \frac{1}{12n} + \sum_{j=1}^{n}\left[F(X_{(j)}) - \frac{2j-1}{2n}\right]^2$$

where $X_{(1)} \leq \ldots \leq X_{(n)}$ would be the variational series based on the sample $X_1,\ldots,X_n$. According to the ω^2 test with significance level α, the hypothesis H_0 is rejected whenever $\omega_n^2 \geq \omega_\alpha^2$, where ω_α^2 is the upper α-quantile of the distribution of ω^2, (i.e. $P\{\omega^2 < \omega_\alpha^2\} = 1 - \alpha$.) A similar constructed test, based on the statistic $\omega_n^2\,[(1 - F(x))/F(x)]$ may also render powerful explanatorial, clinical, field or remote specified , time series depnndent, YFV-related predictors.

The asymptotic distribution can be further characterized in several different ways in an explanatorial, endemic, transmission-oriented,geo-spatiotemporal, geosampled, eco-epidemiological, forecasting, clinical, field or remote-specified, LULC, YFV-related, risk model. First, the central limit theorem states that pointwise, $\hat{F}_n(t)$ has asymptotically normal distribution with the standard $\sqrt{n}$ rate of convergence $\sqrt{n}(\hat{F}_n(t) - F(t)) \xrightarrow{d} \mathcal{N}\left(0, F(t)(1 - F(t))\right)$ This result is extended by the Donsker's theorem, which asserts that the empirical process $\sqrt{n}(\hat{F}_n - F)$, viewed as a function indexed by $t \in \mathbb{R}$, converges in distribution in the Skorokhod space $D[-\infty, +\infty]$ to the mean-zero Gaussian process $G_F = B \circ F$, where B is the standard Brownian bridge where the covariance structure of this Gaussian process is $E[\,G_F(t_1)G_F(t_2)] = F(t_1 \wedge t_2) - F(t_1)F(t_2)$.

Brownian bridge is a continuous-time stochastic process $B(t)$ whose probability distribution is the conditional probability distribution of a Wiener process $W(t)$ (a mathematical model of Brownian motion) given the condition that $B(1) = 0$. A Brownian bridge is the result of Donsker's theorem. More precisely: $B_t := (W_t | W_1 = 0)$, $t \in [0, 1]$ in a robust, eco-epidemiological, time series dependent, sylvatic, YFV-related, LULC, explanatorial, clinical, field or remote, geosampled, probabilistic, forecast, vulnerability, ovispoition, capture point, regression model The expected value of the bridge would then be zero, with variance $t(1 - t)$, implying that the most uncertainty probabilities is in the middle of the bridge, with zero uncertainty at the nodes. The covariance of $B(s)$ and $B(t)$ would then be $s(1 - t)$ if $s < t$ The increments in a Brownian bridge are not independent (Griffith 2003).If $W(t)$ is a standard Wiener

process (i.e., for $t \geq 0$, $W(t)$ would then be normally distributed with expected value 0 and variance t, and the increments would be stationary and independent). As such, $B(t) = W(t) - tW(1)$ would be a Brownian bridge for $t \in [0, 1]$ in the model derivatives. Conversely, if $B(t)$ is a Brownian bridge and Z is a standard normal random clinical, field or remote geosampled explanatorial, sylvatic YFV-related probabilsitically regressable variable, then the process $W(t) = B(t) + tZ$ would be a Wiener process for $t \in [0, 1]$.

The Wiener process W_t is characterised by three properties: 1) $W_0 = 0$, 2) The function $t \rightarrow W_t$ is almost surely everywhere continuous and 3) W_t has independent increments with $W_t - W_s \sim N(0, t-s)$ (for $0 \leq s < t$), where $N(\mu, \sigma^2)$ would denote a empirically regressed geo-spatiotemporal, clincial, field or remote geosampled, YFV-related normal distribution with expected value μ and variance σ^2. The last condition means that if $0 \leq s_1 < t_1 \leq s_2 < t_2$ then $W_{t1} - W_{s1}$ and $W_{t2} - W_{s2}$ are independent geosampled, LULC, sub-meter resolution, *Ae. egypti*, oviposition, -specified, probabilistic, random variables, and the similar condition holds for n increments. An alternative characterisation of the Wiener process is the so-called *Lévy characterisation* that says that the Wiener process is an almost surely continuous martingale with $W_0 = 0$ and quadratic variation $[W_t, W_t] = t$ (which means that $W_t^2 - t$ is also a martingale) (Cressie 1993). A third characterisation is that the Wiener process has a spectral representation as a sine series whose coefficients are independent $N(0, 1)$ random variables. This representation can be obtained using the Karhunen–Loève theorem.

An arbovirologist, entomologist or YF experimenter may consider a square, integrable, zero-mean, randomized, explanatorial, empirically regressable, probabilistic process X_t defined over a probability space $(\Omega, F, \mathbf{P})$ in a geo-spectrotemrpoally geosampled, YFV-related, forecasting eco-epidemiological, LULC, probablistic, *Ae. egypti*, oviposition, risk model and indexed over a closed interval $[a, b]$, with covariance function $K_X(s, t)$. By so doing, the geo-spatiotemporal or geo-spectrotemporal, ovispoition, sub-meter resolution, covariate paramter estimators would be orgainzed As $\forall t \in [a, b]$ $X_t \in L^2(\Omega, F, \mathbf{P}), \forall t \in [a, b]$ $\mathbf{E}[X_t] = 0$ and $\forall t, s \in [a, b]$ $K_X(s, t) = \mathbf{E}[X_s X_t]$. Thereafter K_X and a linear operator T_{KX} may be defined in the following way:

$$\begin{cases} T_{K_X} : L^2([a, b]) \rightarrow L^2([a, b]) \\ f \mapsto \int_a^b K_X(s, \cdot) f(s) ds \end{cases}$$

Since T_{KX} is a linear operator, it makes sense to talk about its eigenvalues λ_k and eigenfunctions e_k, which are found solving the homogeneous Fredholm integral equation of the second kind.

$$\int_a^b K_X(s, t) c_k(s) \, ds = \lambda_k c_k(t)$$

Fredholm Equation is an Integral Equation in which the term containing the Kernel Function has constants as integration limits (Hazewinkle 2001).

A closely related form to the Fredholm Equation is the Volterra integral equation which has variable integral limits. A linear Volterra equation of the first kind for probabilisitically regressing an operationizable, geo-spatiotemporally, geosampled probablistic, YFV-related, eco-epidemiological, clinical, field or remote dataset of time series covariate,paramter estimators. In operator theory, and in Fredholm theory, the corresponding equations are called the Volterra operator(Hosmer and Lemeshew 2002), eco-epidemiological, dataset of explanatorial, clinical, field or remote-specified, LULC-related variables could be $f(t) - \int_a^t K(t, s) x(s) \, ds$ where f is a

given function and x is an unknown function to be solved for. A linear Volterra equation of the second kind for the empirical geosampled datset then could then be written $x(t) = f(t) + \int_a^t K(t,s)x(s)\,ds$. A linear Volterra integral equation is a convolution equation if $x(t) = f(t) + \int_{t_0}^t K(t-s)x(s)\,ds$ (Cressie 1993). An inhomogeneous Fredholm equation of the first kind then could written as: $g(t) = \int_a^b K(t,s)f(s)\,ds$ and the problem is, given the continuous kernel function $K(t,s)$, and the function $g(t)$, to find the function $f(s)$ (Hazewinkle 2001). If the kernel is a function only of the difference of its arguments, namely $K(t,s) = K(t-s)$ in the endemic, transmission-oriented, probilistically regressable, explanatorial, clinical, field or remote geosampled, forecast, vulnerability, *Ae. egypti* , oviposition sub-mter resolution LULC model and the limits of integration are $\pm\infty$, then the right hand side of the equation can be rewritten as a convolution of the functions K and f and therefore the solution mey be givenby $f(t) = \mathcal{F}_\omega^{-1}\left[\dfrac{\mathcal{F}_t[g(t)](\omega)}{\mathcal{F}_t[K(t)](\omega)}\right] = \int_{-\infty}^{\infty} \dfrac{\mathcal{F}_t[g(t)](\omega)}{\mathcal{F}_t[K(t)](\omega)} e^{2\pi i \omega t}\,d\omega$ where $\mathcal{F}_t$ and $\mathcal{F}_\omega^{-1}$ are the direct and inverse Fourier transforms respectively.

Let X_t be a zero-mean square integrable explanatorial, time series dependent, stochastic process defined over a probability space (Ω, F, **P**) and indexed over a closed and bounded interval $[a, b]$, with continuous covariance function $K_X(s, t)$ in a geo-spatiotemporally, geosampled, explanatorial, clinical, field or remote, geo-spectrotemporal or geo-spatiotemrporal, ecoe-pidemiological, geosampled, sylvatic YFV-related eco-epidemiological, endemic, transmission-oriented, forecasting, regression-related, African riceland, risk model. By so doing, $K_X(s,t)$ would be a Mercer kernel. To explain Mercer's theorem, for empirically regressable, risk modeling geo-spatiotemporal, geosampled, an ecologist, entomologist or experimenter must first consider an important special. The covariance function K_X satisfies the definition of a Mercer kernel. By Mercer's theorem, there consequently exists a set $\{\lambda_k, e_k(t)\}$ of eigenvalues and eigenfunctions of T_{KX} forming an orthonormal basis of $L^2([a,b])$, and K_X can be expressed as $K_X(s,t) = \sum_{k=1}^{\infty} \lambda_k e_k(s) e_k(t)$. The process X_t can be expanded in terms of the eigenfunctions e_k as: $X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$ where the covariate paramter estimator coefficients (e.g., random variables) Z_k are given by the projection of X_t on the respective eigenfunctions $Z_k = \int_a^b X_t e_k(t)\,dt$. A kernel, in this context, would be a symmetric continuous function that maps $K : [a,b] \times [a,b] \to \mathbb{R}$ where symmetric means that $K(x, s) = K(s, x)$. K is said to be non-negative definite (or positive semidefinite) if and only if $\sum_{i=1}^{n}\sum_{j=1}^{n} K(x_i, x_j)c_i c_j \geq 0$ for all finite explanatorial, georeferenced, clincial, field or remote –related sequences of geosampled sub-mter resolution LULC sample points $x_1, ..., x_n$ of $[a, b]$ and all choices of covariate parameter estimators $c_1, ..., c_n$ (cf. positive definite kernel).

More generally, a Wiener process $W(t)$ for $t \in [0, T]$ can be decomposed into $W(t) = B\left(\dfrac{t}{T}\right) + \dfrac{t}{\sqrt{T}}Z$. Another representation of the Brownian bridge based on the

413

Brownian motion is, for $t \in [0, 1]$ $B(t) = (1-t)W\left(\frac{t}{1-t}\right)$. Conversely, for $t \in [0, \infty]$ $W(t) = (1+t)B\left(\frac{t}{1+t}\right)$. The Brownian bridge may also be represented as a Fourier series with stochastic coefficients, as $B_t = \sum_{k=1}^{\infty} Z_k \frac{\sqrt{2}\sin(k\pi t)}{k\pi}$ where $Z_1, Z_2, \cdots$ are independent identically distributed standard normal random variables (see the Karhunen–Loève theorem).

In the theory of stochastic processes, the Karhunen–Loève theorem is a representation of a stochastic process as an infinite linear combination of orthogonal functions, analogous to a Fourier series representation of a function on a bounded interval. Stochastic processes given by infinite series of this form may be condered for seasonal qualitatively regressively quantitating probabilistically of an operationizable datset of clinical, field or remote-specified, geo-spatiotemporally geosampled, YFV-related parameterizable, *Ae. egypti*, oviposition covariate, estimator datasets. There exist many such expansions of a stochastic process that may be viable for regression-related, geopredictive, probabilistic, eco-epidemiological, risk, modeling if the process is indexed over $[a, b]$. As such, any geopredictive, eco-epidemiological, explanatorial, probabilistic, YFV-related, risk model may render an orthonormal basis of $L^2([a, b])$ when regressively qualiatively quantaitaing an expansion thereof in that form. The importance of the Karhunen–Loève theorem is that it would yield an optimal residually forecasted derivative since during the probabilistic tabulation exercise employing the theorem would yield the optimal unbiased covariate paramter estimator dataset since the output would minimize the total mean squared error.

Moreover, if the probabilsitically regression-related, sylvatic YFV-related process is Gaussian, then the explanatorial, clinical, field or remote geo-spatiotemporally geosampled explanatorial, random variables Z_k would be Gaussian and stochastically independent. This result would generalize the Karhunen–Loève transform. By so doin, an important example of a centered real stochastic process on $[0, 1]$ in a YFV-related eco-epidemiological, time series dependent model output would be the Wiener process. The continuous-time stochastic process $W(t)$ for $t \geq 0$ with $W(0) = 0$ and such that the increment $W(t) - W(s)$ is Gaussian with mean 0 and variance $t - s$ for any $0 \leq s < t$, and increments for nonoverlapping time intervals are independent. Brownian motion (i.e., random walk with random step sizes) is the most common example of a Wiener process. The Karhunen–Loève theorem can be used to provide a canonical orthogonal representation of the forecasted, geo-spatially interpolated, explanatorial, clinical, field or remote geosampled variables. By so doing, the expansion would consist of sinusoidal functions. In contrast to a Fourier series where the coefficients are fixed numbers and the expansion basis that consists of sinusoidal functions (that is, sine and cosine YFV-related functions), the coefficients in the Karhunen–Loève theorem are random variables and the expansion basis depends on the process. In fact, the orthogonal basis functions used in this representation are determined by the covariance function of the process. One can think that the Karhunen–Loève transform adapts to the process in order to produce the best possible basis for its expansion.

In the case of a centered stochastic process $\{X_t\}_{t \in [a, b]}$ (centered means $\mathbf{E}[X_t] = 0$ for all $t \in [a, b]$) satisfying a technical continuity condition, X_t admits a

decomposition $X_t = \sum_{k=1}^{\infty} Z_k e_k(t)$ where Z_k are pairwise, uncorrelated, random variables and the functions e_k are continuous real-valued functions on [a, b] that are pairwise orthogonal in $L^2([a, b])$. It is therefore sometimes said that the expansion is bi-orthogonal since the random coefficients Z_k are orthogonal in the probability space while the deterministic functions e_k are orthogonal in the time domain. The general case of a process X_t that is not centered can be brought back to the case of a centered process by considering $X_t - \mathbf{E}[X_t]$ which is a centered process. The above expansion into uncorrelated probabilistic, geo-spetrotemporally, geosampled, YFV-related, randomized SAS/GIS-geoclassified, variables may be also known as the Karhunen–Loève expansion or Karhunen–Loève decomposition. The empirical version (i.e., with the coefficients computed from a sample) is known as the Karhunen–Loève transform (KLT), principal component analysis, proper orthogonal decomposition (POD)(Cressie 1993), Empirical orthogonal functions (a term used in meteorology and geophysics), or the Hotelling transform.

Hotelling's T-squared distribution is a generalization of the Student's t-distribution in multivariate setting, and its use in statistical hypothesis testing and confidence regions. If $\mathbf{x}_1, \ldots, \mathbf{x}_{n_x} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$ and $\mathbf{y}_1, \ldots, \mathbf{y}_{n_y} \sim N_p(\boldsymbol{\mu}, \mathbf{V})$, with the samples independently drawn from two independent multivariate normal distributions with the same mean and covariance, and we define $\overline{\mathbf{x}} = \frac{1}{n_x} \sum_{i=1}^{n_x} \mathbf{x}_i \quad \overline{\mathbf{y}} = \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$ as the sample means, and $\mathbf{W} = \frac{\sum_{i=1}^{n_x}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})' + \sum_{i=1}^{n_y}(\mathbf{y}_i - \overline{\mathbf{y}})(\mathbf{y}_i - \overline{\mathbf{y}})'}{n_x + n_y - 2}$ as the unbiased pooled covariance matrix estimate, then Hotelling's two-sample T-squared statistic is $t^2 = \frac{n_x n_y}{n_x + n_y}(\overline{\mathbf{x}} - \overline{\mathbf{y}})' \mathbf{W}^{-1}(\overline{\mathbf{x}} - \overline{\mathbf{y}}) \sim T^2(p, n_x + n_y - 2)$ and it can be related to the F-distribution by $\frac{n_x + n_y - p - 1}{(n_x + n_y - 2)p} t^2 \sim F(p, n_x + n_y - 1 - p).$

By employing the Hotellings T-squared distribution a probabilistic, canonical correlation analyses can be performed in SAS for an empirical, explanatorial dataset of clinical, field or remote geo- spatiotemporal or geo-spectrotemrporal, geosampled sylvatic, YF-related LULC sub-meter resolution, oviposition analyses. If the vector $_p\mathbf{d}_I$ is Gaussian multivariate-distributed with zero mean and unit covariance matrix $N(_p0_I, _pI_p)$ and $_m\mathbf{M}_p$ is a p x p matrix with a Wishart distribution with unit scale matrix and m degrees of freedom $W(_pI_p, m)$ then $m(_I\mathbf{d}'\,_p\mathbf{M}^{-1}\,_p\mathbf{d}_I)$ has a Hotelling T^2 distribution with dimensionality parameter p and m degrees of freedom. In statistics, the Wishart distribution is a generalization to multiple dimensions of the chi-squared distribution, or, in the case of non-integer degrees of freedom, of the gamma distribution.

If the notation $T_{p,m}^2$ is used to denote a geo-spatiotemrpoally geosampled, explanatorial, clincial, field or remote, YFV randomized LULC, variable having Hotelling's T-squared distribution with parameters p and m then, if a random variable X has Hotelling's T-squared distribution, $X \sim T_{p,m}^2$ then $\frac{m - p + 1}{pm} X \sim F_{p,m-p+1}$ where $F_{p,m-p+1}$ is the F-distribution with parameters p and $m-p+1$. Hotelling's T-squared statistic is a generalization of Student's t statistic that is used in multivariate hypothesis testing, and is defined as follows. Let $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a p-variate normal distribution with location $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be n independent

random variables, which may be represented as $p \times 1$ column vectors of real numbers. Define $\bar{x} = \frac{x_1 + \cdots + x_n}{n}$ to be the sample mean. It can be shown that $n(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu) \sim \chi^2_p$ where χ^2_p is the chi-squared distribution with p degrees of freedom. To show this use the fact that $\bar{x} \sim \mathcal{N}_p(\mu, \Sigma/n)$ and then derive the characteristic function of the random variable $y = n(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu)$. This may be probabilistically, geo-spatiotemporally employed as below,

$$\phi_y(\theta) = E\, e^{i\theta y} = E\, e^{i\theta n(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu)} = \int e^{i\theta n(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu)} (2\pi)^{-\frac{p}{2}} |\Sigma/n|^{-\frac{1}{2}} e^{-\frac{1}{2}n(\bar{x} - \mu)'\Sigma^{-1}(\bar{x} - \mu)} \, dx_1 \dots dx_p$$

$$= \int (2\pi)^{-\frac{p}{2}} |\Sigma/n|^{-\frac{1}{2}} e^{-\frac{1}{2}n(\bar{x} - \mu)'(\Sigma^{-1} - 2i\theta\Sigma^{-1})(\bar{x} - \mu)} \, dx_1 \dots dx_p = |(\Sigma^{-1} - 2i\theta\Sigma^{-1})^{-1}/n|^{\frac{1}{2}} |\Sigma/n|^{-\frac{1}{2}} \int (2\pi)^{-\frac{p}{2}} |(\Sigma^{-1} - 2i\theta\Sigma^{-1})^{-1}/n|^{-\frac{1}{2}} e^{-\frac{1}{2}n(\bar{x} - \mu)'(\Sigma^{-1} - 2i\theta\Sigma^{-1})(\bar{x} - \mu)} \, dx_1 \dots dx_p,$$

$$= |(I_p - 2i\theta I_p)|^{-\frac{1}{2}} = (1 - 2i\theta)^{-\frac{p}{2}}.$$ However, Σ is often unknown and we wish to do hypothesis testing on the location μ.

An eco-epidemiological, probabilsitic, explanatorial, empirical datatset of clinical, field or remote-geosampled, YFV-related, predictive, sylvatic *Ae. aegypti,* risk model may be defined in SAS as $W = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'$ to be the sample covariance. By so doing, it can be shown the W is positive-definite and as such (n-1) W would also be positive defintIt can be shown that W is positive-definite and $(n-1)W$ follows a *p*-variate Wishart distribution with *n*−1 degrees of freedom. The Wishart distribution is a generalization of the univariate chi-square distribution to two or more variables. It is a distribution for symmetric positive semidefinite matrices, typically covariance matrices, the diagonal elements of which are each chi-square random variables (Crtessie 1993). In the same way as the chi-square distribution can be constructed by summing the squares of independent, identically distributed, zero-mean univariate normal random variables, a time series dependent, endmember,

YFV-related,LULC, Wishart distribution can be constructed by summing the inner products of independent, identically distributed, zero-mean multivariate normal random vectors. The Wishart distribution is often used as a model for the distribution of the sample covariance matrix for multivariate normal random data, after scaling by the sample size (Griffith 2003)..

Hotelling's *T*-squared statistic may be then probabilistically empirically defined to be $t^2 - n(\bar{x} - \mu)'W^{-1}(\bar{x} - \mu)$ and, also from above, $t^2 \sim T^2_{p,n-1}$ i.e. $\frac{n-p}{p(n-1)} t^2 \sim F_{p,n-p}$, where $F_{p,n-p}$ would be the *F*-distribution with the geosampled explanatorial, parameter estimators *p* and *n*−*p*. In order to calculate a *p* value, multiply the t^2 statistic by the above constant and use the *F*-distribution. The non-null distribution of this statistic is the noncentral F-distribution (the ratio of a non-central Chi-squared random variable and an independent central Chi-squared random variable) $\frac{n_x + n_y - p - 1}{(n_x - n_y - 2)p} t^2 \sim F(p, n_x + n_y - 1 - p; \delta)$, with $\delta = \frac{n_x n_y}{n_x + n_y} \nu' V^{-1} \nu$, where ν is the difference vector between the population means.

The set of all càdlàg functions from *E* to *M* is often denoted by *D*(*E*; *M*) (or simply *D*) and is called Skorokhod space (Cressie 1993).Skorokhod space can be assigned a topology that, intuitively allows a ecologist, entomologist or experimenter to "wiggle space and time a bit" in an explanatorial, geo-spatiotemporal, geosampled, eco-epidemiological, LULC, clinical, field or remote-specified, endemic, transmission-oriented,YFV-related, forecasting, risk model. For simplicity, the model paramter estimator may take *E* = [0, *T*] and *M* = R^n However, an analogue

of the modulus of continuity, $\varpi'_f(\delta)$ must be defined in the forecasting, risk model . For any $F \subseteq E$, set $w_f(F) := \sup_{s,t \in F} |f(s) - f(t)|$ and, for $\delta > 0$, defining the càdlàg modulus to be $\varpi'_f(\delta) := \inf_{\Pi} \max_{1 \le i \le k} w_f([t_{i-1}, t_i])$, where the infimum runs over all partitions $\Pi = \{0 = t_0 < t_1 < \ldots < t_k = T\}$, $k \in \mathbb{N}$, with $\min_i (t_i - t_{i-1}) > \delta$ may render optimal, estimators (Cressie 1993). This definition would make sense for non-càdlàg f in a geospatiotemporal, geosampled, YFV-related, eco-epidemiological, geopredictive, risk model (just as the usual modulus of continuity makes sense for discontinuous functions) where it can be shown that f is càdlàg if and only if $\varpi'_f(\delta) \to 0$ as $\delta \to 0$. Now let Λ denote the set of all strictly increasing, continuous bijections from E to itself (these are "wiggles in time") in the model. Then let $\|f\| := \sup_{t \in E} |f(t)|$ denote the uniform norm on functions on E. Therefater by defining the Skorokhod metric σ on D by $\sigma(f, g) := \inf_{\lambda \in \Lambda} \max\{\|\lambda - I\|, \|f - g \circ \lambda\|\}$, where $I: E \to E$ is the identity function, the "wiggle" intuition, $\|\lambda - I\|$ would measure the size of the "wiggle in time", and $\|f - g \circ \lambda\|$ would measure the size of the "wiggle in space". It thus may be shown that the Skorokhod metric is indeed a metric in a robust, interpolatable, endemic, transmission-oriented, explanatorial, time series dependent, LULC, sylvatic ,YFV-related probabilsitically regression-oriented, forecasting, eco-epidemiological, risk model. The topology Σ generated by σ in the YFV model would be the Skorokhod topology on D.

If the space C of a continuous dataset of empirically regressable, YFV-related, geosampled, probabilistic, endmeber,LULC, time series dependent functions on E is a subspace of D , then the Skorokhod topology may be relativized to C in the model probabilistic derivatives as it would coincide with the uniform topology . It may be shown employing convergence of regression-related, probability measures that, although D is not a complete space with respect to the Skorokhod metric σ, there is a topologically equivalent metric σ_0 with respect to which D is complete in the geo-spatiotemporal, geosampled, endemic, transmission-oriented, YFV-related geopredictive, eco-epidemiological, probabilistic, LULC, risk model (see Billingsley 199), With respect to either σ or σ_0, D is a separable space (Hazewinkle 2001). Thus, Skorokhod space is a Polish space. In the mathematical discipline of general topology, a Polish space is a separable completely metrizable topological space; that is, a space homeomorphic to a complete metric space that has a countable dense subset (Cressie 1993). Under the Skorokhod topology and pointwise addition of functions, D would not be a topological group in a explanatorial, clincial, field or remote-specified, LULC, time series dependent, risk model as can be seen by the following example:Let $E = [0, 2)$ be the unit interval and take $f_n - \chi_{[1-1/n, 2)} \in D$ to be a sequence of characteristic functions. Despite the fact that $f_n \to \chi_{[1,2]}$ in the Skorokhod topology in the YFV-related model, the sequence $f_n - \chi_{[1,2)}$ would not converge to 0.

By an application of the Arzelà–Ascoli theorem, an ecologist, entomologist, or experimenter can show that a sequence $(\mu_n)_{n=1,2,\ldots}$ of probability measures on Skorokhod space D in a time series, forecasting, explanatorial, clinical, field or remote, YFV, LULC, eco-epidemiological, risk model is tight if and only if both the following conditions are met: $\lim_{a \to \infty} \limsup_{n \to \infty} \mu_n(\{f \in D \mid \|f\| \ge a\}) = 0$, and $\lim_{\delta \to 0} \limsup_{n \to \infty} \mu_n(\{f \in D \mid \varpi'_f(\delta) \ge \varepsilon\}) = 0$ for all $\varepsilon > 0$.
The Arzelà–Ascoli theorem is a fundamental result of mathematical analysis giving necessary

and sufficient conditions to decide whether every sequence of a given family of real-valued continuous functions defined on a closed and bounded interval has a uniformly convergent subsequence (Hazewinkle 2001). The main condition in the risk model eco-epidemiological, probabilistic, residualized, regressed derivatives would then be the equicontinuity of the family of functions. The theorem is the basis of many proofs in mathematics, including that of the Peano existence theorem in the theory of ordinary differential equations, Montel's theorem in complex analysis, and the Peter–Weyl theorem in harmonic analysis. Based on the Peano existence theorem if an ecologist, entomologist or experimenter lets D be an open subset of $\mathbb{R} \times \mathbb{R}$ with $f: D \to \mathbb{R}$ a continuous function and $y'(x) = f(x, y(x))$ a continuous, explicit first-order differential equation defined on D, then every initial value problem $y(x_0) = y_0$ for f with $(x_0, y_0) \in D$ has a local solution $z: I \to \mathbb{R}$ where I is a neighbourhood of x_0 in $\mathbb{R}$, such that $z'(x) = f(x, z(x))$ for all $x \in I$. Suppose that $\mathcal{F}$ is a family of meromorphic functions on an open set D in a then is $z_0 \in D$ is such that $\mathcal{F}$ is not normal at z_0, and $U \subset D$ is a neighborhood of z_0, and $\bigcup_{f \in \mathcal{F}} f(U)$ would be dense in the complex plane in a robust, empirically regressable, geo-spatiotempoal, geosampled, explanatorial, clinical, field or remote geosampled, YFV-related probabilsitic, eco-epidemiological, LULC, risk model based on the Montel Theoreum. In mathematics, the Peter–Weyl theorem is a basic result in the theory of harmonic analysis, applying to topological groups that are compact, but are not necessarily abelian (Cressie 1993).

Alternatively, a sequence $\{f_n\}_{n \in \mathbf{N}}$ of continuous, geosampled, explanatorial, clinical, field or remote specified functions on an interval $I = [a, b]$ in an eco-epidemiological, YFV-related, probabilsitic, explanatorial, clinical, field or remote-specified, LULC, risk model may be uniformly bounded if there is a number M such that $|f_n(x)| \leq M$ for every function f_n belonging to the sequence, and every $x \in [a, b]$. By so doing, the sequence would be equicontinuous if, for every $\varepsilon > 0$, there exists $\delta > 0$ such that $|f_n(x) - f_n(y)| < \varepsilon$ whenever $|x - y| < \delta$ for all functions f_n in the algrothmic sequence. Succinctly, a sequence in an eco-epidemiological, YFDV-related, probabilistic, explanatorial, clinical, field or remote-specified, geospecified, LULC, predictive, *Ae .aegypti,* ovisposition, eco-epidemiological, risk model is equicontinuous if and only if all of its elements admit the same modulus of continuity. In simplest terms, the theorem can be stated as follows: Consider a sequence of real-valued continuous functions $\{f_n\}_{n \in \mathbf{N}}$ in a eco-epidemiological, explanatorial, geo-spatiotemporal, YFV-related, forecasting, risk model defined on a closed and bounded interval $[a, b]$ of the real line. If this sequence is uniformly bounded and equicontinuous, then there exists a subsequence (f_{nk}) that converges uniformly. The converse is also true, in the sense that if every subsequence of $\{f_n\}$ itself has a uniform convergent subsequence, then $\{f_n\}$ is uniformly bounded and equicontinuous.

Let $\mathcal{D} = \mathcal{D}[0, 1]$ be the space of real-valued functions x on $[0, 1]$ that are right-continuous and have left-hand limits, i.e. $x(t+) = \lim_{s \downarrow t} x(s)$ exists, $x(t+) = x(t)$ for all $0 \leq t < 1$, $x(t-) = \lim_{s \uparrow t} x(s)$ exists for all $0 < t \leq 1$. In probabilistic literature, such a function is also said to be a cadlag function, "cadlag" being an acronym for the French "continu à droite, limites à gauche". Introducing a norm on $\mathcal{D}$ by setting $\|x\| = \sup_{0 \leq t \leq 1} |x(t)|$, then $\mathcal{D}$ becomes a Banach space, but it is easy to see that it is non-

separable (cf. also Separable space). The Banach space is a complete normed vector space. The function spaces introduced by D. Hilbert, M. Fréchet and F. Riesz between 1904 and 1918 served as the starting point for spaces of strong and weak convergence, compactness, linear functional, linear operator etc., A Banach space is a complete vector space B with a norm $\|\cdot\|$ (Hazewinkle 2001). Two norms $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ are called equivalent if they give the same topology, which is equivalent to the existence of constants c and C such that $\|v\|_{(1)} \le c \|v\|_{(2)}$ and $\|v\|_{(2)} \le C \|v\|_{(1)}$ hold for all v (Cressie 1993) In the finite-dimensional case, all norms are equivalent. An infinite-dimensional space in an explanatorial, interpolatable, geo-spatiotemporally, geosampled, YFV-related , clinical, field or remote, ArcGIS geoclassified LULC , forecasting, eco-epidemiological, risk model can have many different norms. A basic example is n-dimensional Euclidean space with the Euclidean norm. Usually, the notion of Banach space is only used in the infinite dimensional setting, typically as a vector space of functions(Griffith 2003). For example, the set of continuous functions on closed interval of the real line with the norm of a function f given by
$$\|f\| = \sup_{x \in \mathbb{R}} |f(x)|$$
is a Banach space in a, explanatorial, YFV-related LULC risk model , where $\sup$ denotes the supremum. On the other hand, the set of continuous functions on the unit interval $[0, 1]$ with the norm of a function f given by
$$\|f\| = \int_0^1 |f(x)|\, dx$$
is not a Banach space in the geopredictive, risk model because it is not complete. For instance, the

$$f_n = \begin{cases} 1 & \text{for } x \le \dfrac{1}{2} \\ \dfrac{1}{2}n + 1 - nx & \text{for } \dfrac{1}{2} < x \le \dfrac{1}{2} + \dfrac{1}{n} \\ 0 & \text{for } x > \dfrac{1}{2} + \dfrac{1}{n} \end{cases}$$

Cauchy sequence of functions would not converge to a continuous function in the model. A sequence $a_1, a_2, \ldots$ such that the metric $d(a_m, a_n)$ satisfies $\lim_{\min(m,n)\to\infty} d(a_m, a_n) = 0.$ Cauchy sequences in the rationals do not necessarily converge, but they do converge in the reals(empirical, operationizable datset of explanatorial, clinical, field or remote ,LULC, geo-spatiotemporally geosampled, covariate paramter estimator coefficients).These discrete interger values can be defined using either Dedekind cuts or Cauchy sequences. Dedekind cut is a partition of the rational numbers into two non-empty sets A and B, such that all elements of A are less than all elements of B, and A contains no greatest element (Hazewinkle 2001) Hilbert spaces with their norm given by the inner product are examples of Banach spaces in the explanatorial, sylvatic YFV-reltaed forecasting LULC, African, expanding, riceland, ovispoition, vulnerability model . While a Hilbert space is always a Banach space, the converse need not hold in a robust, YFV-related probabilsic risk model. Therefore, it is possible for a Banach space not to have a norm given by an inner product. For instance, the supremum norm cannot be given by an inner product in the model.

Examples of inner product spaces include: 1. The real numbers $\mathbb{R}$, where the inner product is given by $(x, y) = x\,y$. 2. The Euclidean space $\mathbb{R}^n$, where the inner product is given by the dot product $((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = x_1 y_1 + x_2 y_2 + \cdots x_n y_n$ Inner product is a generalization of the dot product (Cressie 1993). In a vector space, it is a way to multiply vectors together, with the result of this multiplication being a scalar. The vector space of real functions

whose domain is an closed interval $[a, b]$ with inner product $\langle f, g \rangle = \int_a^b f g \, dx.$ The dot product can be defined for two vectors $\mathbf{X}$ and $\mathbf{Y}$ by $\mathbf{X} \cdot \mathbf{Y} = |\mathbf{X}| |\mathbf{Y}| \cos \theta$, where θ is the angle between the vectors and $|\mathbf{X}|$ is the norm. It follows immediately that $\mathbf{X} \cdot \mathbf{Y} = 0$ if $\mathbf{X}$ is perpendicular to $\mathbf{Y}$ (Hazewinkle 2001). The dot product therefore would render the geometric interpretation in a explanatorial, geo-spatiotemporally, geosampled YFV-related, forecasting, risk model as the length of the projection of $\mathbf{X}$ onto the unit vector $\hat{\mathbf{Y}}$ when the two vectors are placed so that their tails coincide. By writing $A_x = A \cos \theta_A$, $B_x = B \cos \theta_B$, $A_y = A \sin \theta_A$, $B_y = B \sin \theta_B$, into the regression equation for qualitatively quantiating an empirical geo-spatiotemporally, geosampled, dataset of georeferencable, optimizable, YFV-related, explanatorial, clinical, field or remote-specified, LULC covariate, parameter estimators, it follows that the model residually forecasts would yield $\mathbf{A} \cdot \mathbf{B} = A B \cos (\theta_A - \theta_B) = A B (\cos \theta_A \cos \theta_B + \sin \theta_A \sin \theta_B)$ which then was equalivent to

$A \cos \theta_A B \cos \theta_B + A \sin \theta_A B \sin \theta_B = A_x B_x + A_y B_y$. So, in general, $\mathbf{X} \cdot \mathbf{Y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + \ldots + x_n y_n$. yi. This may be written succinctly employing Eistein summations in Wolfram Language.

In Jacob et al. 2012 a Possionian with a non-homogenopus gamma distributed mean was constructed using Wolfram language for regressively qualitatively quantitating district-level (see Jacob et al. 2012). Specific , district –level prevalence measures , however were forecasted using probablistic autorgressive matrix specifications. This model was written very succinctly using Einstein summation notation as $\mathbf{X} \cdot \mathbf{Y} = x_i y_i$. The dot product was implemented in the Wolfram Language as Dot[a, b], or simply by using a period, a . b. The dot product is commutative $\mathbf{X} \cdot \mathbf{Y} = \mathbf{Y} \cdot \mathbf{X}$, and distributive $\mathbf{X} \cdot (\mathbf{Y} + \mathbf{Z}) = \mathbf{X} \cdot \mathbf{Y} + \mathbf{X} \cdot \mathbf{Z}$. The associative property was meaningless for the dot product because $(\mathbf{a} \cdot \mathbf{b}) \cdot \mathbf{c}$ in the malarial model as is not defined since $\mathbf{a} \cdot \mathbf{b}$ is a scalar and therefore cannot itself be dotted. However, it does satisfy the property $(r\mathbf{X}) \cdot \mathbf{Y} = r(\mathbf{X} \cdot \mathbf{Y})$ for r a scalar. The derivative of a dot product of vectors is $\frac{d}{dt}[\mathbf{r}_1(t) \cdot \mathbf{r}_2(t)] = \mathbf{r}_1(t) \cdot \frac{d\mathbf{r}_2}{dt} + \frac{d\mathbf{r}_1}{dt} \cdot \mathbf{r}_2(t).$ Case as count were used as a response explanatorial, geo-spatiotemporal geopredictive variables (e.g., meterological data, distribution of health centers, etc) for remotely targeting districts that had higher prevalence rates from 2006 to 2010 in Uganda. Results from both a Poisson and a negative binomial (i.e., a Poisson random variable with a gamma distrusted mean) revealed that the covariates rendered from the model were significant, but furnished virtually no predictive power.

Importantly, the dot product is invariant under rotations $\mathbf{A}' \cdot \mathbf{B}' = A_i' B_i' = a_{ij} A_j a_{ik} B_k = (a_{ij} a_{ik}) A_j B_k = \delta_{jk} A_j B_k = A_j B_j = \mathbf{A} \cdot \mathbf{B}$, where Einstein summation has been used. The dot product is also called the scalar product (Cressie 1993). In the scalar product context, the probabilistically regressable values are usually written (a, b). By so doing, the dot product in a explanatorial, geo-spatiotemrpoal, geosampled, , YFV-related LUCL, forecasting, eco-epidemiological, risk model may be defined for tensors A and B by $A \cdot B \equiv A^\alpha B_\alpha$. Further, for four-vectors a_μ and b_μ, a YFV-related, forecasting, eco-epidemioloigical, risk model may be defined by $a_\mu \cdot b_\mu = a_\mu b^\mu = a^0 b^0 - a^1 b^1 - a^2 b^2 - a^3 b^3 = a^0 b^0 - \mathbf{a} \cdot \mathbf{b}$, where $\mathbf{a} \cdot \mathbf{b}$ is the usual three-dimensional dot product. More precisely, for a real vector space, an product $\langle \cdot, \cdot \rangle$ would satify the following four properties. Let u, v, and w be vectors and α be a scalar, then: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$, $\langle \alpha v, w \rangle = \alpha \langle v, w \rangle$, $\langle v, w \rangle = \langle w, v \rangle$, $\langle v, v \rangle \geq 0$ and equal if and only if $v = 0$

(Hazewinkle 2001). The fourth condition in the list above is known as the positive-definite condition. Related thereto, note that some authors define an inner product to be a function $(\cdot, \cdot)$ satisfying only the first three of the above conditions with the added (weaker) condition of being (weakly) non-degenerate (i.e., if $(v, w) = 0$ for all w, then $v = 0$) (See Griffith 2003). In such literature, functions satisyfing all the above mentioned conditions in a explanatorial, geo-spatiotemporal, geosampled, sylvatic, oviposition, sub-meter resolution, YFV-related, forecasting, eco-epidemiological, vulnerability , risk model would rely on positive –definite inner products (see Ratcliffe 2006). Unfortunately inner products in a YFV-risk model may be indefinite. These difference may be realized in a regression-based weighted matrix in an ArcGIS cyberenvironment. By so doing, the magnitude of psotive definite " norms" would be qualitatively quantiated which would then subsequently yield absolute covariate parameter estimators associated with time series prevelance, for example.

In some instance, the Lorentizan inner products may be an indefinite inner product in a explanatorial, clinical, field or remote geo-spatiotemporally, geosampled, hyperendemic, YFV-related, forecasting, risk model. The four-dimensional Lorentzian inner product is used as a tool in special relativity, namely as a measurement which is independent of reference frame and which replaces the typical Euclidean notion of distance. The standard Lorentzian inner product on $\mathbb{R}^4$ is given by $-dx_0^2 + dx_1^2 + dx_2^2 + dx_3^2$ i.e., for vectors $v = (v_0, v_1, v_2, v_3)$ and $w = (w_0, w_1, w_2, w_3)$, $(v, w) = -v_0 w_0 + v_1 w_1 + v_2 w_2 + v_3 w_3$. $\mathbb{R}^4$ endowed with the metric tensor induced by the Lorentzian inner product which is known as Minkowski space and is denoted $\mathbb{R}^{1,3}$ (Cressie 1993). The Lorentzian inner product on $\mathbb{R}^4$ is nothing more than a specific case of the more general Lorentzian inner product $(\cdot, \cdot)$ on n-dimensional Lorentzian space with metric signature $(1, n-1)$ (Griffith 2003). In this more general environment, the inner product of two vectors $x = (x_0, x_1, \ldots, x_{n-1})$ and $y = (y_0, y_1, \ldots, y_{n-1})$ may take on the form $(x, y) = -x_0 y_0 + x_1 y_1 + \cdots + x_{n-1} y_{n-1}$. when qualitatively quantitating an eco-epidemiological dataset of explanatorial, geo-spatiotemporally, geosampled, sylvatic YFV-related, time series dependnet, geopredictive, georeferencable, interpolatable covariate, parameterizable, regression, estimators. The Lorentzian inner product of two such vectors is sometimes denoted $x \circ y$ to avoid the possible confusion of the angled brackets with the standard Euclidean inner product (Ratcliffe 2006). Analogous presentations can also be made if the equivalent metric signature $(n-1, 1)$ (i.e., $(3, 1)$ for Minkowski space) is employed when qualitatively regressively quantiating an empirical operationizable, datset of geo-spatiotemporally geosampled , clinical, field or remote –specified covariate, paramter estimator coefficient values.

For a four-vector $x = (x_0, x_1, x_2, x_3)$ in Minkowski space, a explanatorial, dataset of geo-spatiotemporally, geosampled YFV-related, geoclassified,LULC, geopredictive variables x_1, x_2, and x_3 can be thought of as space variables with x_0 as the time variable. A four-dimensional pseudo-Euclidean space of signature (1,3) , suggested by H. Minkowski (1908) as a geometric model of space-time in the special theory of relativity (see Cressie 1993). Corresponding to each event there is a point of Minkowski space, three coordinates of which represent its coordinates in the three-dimensional space; the fourth coordinate is ct , where c is the velocity of light and t is the time of the event. The space-time relationship between two events is characterized by the so-called square interval: s 2 =c 2 (Δt) 2 −(Δx) 2 −(Δy) 2 −(Δz) 2 .

A vector space together with an inner product on it is called an inner product space. This definition also applies to an abstract vector space over any field in an explanatorial, geo-spatiotemporally, geosampled, YFV-related, geopredictive, eco-epidemiologically, forecasting, risk model. When given a complex vector space in a forecasting, risk model, the third property above is usually replaced by $(v, w) = \overline{(w, v)}$, where $\bar{z}$ refers to complex conjugation. This inner product property in a YFV risk model would be a Hermitian inner product and a complex vector space with a Hermitian inner product (i.e., a Hermitian inner product space). Every inner product space is a metric space(Cressie 1993). The metric would be given by $g(v, w) = (v - w, v - w)$. If this process results in a complete metric space in the YFV risk model it would be a Hilbert space. What's more, every inner product in the model naturally would induce a norm of the form $|x| = \sqrt{(x, x)}$, whereby it follows that every inner product space in the model probabilsically forecasted explanatorial, LULC-related clinical, field or remote geo-spatiotemporally, geosampled derivatives would also naturally be a normed space. Inner products in a explanatorial, YFV-related , geopredictive, eco-epidemiological, forecasting, risk model which fail to be positive-definite yield "metrics" - and hence, "norms" that may thus actually render something different due to the possibility of failing their respective positivity conditions. For example, n-dimensional Lorentzian Space (i.e., the inner product space consisting of $\mathbb{R}^n$ with the Lorentzian inner product) comes equipped with a metric tensor of the form $(ds)^2 = -dx_0^2 + dx_1^2 + \cdots + dx_{n-1}^2$ and a squared norm of the form $|v|^2 = -v_0^2 + v_1^2 + \cdots + v_{n-1}^2$ for all vectors $v = (v_0, v_1, \ldots, v_{n-1})$. In particular, one can have negative infinitesimal distances and squared norms, as well as nonzero vectors whose vector norm is always zero. As such, the metric (respectively, the norm) fails to actually be a metric (respectively, a norm) in the YFV-related risk model, though they usually are still called such when no confusion may arise.

Several generalizations of the Skorokhod space are worth mentioning. Instead of real-valued probabilsitically regresssable, explanatorial, operationizable, YFV-related, risk model functions on $[0, 1]$ it is possible to consider functions defined on $[0, \infty)$ and taking values in a metric space E. The space of cadlag functions obtained in this way is denoted by $\mathcal{D}_E[0, \infty)$ and if E is a Polish space, then $\mathcal{D}_E[0, \infty)$, with the appropriate topology, is also a Polish space, where these spaces are treated systematically. mathematical discipline of general topology, a Polish space is a separable completely metrizable topological space; that is, a space homeomorphic to a complete metric space that has a countable dense subset (Cressie 1993). Another generalization is obtained when the one-dimensional parameter t (often regarded as "time") is replaced by multi-variate variable $t = (t_1, \ldots, t_k)$. Let $[0, 1]^k$ denote unit cube in $\mathbb{R}^k$. It may be possible to introduce the space $\mathcal{D}[0, 1]^k$ of cadlag functions on $[0, 1]^k$ and a Skorokhod topology on it in a optimizable, explanatorily interpolatable, clinical, field or remote geo-spatiotemporally, geosampled, sylvatic, YFV-related, probabilistic,ovispoition, eco-epidemiological, risk model for rendering robust stochastically/detrministically interpolatable, covariate, parameter estimators.

The uniform rate of convergence in Donsker's theorem can be quantified by the result known as the Hungarian embedding: $\limsup_{n \to \infty} \frac{\sqrt{n}}{\ln^2 n} \left\| \sqrt{n}(\hat{F}_n - F) - G_{F,n} \right\|_\infty < \infty,$ in a -spatiotemporally, geosampled, sylvatic, YFV-related, probabilistic, risk model. In theory of probability, the Komlós–Major–Tusnády approximation (also known as the KMT

approximation, the KMT embedding, or the Hungarian embedding) is an approximation of the empirical process by a Gaussian process constructed on the same probability space (Cressie 1993). A probability space in a geosampled, geo-spatiotemproal, entomological-related, explanatorial, endemic, transmission-oriented, forecasting, probabilistic, risk model consists of three parts: 1) A sample space, Ω, which is the set of all possible outcomes.2)A set of events $\mathcal{F}$, where each event is a set containing zero or more outcomes (predicted geolocations of prolific immature habitats);and, 3)The assignment of probabilities to the events; that is, a function P from events to probabilities (e.g., tillering LULC agroecosystem to highly productive immature *Ae egypti* habitats)

In mathematics, fuzzy measure theory considers generalized measures in which the additive property is replaced by the weaker property of monotonicity. Let $\mathbf{X}$ be a universe of discourse, $\mathcal{C}$ be a class of subsets of $\mathbf{x}$, and $E, F \in \mathcal{C}$. A function $g : \mathcal{C} \to \mathbb{R}$ where 1) $\emptyset \in \mathcal{C} \Rightarrow g(\emptyset) = 0$ and 2) $E \subseteq F \Rightarrow g(E) \leq g(F)$ is called a fuzzy measure. A fuzzy measure is called normalized or regular if $g(\mathbf{X}) = 1$. The central concept of fuzzy measure theory is the fuzzy measure which was introduced by Choquet in 1953 and independently defined by Sugeno in 1974 in the context of fuzzy integrals. In mathematics, the Sugeno integral, named after M. Sugeno, is a type of integral with respect to a fuzzy measure.Let (X, Ω) be a measurable space and let $h : X \to [0, 1]$ be an Ω-measurable function in an explanatorial, clinical, field or remote- geosclassifed, LULC-oriented, endemic, transmission-related, geo-spatiotemporally, geosampled, YFV, probabilistic, risk model .The Sugeno integral over the crisp may be then set $A \subseteq X$ of the function h with respect to the fuzzy measure g which may be defined by:

$$\int_A h(x) \circ g = \sup_{E \subseteq X} \left[\min \left(\min_{x \in E} h(x), g(A \cap E) \right) \right] = \sup_{\alpha \in [0,1]} \min \left(\alpha, g(A \cap F_\alpha) \right)$$

where $F_\alpha = \{x | h(x) \geq \alpha\}$.The Sugeno integral over the fuzzy set $\tilde{A}$ of the function h with respect to the fuzzy measure g may then be defined by:

$$\int_A h(x) \circ g = \int_X [h_A(x) \wedge h(x)] \circ g$$

where $h_A(x)$ is the membership function of the fuzzy set $\tilde{A}$.There exists a number of different classes of fuzzy measures including plausibility/belief measures; possibility/necessity measures; and probability measures which are a subset of classical measures(Hazewinkle 2001).

Alternatively, the rate of convergence of $\sqrt{n}(\hat{F}_n - F)$ can also be quantified in terms of the asymptotic behavior of the sup-norm of this expression in an explanatorial, clinical, field or remote geosampled, probabilistically, empricaly regressable, geo-spatiotemporally, geosampled, YFV-related, probabilistic, risk model. Number of results exist in this venue, for example the Dvoretzky–Kiefer–Wolfowitz inequality can provide bound on the tail probabilities of $\sqrt{n}\|\hat{F}_n - F\|_\infty$: $\Pr\left(\sqrt{n}\|\hat{F}_n - F\|_\infty > z \right) \leq 2e^{-2z^2}$ in an empirical dataset of eco-epidemiological, time series dependent, forecastable, explanatorial, interpolatable, clinical, field or remote geosampled, YFV-related variables. In the theory of probability and statistics, the Dvoretzky–Kiefer–Wolfowitz inequality predicts how close an empirically determined distribution function will be to the distribution function from which the empirical samples are drawn. It is named after Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz, who in 1956 proved[the inequality with an unspecified multiplicative constant C in front of the exponent on the right-hand side. In 1990, Pascal Massart proved the inequality with the sharp constant $C = 1$, confirming a conjecture due to Birnbaum and McCarty. Given a natural number n, let X_1, X_2, …, X_n be real-valued

independent and identically distributed , probabilistically regressable, YFV-related, empirical dataset of LULC explanators with distribution function $F(\cdot)$. Let F_n denote the associated empirical distribution function defined by $$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le x\}}, \qquad x \in \mathbb{R}.$$ in the risk model probabistic, risk-related output.The Dvoretzky–Kiefer–Wolfowitz inequality would then bound the probability that the random function since F_n would differ from F by more than a given constant $\varepsilon > 0$ anywhere in the forecasted derivative dataset. More precisely, there is the one-sided estimate $\Pr\left(\sup_x (F_n(x) - F(x)) > \varepsilon\right) \le e^{-2n\varepsilon^2}$ where every $\varepsilon > \sqrt{\frac{1}{2n}\ln 2}$ also implies a two-sided probabilistic estimate $\Pr\left(\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| > \varepsilon\right) \le 2e^{-2n\varepsilon^2}$ for every $\varepsilon > 0$. The explanatorial, geo-spatiotemporally or geo-spatiotemporally geosampled, sylvatic YFV-related, probabilistic, geopredictive, eco-epidemiological, risk model would be based on the Glivenko–Cantelli theorem by quantifying the rate of convergence as n tends to infinity in the risk model derivatives. By so doing, the model resdiual probabilty estimates may be also regressively tabulates by employing the tail probabilities of the Kolmogorov–Smirnov statistic.

The Kolmogorov has shown that if the cdf F is continuous, then the expression $\sqrt{n}\|\hat{F}_n - F\|_\infty$ converges in distribution to $\|B\|_\infty$, which has the Kolmogorov distribution that does not depend on the form of F.Another result, which follows from the law of the iterated logarithm, is that $\limsup_{n\to\infty} \frac{\sqrt{n}\|\hat{F}_n - F\|_\infty}{\sqrt{2\ln\ln n}} \le \frac{1}{2},$ and $\liminf_{n\to\infty} \sqrt{2n\ln\ln n}\|\hat{F}_n - F\|_\infty = \frac{\pi}{2},$ These inequalities follow from the case where F corresponds to be the uniform distribution on [0,1] in view of the fact that F_n may have the same explanatorial, probabilistic, eco-epidemiological, time series dependent, forecastable, explanatorily interpolatable, LULC-related, clinical, field or remote geosampled, YFV-related regressed distributions as $G_n(F)$ where G_n is the empirical distribution of U_1, U_2, …, U_n, where there are independent and Uniform(0,1)s, and where $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \stackrel{d}{=} \sup_{x\in\mathbb{R}} |G_n(F(x)) - F(x)| \le \sup_{0 \le t \le 1} |G_n(t) - t|,$ with equality can occur in the model outputs but if and only if F is continuous.

Bootstrapping is the practice of estimating properties of a probabilistic explanatorial, geo-spatiotemroral, predictor (such as its variance) by measuring those properties when sampling from an approximating distribution. One standard choice for approximating an explanatorial explanatorial, clincial, field or remotegeospatiotemrpoally geosampled, YFV-related, stochastically/deterministically interpolated LULC distribution is the empirical distribution function for. If a set of probabilistic, sylvatic,YFV-related observations is detremined to be to be from an independent and identically distributed population, a number of resamples with replacement of the observed dataset (and of equal size to the observed dataset) may be validated in SAS .These data may also be used for constructing hypothesis tests or, as an alternative to statistical inference based on the assumption of a parametric model especially if any model assumptions are in doubt, or where parametric inference is impossible or requires complicated formulas for the calculation of standard errors.

Thereafter, by derivation of the moment conditions and optimal weighting matrix without distributional assumptions for the GM estimation procedure of the selected probabilsic geosampled, autoregressive LULC-related, explanatorial, clinical, field or remote geoclassifed,

424

YFV, sub-meter resolution, covariate parameter estimators in the Gulu eco-epidemiological study site, we regressively quantitated the disturbance process while defining a generalized two-stage least square estimator for the seasonal geosampled, YFV-related, regression parameters. We proved consistency of a GM estimator and provided Monte Carlo evidence for geospatially adjusting the empirically probabilistically regressed dataset of the explanatorial, clinical, field or remote geoclassified, georeferenced explanatorial, geopredictive variables of seasonal-sampled YFV-related covariate, parameter estimators in the agro-village complex epidemiological study site . We noted that the asymptotic distribution of the probabilistic, residualized YFV-related time series autocovariate matrices from a fitted multivariate, geopredictive, explanatorial, reduced, rank a model and a portmanteau test was extendable to a reduced rank model.

Traditionally, hierarchical generalized, empirically regressable, time serie dependent, functional data are widely seen in complex YFV-related studies where subunits are nested within units, which in turn are nested within treatment groups. A generalized framework of functional mixed effects explanatorial, geo-spatiotemporal YFV-related , explanatorial, clinical, field or remote, geoclassified, LULC forecastable, covariate parameter estimator may be generated but the model derivatives may be serially correlated. Alternatively,Penalized splines may be employed to model the functions of functional mixed effects model for such data where within-unit and within-subunit variations may be modeled through separate sets of principal components. By so doing, the subunit level explanatorial geo-spatiotemporal, time series functions may be allowed to be correlated. Penalized splines are used to model both the mean functions and the principal components functions, where roughness penalties are used to regularize the spline fit(Cressie 1993).

An expectation maximization (EM) algorithm may be also developed to fit a explanatorial, geo-spatiotemporal or geo-spectrotemporal, eco-epidemiologically geosampled, sylvatic YFV-related model, clinical, field or remote covariate, parameter estimators, while the specific covariance structure of the model may be utilized for computational efficiency to avoid storage and inversion of large matrices. Our dimension reduction with principal components provides an effective solution to the difficult tasks of modeling the covariance kernel of a random function and modeling the correlation between functions. The proposed methodology may be illustrated in SAS using simulations and an empirical dataset from any clincial, field or remote specified, YFV-related LULC study

The explantorial, clinical, field and remote geo-spatiotemrpoally , geosampled, sylvatic YFV-related risk eco-epidemiological LULC, African, expanding, Riceland, agro-irrigated model adequacy was based on the residual autocovariance generated from the filtering exercises. Thereafter, a convenient asymptotic approximation for geoprediction of the statically important, empirical probabilistically dataset of the regressed, YFV-related, covariate, parameter estimators, regression coefficients was determined using a mean square error matrix. This model employed the reduced rank forest canopy to agro-village complex, LULC-related, geopredictive, probabilsitic estimators for calculating the higher-order autoregressive functions. These estimators were then re-computed employing variance-standardized, normalized covariances for quantitating the observed time series dependent, YFV-related uncertainty probabilities in a structural-equation model fit.

Interestingly for the time series explanatorial, YFV-related, risk based analysis, two well-known versions of a portmanteau test were employed for testing for latent autocorrelation in ArcGIS. These tests were based on whether any geocalssified LULC group of autocorrelations of residual time series were different from zero in the geo-spatiotemrpoally, geosampled, eco-epidemiological empirical dataset of forecasted derivatives. The Ljung–Box test, which is an improved version of the Box–Pierce test, having been devised at essentially the same time which may identify seemingly trivial simplifications omitted in any optimizing algorithm (www.esri.com). The Ljung–Box test is a type of statistical test of whether any of a group of autocorrelations of a explanatory time series are different from zero(Box and Jenkins 1985). Instead of testing randomness at each distinct lag, the test evaluated the "overall" randomness based on a number of lags, and was therefore a portmanteau test .

The Ljung–Box test in ArcGIS was defined as follows for constructing the seasonal, robust, YFV-related, explanatorial, probabilistic, eco-epidemiological, risk model where H_0 was the geosampled, YFV-related, data which were the independently distributed (i.e. the correlations in the population from which the sample is taken are 0, so that any observed YFR-related correlations in the data result from randomness of the sampling process). H_a then was the data not independently distributed. The test statistic was then $Q = n\,(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$ where n was the seasonal empirical sampled, sylvatic, YFV-related parameter estimator dataset size, $\hat{\rho}_k$ was the sample autocorrelation at lag k, and h was the number of lags being tested. For significance level α, the critical region for rejection of the hypothesis of randomness then in our seasonal geopredictive, endemic, transmission-oriented, YFV-related, eco-epidemiological, risk model was $Q > \chi^2_{1-\alpha,h}$ where $\chi^2_{1-\alpha,h}$ was the α-quantile of the chi-squared distribution with h degrees of freedom.

We employed the Ljung–Box test for constructing a robust ARIMA employing the geosampled, clinical, field and remote-geosampled YFV-related explanatory model parameter estimator employing the empirical sampled field and remote covariates. Note that it was applied to the residuals of a fitted ARIMA model, not the original series, thus in our applications the hypothesis actually being tested was that the time series YFV-related model residuals from the ARIMA model that had no latent autocorrelation coefficients. When testing the residuals of an estimated ARIMA model, the degrees of freedom need to be adjusted to reflect the parameter estimation. For example, for an ARIMA(p,0,q) model, the degrees of freedom should be set to $m - p - q$ (see Box and Jenkins 1985).The Box-Pierce test uses the test statistic which was given , by $Q_{\mathrm{BP}} = n \sum_{k=1}^{h} \hat{\rho}_k^2,$ which used the critical regions as defined above. Our simulation ARIMA study showed that the Ljung–Box statistic can model explanatory geopredictive YFR-related time series data.

To model a series of monthly counts of new seasonal, YRV episodes in Gulu, GSARIMA models and GARIMA models with a deterministic seasonality component were developed. GSARIMA and GARIMA models are an extension of the class of GARMA models and are suitable of parsimonious, forecast, risk modeling non-stationarity in seasonal time series count data with binomial conditional distributions(see Cressie 1993). The seasonal, YFV-related, time

series dependent, probabilistic, explanatorial, risk models were presented with a choice of identity link function or logarithmic link functions. A choice between two transformation methods were then applied to deal with the zero-valued, explanatorial, geopredictive, time series dependent, probabilistic, sylvatic, YFV-related, covariate, parameter estimator in the explanatorial, clinical, field and remote-geosampled, time series, empirical dataset using a threshold parameter. When a seasonal vector, arthropod-related, explanatorial, geopredictive, eco-epidemiological, risk model, count variable, time series has many endemic transmission-oriented, risk-based observations of zero, both transformation methods and several threshold parameters should be explored in order to find the best fitting model (Jacob et. al.2005b).

Bayesian GSARIMA and GARIMA models were then applied to the seasonal, YFV-related, case count, time series data geosampled at the Gulu epidemiological study site. GARMA models related the time series geosampled, YFV-related, explanatorial, observational, clinical, field and remote observational predictors and ARMA components to a transformation of the mean parameter of the data distribution (λ_t), via a link function. In this research, a log link function ensured that λ_t was constrained to the domain of positive geosampled, endemic, transmission-oriented, explanatorial, covariate, parameter estimator, coefficient values. Both a GSARIMA and a GARIMA model with a deterministic seasonality component were then selected, based on different criteria. The sylvatic YFV-related GARIMA model with deterministic seasonality showed a lower DIC, but the YFV-related GSARIMA model had a lower mean absolute relative error and required fewer clinical, field, or remote geosampled, coavariate, parameter estimators, for enabling the quantitation process.

Further, Bayesian modeling allowed for the analysis of the posterior geopredictive seasonal explanatorial, YFV-related, probabilistic distributions. The performance of the selected negative binomial model was then compared with that of a Gaussian version of the model on Box-Cox transformed field and remote-sampled data. These distributions did not perfectly mirror the distribution of the residual forecasted derivatives for either model. This may have been an indication that the assumptions about the underlying seasonal geopredictive explanatorial, YFV-related risk-based, geoclassified, sylvatic, LULC distributions were not entirely appropriate for either endemic, transmission-oriented, probabilistic, eco-epidemiological, risk model. However, analysis of the regression, YFV-related, geo-spatiotemporal residuals showed that the posterior geopredictive distributions were much better for the negative binomial GSARIMA geopredictive, YFV-related, explanatorial, geo-spatiotemporal, LULC-related, eco-epidemiological, risk model than for its Gaussian version on the transformed data when counts were low. Both models accounted for latent autocorrelation in the geosampled explanatorial, endemic, transmission-oriented, risk-related, YFV-related, clincial, field or remote data, but the negative binomial model had an 6% better MARE than the Gaussian version on transformed data (0.314 *vs.* 0.478).

For many explanatorial, YFV-related, probabilistic, linear models, a MLE can be found as an explicit function of the observed data $x_1, ..., x_n$ (e.g., geosampled YFV-related clinical, field or remote variables). For many other models, however, no closed-form solution to the maximization problem is known or available, and an MLE has to be found numerically using optimization methods in SAS. For some problems, there may be multiple estimates that maximize the likelihood. For other problems, no maximum likelihood estimate exists (meaning that the log-likelihood function increases without attaining the supremum value).

Here, the MLE coincided with the most probable Bayesian estimator given a uniform prior distribution LULC-related , geo-spatiotemporal, geosampled, covariate, paramter estimators estimators Indeed, the maximum a posteriori estimate was the parameter θ that maximized the probability of θ given the geosampled, explanatorial,clinical, field or remote data as defined by the Bayes' theorem: $P(\theta|x_1, x_2, \ldots, x_n) = \dfrac{f(x_1, x_2, \ldots, x_n|\theta)P(\theta)}{P(x_1, x_2, \ldots, x_n)}$ where $P(\theta)$ was the prior distribution for the parameter θ and where $P(x_1, x_2, \ldots, x_n)$ was the probability of the geosampled clincial, field or remote geosampled data averaged over all parameters. Since the denominator was independent of θ, the Bayesian estimator wasobtained by maximizing $f(x_1, x_2, \ldots, x_n|\theta)P(\theta)$ with respect to θ. We assumed that the prior $P(\theta)$ was a uniform distribution. The Bayesian estimator was obtained by maximizing the likelihood function [i.e., $f(x_1, x_2, \ldots, x_n|\theta)$]. Thus the Bayesian estimator coincided with the MLE for a uniformed, YFV-related, explanatorial, geo-spatiotemporal, prior distribution $P(\theta)$. We found that the MLE was an extreme sylvatic, YFV-rerlated, probabilistic, geoclassified sub-meter resolution, eco-georeferenceable, LULC estimator which was obtained by maximizing a function of θ, where the objective function (c.f., the loss function) was quantiated as $\hat{\ell}(\theta|x) = \dfrac{1}{n}\sum_{i=1}^{n}\ln f(x_i|\theta),$ and where the sample analogue of the expected log-likelihood was quantiated by $\ell(\theta) = \mathbb{E}[\ln f(x_i|\theta)]$, where this expectation was taken with respect to the true LULC density [i.e., $f(\cdot|\theta_0)$].

Maximum-likelihood estimators have no optimum properties for finite samples, in the sense that when evaluated on finite samples, other estimators have greater concentration around the true parameter-value(Cressei 1993) However, like other estimation methods, ML estimation possess a number of attractive limiting properties for robust, geo-spatiotemporal, YFV-related explanatorial, forecast, eco-epidemiological, vulnerability, risk modeling.For example, the sample size of an empirical geosampled, eco-epidemiological explanatorial, clinical, field or remote geosampled, geoclassified, LULC dataset may increase to infinity, where sequences of MLEs have these properties:1)Consistency whereby the sequence of MLEs converges in probability to the value being estimated;and, 2)Asymptotic normality: as the sample size increases, the distribution of the MLE tends to the Gaussian distribution with mean θ and covariance matrix equal to the inverse of the Fisher information matrix.Efficiency, (i.e., it achieves the Cramér–Rao lower bound when the YFV-related sample size tends to infinity). This means that no consistent, covariate parameter estimator has lower asymptotic mean squared error than the MLE (or other geosampled, explanatorial, clinical, field or remote estimators attaining this bound) in any geo-spatiotemrpoally, geosampled LULC-related YFV, , forecasting, regression-related probabilsitic, geo-spatiotemporal, eco-epidemiological, risk model.Second-order efficiency may be applicable to post correction bias quantitation in the forecst derivatives. Since the MLE is consistent regardless, having a sufficiently large number of explanatorial, clinical, field or remote geosampled, sylvatic, YFV-related observations n, it is possible to find the value of θ_0 with arbitrary precision. This means that as n goes to infinity a geo-spatiotemporally, geosampled, probabilistic,explanatorial, clinical, field or remote-specified,YFV-related, geoclassified LULC covariate, paramter estimator [i.e., $\tilde{\theta}$] will converge in probability to its true value:$\tilde{\theta}_{\text{mle}} \xrightarrow{p} \theta_0$.

The fact that the cumulative distribution functions did not perfectly match the diagonal indicated in the YFV-related LULCeco-epidemiolgical, forecasting risk model suggested that there is room for improvement, through modeling a more complex latent autocorrelation structure through time varying SARIMA parameters and through the inclusion of more seasonal-sampled covariate, parameter estimators. It is also however possible that assuming an underlying negative binomial distribution is not entirely appropriate for YFV-related LULC risk modelling. In the latter case, the DIC, would be based on the assumption which may have a less value than the MARE for robustly comparing explanatorial, geopredictive, YFV-related, endemic, transmission-oriented, risk model, residually forecasted derivatives. Fortunately, MARE does not depend on the assumption of a true underlying distribution, thus it would be easier for implementing IVM.

Importantly, lagged, YFV-related, probabilsic observations employed as seasonal-geosampled, geoclassified, LULC covariate parameter estimators in PROC ARIMA were logarithmically transformed, which unfortunately was not possible for observations with a value of zero. To circumvent this problem, we added a small constant to the geosampled, clinical, field and remote geo-spatiotemporal, geoclassified, YFV data for the zero-valued, probabilistic, parameter estimator, covariate coefficients. Grunwald and colleagues (2000) considered a conditional linear autoregressive (CLAR) model with an identity link function. In order to ensure a positive λ_t, restrictions we used the explanatorial, geopredictive, endemic, transmission-oriented, risk-related, regression-based, YFV-related,LULC, covariate, parameter estimators as a variant of the GARMA model. We then constructed a robust, generalized, linear, autoregressive, moving average model.

Heinen (2003) proposed a class of explanatorial, predictive, autoregressive, conditional Poisson (ACP) models which may allow for over and under dispersion in the marginal distribution of seasonal, geosampled explanatorial, geopredictive, YFV-related, clinical, field and remote-specified, geosampled empirical data. Another class of Poissonized probability models with autocorrelated error structure that uses "binomial thinning are called integer-valued autoregressive (INAR) models". INAR models may be theoretically extended to moving average (INMA) and INARMA seasonal, probabilsitic,empriically regressable, YFV-related, time series dependnet, explanatorial, geopredictive, risk models in ArcGIS.

An alternative, seasonal, YFV-related, parameter-driven, ArcGIS-related, probabilsitic, risk modeling approach may assume an autoregressive process on time specific random effects as introduced in the mean structure, using a logarithmic link function. Such a model is sometimes called a stochastic autoregressive mean (SAM) model and has frequently been applied in Bayesian geo-spatiotemporal, hierarchical, generalized, probabilsic modeling. This model can compare various regression models for time series of counts which can account for discreetness, overdispersion and serial correlation. Besides observation- and covariate parameter estimator-driven YFV-related,probabilsitic, eco-epidemiological, risk models based upon corresponding conditional Poissonian distributions, a ecologist or data analyst may also consider a dynamic ordered probit model as a flexible specification to capture the salient explanatorial georferncable, explanatorial, clinical, field or remote geosampled, eco-epidmiological, data, feature attributes based on time series of counts. For all sylvatic YFV-related, LULC, forecast, vulnerability,risk models, appropriate efficient estimation procedures may be generated in

SAS(e.g., PROC MCMC). For covariate parameter estimator-driven, YFV-related LULC specifications this requires Monte Carlo procedures like simulated Maximum likelihood or Markov Chain Monte-Carlo. The methods including corresponding diagnostic tests may then be illustrated with, YFV-related G(S)ARIMA, geoclassified LULC, risk models. The residually forecasted derivatives may be particularly useful in the drive towards YFV-related elimination, but could also be applied to other vector arthropod-related diseases. Although building and fitting Bayesianized,geo-spatiotemproal or geo-spectrotemporal, eco-epidemiological, vulnerability, endmeic, YFV-related explanatorial, eco-epidemiological, forecast-oriented, GSARIMA-related, endemic transmission-oriented, geoclassified LULC, risk models is laborious, they may provide more realistic distributions when quantitating time series of LULC counts than do Gaussian methods on transformed data, especially when counts are low.

We then employed a binary, geographic weighted, geo-spatiotemproal matrix based on a planar graph for geo-spatiotemporally quantitating the time series dependent, YFV-related, explanatorial, endemic, transmission-oriented, geopredictor variables. Seasonal explanatorial, geopredictive, sylvatic YFV-related, spatial autocorrelation devices were constructed from geographic weights matrices in ArcGIS to capture the covariation among the geosampled, endemic, transmission-oriented, covariate LULC-derived, parameter estimator, coefficient values of one or more random variables that were associated with the configuration of the LULC areal units. We generated a Moran's scatterplot in ArcGIS employing the time series LULC changes which were then transformed to robust explanatorial, geopredictive, Cartesian coordinates for representing the forest canopy to agro-village complex LULC changes based on geo-spatiotemporal changes in ,precipitation and human population statistics at the Gulu epidemiological study site to display probabilistically regressed values for the ecogeoreferenced, eco-epidemiological, wavelength, frequency, YFV-related, geoclassified LULC, data feature attributes. We employed a standardized versus summed nearby standardized values format in ArcGIS whose associated bivariate regression slope coefficient was used to generate the unstandardized Moran's Coefficient (MC).

Thereafter, we employed a datset of orthogonal spatial filter eigenvectors in ArcGIS to determine if the time series explanatorial, geopredictive, YFV-related, endemic transmission-oriented, LULC-specified, data, feature attributes represented positive or negative autocorrelation. Autocorrelation can quantify time series cluster-based varying and constant explanatory georefernced geopredictive covariates (Griffith 2003). We quantitated the residual heteroskedastic, geoclassified, LULC geoparameters in the spatial filter hierarchical, autoregressive, sylvatic YFV cluster-based, endemic, transmission-oriented, eco-epidemiological, geoclassified LULC, risk-related forecasting African, ricland agro-irrgated village models. Heteroskedascity occurs when the standard deviations of a sampled variable monitored over a specific amount of time is non-constant (Hosmer and Leneshew 2000). Non-quantitation of latent, heteroskedatic, explanatorial, clinical, field or remote-specified parameters in a time series explanatorial, geopredictive, vector insect larval habitat, geoclassified LULC, distribution model can propagate errorresdiuals (Jacob et al. 2009d, Jacob et al. 2010b).

Modeling non-constant variance, or heteroscedasticity in a seasonal, YFV-related explanatorial, LULC, geopredictive, endemic, transmission-oriented, risk model can improve the efficiency of estimates of the covariate, parameter estimators associated with the mean of a series

while providing insight into the volatility of the series. SAS/ETS software provides capability to do linear and nonlinear time series dependent, YFV-related, geoclassified, LULC, explanatorial regression with heteroskedatic models using the AUTOREG and MODEL procedures, respectively. This example makes use of the MODEL procedure. One of the key assumptions of regression analysis is that the variance of the errors is constant across observations (Hosmer and Lemeshew 2000). Unfortunately, this assumption is often violated when forecast, eco-epidemiological, geo-spatiotemporal,risk modeling time series or panel-related, seasonal, vector arthropod-related, time series dependent, geoclassified, LULC, endemic, transmission-oriented, clinical, field and remote geosampled data, resulting in inefficient covariate, parameter estimatators and inaccurate forecast error variance.

$$
\begin{aligned}
g(y_t, x_t, \theta) &= \varepsilon_t \\
\varepsilon_t &= H_t \times \xi_t \\
H_t &= \begin{bmatrix} \sqrt{h_{t,1}} & 0 & \cdots & 0 \\ 0 & \sqrt{h_{t,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{h_{t,s}} \end{bmatrix} \\
h_t &= g(y_t, x_t, \phi)
\end{aligned}
$$

Consider the following generalized, model: where

$\xi_t \sim N(0, \Sigma)$.

Suppose a seasonal, sylvatic, time series, YFV-related, endemic, transmission-oriented, prognosticative, explanatorial, time series dependent, eco-epidemiological, geoclassified LULC, risk model is homoscedastic, [i.e.,h(t)=1]. If a probabilistic, geo-spatiotemporal, geosampled, YFV-related risk model is heteroscedastic with known form a ecologist or data analyst can improve the efficiency of the geosampled geoparameter probabilistically regressed, estimates by performing a weighted regression.

The weight variable, using this notation, is $\dfrac{1}{\sqrt{h_t}}$ if the residually forecasted errors for an explanatorial, robust, geopredictive, YFV-related, geoclassified LULC, risk model are heteroscedastic and the functional form of the variance is known, the model for the variance can be estimated along with the regression function. Thereafter, to specify a functional form for the variance, a ecologist, entomologist or data analyst may assign the function an H.var variable where var is the equation variable in AUTOREG. Thus, for instance, if the researchist desires to estimate the scale parameter for the variance of a simple OLS time series dependent, explanatorial, predictive, vulnarability, sylvatic YFV endemic transmission-oriented, probabilistic, regression-based, risk model $y = a + b * x$ can be specified in SAS as:

```
proc model data=s;
   y = a + b * x;
   h.y = sigma**2;
 fit y;
```

In so doing, the above YFV-related, predictive, risk model would have a constant variance, sigma**2. PROC MODEL estimates the constant-variance model by default; therefore the h.y equation is unnecessary in this case. Thereafter, a ecologist, entomologist or researchist may consider the same model with the following functional form for the variance: $h_t = \sigma^2 x^{\alpha}$ This seasonal, YFV-related, probabilsitic, geo-spatiotemporal geosampled, forecast, eco-epidemiological, geoclassified LULC, risk model may be written as

```
proc model data=s;
  y = a + b * x;
  h.y = sigma**2 * x**alpha;
fit y;
```

In addition to estimating the parameters, a and b, the MODEL procedure may also estimates the time series dependent, YFV-related parameters sigma and alpha, which may then be determined to be associated with the error variance.

The following example may be then employed for time series dependent, YFV-related, probabilistic, geoclassified LULC, data creation from an average of daily rainfall figures from January 1961 to March 2000. A simple, commonly used rate model is the Vasicek model:

$$rate_t = rate_{t-1} + \kappa \times (\theta - rate_{t-1}) + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

The following PROC MODEL statements may then be employed to fit a geo-spatiotemporal, geosampled, explanatorial, clinical, field or remote geosampled, YFV-related, endemic, probabilistic, geoclassified LULC, risk model as:

```
proc model data=weeklyraindall rates;
  ffrate = lag(ffrate) + kappa  * (theta - lag(ffrate));
  lag_ffrate = lag( ffrate );
  label kappa = "Speed of Mean Reversion";
  label theta = "Long term Mean";
fit ffrate / fiml breusch=( lag_ffrate ) out=resid outresid;
run;
```

The test for heteroscedasticity then would be positive. To correct for the heteroscedasticity, a researchist could use a time series dependent, geo-spatiotemproal, variance model. Another popular model for this type of data is the Cox Ingersoll Ross model, which could account for the heteroscedastic, *Ae aegypti*, oviposition, sub-meter resolutiuon, LULC, covariate parameter estimator employing $$rate_t = rate_{t-1} + \kappa \times (\theta - rate_{t-1}) + \epsilon_t \quad \epsilon_t \sim N(0, \sigma^2 rate_{t-1})$$

The following PROC MODEL statements could then be employed to fit this probilistic, eco-epidemiological,YFV-related, endemic, transmission-oriented, geoclassified LULC, risk model:

```
proc model data=weekly rainfall rates;
  ffrate = lag(ffrate) + kappa * (theta - lag(ffrate));
  h.ffrate = sigma**2 * lag(ffrate);
```

```
   label kappa = "Speed of Mean Reversion";
   label theta = "Long term Mean";
   label sigma ="Constant part of variance";
 fit ffrate / fiml out=resid outresid;
run;
```

Interestingly, the dependency on the lag of FFRATE has been removed, by SAS but the residuals would still remain heteroscedastic. A GARCH (1,1) related YFV endemic transmission-oriented risk-related, geoclassified LULC, forecasting, risk model may then be considered using

$$rate_t = rate_{t-1} + \kappa \times (\theta - rate_{t-1}) + \varepsilon_t; \quad h_t = arch_0 + arch_1 \times \varepsilon_{t-1}^2 + garch_1 \times h_{t-1}$$

$$\varepsilon_t \sim N(0, h_t)$$

. Residuals from a Cox Ingersoll Ross YFV-related, explanatorial, geopredictive, regression-based, eco-epidemiological, probabilistic, geoclassified LULC, risk model could then be constructed employing the PROC MODEL statements as by fit a GARCH(1,1) model:

```
 proc model data=weekly rainfall rates;
   ffrate = lag(ffrate) + kappa * (theta - lag(ffrate));
   if ( _OBS_ = 1 ) then
   h.ffrate = arch0 + arch1 * mse.ffrate + garch1 * mse.ffrate;
   else
   h.ffrate = arch0 + arch1 * zlag(resid.ffrate**2)
         + garch1 * zlag(h.ffrate);
 fit ffrate / fiml out=resid outresid;
run;
```

The residuals from this estimation may then be employed with other clinical, field or remote geosampled covariate parameter estimators(e.g., time series forest canopy to agro-village complex LULC prognosticators) in order to generate unbiased stochastically and/or deterministically explanatorily interpotable covariate,parameter estimators in the forecasted geospatialized, sylvatic YFV, clinical, field or remote-specified, residualized, ovispoition, sub-meter resolution, propogagtional, probabilsitic derivatives.

The linearized probabilsitic, regression seasonal, explanatorial, geopredictive, YFV-related, endemic, transmission-oriented, geoclassified LULC, forecasting, eco-epidemiological, risk model $Y = X'\beta + U$, where X is the vector of explanatorial, geopredictive, endemic, transmission-oriented, risk related, clinical, field or remote-specified, geopredictive variables and β was a $k \times 1$ column vector of geoparameters to be estimated (e.g., NDVI measured time series ArcGIS quantitated, covariate, parameter estimator, geoclassified, LULC coefficients) .

We employed the OLS estimator $\tilde{\beta}_{OLS} = (X'X)^{-1}X'Y$, where X denoted the matrix of stacked X_i' explanatorial, georeferncable, geoclassified LULC, clinical, field and remote geosampled time series dependent, explanatorial, sylvatic YFV-related geoclassified, covariate, paramter estimator, coefficient values observed in the data geosampled at the Gulu

epidemiological study site. Since the sample errors had equal variance σ^2 and were uncorrelated, then the least-squares estimate of β was the best linear unbiased estimator and its variance was easily estimated with $v_{OLS}[\hat{\beta}_{OLS}] = s^2(X'X)^{-1}, s^2 = \dfrac{\sum_i \hat{u}_i^2}{n-k}$ where $\hat{u}_i$ was the probabilistic, regression residuals. When the assumptions of $E[uu'] = \sigma^2 I_n$ are violated, the OLS estimator loses its desirable properties (Hosmer and Lemeshew 2000). Indeed, $V[\hat{\beta}_{OLS}] = V[(X'X)^{-1}X'Y] = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$ where $\Sigma = V[u]$ appeared in the forecasted clinical, field or remote, geopredictive variables. While the OLS point estimator remained unbiased in our seasonal, geopredictive, probabilsitic, YFV-related, endemic, transmission-oriented, geoclassified LULC, risk model, it may not have been the optimal in the sense of having minimum mean square error in the residually forecasted derivatives. The OLS variance estimator $v_{OLS}[\hat{\beta}_{OLS}]$ does not provide a consistent estimate of the variance of the OLS estimates (Greene 1998).

Since the probabilistically quantiated regression errors in our time series dependent, explanatorial, geopredictive, probabilsitic, YFV-related, endemic, transmission-oriented, geoclassified, LULC, risk model u_i were independent, but had distinct variances σ_i^2, we assumed that $\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ may be optimally estimated with $\hat{\sigma}_i^2 = \hat{u}_i^2$. This same assumption provided White's (1980) estimator, often referred to as heteroscedasticity-consistent estimator): $v_{HCE}[\hat{\beta}_{OLS}] = \frac{1}{n}(\frac{1}{n}\sum_i X_i X_i')^{-1}(\frac{1}{n}\sum_i X_i X_i' \hat{u}_i^2)(\frac{1}{n}\sum_i X_i X_i')^{-1} = (X'X)^{-1}(X' \text{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2)X)(X'X)^{-1}$ where X denotes a matrix of stacked X_i' clinical, field and remote-sampled coefficient values from the empirically, geo-spatiotemporal or geo-spectrotemporal, eco-epidemiologically geosampled datasets. The estimator we employed was from then derived in terms of the generalized method of moments (GMM).

In our seasonal, geopredictive, vector arthropod-related, endemic, transmission-oriented, risk based data analyses, the GMM was a generic method for the geosampled, clincial, field and remote time series dependent, geopredictive, explanatorial, estimating parameters. Usually GMM applied in the context of semiparametric, seasonal, YFV-related, explanatorial geopredictive, endemic, transmission-oriented, geoclassified LULC, risk models employ a parameter of interest that is finite-dimensional, whereas the full shape of the distributional function of the geosampled data may not be known, and therefore the MLE is not applicable. In statistics a semiparametric model is a model that has parametric and nonparametric components.(Hosmer and Lemeshew 2000). Our seasonal, probabilsitc, sylvatic, YFV-related, explanatorial, geopredictive, endemic, transmission-oriented, geoclassified LULC, risk model was then a collection of distributions: $\{P_\theta : \theta \in \Theta\}$ indexed by a parameter θ. A parametric explanatorial, YFV-related, probabilistic, time series seasonal, endemic, transmission-oriented, risk model is one in which the indexing parameter is a finite-dimensional vector (in k-dimensional Euclidean space for some integer k(forest-canopy to agro-village complex LULC change); (i.e. the set of possible sampled values for θ would be a subset of $\mathbb{R}^k$, or $\Theta \subset \mathbb{R}^k$). In this case a could remark that θ is finite-dimensional in a robust, seasonally probabilictic, YFV-related, endemic, transmission-oriented, eco-epidemiologica, risk model. In nonparametric models, the set of possible values of the parameter θ is a subset of some space, not necessarily

finite dimensional (Greene 1998). For instance, a researchist might consider a datset of all seasonal, YFV-related, spatial distributions with mean 0. Such spaces would then be vector spaces with topological structure, but may not be finite dimensional as vector spaces. Thus, $\Theta \subset \mathbb{F}$ for some possibly infinite dimensional space $\mathbb{F}$ may conduct a powerful covariate, parameter estimation. In semiparametric models, the parameter has both a finite dimensional component and an infinite dimensional component (often a real-valued function defined on the real line)(Hosmer and Lemeshew 2000). Thus, the parameter space Θ in a semiparametric, seasonal, YFV-related, explanatorial, geopredictive, eco-epidemiological, probabilistic, risk model would satisfy $\Theta \subset \mathbb{R}^{k} \times \mathbb{F}$, where $\mathbb{F}$ is an infinite dimensional non-Euclidean space.

Parametric and semi-parametric, seasonal, vector arthropod-related explanatory geopredictive, endemic, transmission-oriented, wavelength, submeter resolution, LULC risk models often use smoothing or kernels (Jacob et al. 2005b, Griffith 2005). In statistics and image processing, to smooth an empirical dataset of geo-spatiotermpoally geosampled probablsitically regressed, LULC estimators is to create an approximating function that attempts to capture important patterns in the data, while leaving out noise or other fine-scale structures/rapid phenomena (www.esrei.com). In smoothing, seasonal empirically, probabilistically regressed, YFV-related, geoclassified LULC, data points of a signal may be modified so that individual geo-spatiotemporally, geosampled points presumably because of noise are reduced, and points that are lower than the adjacent geosampled points are increased leading to a smoother signal. Smoothing may be used in important ways that can aid in seasonal, sylvatic YFV-related, geoclassified LULC, predictive, eco-epidemiological, risk-based, time series, data analysis including 1) being able to extract more information from the data as long as the assumption of smoothing is reasonable and (2) being able to provide a risk-based analyses that are both flexible and robust. Many different algorithms may be used in smoothing of empirical geosampled, YFV-related clinica, field and remote-specified, explanatorial covariate parameter estimators. Data smoothing for regressively quantitating vector arthropod-related probabilistically geosampled, covariate parameter estimator significance is typically done through the simplest of all density estimators, the histogram (see Jacob et al. 2005b, Griffith 2005).Other higher level smoothing techniques may also be employed for deriving optimal, unbiased, asymptotical, YFV-related, stochastic/deterministic, covariate parameter estimators. For example, use of an interpolating spline may fit a smooth curve exactly through a time series dataset of geo-spatiotemporal or geo-spectro0temporal, geosampled, YFV-related empirically regressable, explanatorial, clinical, field or remote, geoclassified, LULC, seasonally sensitive points.

In mathematics, a spline is a sufficiently smooth polynomial function that is piecewise-defined, and possesses a high degree of smoothness at the places where the polynomial pieces connect (which are known as knots) (Cressie 1993). In interpolating problems, spline interpolation is often referred to as polynomial interpolation because it yields similar results, even when using low-degree splines, to interpolating with higher degree polynomials while avoiding instability due to Runge's phenomenon. In the mathematical field of numerical analysis, Runge's phenomenon is a problem of oscillation at the edges of an interval that occurs when using polynomial interpolation with polynomials of high degree over a set of equispaced interpolation points (Hosmer and Lmeshew 2002).

In computer graphics splines are popular curves because of the simplicity of their construction, their ease and accuracy of evaluation, and their capacity to approximate complex shapes through curve fitting and interactive curve design. The most commonly used splines in vector arthrpos-related forecast , vulnerability, eco-epidemiological, risk models are cubic spline, cubic B-spline and cubic Bézier spline. They are common, in particular, in spline interpolation simulating the function of splines. The term spline is adopted from the name of a flexible strip of metal commonly used by draftsmen to assist in drawing curved lines. Splines are curves, which are usually required to be continuous and smooth(Cressie 1993). Splines are usually defined as piecewise polynomials of degree n with function values and first $n\text{-}1$ derivatives that agree at the points where they join. The abscissa values of the join points are called knots. The term "spline" is also used for polynomials (splines with no knots) and piecewise polynomials with more than one discontinuous derivative. As such, YFV-related explanatorial splines with no knots would be generally smoother than splines with knots, which would be subsequently generally smoother than splines with multiple discontinuous derivatives. Splines with few knots are generally smoother than splines with many knots(Cressie 1993); however, increasing the number of knots usually increases the fit of the spline function in LULC data (Jacob et al. 2011). Knots give the curve freedom to bend to more closely follow the data (Griffith 2003).

Thus, a spline in a explanatorial, clinical, field or remote-specified, geo-spatiotemporally geosampled, probabilistic, piecewise-polynomial real function $S : [a, b] \to \mathbb{R}$ on an interval [a,b] composed of k subintervals $[t_{i-1}, t_i]$ with $a = t_0 < t_1 < \cdots < t_{k-1} < t_k = b$. The restriction of S to an interval i is a polynomial $P_i : [t_{i-1}, t_i] \to \mathbb{R}$, so that $S(t) - P_1(t)$, $t_0 \le t < t_1 S(t) = P_2(t)$, $t_1 \le t < t_2 S(t) = P_k(t)$, $t_{k-1} \le t \le t_k$. The highest order of the polynomials $P_i(t)$ is said to be the order of the spline S. The spline is said to be uniform if all subintervals are of the same length, and **non-**uniform otherwise (Cressie 1993). The idea then in a sylvatic YFV-related, ovispoition eco-epidemiological, geoclassified, LULC, probabilsitic, African, ricland, agro-irrigated, village-level, risk model is to choose the polynomials in a way that guarantees sufficient smoothness of S. Specifically, for a spline of order n, S is required to be both continuous and continuously differentiable to order n-1 at the interior points t_i: for $i = 1, \ldots, k - 1$ and $j = 0, \ldots, n - 1 P_i^{(j)}(t_i) = P_{i+1}^{(j)}(t_i)$ The method requires that a certain number of moment conditions were specified for the model. These moment conditions are functions of the model parameters and the data, such that their expectation is zero at the true values of the parameters. The GMM method then minimizes a certain norm of the sample averages of the moment conditions (Griffith 2003). The GMM estimators are known to be consistent, asymptotically normal, and efficient in the class of all estimators that don't use any extra information aside from that contained in the moment conditions.

Note in the literature, the covariance matrix $\hat{\Omega}_n$ of the $\sqrt{n}$-consistent limiting distribution: $\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \Omega)$, where $\Omega = E[XX']^{-1} Var[Xu] E[XX']^{-1}$, and the forecasting equation $\hat{\Omega}_n = (\frac{1}{n} \sum X_i X_i')^{-1} (\frac{1}{n} \sum X_i X_i' \hat{u}_i^2)(\frac{1}{n} \sum X_i X_i')^{-1} = n(X'X)^{-1}(X' \operatorname{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2)X)(X'X)^{-1}$ Thus,

$$\hat{\Omega}_n = n \cdot v_{HCE}[\hat{\beta}_{OLS}] \quad \widehat{Var[Xu]} = \frac{1}{n} \sum_i X_i X_i' \hat{u}_i^2 = \frac{1}{n} X' \operatorname{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2) X$$

may be revealed when

436

geo-spatiotemporally, empirically regressing, probabilsitic, clincial, field or remote gesampled time series dependnet, forecasters. Precisely which covariance matrix is to be employed in a robust, YFV-related, explanatorial, seasonal, geopredictive, endemic, transmission-oriented, probabilistic, eco-epidemiological, risk model then should be a matter of the context. Alternative estimators have been proposed in MacKinnon an d White (1985) that correct for unequal variances of regression residuals due to different leverage. Unlike the asymptotic White's estimator, these estimators may unbiased when the sylvatic, YFV-related, time series, eco-georferenceable, ecoe-pidemiological, sub-meter, resol;ition grod-stratfiied, LULC data are homoscedastic.

Another approach that may be employed for future YFV-related geocalssified LULC research is g an ArcGIS-based Land Information Surface (LIS) model. These models are designed explicitly for soil moisture LULC estimation. The LIS model is very customizable with the ability to choose many different inputs for geo-spatiotemporal, geosampled covariate, parameter estimators (e.g., elevation, soil types, and land use classification) and other related data (i.e., radiation and meteorological fields including precipitation updated hourly or 3-hourly). LIS has the ability to run several "tiles" within a QuickBird, digitized ,grid cell that has different geoclassified, geo-spatiotemporal,seasonal LULC classifications, (Jacob et al. 2010b), so even if a digitized cell classified at a riceland agroecosystem LULC study site was for example 5% urban, that portion could still be monitored. In previous research, Jacob et al. (2010b) generated an LIS map using field and QuickBird-sampled explanatorial, probabilistic predictors of *Cx. pipiens/restuans* for 15 larval habitats in Urbana/Champaign, Illinois USA (Figure 17). The LIS framework may be thus used to provide information on surface soil moisture conditions related to the sampled *Aedes* and wild monkey habitats at the Gulu study site, for example. In these models the configuration, may be based on a Land Data Assimilation System (NLDAS) forcing data (1/8- degree, hourly) up to 3 days before the sampling day and a Global Land Data Assimilation System (GLDAS) forcing data (1/2 degree, 3-hourly) up to 12 hours previous to the time of sampling. Additionally, a researchist may augment YFV-related, time series, explanatorial model covariate, paramter estimators with extra georeferencable, explanatorial, geopredictive, time series dependent, LULC variables at each grid point to quantify such data as: (1) depth in an open artificial water container (taking into account precipitation and evaporation influenced by temperature, humidity, winds, and radiation), (2) water depth in a shaded water container such as a waste tire (similar to the first, but with little or no solar radiation), and (3) potential standing water on the ground, assuming an area with suitable LULC topography exists inside the QuickBird grid cell with no drainage by runoff. This would provide monitoring of potential conditions favorable to *Ae egypti* and other YFV-related mosquito outbreaks. Thereafter, all generated model covariate paramter estimators may be analyzed using various spatial statistical algorithms.

A hierarchical, Bayesian, generalized, probabilistic, orthogonal matrix generated in PROC MCMC optimal seasonal sylvatic ,YFV-related, explanatorial interpolatable predictors by quantitating the asymptotical variance distribution in the geosampled data for identifying prolific, geo-spatiotemporal geosampled LULC count data. Currently, interest is emerging within bioinformatics to use Bayesian statistics for building geo-spatiotemporal, geosampled, vector insect larval habitat, eco-epidemiological, forecasting,risk-related, distribution models of various kinds. As such in a ecologist, entomologist or data anlyst could construct Bayesian, probablistic,

estimation matrices based on the geosampled *Aedes* and wild monkey habitat explanatory predictor covariate, paramter estimators, for example. Often, when constructing Bayesian frameworks for quantifying vector mosquito larval habitat data there is a requirement of a Box-Cox type of power transformation and a Markov Chain Monte Carlo (MCMC) algorithm for enabling the use of non-normal probability models with a spatial autoregressive specification but in a hierarchical model context. For example, in Jacob et al. (2011a), a Bayesian model was generated using WinBUGs to quantify stochastic error propagation in multiple datasets of *An. gambiae s.l.* explanatory predictor, covariate, paramter estimator, geo-spatiotemporally geosampled in Malindi and Kisumu, Kenya. WinBUGS is statistical software for Bayesian analysis using MCMC methods based on the BUGS (Bayesian inference Using HYPERLINK (http:www.mrc-bsu.cam.ac.uk/bugs/winbugs/content shtml).

In Bayesian settings, geosampled explanatorial, prognosticative, covariate paramter estimator Depth of habitat was the most important variable associated with productive habitats based on geo-spatiotemporal, geosampled, count data. Thus for future research efforts researchist could incorporate multiple Bayesianistic LULC models for determining probabilistic, time series predictors that are associated with the sampled *Aedes* mosquitoes and wild monkey habitats at the Gulu study site, for example. The assumption would be that the Bayesian LULC eco-epidmeiological, risk-related analysis may enable researchers to estimate both conditional autoregressive and spatial filter hierarchical, generalized linear model specifications employing the geo-spatiotemporal, geosampled, YFV-related, LULC mosquito data and wild monkey habitat explanatorial covariate, parameter estimators. Spatial filter eigenvectors can then be employed to quantitate, probabilistic, uncertainty estimates in the YFV-related geocalssied LULC distribution models by geospatially adjusting the georeferenced, geosampled explanatorial, clinical,field or remote data in SAS/GIS®

SAS/GIS® software (HYPERLINK "http://www.sas.com/products/gis/" http://www.sas.com/products/gis/).provides an interactive GIS within the SAS System for data analyses. SAS/GIS® and SAS STAT® software may thus be employed for determining important probability values of regressed explanatorial clinical, field or remote geoclassified, tiem series dependent, covariate, paramter estimators associated to productive *Aedes* and wild monkey habitats at the Gulu riceland agroecosystem, eco-epidemiological study site. SAS STAT® provides data analysis in downloadable, experimental versions of three procedures for SAS 9.3® on Windows GENMOD, LIFEREG, and PHREG. Fortunately, the new BAYES statement in these procedures can generate a Bayesian eco-epidemiological, sylvatic, YFV-related LULC , forecasting, risk-related analyses and inference capability in generalized linear models. This version is named BGENMOD and contains the full functionality of the original procedures. Spatial modeling in SAS/GIS® and SAS STAT software allows errors of estimations to be quantified, making it possible to assess the precision of a vector insect larval distribution habitat, eco-epidemiological, probabilistic, risk map and relative importance of factors associated with abundance (Jacob et al. 2011a, Jacob et al. 2010b, Jacob et al. 2009d).

Further an ecologist, entomologist or data analyst may also include time series dependent, explanatorial, clinical, field and remote-geo-spatiotemporally geosampled ArcGIS, LULC-related, covariate parameter estimators and seasonal meteorological coefficient estimates from

geosampled *Aedes* mosquito and monkey cooperative breeding areas and communal resting sites to define expectations for prior distributions in WinBUGS®. The BUGS (Bayesian inference Using Gibbs Sampling) project is concerned with flexible software for the Bayesian analysis of complex statistical models using MCMC specifications (Gilks 1996). In WinBugs®, MCMC chains may be generated for the geosampled, explanatorial, clinical, field and remote environmental variables from each potential *Aedes* mosquito habitats sampled at an ricland agroecosystem, geoclassified LULC, eco-epidemiological interventional study site. Thereafter, the WinBUGS Deviance Information Criterion (DIC) tool can be employed to obtain mean posterior deviance values and to construct improvement of fit tables and deviation statistics to identify the best fitting *Aedes* habitat models. By specifying explanatorial, clinical, field or remote geo-spatiotemporal, sylvatic, YFV-related covariate, paramter estimator, coefficients in a Bayesian framework in WinBUGS, a researchist may then account for the explanatorial, predictive, covariate, parameter estimators and determine which ones(s) are significantly associated with a particular ArcGIS derived, time series, geo-spatiotemporal or geo-spectrotemporal, eco-epidemiologically geoclassified LULC variables.

Finally with the advent of satellite remote sensing products even higher spatial resolution data may be employed than QuickBird 0.61m visible and NIR data to identify and quantitate YVR-related LULC and other seasonal explanatorial geopredictors. For example, in Jacob et al. (2011d) geospatially lagged and simultaneous autoregressive, eco-epidemiological, explanatorial, geo-spatiotemporal, risk models based on multiple predictor variables of immature WNV mosquito vector *Cx. quinquefacistus* and Worldview 1 (WV-1) 5m visible and NIR data was employed to help implant a remote larval habitat-based surveillance system in Trinidad and Tobago. Initially, the authors used Geomatica Ortho Engine® v. 10.2 for extracting a 3-Dimensional digital elevation model (DEM) from the WV-1 raw imagery. Results of the DEM analyses indicated a statistically significant inverse linear relationship between total geosampled *Cx. quinquefacistus* immature data and elevation (m) ($R^2 = 0.439$; $p < 0.0001$), with a standard deviation of 10.41. Additional field, geosampled information was derived using data from an orthogonal digitized grid-matrix constructed in an ArcGIS and overlaid onto the WV-1 data. A unique identifier was placed in the centroid of each grid cell. Univariate statistics and Poisson regression models were then generated using the georeferenced covariate parameter estimators in SAS/GIS®. Coefficient estimates were also used to define expectations for prior distributions in a Bayesian estimation matrix using MCMC specifications which revealed that the covariate , paramter estimator Depth of the geosampled habitat and Distance to the nearest house were important geopredictors related to productive habitats based on egg-raft count data. Therefore, a ecologist, entomologist or data analyst can conduct a robustifiable geo-spatiotemporal,probabilistic, regression-related, residual, trend analyses employing autocorrelation indices linked to tabular data in SAS PROCLMIXED® , for instance along with the *Aedes* habitat count data in an ArcGIS® geodatabase. An Ordinary kriged-based interpolator may then then be used in Geostatistical Analyst Extension of ArcGIS based on adjusted Bayesian estimates to identify unknown and unsampled prolific *Aedes* aquatic larval habitats. The ability to automate tasks and to run the complex spatial autocorrelation and the Bayesian probabilsic, estimation models may thus aid in developing IVM control strategies for YFV at the Gulu riceland, agro-ecosystem, study site. The resulting geoclassified LULC, explanatorial, clinical, field or remote-related, space-time models may then be employed to forecast sites of

immature seasonal *Aedes* mosquito abundance and cooperative breeding areas and communal resting sites throughout the eco-epidemiological, study site.

Thereafter, C++ models can be generated to geopredict high and high YFV infection rates employing *Aedes* mosquito abundance data, habitats meteorological variables, ArcGIS-derived, LULC covariate, parameter estimators and others. Randomized areas in the eco-epidemiological, study site will be used to validate the characterization and prediction tools created. Thereafter these metrics may be used to validate risk of yellow fever transmission by rating variables from 1 to 3 for: 1) continued above average temperatures, 2) continued above average rainfall, 3) large vector population, 4) suspected cases for humans; and, 5) virus activity in nearby areas to the study sites. Ratings of 1 will indicate low to average risk, ratings of 2 will indicate moderate risk and ratings of 3 will indicate high risk

Although in this research we did generate a C++ model based on ArcGIS-derived LULC sub-meter resolution, ovispoition, parameter estimators, meteorological variables and previous case data, entomological factors associated with zoonotic and epidemic transmission factors would create a more robust model. This would be vital for the development and implementation of a YF surveillance system. In order to accomplish this take however it would be necessary to understand and define the relationship of the virus with the mosquitoes species associated with both the zoonotic and epidemic cycle as well as in non-human primates within an eco-epidemiological interventional study site. Understanding the cycle of YFV in an interventional site can provide the basis for YFV surveillance, epidemic risk prediction and control ().

Unfortunately, the zoonotic cycle or maintenance cycle in most areas of Gulu much like the rest of Africa is poorly understood. Thus, it would be vital to initially identity and establish as much as information as possible on the temporal distribution and abundance of mosquito species associated within different LULCs (e.g., forest, savannah, peri-domestic and urban ecosystems) at sentinel study sites. In so doing, the principal non-human and human hosts associated with specific *Aedes* mosquitoes may be identified at a particular site by geo-spatiotemporal distribution employing mosquito blood meal analysis and through identification of sylvatic YFV in mosquito pools. Thereafter, C++ geopredictive,risk models may be constructed employing ArcGIS-derived, LULC-derived, parameter estimators and meteorological variables at the sentinel sites

An IVM may be defined in C++ and ArcGIS for optimal targeting *Aedes* mosquitoes for vector control which can include: 1) evidence-baseddecision-making, 2) integrated approaches 3), collaboration within the health sector and with other sectors, 4) advocacy, social mobilization, and legislation, and 5) capacity-building. In 2004, the WHO adopted IVM globally for the control of all vector-borne diseases. Important recent progress has been made in developing and promoting IVM for national malaria control programmes in Africa at a time when successful malaria control programmes are scaling-up with insecticide-treated nets (ITN) and/or indoor residual spraying (IRS) coverage. While interventions using only ITNs and/or IRS successfully reduce transmission intensity and the burden of malaria in many situations, it was not clear if these interventions alone will achieve those critical low levels that result in other vector arthropod elimination. Despite the successful employment of comprehensive integrated

malaria control programmes, further strengthening of vector control components through IVM for Yellow fever has not been actualized especially during the "end-game" scenarios where control efforts are supposed to go from low transmission situations to sustained local and country-wide elimination. To meet this need and to ensure sustainability of control efforts, YFV control programmes should strengthen their capacity to use C++ ,ArcGIS and spatial statistical data for decision-making with respect to evaluation of current vector control programmes, employment of additional vector control tools in conjunction with, case-detection and treatment strategies, to determine how much and what type of vector control and interdisciplinary input are required to achieve yellow fever elimination. Similarly,on a global scale, there is a need for continued research to identify and evaluate new tools for geospatial *Aedes* immature habitat control that can be integrated with existing biomedical strategies in C++ and ArcGIS within national Yellow fever control programmes.

In conclusion results from both a Poisson and a negative binomial regression (i.e., a Poisson random variable with a gamma distrusted mean) revealed that the village-level seasonal-sampled YFV-related explanatory geopredictive explanatory covariate coefficients were highly significant, but furnished virtually no predictive power. Importantly, taking the Poisson mean as a gamma distributed random variable leas to the negative binomial regression model for deriving various forms of mean-variance relationship, in particular both linear and quadratic, depending on assumptions about the gamma mixing distribution YFV-related, probabilsic, LULC geoclassified parameter estimators. In other words, the sizes of the population denominators were not sufficient to result in statistically significant relationships, while the detected relationships were inconsequential. Thereafter, a linear mean-variance negative binomial model was obtained by allowing the gamma shape parameter to vary across the seasonal, geosampled, eco-epidemiological, clinical, field or remote-specified, YFV-related, probabilistic observations which kept the scale parameter constant, whereas the quadratic form arose from taking the shape parameter as constant and letting the scale vary.

We then generated the normal Q-Q plot for the normalized randomized quantile residuals of the $\text{SARIMA}(3',1,0) \times (1,1,0)_{12}$ model, for which the distribution was slightly leptokurtic. A plot of these normalized randomized quantile residuals against time appears a random scattered at first sight, but upon closer inspection, extreme residuals occured more often during periods with stronger relative changes. This was because the residuals, ε_t, were positively correlated with a relative change in YFV cases, where the linear regression line expressed $\varepsilon_t = 1.85 \log\left(\frac{y_t}{y_{t-1}}\right) + 0.22$, $R^2 = 0.93$. We constructed a spatial autocorrelation plotin SAS/GIS.The MC was 0.068 while the GR was).791 We then constructed a C++ model. Rainfall and an LULC variables were employed to represent rainfall for specific years for quantitating interface distance between human population and forest regions respectively. YFV-related cases represented the number of yellow fever cases reported for a particular year. From these input, we created a formula of form as below, Z = a0 + a1*X + a2*Y and YFCases = a0 + a1*Rainfall + a2*LULC. We then determined the values of constants a0 a1 and a2. We probabilistically prognosticated the value of the sylvatic YFV-related cases without considering effect of human population in initial stage and then later we added the effect of population over yellow fever prevalence. To determine a relationship between our YFV-related explanatory geoclassified LULC, endemic ,transmission-oriented, data, we also created 3-D regression. We used Least square regression plane of the

form Z = a0 + a1*X + a2*Y that provided the best fit for the input, eco-epidemiological, data points geo-spatiotemporally geosampled at the Gulu study site. Only those values were used which minimized the following least square value function: F(a0, a1, a2) = (Zi - a0 - a1*X - a2*Y)². The solution of this formula was found with the matrices with the system F/a0 = 0 , F/a1 = 0 , F/a2 = 0 . The matrix was then used to solve the constants is as follows: Solving these equation provided the unique values of constants. In our case these covariate paramter estimator coefficient values were as follows: a0 = 60.9239, a1 = 0.1957 a2= -65.4218. So the formula was written as follows: YFCases = 60.9239 + 0.1957 * Rainfall - 65.4218 * LULC. These YF case values predicted does not consider effect of population on it. So to consider population data on this prediction, we can re-write the formula as below: YFCases = (1+PI_Factor) * [60.9239 + 0.1957 * Rainfall - 65.4218 * LULC]. We noted from the C++ model output that as the human population grew yellow fever prevalence also grew linearly due to the variation in LULC from agro-village complexes to canopy forested zones and their boundaries. Anthropogenic drivers of ecosystem disturbance via human population growth and urbanization can lead to specific changes in bioecology that may lead to transfer of the YF virus pathogen via mosquitoes to increasing numbers of highly concentrated and susceptible human populations Real time, temporal, vulnerability,eco-epidemiological, *Ae.egypti* LULC, risk maps can aid in implementing IVM .

References:

[1] AB Gilroy and L. Bruce-Chwatt (1945). "Mosquito-control by swamp drainage in the coastal belt of NigeriaHR Grubb, Croydon."

[2] Albanese, G., et al. (1996). Collected papers of Giacomo Albanese. Kingston, Ont., Queens's University.

[3] Anselin, L. (1988). "Spatial Auto-Correlation - Odland,J." Journal of Regional Science **28**(4): 589-592.

[4] Anseline, P. F. (1995). "Crohn's disease in the Hunter Valley region of Australia." Aust N Z J Surg **65**(8): 564-569.

[5] Backenson, P. W., D. J.; Eidson, M.; Smith, P. F.; Kramer, L. D.; Morse, D. L.; Tucker, C. J.; Myers, M. F.; Hay, S. I.; Rogers, D. J. (2002). "Mapping of West Nile Virus Risk in the Northeast United States Using Multi-temporal Meteorological Satellite Data." American Geophysical Union(Spring Meeting 2002): abstract #B42B-05.

[6] Bailey, A. M. C., R. E. (2001). "On the Random Character of Fundamental Constant Expansion." Experimental Mathematics **10**(2): 175-190.

[7] Bailey D. H., B. J. M., Crandall R. E., Pomerance C. (2003). On the binary expansions of algebraic numbers.

[8] BG, J., et al. "A Bayesian Poisson specification with a conditionally autoregressive prior and a residual Moran's Coefficient minimization criterion for estimating leptokurtic distributions in regression-based Multi-Drug Resistant Tuberculosis treatment protocols " In Press.

[9] Bivand, R. S. (1984). "Regression Modeling with Spatial Dependence: An Application of Some Class Selection and Estimation Methods." Geographical Analysis **16**(1): 25-37.

[10] Borwein, J., et al. (2004). Experimentation in mathematics: Computational paths to discovery. DH Bailey and JM Borwein," Experimental Mathematics: Examples, Methods and Implications," Notices, Citeseer.

[11]   Box, G. E. P. and G. M. Jenkins (1970). Time series analysis; forecasting and control. San Francisco,, Holden-Day.

[12]   Brooker, S., et al. (2002). "Tools from ecology: useful for evaluating infection risk models?" Trends Parasitol **18**(2): 70-74.

[13]   Brownstein, J. S., et al. (2002). "Spatial Analysis of West Nile Virus: Rapid Risk Assessment of an Introduced Vector-Borne Zoonosis." Vector-Borne and Zoonotic Diseases **2**(3): 157-164.

[14]   Catalan, E. (1883). Sur la circonférence des neuf points. Nouvelles annales de mathématiques, Gauthier-Villars.

[15]   Cooke, W. H., 3rd, et al. (2006). "Avian GIS models signal human risk for West Nile virus in Mississippi." Int J Health Geogr **5**: 36.

[16]   Coppinger, P. F. (1996). "The Hospital's Dilemma." New Solut **6**(2): 51-60.

[17]   Cressie, N. A. C. (1993). Statistics for spatial data. New York, Wiley.

[18]   Crooke, P. S., et al. (2006). "Estrogens, enzyme variants, and breast cancer: a risk model." Cancer Epidemiol Biomarkers Prev **15**(9): 1620-1629.

[19]   Crosskey, R. W. (1960). "Taxonomic Study of the Larvae of West African Simuliidae (Díptera: Nematocera) with Comments on the Morphology of the Larval Black-Fly Head." Bulletin of the British Museum **10**(1): 1-74.

[20]   D. Hilbert, S. C.-V. (1999). Geometry and the Imagination, Chelsea.

[21]   Dale, P. E., et al. (1998). "An overview of remote sensing and GIS for surveillance of mosquito vector habitats and risk assessment." Journal of Vector Ecology **23**(1): 54-61.

[22]   Diuk-Wasser, M. A., et al. (2007). "Effect of rice cultivation patterns on malaria vector abundance in rice-growing villages in Mali." Am J Trop Med Hyg **76**(5): 869-874.

[23]   Fernandez, A. G., et al. (2004). "Combinacion de atmosfera modificada con nitrogeno y baja humedad relativa como metodo de secado y para inactivacion de mohos en documentos contaminados. (Spanish)." Ciencias de la Informacion **35**(3): 25-29.

[24]   Fotheringham, A. S., C. Brunsdon, and M. Charlton (2002). "Geographically weighted regression: the analysis of spatially varying relationships."

[25]   Getis, A. and D. A. Griffith (2002). "Comparative spatial filtering in regression analysis." Geographical Analysis **34**(2): 130-140.

[26]   Getis, A. and J. K. Ord (1992). "The Analysis of Spatial Association by Use of Distance Statistics." Geographical Analysis **24**(3): 189-206.

[27]   Gilks, C. (1996). "Disease in children with HIV infection in Abidjan. Authors should have emphasised lack of tuberculosis in HIV positive children." BMJ **312**(7041): 1302-1303.

[28]   Glantz, S. A., and B. K. Slinker (2001). A Primer of Applied Regression and Analysis of Variance. New York, USA, McGraw-Hill.

[29]   Griffith, D. A. (2003). Spatial autocorrelation on spatial filtering. Gaining understanding through theory and scientific visualization, Springer-Verlag.

[30]   Griffith, D. A. (2005). "A comparison of six analytical disease mapping techniques as applied to West Nile Virus in the coterminous United States." International Journal of Health Geography **4**: 18-26.

[31]   Griffith, D. A. and P. R. Peres-Neto (2006). "Spatial Modeling in Ecology: The Flexibility of Eigenfunction Spatial Analyses." Ecology **87**(10): 2603-2613.

[32]   Grunwald, P. (2000). "Model Selection Based on Minimum Description Length." J Math Psychol **44**(1): 133-152.

[33]   Gu, W. and R. J. Novak (2005). "Habitat-based modeling of impacts of mosquito larval interventions on entomological inoculation rates, incidence, and prevalence of malaria." Am J Trop Med Hyg **73**(3): 546-552.

[34]   Haight, F. A. (1967). Handbook of the Poisson Distribution. New York, John Wiley & Sons.

[35]   Haining, R. (1991). "Bivariate Correlation with Spatial Data." Geographical Analysis **23**(3): 210-227.

[36]   Hardy, G. and S. Ramanujan (1999). Twelve lectures on subjects suggested by his life and work, AMS Chelsea Publishing Company, New York.

[37]   Hay, N. A., et al. (1997). "A nonswarming mutant of Proteus mirabilis lacks the Lrp global transcriptional regulator." J Bacteriol **179**(15): 4741-4746.

[38]   Hay, S. I., et al. (2000). "Annual Plasmodium falciparum entomological inoculation rates (EIR) across Africa: literature survey, internet access and review." Trans R Soc Trop Med Hyg **94**(2): 113-127.

[39]   Hayasaka, T., et al. (1998). Microwave remote sensing of the atmosphere and environment : 15-17 September 1998, Beijing, China. Bellingham, Wash., USA, SPIE.

[40]   Heinen, C., et al. (2003). "Flow of newtonian/non-newtonian fluids in a bundle of tubes and in a packing of beads by MRI." Magn Reson Imaging **21**(3-4): 377-379.

[41]   Helfenstein, U. (1986). "Box-Jenkins modelling of some viral infectious diseases." Stat Med **5**(1): 37-47.

[42]   Henderson, B. E., et al. (1968). "Yellow fever immunity surveys in northern Uganda and Kenya and eastern Somalia, 1966-67." Bull World Health Organ **38**(2): 229-237.

[43]   Hinde, P. R. and A. J. Mturi (1996). "Social and economic factors related to breast-feeding durations in Tanzania." J Biosoc Sci **28**(3): 347-354.

[44]   Honsberger, R. (1976). Mathematical GEMS II. United States of America, The Mathematical Association of America.

[45]   Hosmer, D. and S. Lemeshow (2002). Applied, New York: John Wiley and Sons.

[46]   Howard, W. A. (1969). Remote sensing of the urban environment, a selected bibliography. Monticello, Ill.,, Council of Planning Librarians.

[47]   Ihaka, R. (1998). "R: Past and future history." COMPUTING SCIENCE AND STATISTICS: 392-396.

[48]   Jacob, B. G., K. L. Arheart, D. A. Griffith, C. M. Mbogo, A. K. Githeko, J. L. Regens, J. I. Githure, R. Novak, and J. C. Beier (2005). "Evaluation of environmental data for identification of Anopheles (Diptera: Culicidae) aquatic larval habitats in Kisumu and Malindi, Kenya." J Med Entomol **42**(5): 751-755.

[49]   Jacob, B. G., D. A. Griffith, E. J. Muturi, E. X. Caamano, J. I. Githure, and R. J. Novak (2009). "A heteroskedastic error covariance matrix estimator using a first-order conditional autoregressive Markov simulation for deriving asympotical efficient estimates from ecological sampled Anopheles arabiensis aquatic habitat covariates." Malaria Journal **8**(1): 216-225.

[50]   Jacob, B. G., J. T. Gunter, E. J. Muturi, E. X. Caamano, J. I. Githure, J. L. Regens, and R. J. Novak (2010). "Quantifying Stochastic Error Propagation in Bayesian Parametric Estimates of Anopheles gambiae s.l. aquatic habitats." International journal of remote sensing **11**: 67-78.

[51]   Jacob, B. G., D. A. Grifiith, J. M. Mwangnagi, C. Mbogo, and R. J. Novak (2010). "Uniform Convergence of Ergodic Markov Chains Using Gaussian Quadratures in SAS

PROC NLMIXED for Calculating Marginal Likelihoods in Space Time-Varying Coefficients Of Urban Anopheles gambiae s.l. aquatic Habitats." Acta Paristology of China **14**(3): 41-53.

[52]	Jacob, B. G., D. D. Chadee, and R. J. Novak (2011). "Adjusting second moment bias in eigenspace using Bayesian empirical estimators, Dirichlet tessellations and Worldview 1 data for predicting Culex quinquefasciatus in Trinidad." Journal of Geographic Information Systems **2**: 244-274.

[53]	Jacob, B. G., et al. (2011). "A Cartographic Analysis Using Spatial Filter Logistic Model Specifications for Implementing Mosquito Control in Kenya." Urban Geography **32**(2): 363-377.

[54]	Jacob, B. G., et al. (2011). "Geomapping generalized eigenvalue frequency distributions for predicting prolific Aedes albopictus and Culex quinquefasciatus habitats based on spatiotemporal field-sampled count data." Acta Trop **117**(2): 61-68.

[55]	Jacob, B. G., et al. (2007). "Environmental abundance of Anopheles (Diptera: Culicidae) larval habitats on land cover change sites in Karima Village, Mwea Rice Scheme, Kenya." American Journal of Tropical Medicine and Hygiene **76**(1): 73-80.

[56]	Jacob, B. G., et al. (2008). "Hydrological modeling of geophysical parameters of arboviral and protozoan disease vectors in Internally Displaced People camps in Gulu, Uganda." Int J Health Geogr **7**: 11.

[57]	Jacob, B. G., et al. (2006). "A grid-based infrastructure for ecological forecasting of rice land Anopheles arabiensis aquatic larval habitats." Malaria Journal **5**(1): 41-47.

[58]	Jacob, B. G., et al. (2007). "Remote and field level quantification of vegetation covariates for malaria mapping in three rice agro-village complexes in Central Kenya." Int J Health Geogr **6**(1): 21.

[59]	Jacob, B. G., et al. (2012). "Quasi-likelihood techniques in a logistic regression equation for identifying Simulium damnosum sl. larval habitats intra-cluster covariates in Togo." Geo-spatial Information Science **15**(2): 117-133.

[60]	Jacob, B. G., et al. (2013). "Validation of a Remote Sensing Model to Identify Simulium damnosum sl Breeding Sites in Sub-Saharan Africa." PLoS neglected tropical diseases **7**(7): e2342.

[61]	Jacob, B. G., et al. (2003). "Occurrence and distribution of Anopheles (Diptera : Culicidae) larval habitats on land cover change sites in urban Kisumu and urban Malindi, Kenya." J Med Entomol **40**(6): 777-784.

[62]	Jacob, B. G., et al. (2006). "Spatially targeting Culex quinquefasciatus aquatic habitats on modified land cover for implementing an Integrated Vector Management (IVM) program in three villages within the Mwea Rice Scheme, Kenya." Int J Health Geogr **5**: 18-27.

[63]	Jensen, J. R. (2005). Introductory digital image processing a remote sensing perspective. Upper Saddle River, N.J., Prentice Hall.

[64]	John, N., et al. (1992). "Applied statistics." Allyn & Bacon publication Inc: 1102-1210.

[65]	Kachitvichyanukul, V. and B. W. Schmeiser (1988). "Binomial random variate generation." Communications of the ACM **31**(2): 216-222.

[66]	Kalluri, U. C., et al. (2007). "Genome-wide analysis of Aux/IAA and ARF gene families in Populus trichocarpa." BMC Plant Biol **7**: 59.

[67]	Keating, J., et al. (2004). "Characterization of potential larval habitats for Anopheles mosquitoes in relation to urban land-use in Malindi, Kenya." Int J Health Geogr **3**(1): 9.

[68]	Lachman, A. S., et al. (1985). "Prevalence of Mitral-Valve Prolapse in Non-Care-Seeking

Adolescents." Clinical Research **33**(3): A757-A757.

[69]   Lawless, J. F. (1987). "Negative binomial and mixed Poisson regression." Canadian Journal of Statistics **15**(3): 209-225.

[70]   Lefsky, M. A., et al. (1999). "Lidar remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests." Remote sensing of environment **70**(3): 339-361.

[71]   Lemeshow, S. and D. Hosmer (2000). Applied Logistic Regression. Hoboken, NJ, USA., John Wiley & Sons, Inc.

[72]   Levine, R. S., et al. (2004). "Geographic and ecologic distributions of the* Anopheles gambiae* complex predicted using a genetic algorithm."

[73]   Liang, X. Z., et al. (2005). "Development of land surface albedo parameterization based on Moderate Resolution Imaging Spectroradiometer (MODIS) data." Journal of Geophysical Research: Atmospheres (1984–2012) **110**(D11).

[74]   Linthicum, K. J., et al. (1987). "Detection of Rift Valley fever viral activity in Kenya by satellite remote sensing imagery." Science **235**(4796): 1656-1659.

[75]   Loader, C. (2000). Fast and Accurate Comupation of binomial Probabilities. Springer.

[76]   Lukevit?s?, E. I. A., et al. (1992). VII-th International Conference on the Organometallic and Coordination Chemistry of Germanium, Tin, and Lead, Riga Latvia, September 20-25, 1992 : abstracts. Riga, Latvia, Institute of Organic Synthesis, Latvian Academy of Sciences.

[77]   MacKinnon, J. G. and H. White (1985). "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties." Journal of Econometrics **29**(3): 305-325.

[78]   Mahler, K. (1929). "Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen." Mathematische Annalen **101**(1): 342-366.

[79]   McDonald, I. R., et al. (2008). "Molecular ecology techniques for the study of aerobic methanotrophs." Appl Environ Microbiol **74**(5): 1305-1315.

[80]   Muturi, E. J., et al. (2008). "Environmental factors associated with the distribution of Anopheles arabiensis and Culex quinquefasciatus in a rice agro-ecosystem in Mwea, Kenya." Journal of Vector Ecology **33**(1): 56-63.

[81]   Mwangangi, J. M., et al. (2007). "Environmental covariates of Anopheles arabiensis in a rice agroecosystem in Mwea, Central Kenya." J Am Mosq Control Assoc **23**(4): 371-377.

[82]   Neter, J., et al. (1990). "Applied statistical models." Boca Raton, FL: CRC Lewis Publishers.

[83]   Nielsen, C. F., et al. (2008). "Risk factors associated with human infection during the 2006 West Nile virus outbreak in Davis, a residential community in northern California." Transaction of the Royal Society of Tropical Medicine and Hygiene.(102): 817-822.

[84]   Nielsen, L. M. (1921). Tycho Brahes bogtrykkeri paa Hveen. Uppsala,, Almqvist & Wiksells boktryckeri a.-b.

[85]   Patz, J. A., et al. (2000). "The effects of changing weather on public health." Annual review of public health **21**(1): 271-307.

[86]   Pearson, K. (1922). "Tables of the incomplete gamma-function." Nature **110**: 669.

[87]   Peterson, A. T., et al. (2004). "Ecologic and geographic distribution of filovirus disease." Emerg Infect Dis **10**(1).

[88]   Peterson, A. T., et al. (2002). "Ecologic niche modeling and potential reservoirs for Chagas disease, Mexico." Emerg Infect Dis **8**: 662-667.

[89]   Roberts, D. R. and M. H. Rodriguez (1994). "The environment, remote sensing, and malaria control." Ann N Y Acad Sci **740**(1): 396-402.

[90]   Rogers, D. J., et al. (2002). "Satellite imagery in the study and forecast of malaria." Nature **415**(6872): 710-715.

[91]   Schliessmann, D., et al. (1973). "Drainage and larviciding for control of a malaria focus in Haiti." Mosquito News **33**(3).

[92]   Serie, C., et al. (1968). "Etudes sur la fièvre jaune en Ethiopie. I. Introduction-symptomatologie clinique amarile." Bull World Health Organ **38**(6): 835.

[93]   Shanks, C. A., et al. (1993). "A pharmacokinetic-pharmacodynamic model for quantal responses with thiopental." Journal of pharmacokinetics and biopharmaceutics **21**(3): 309-321.

[94]   Sharma, V. and K. Mehrotra (1986). "Malaria resurgence in India: a critical study." Soc Sci Med **22**(8): 835-845.

[95]   Singh, A. (1989). "Review Article Digital change detection techniques using remotely-sensed data." International journal of remote sensing **10**(6): 989-1003.

[96]   Sokal, R. R., et al. (1998). "Local spatial autocorrelation in a biological model." Geographical Analysis **30**(4): 331-354.

[97]   Sondow, J. (2003). "An Infinite Product for e^ gamma via Hypergeometric Formulas for Euler's Constant, gamma." arXiv preprint math/0306008.

[98]   Sondow, J. and W. Zudilin (2006). "Euler's constant, q-logarithms, and formulas of Ramanujan and Gosper." The Ramanujan Journal **12**(2): 225-244.

[99]   Spanier, J. and K. B. Oldham (1987). An atlas of functions, Taylor & Francis/Hemisphere.

[100]   Stegun, I. A. (1972). "Legendre functions." Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables: 331-353.

[101]   Stehman, S. V., et al. (2000). "Combining accuracy assessment of land-cover maps with environmental monitoring programs." Environmental Monitoring and Assessment **64**(1): 115-126.

[102]   Steinhaus, H. (1999). Mathematical snapshots, Courier Dover Publications.

[103]   Story, M. and R. G. Congalton (1986). "Accuracy assessment-A user\'s perspective." Photogrammetric Engineering and remote sensing **52**(3): 397-399.

[104]   Thomson, M. C., et al. (2004). "Impact of climate variability on infectious disease in West Africa." EcoHealth **1**(2): 138-150.

[105]   Tiefelsdorf, M. and B. Boots (1997). "A note on the extremities of local Moran's Iis and their impact on global Moran's I." Geographical Analysis **29**(3): 248-257.

[106]   Titchmarsh, E. C. (1986). The theory of the Riemann zeta-function, Oxford University Press.

[107]   Toe, L., et al. (1997). "Onchocerca volvulus: comparison of field collection methods for the preservation of parasite and vector samples for PCR analysis." Bull World Health Organ **75**(5): 443.

[108]   Utzinger, J., et al. (2002). "The economic payoffs of integrated malaria control in the Zambian copperbelt between 1930 and 1950." Tropical Medicine & International Health **7**(8): 657-677.

[109]   Vallee-Poussin, M. d. l. (1898). "Sur les valeur moyennes de certaines fonctions arithmetiques." Ann. Soc. Sci.(22): 84-90.

[110]   Walker, K. and M. Lynch (2007). "Contributions of Anopheles larval control to malaria

suppression in tropical Africa: review of achievements and potential." Medical and veterinary entomology **21**(1): 2-21.

[111] Waring, R. H., et al. (1995). "Imaging radar for ecosystem studies." BioScience: 715-723.

[112] Watson, E. S. (1966). Differential microcalorimeter, Google Patents.

[113] Wilkinson, K. W., et al. (1997). "Crystallization of cytochrome b562 from Erwinia chrysanthemi." Acta Crystallogr D Biol Crystallogr **53**(Pt 2): 197-199.

[114] Wood, D. J. (1991). "Corporate social performance revisited." Academy of management review **16**(4): 691-718.