International Journal of Geographic Information System
Vol. 1, No. 1, November 2013, PP: 01 - 143
Available Online at http://acascipub.com/Journals.php

Research article

Heuristically optimizing logarithmically transformed mean zero Gaussian vectors in PROC ARIMA using a random deviation from an intercept term and a normal frequency distributed Autoregressive Integrated Moving Average Time Series for forecasting malarial regressors in Uganda

Benjamin G. Jacob[1], Daniel A. Griffith[2], Semiha Caliskan[1], Dissanayake Gunawardena[3], Robert J. Novak[1]

[1] Global Infectious Disease Research Program, Department of Public Health, College of Public Health, University of South Florida, 3720 Spectrum Blvd, Suite 304, Tampa, Florida, USA 33612 Tel: (813) 974-2311 Fax : (813) 974-4718 bjacob1@health.usf.edu, rnovak@health.usf.edu, tunnasch@health.usf.edu

[2]Political and Policy Sciences, The University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080-3021 Tel: (972) 883-4932 Fax: (972) 883-6436 dagriffith@utdallas.edu

[3]USAID Presidents Malaria incentive (PMI), Uganda

Abstract

Lagged dependent variables have routinely been used in district-level geopredictive malaria-related regression analysis to provide robust estimates of the effects of independent variables. However some research argues that employing such variables in these regressions produces negatively biased coefficient estimates. These concerns may be easily resolved by specifying a regression model that accounts for autocorrelation in the error term in a geopredictive district-level malaria-related risk model In this research we constructed multiple linear and non-linear geopredictive autoregressive district-level malaria-related hyperendemic transmission oriented risk models in PROC ARIMA using multiple malarial-related field ,clinical(e.g. prevalence rates) and remote sampled (e.g. Normalized Vegetation Difference index (NDVI)] estimators from 2006 to 2010, in Uganda. We did so to quantitate latent autocorrelation and other non-normal residuals in the regression forecasts targeting important district-level covariates. Initially, a Poisson and a negative binomial (i.e., a Poisson random variable with a gamma distrusted mean) regression was constructed in PROC REG employing the sampled estimators which revealed that the covariate coefficients and their marginal probabilities derived from the district-level risk model were significant but, the forecasts had no predictive power. Inclusion of indicator variables denoting the time sequence and the district geolocational spatial structure was then performed with Thiessen polygons in ArcGIS. The data was then

exported into an SAS/GIS eigenfunction decomposition spatial filter algorithm. The outputs however failed to reveal any unbiased estimators. Thereafter, an Autoregressive Integrated Moving Average (ARIMA) Time Series model was constructed in PROC ARIMA which rendered a conspicuous first-order temporal residual spatial structure. A random effects term was then specified using the sampled coefficients. This random effects term displayed no latent uncertainty autocovariate effects. The model's forecasted residual error variance however, implied a substantial variability in the district-level regressed seasonal prevalence rates. Thereafter, a series of digital elevation models (DEM) was constructed in ArcGIS which spatially adjusted the non-linear derivatives from the ARIMA model. A final risk model was then calculated as: μ= exp [a + re+ LN (population)], Y ~Poisson(μ) +DEM (zonal statistic). The mixed-model estimation results included: a = -3.1876 re ~ n (0, s^2) mean re = -0.0010 s^2 = 0.2513 where P(S-W) = 0.0005 and the Pseudo-R^2 = 0.3103.

Keywords: SAS/GIS, Autoregressive Integrated Moving Average (ARIMA), ArcGIS, Poisson, QuickBird, residual autocorrelation

Introduction

Routinely, time-series models have been constructed in IBM® SPSS® using seasonal district-level geopredictive malarial prevalence data as a dependent variable to geographically forecast seasonal case distribution data. SPSS is a computer program employed for survey authoring and deployment (i.e., SPSS Data Collection), data mining (e.g., SPSS Modeler), text analytics, and deployment for batch and automated scoring services (www-01.ibm.com/software/analytics/spss/). One of the primary formats for a malarialogist/experimenter to analyze seasonal hyperendemic transmission oriented field-sampled data files in SPSS is by using Data View, so that each row of a data sample sheet (e.g.EXCEL file) can be viewed as a source of field/clinical/remote sampled data attributes and each column as a predictor variable (e.g., characteristic or property of each data source). Typically, the malarialogist/experimenter enters the district-sampled data and edits it in SPSS after establishing the names and other properties of the sampled explanatory hyperendemic transmission oriented covariate coefficients in the empirical dataset using Variable View. He or she thereafter, routinely clicks on the Variable View tab to define the names and other properties of each sampled variable in the dataset. In this fashion, each district-level time series malarial-related hyperendemic transmission oriented geopredictive variables would be represented as a row, and various properties of the variable would be represented as columns allowing the malarialogist/experimenter to change the properties of the existing field/clinical/remote-sampled data attributes or, to establish properties for new sampled variables.

The two basic types of district-level time series SPSS malaria-related geopredictive hyperendemic transmission oriented seasonal variables are numeric and string. Numeric variables may only have numbers assigned (e.g., district-level prevalence rates). String variables may contain letters or numbers but, even if a string variable happens to contain only numbers, numeric operations conducted on an empirical dataset of district-level time series field/clinical/remote-sampled hyperendemic transmission oriented variables will not be permitted in SPSS (e.g., finding the mean, variance, standard deviation, etc...). If a sampled district-level geopredictive time series hyperendemic transmission oriented numeric variable is selected, the malarialogist/experimenter can then just then click in the width box or, the decimal box in the database to change the default values characters reserved for displaying sampled numbers with multiple decimal places. For whole numbers, the decimals can even be dropped down to 0.

Alternatively, if a malarialogist/experimenter chooses a string district-level malaria-related time series geopredictive hyperendemic transmission oriented variable, SPSS can quantitate how much "room" to leave in the memory for of each sampled explanatory field/clinical/remote sampled covariate coefficient measurement value for indicating the number of characters to be allowed for data entry in the string variable. The width of the district-level geopredictive variable would thereafter be the number of characters SPSS will allow to be entered for the sampled variable. If it is a numerical district-level field/clinical/remote sampled hyperendemic transmission oriented value and has decimals (e.g., depth of a particular sampled malaria-related mosquito habitat), the total width grid cell will

include a spot for each decimal, as well as one for the decimal point. The malarialogist/experimenter may then change a width of a data numerical entry by clicking in the width cell for the desired explanatory covariate coefficient value or type in a new number or, use the arrow keys at the edge of the cell. If more decimals have been entered or computed by SPSS, the additional district-level malaria-related seasonal geopredictive information will be retained internally but, not displayed on screen. For whole field/clinical/remote sampled hyperendemic transmission oriented numbers, the malarialogist/experimenter may choose to even reduce the number of decimals to zero for regressing the ecological empirical datasets parsimoniously.

In SPSS the label of a seasonal geopredictive field/clinical/remote sampled malaria-related hyperendemic transmission oriented variable then would be a string of text to identify what a district-level variable actually statistically represents. Unlike the name, the label is limited to 255 characters and may contain spaces and punctuation. (http://my.ilstu.edu/~mshesso/SPSS/data).For instance, if there is a district-level sampled georefernced hyperendemic transmission oriented geopredictive variable for each question on a field-sample sheet, a malarialogist/experimenter could type the question (e.g., What is the district-level weekly rainfall rate?) as the field/clinical/remote variable label in SPSS. Although the variable label will explain what the sampled district level malarial-related geopredictive time series explanatory field/clinical/remote sampled hyperendemic transmission oriented variable linearly represents, for categorical data (e.g., discrete data of both nominal and ordinal levels of measurement), commonly the information required for constructing a robust malarial-related risk model would be based on which explanatory hyperendemic transmission oriented covariate coefficient values represent which field-sampled categories. To indicate how these numbers are assigned in SPSS, a malarialogist/experimenter would then add labels to specific seasonal-sampled hyperendemic transmission oriented covariate coefficient measurement values by clicking on the box in the values cell. The real value of the district-level field/clinical/remote sampled hyperendemic transmission oriented labels could then be seen in the Data View by clicking on the "toe tag" icon in the tool bar which would then subsequently switch between the numeric values and their labels in the classified dataset.

Importantly, even though there will be some numerical codes recorded in SPSS for each empirical-sampled district-level time series field/clinical/remote sampled malaria-related hyperendemic transmission oriented data attribute, SPSS can be signaled to treat the sampled data as missing. For example, SPSS could simply display a single sampling period (e.g., SYSTEM MISSING data). After clicking on the ... button in the missing cell and then declaring "9", "99", and "999", SPSS would then treat the district-level sampled field/clinical/remote malaria-related hyperendemic transmission oriented geopredictive variables as missing (i.e., these values will be ignored). The columns property would then tell SPSS how wide the column should be for each sampled district-level variable. The column size would then indicate how much space is allocated rather than the degree to which it is filled. Routinely, the alignment property would indicate whether the district-level field/clinical/remote sampled malaria-related information in the Data View should be left-justified, right-justified, or centered. Thereafter, the Measure property would indicate the level of the sampled explanatory hyperendemic transmission oriented covariate coefficient measurement values. Since SPSS does not differentiate between interval and ratio levels for variable measurements, both of these district-level malaria-related seasonally quantitative field/clinical/remote sampled hyperendemic transmission oriented variable types would then be lumped together as "scale". Nominal and ordinal levels of the measurements however, would be differentiated in the empirical dataset.

Additionally, in SPSS, independent time series district-level geopredictive malaria-related explanatory field/clinical/remote sampled hyperendemic transmission oriented explanatory covariate coefficient dataset specified on the Variables tab can be explicitly also included in any seasonal estimated model. This is in contrast to the Expert Modeler where the independent variables would only be included, if they have a statistically significant relationship with the dependent variable (e.g., district-level malarial prevalence rates). Fortunately, SPSS will allow entry of multiple district-level sampled malarial-related time series hyperendemic transmission oriented variables into a regression in blocks, prior to the stepwise regression. If the malarialogist/experimenter does not block the filed/clinical/remote independent variables or, uses stepwise regression, a column will be created listing all of the independent variables specified. This column would then specify the method that SPSS will use to run the regression.

Routinely, time series district-level field/clinical/remote sampled hyperendemic transmission oriented exploratory observational geopredictors are added to a malaria-related risk model in a stepwise fashion. The geopredictor is then tested to determine levels of variance in the dependent variable (e.g., district-level stratified prevalence rates) that occur simply due to chance. A malarialogist/experimenter would then continue to add more hyperendemic transmission oriented geopredictors to the model which in most circumstances would improve the ability of the explanatory coefficients to explain the dependent variable, although some of this may cause an increase in R^2 simply due to chance variation in that particular sample. The adjusted R^2 in SPSS may then yield a more honest value to estimate the R^2 for the empirical sampled district-level field/clinical/remote sampled estimator dataset. Adjusted R^2 can then be computed using the formula 1 - ((1 - Sq.)(N - 1)/ (N - k - 1). Commonly in this formula when the number of regressable district-level time series empirical sampled hyperendemic transmission oriented observations is small and the number of exploratory predictors is large, there will be a much greater difference between R^2 and adjusted R^2 as the ratio of (N - 1) / (N - k - 1) will be much greater than 1. By contrast, when the number of district-level time series field/clinical/remote seasonal-sampled hyperendemic transmission oriented observations is very large in an empirical malaria-related empirical ecological dataset compared to the number of district-level sampled exploratory predictors, the value of R^2 and adjusted R^2 will be much closer as the ratio of (N - 1)/(N - k - 1) will approach 1. Thereafter, routinely a standard error of the estimate would be rendered in the SPSS Annotated SPSS Output. The standard error of a robust geopredictive district-level field/clinical/remote sampled malaria-related model hyperendemic transmission oriented residual forecast estimate, [i.e., the root mean square error,(RSME)] would then be the standard deviation of the error term, which is the square root of the Mean Square Residual (or Error)(see Jacob et al. 2005b).

Alternatively, a malarialogist and/or an experimenter could specify a custom exponentially weighted autoregressive integrated moving average (ARIMA) or exponential smoothing in SPSS for constructing a hyperendemic transmission oriented robust geopredictive district-level malaria-related time series regression model. Fortunately, the ARIMA time series models form a general class of linear models which are widely used in autoregressive risk modeling for forecasting time series. The purpose of ARIMA methods for time series district-level malaria-related risk modeling then would be to fit a stochastic randomly determined district-level geopredictive model to a given set of time series district-level field/clinical/remote sampled hyperendemic transmission oriented data attributes, such that the model can closely approximate the process that is actually generating the data. Given a time series of district-level malaria-related data attributes $X_t$where $t$is an integer index and the X_t are the sampled explanatory hyperendemic transmission oriented covariate coefficient values, then an ARIMA(p' ,q) model can be provided 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, $\alpha_i$are the parameter estimators of the autoregressive part of the model and $\theta_i$are the geopredictive estimators of the moving average part and $\varepsilon_t$are the forecasted error terms. These error terms $\varepsilon_t$are generally assumed to be independent, identically distributed (i.d.d) variables sampled from a normal distribution with zero mean. If then a malarialogist/experimenter assumes now that the polynomial $\left(1-\sum_{i=1}^{p'}\alpha_i L^i\right)$has a unitary root of multiplicity d, then it can be rewritten as: $\left(1-\sum_{i=1}^{p'}\alpha_i L^i\right)=\left(1-\sum_{i=1}^{p'-d}\phi_i L^i\right)(1-L)^d$. An ARIMA($p$,$d$,$q$) process in a time series geopredictive malaria-related district-level risk model would express this polynomial factorization property with $p=p'-d$, which subsequently thereafter could be described by: $\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t=\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$. By so doing the geopredictive district-level risk model may be thought as a particular case of an ARMA($p+d$,q) process having the autoregressive polynomial with d unit roots. The model can then be generalized as follows $\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t=\delta+\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$ (see Box and Jenkins 1976)

By so doing, an ARIMA(p,d,q) process in a district-level geopredictive time series field/clinical/remote sampled malaria-related risk model with drift $\delta/(1-\Sigma\varphi_i)$ can be robustly constructed.Thereafter, a district-level geopredictive ARIMA (p,d,q) risk model of the time series $\{x_1, x_2, \dots\}$ may be employed to quantitate empirical ecological georeferenced malaria-related field/clinical/remote sampled hyperendemic oriented observational exploratory predictors by employing $\Phi_p(B)\Delta^d x_t = \Theta_q(B)\epsilon_t$ where B is the backward shift operator, $Bx_y=x_{y-1}$, $\Delta = 1 - B$ is the backward difference and where Φ_p and Θ_q are polynomials of order p and q, respectively. In SPSS, ARIMA (p,d,q) models are the product of an autoregressive part [e.g.AR(p)] $\Phi_p = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, an integrating part $I(d) = \Delta^{-d}$ and a moving average MA(q) part $\Theta_q = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ respectively (see Box and Jenkins, 1976). Before undertaking these steps, however, a malarialogist/experimenter must be certain that the geopredictive time series is stationary in the ecological empirical dataset parameter estimators. That is, in the district-sampled datasets the covariance between any two seasonal sampled field/clinical/remote explanatory hyperendemic transmission oriented covariate coefficient measurement values of the time series must be dependent upon only the time interval between those particular sampled values and not on their absolute geolocation in time. A district-level malaria-related geopredictive ARIMA model can then be viewed as a "cascade" of two models: The first would be non-stationary: $Y_t = (1 - L)^d X_t$ while the second would be wide-sense stationary: $\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) Y_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$. where explanatory hyperendemic transmission oriented forecasts are routinely made for the process Y_t.

Some authors, including Jacob et al. (2013b), and Griffith (2005) employed a different convention for risk assessing multiple geopredictive autoregressive vector arthropod-related coefficients in time series models. For instance, in their models district-sampled explanatory covariate coefficient non-constant variance estimates were removed from empirical datasets by performing natural log transformations. Thereafter, the authors removed the trend in the sampled series by quantitating first difference values in the empirical dataset. If very large autocorrelations were then observed at lags spaced n periods apart, the authors determined that they had evidence of periodicity in the forecasted estimates. The objective of the identification stage then was to identify the autocorrelation uncertainty coefficients throughout seasonal differencing at a selected sample period and then rigorously quantitate any residual error variance in the forecasts employing an eigenfunction decomposition algorithm. By so doing, the authors allowed all the polynomials from the models involving the lag operator to appear in a similar form throughout the residually forecasted estimates.

Similarly, a time series district-level malaria-related geopredictive malarial-related SPSS derived ARIMA risk model could be written as $\left(1 + \sum_{i=1}^{p} \phi_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$. Then the models can, after choosing p and q, be fitted by least squares regression to determine the seasonal-sampled explanatory district-level field/clinical/remote sampled hyperendemic transmission oriented covariate coefficient statistical significance. Thereafter, the exact likelihood could be computed via a state-space representation of the ARIMA process, and the innovations and their variance could then be found by a Kalman filter.

The Kalman filter, also known as linear quadratic estimation (LQE), is an algorithm that employs a series of measurements observed over time, containing noise (e.g., district-level geopredictive malaria-related empirical random variations) and other inaccuracies, while simultaneously producing estimates of unknown variables that tend to be more precise than those based on a single measurement alone. More formally, the Kalman filter operates recursively on streams of noisy input data to produce a statistically optimal estimate of the underlying system

state.The initialization of the differenced ARIMA process employs stationarity and is based on Gardner et al. (1980). For a differenced process the non-stationary components in a malaria-related geopredictive autoregressive district-level risk model may be given by a diffuse prior controlled by Kappa. These methodologies can be defined as the prior variance computated by a multiple of the innovations variance tabulated from an empirical ecological dataset of field/clinical/remote sampled regressors used to construct differenced district-level hyperendemic transmission oriented risk models. District-level malaria- related field/clinical/remote sampled time series geopredictive observations which are still controlled by the diffuse prior, as determined by having a specific Kalman gain (e.g. 1e4), can then be excluded from the likelihood calculations. For ARIMA models with differencing, the differenced series will follow a zero-mean ARMA model.

If a 'dreg' term is included, in SPSS a linear regression (with a constant term if 'include mean' is true) will be fitted with a geopredictive time series ARMA model for the error term. The differenced series variance matrix for the ARIMA models will then follow a zero-mean ARMA model. If a 'xreg' term is also included in SPSS, a linear regression (with a constant term if 'include.mean' is true) will be subsequently fitted with an ARMA model for the error term. The variance matrix of the estimates may then be found from the Hessian of the log-likelihood, estimates. By so doing, the estimators would subsequently minimize the error term in the district-level dataset of regressed residually forecasted explanatory hyperendemic transmission oriented covariate coefficient estimates.

As such, a seasonal ARIMA-related SPSS derived district-level geopredictive malaria-related regression-based risk model then would simply be an ARIMA(*p*,*d*,*q*) model where the sampled parameters *p*, *d*, and *q* are non-negative integers. These integers would then be related to the order of the AR, integrated and MA parts of a robust geopredictive district-level malaria-related regression-based hyperendemic transmission–oriented risk model residually forecasted components respectively. SPSS could then combine serially correlated methods in the AR and MA into a composite model of the time series for deriving statistically significance of each sampled district-level explanatory hyperendemic transmission–oriented covariate. The risk model residually forecasted estimates in SPSS could then be additionally regressed for quantitating any latent autocorrelation error coefficients and partial autocorrelation uncertainty error coefficient estimates in the district-level time series forecasted geopredictive malarial data attributes.

Partial autocorrelations measure the degree of association between various lags when the effects of other lags are removed (Griffith 2003). If the autocorrelation between Y_t and Y_{t-1} in a district-level geopredictive time series malaria-related risk model is significant, this would signify a similar significant autocorrelation between Y_{t-1} and Y_{t-2}, as they would just one period apart in the autocovariate error matrix. Since both Y_T and Y_{t-2} would be both correlated with Y_{t-1} in the district-level risk model, they would also be correlated with each other. Therefore, by removing the effect of Y_{t-1}, a malarialogist/exprimenter could measure the true correlation between Y_t and Y_{t-2}.Additionally, a partial autocorrelation coefficient of order k can be determined by regressing the sampled time series geopredictive explanatory district-level field/clinical/remote sampled malaria-related hyperendemic trasnmission oriented covariate coefficient measurement values by its lagged values employing $Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \cdots + b_k Y_{t-k}$ (see Box and Jenkins 1976). This form of a seasonal malaria-related regression-based equation would be an (AR) one, since its independent variables would be time-lagged values of the dependent variable. This multiple regression could then identify the partial autocorrelation (i.e,AK) in the risk model district-level field/clinical/remote malarial related residual forecasts. If the malarialogist/experimenter regresses Y_T only against Y_{t-1} in the district-level risk model, then a value for α_1. would be rendered. If Y_t against both Y_{t-1} and Y_{t-2} are regressed in the model, then the values for both α_1 and α_2 would be derived. These partial autocorrelation district-level malaria–related geopredictive uncertainty coefficients can be plotted in SPSS. This plot would be a partial autocorrelation function of the regressed district-level malaria-related explanatory field/clinical/remote sampled hyperendemic transmission oriented covariate coefficients.

Additionally, a malarialogist/experimenter may find the autocorrelation and partial autocorrelation plots in SPSS very helpful for residually quantitating latent forecasted field/clinical/remote sampled uncertainty error estimators in any seasonal malarial related district-level empirical dataset. For instance, the Forecasting optional add-on module in SPSS can provide multiple residual analytic techniques for identifying non-normal seasonal-sampled district-level explanatory hyperendemic transmission-oriented uncertainty estimators. This would include creation of summary

plots across time series parameter estimator model outputs including histograms of stationary *R*-square, R^2, root mean square error (RMSE), mean absolute percentage error MAPE, maximum absolute error (MaxAE), maximum absolute percentage error (MaxAPE), and normalized Bayesian information (BIC) criterion with box plots of residual autocorrelations and partial autocorrelations.

The Bayesian information criterion (BIC) (Schwarz, 1978) or Schwarz criterion (also SBC, SBIC) is a criterion for model selection among a finite set of models. The criterion was derived to serve as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model. Although the original derivation measures that the observed data as i.d.d. arising from a probability distribution in a regular exponential family, BIC has been traditionally employed in a much larger scope of model selection. To better justify the widespread applicability of BIC, a malarialogist/experimenter may, for example, choose to derive the information-theoretic criterion in a very generalized district-level geopredictive framework, one that does not assume any specific form for the likelihood function, but only requires that it satisfies certain non-restrictive regualrity conditions. For instance , a Bayesian information criterion for singular district-level time series geopredictive malaria-related risk models may be proposed. The malarialogist/experimenter may then consider approximate Bayesian model choice for model selection problems that involve models whose Fisher-information matrices may fail to be invertible along other competing district-level malaria-related submodels.

In mathematical statistics and information theory, the Fisher information is the variance of the score, or the expected value of the observed information (Edgeworth 1908).. In Bayesian statistics, the asymptotic distribution of the posterior mode depends on the Fisher information and not on the prior (according to the Bernstein–von Mises theorem, which was anticipated by Laplace for exponential families). The role of the Fisher information in the asymptotic theory of MLE was emphasized by the statistician R.A. Fisher (following some initial results by F. Y. Edgeworth). The Fisher information is also used in the calculation of the Jeffreys prior, which is used in Bayesian statistics. The Jeffreys prior, , is a non-informative (objective) prior distribution on parameter space that is proportional to the square root of the determinant of the Fisher information: $p\left(\vec{\theta}\right) \propto \sqrt{\det \mathcal{I}\left(\vec{\theta}\right)}$. It has the key feature that it is invariant under reparameterization of the parameter vector $\vec{\theta}$. This makes it of special interest for use with scale parameters The Fisher-information matrix is used to calculate the covariance matrices associated with maximum-likelihood estimates. It can also be used in the formulation of test statistics, such as the Wald test (Frieden 2004).

Thus, Fisher information would be a way of measuring the amount of information that a sampled geopredictive malaria-related district-level hyperendemic transmission oriented observable random variable *X* carries about an unknown parameter θ upon which the probability of *X* would depend. The probability function for *X*, which would also be the likelihood function for θ in the malaria-related risk model would then be a function $f(X; \theta)$ as it would be the probability mass (or probability density) of the sampled district-level random variable *X* conditional on the value of θ. The partial derivative with respect to θ of the natural logarithm of the likelihood function in the risk model residual forecasts then would be based on the score. Under certain regularity conditions, it may be shown that the first moment of the score in a geopredictive district-level malaria-related risk model (that is, its expected value) is 0:

$$\mathrm{E}\left[\left.\frac{\partial}{\partial\theta}\log f(X;\theta)\right|\theta\right] = \mathrm{E}\left[\left.\frac{\frac{\partial}{\partial\theta}f(X;\theta)}{f(X;\theta)}\right|\theta\right] = \int \frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)} f(x;\theta)\,dx =$$

[e.g.

$$= \int \frac{\partial}{\partial\theta} f(x;\theta)\,dx = \frac{\partial}{\partial\theta}\int f(x;\theta)\,dx = \frac{\partial}{\partial\theta}\,1 = 0.$$

].The second moment (i.e., the Fisher information) would then be

$$\mathcal{I}(\theta) = \mathrm{E}\left[\left.\left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2\right|\theta\right] = \int \left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2 f(x;\theta)\,dx\,,$$

where, for any

field/clinical/remote sampled hyperendemic transmission oriented value of θ, the expression E[...|θ] would denote the conditional expectation over the sampled district-level values for X with respect to the probability function $f(x; \theta)$ given θ. Note that $0 \leq \mathcal{I}(\theta) < \infty$ (see Cressie 1993). District-level malaria-related sampled random variable carrying high Fisher information would then imply that the absolute value of the score is high. The Fisher information is not a function of a particular observation, as the random variable X has been averaged out (Gilks 1996) .Since the expectation of the optimal score would be zero for a time series malaria-related geopredictive risk model, the Fisher information would also t be he variance of the score. If $\log f(x; \theta)$ is twice differentiable with respect to θ in the geopredictive risk model and under certain regularity conditions, then the Fisher information may also be written $\mathcal{I}(\theta) = -\operatorname{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\middle|\theta\right]$, since $\frac{\partial^2}{\partial\theta^2}\log f(X;\theta) = \frac{\frac{\partial^2}{\partial\theta^2}f(X;\theta)}{f(X;\theta)} - \left(\frac{\frac{\partial}{\partial\theta}f(X;\theta)}{f(X;\theta)}\right)^2 = \frac{\frac{\partial^2}{\partial\theta^2}f(X;\theta)}{f(X;\theta)} - \left(\frac{\partial}{\partial\theta}\log f(X;\theta)\right)^2$ as and $\operatorname{E}\left[\frac{\frac{\partial^2}{\partial\theta^2}f(X;\theta)}{f(X;\theta)}\middle|\theta\right] = \ldots = \frac{\partial^2}{\partial\theta^2}\int f(x;\theta)\,dx = \frac{\partial^2}{\partial\theta^2}\,1 = 0.$ Thus, the Fisher information would essentially be the negative of the expectation of the second derivative with respect to θ of the natural logarithm of f in a robust geopredictive malaria-related hyperendemic transmission oriented risk model Information may be seen to be a measure of the "curvature" of the support curve near the maximum likelihood estimate of θ. A "blunt" support curve (i.e., one with a shallow maximum) would have a low negative expected second derivative, and thus low information; while a sharp one would have a high negative expected second derivative and thus high information. Information is additive, in that the information yielded by two independent experiments is the sum of the information from each experiment separately: $\mathcal{I}_{X,Y}(\theta) = \mathcal{I}_X(\theta) + \mathcal{I}_Y(\theta)$. (see Frieden 2004). This result follows from the elementary fact that if malarial-related random variables are independent, the variance of their sum is the sum of their variances. Hence, the information in a random sample of size n is n times that in a sample of size 1 (if sampled district-sampled malaria-related observations are i.d.d.).The information provided by a sufficient statistic is the same as that of the sample X. This may be seen by using Neyman's factorization criterion in SPSS or SAS/GIS for a sufficient statistic.

Fisher's factorization theorem or factorization criterion provides a convenient characterization of a sufficient statistic. For example, if the probability density function (pdf) is $f_\theta(x)$ in a malaria-related model, then T is sufficient for θ if and only if nonnegative functions g and h can be found such that. the density f can be factored into a product such that one factor, h, does not depend on θ and the other factor, which does depend on θ, depends on x only through T(x). Thus, if T(X) is sufficient for θ in a geopredictive malaria-related district-level risk model , then $f(X;\theta) = g(T(X),\theta)h(X)$ for some functions g and h. The equality of information then follows from the following fact: $\frac{\partial}{\partial\theta}\log\left[f(X;\theta)\right] = \frac{\partial}{\partial\theta}\log\left[g(T(X);\theta)\right]$ which follows from the definition of Fisher information, and the independence of h(X) from θ. More generally, if T = t(X) is a statistic in a geopredictive sampled hyperendemic transmission oriented malaria-related model, then $\mathcal{I}_T(\theta) \leq \mathcal{I}_X(\theta)$ with equality if and only if T is a sufficient statistic. If $X(\mathbf{x}) = X(x_1, x_2, \ldots, x_n)$ is a random vector in $\mathbb{R}^n$ and l $f_X(\mathbf{x})$ is a probability distribution on X with continuous first and second order partial derivatives, the Fisher information matrix of X would be the $n \times n$ matrix J_X whose (i, j)th entry would be given by $(J_X)_{i,j} = \left\langle \frac{\partial \ln f_X(\mathbf{x})}{\partial x_i} \frac{\partial \ln f_X(\mathbf{x})}{\partial x_j} \right\rangle = \int_{\mathbb{R}^n} \frac{\partial \ln f_X(\mathbf{x})}{\partial x_i} \frac{\partial \ln f_X(\mathbf{x})}{\partial x_j} f_X(\mathbf{x})\, d^n\mathbf{x}.$ (Papathanasiou, 1993).

The formula for the BIC is: $-2 \cdot \ln p(x|M) \approx \mathrm{BIC} = -2 \cdot \ln \hat{L} + k \ln(n)$ (Akaike 1974). Under the assumption that the model errors or disturbances are i.d.d according to a normal distribution and that the boundary condition of the derivative of the log likelihood with respect to the true variance is zero, this becomes $\mathrm{BIC} = n \cdot \ln(\widehat{\sigma_e^2}) + k \cdot \ln(n)$ based on an additive constant, which depends only on n and not on the

model where $\widehat{\sigma_e^2}$ is the error variance. The error variance in the time series geopredictive seasonal district-level autoregressive malarial-related risk model would then be defined as $\widehat{\sigma_e^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2.$

Commonly singular models do not obey the regularity conditions underlying the derivation of BIC and the penalty structure in BIC generally does not reflect the frequentist large-sample behavior of their marginal likelihood. While large-sample theory for the marginal likelihood of singular geopredictive district-level malaria-related models has been developed recently , the resulting approximations still are highly dependent on the true sampled parameter estimator value which can lead to a paradox of circular reasoning. Guided by examples such as determining the number of components of mixture malaria-related district-level risk models, the number of factors in latent factor models or the rank in reduced-rank regression may instead be proposed as a resolution to this paradox for rendering a practical extension of BIC for singular district-malaria-related geopredictive risk model selection problems. The model however would be based, in part, on the likelihood function of the residually forecasted estimates and thus it would be closely related to the Akaike information criterion (Akaike 1974).

The AIC is a measure of the relative quality of a statistical model, founded on information entropy,for a given set of data which is quantitated by the trade-off between the goodness of fit of the model and the complexity of the model (Akaike, 1974),).When fitting seasonal district-level geopredictive field/clinical/remote malaria-related risk models, it is possible to increase the likelihood by adding sampled explanatory hyperendemic transmission oriented estimators but, doing so may result in overfitting (see Jacob et al. 2011b, Jacob et al. 2009d). In relevance to chi-squared(χ^2) fitting for district-level geopredictive risk modeling, if a malarialogist/experimenter wishes to select amongst competing models where the likelihood functions assume that the underlying errors are normally distributed with mean zero and independent, a χ^2 model fitting may be employed .For χ^2 fitting for the geopredictive malaria-related district-level model, the likelihood would then be given by $L = \prod_{i=1}^{n}\left(\frac{1}{2\pi\sigma_i^2}\right)^{1/2}\exp\left(-\sum_{i=1}^{n}\frac{(y_i - f(\mathbf{x}))^2}{2\sigma_i^2}\right) \therefore \ln(L) = \ln\left(\prod_{i=1}^{n}\left(\frac{1}{2\pi\sigma_i^2}\right)^{1/2}\right) - \frac{1}{2}\sum_{i=1}^{n}\frac{(y_i - f(\mathbf{x}))^2}{\sigma_i^2} \therefore \ln(L) = C - \chi^2/2,$ where *C* would be a constant independent of the risk model, and dependent only on the use of particular sampled explanatory time series hyperendemic transmission-oriented data points.(i.e. those points that does not change if the data does not change).The AIC can then be given by $AIC = 2k - 2\ln(L) = 2k - 2(C - \chi^2/2) = 2k - 2C + \chi^2$. As only differences in AIC are meaningful, the constant *C* can then be ignored, allowing the malarialogist/experimenter to take $AIC = \chi^2 + 2k$ for model comparisons. Another convenient form arises also if the σ_i are assumed to be identical and the residual sum of squares (RSS) is available. Then a malarialogist/experimenter would achieve AIC = n ln(RSS/n) + 2k + C, where again C can be ignored in model comparisons. Fortunately, both BIC and AIC can resolve this problem by introducing a penalty term for the number of parameter estimators in the risk model.

Penalized regression methods in SPSS for simultaneous variable selection and coefficient estimation, especially those based on the lasso of Tibshirani (1996), have received a great deal of attention in recent years, mostly through frequentist models. Properties such as consistency in district-level time series geopredictive malaria-related risk-based data attributes have been studied, and are achieved by different lasso variations (Jacob et al. 2009d). Within such an SPSS derived autoregressive district-level risk related model framework, a malarialogist/experimenter may look at a fully Bayesian formulation which may then reveal flexiblity enough to encompass most versions of the lasso that have been previously considered in statistical and ArcGIS literature.

The advantages of the hierarchical Bayesian formulations for quantitating district-level time series malaria-related model geopredictive parameter estimators in SPSS would then be many. For instance, Bayesian Network Model Nuggets may be able to accommodate and quantitate multiple geopredictive time series empirical sampled explanatory field/clinical/remote hyperendemic transmission oriented covariate coefficients efficiently in a probabilistic directed acyclic graphical model. This probabilistic graphical model can represent a empirical dataset

of district-level t geopredictive time series hyperendemic transmission oriented random variables and their conditional dependencies via a directed acyclic graph (DAG).

In ArcMap™ software a DAG can be defined as a directed district-level geopredictive malaria-related graph with no directed cycles. That is, it may be formed by a collection of vertices and directed edges, each edge connecting one vertex to another, such that there is no way to start at some vertex v and follow a sequence of edges that eventually loops back to v again. Each directed acyclic district-level geopredictive malaria risk -related graph would then give rise to a partial order $\leq$ on its vertices, where $u \leq v$ occur when there exists a directed path from u to v in the DAG. However, many different DAGs may give rise to this same reachability relation in a district-level geopredictive malaria-related risk model. For example, a DAG with two edges $a \rightarrow b$ and $b \rightarrow c$ in a district-level malaria-related risk model output would have the same reachability as the graph with three edges $a \rightarrow b$, $b \rightarrow c$, and $a \rightarrow c$. Further, if G is a DAG in the risk model, its transitive reduction would then be the graph with the fewest edges which then would represent the same reachability as G, and its transitive closure could then be the district-level graph with the most edges that represents the same reachability.

Further, in ArcMap™ the transitive closure of G would have an edge $u \rightarrow v$ for every related pair $u \leq v$ of distinct sampled hyperendemic transmission oriented elements in the reachability relation of G, and may therefore be thought of as a direct translation of the reachability relation $\leq$ into graph-theoretic terms. Thus, every partially ordered district-level seasonal malaria-related empirical dataset may be translated into a DAG in such a manner. If a DAG G then represents a partial order $\leq$ in the district-level geopredictive risk model then the transitive reduction of G would be a subgraph of G with an edge $u \rightarrow v$ for every pair in the covering relation of $\leq$. In such circumstances, transitive reductions in ArcMap™ would be a useful in visualizing the partial orders of the geosampled SPSS derived district-level malarial attributes they represent, because they would have fewer edges than other geopredictive graphs representing the same orders and this would lead to simpler graph drawings. A Hasse diagram, for instance of a partial order may be generated which is a drawing of the transitive reduction in which the orientation of each edge is shown by placing the starting vertex of the edge in a lower position than its ending vertex (see Griffith 2003).

Fortunately, every directed SPSS constructed time series district-level malaria-related geopredictive ArcMap™ acyclic graph would have a topological ordering. This is an ordering of the vertices such that the starting endpoint of every edge in the malaria-related risk model would occur earlier in the ordering than the ending endpoint of the edge. In general, this ordering is not unique for district-level malaria-related geopredictive risk models; a DAG has a unique topological ordering, if and only if, it has a directed path containing all the vertices, in which case the ordering is the same as the order in which the vertices appear in the path (see Cressie 1993). The family of topological orderings of a DAG is the same as the family of linear extensions of the reachability relation for the DAG, so any two graphs representing the same partial order have the same set of topological orders (Griffith 2003). Topological sorting for district-level time series malaria- related geopredictive risk modeling is the algorithmic problem of finding topological orderings; it can be solved in linear time (see Jacob et al. 2009d). It may also then be also possible to check whether a given directed graph is a DAG in linear time, for a robust geopredictive malaria-related district-level model by attempting to find a topological ordering and then testing whether the resulting ordering is valid.

Importantly, some algorithms become simpler when used on DAGs in ArcMap™ instead of general graphs, based on the principle of topological ordering. For instance, it may be possible to find shortest paths and longest paths from a given starting vertex in DAGs in linear time for a SPSS derived district-level geopredictive malaria-related model by processing the vertices in a topological order, and calculating the path length for each vertex in ArcMap™ to be either the minimum or maximum length obtained via any of its incoming edges. In contrast, for arbitrary district-level malaria-related risk graphs the shortest path may require slower algorithms such as Dijkstra's algorithm as longest paths in arbitrary geopredictive graphs are hard to find.

Dijkstra's algorithm is an algorithm for finding a graph geodesic, (i.e., the shortest path between two graph vertices in a graph). It functions by constructing a shortest-path tree from the initial vertex to every other vertex in a graph(ArcMap™ malaria-related geopredictive district-level time series).The algorithm is implemented as

Dijkstra[*g*] in the SPSS package`. The worst-case running time for the Dijkstra algorithm on a graph with *n* nodes and *m* edges is $O(n^2)$ as it allows for directed cycles. The algorithm will find the shortest paths from a source node *S* to all other nodes in time series malaria-related district-level graph. For a robust malaria-related geopredictive model this may be illustrated as $O(n^2)$for node selection and $O(m)$for distance updates. While $O(n^2)$is the best possible complexity for dense malaria time series district-level graphs, the complexity can be improved significantly for sparse graphs in SAS/GIS With slight modifications, Dijkstra's algorithm can be also used as a reverse algorithm in the database which can help maintain minimum spanning trees for the sink node in a district-level malaria-related geopredictive risk model. With further modifications, it can be extended to become bidirectional. The bottleneck in Dijkstra's algorithm in SAS/GIS is node selection (www.sas.edu).

DAG representations of partial orderings have many applications in district-level geopredictive time series malaria-related model in scheduling problems for systems of tasks with ordering constraints. For instance, a DAG may be employed to describe the dependencies between cells of a field/clinical/remote sampled district-level spreadsheet. Further, if one cell is computed by a formula involving the sampled hyperendemic transmission oriented value of a second cell, a DAG edge may be drawn from the second cell to the first one. If the input district-level geopredictive explanatory hyperendemic transmission covariate coefficient values to the spreadsheet change, all of the remaining sampled values of the spreadsheet may be recomputed with a single evaluation per cell, by topologically ordering the cells and re-evaluating each cell in an ordered format. Dependency graphs without circular dependencies form directed acyclic graphs (Cressie 1993).

Thus, employing DAG, a Bayesian network could represent the probabilistic relationships between regressed seasonal sampled statistically significant georeferenced explanatory hyperendemic transmission oriented covariate coefficients and geolocations of district-level disease transmission. Formally, Bayesian networks are directed acyclic graphs whose nodes represent random variables in the Bayesian sense: they may be observable quantities, latent variables, unknown parameters or hypotheses (Griffith 2003). Edges may then represent conditional dependencies in a district-level geopredictive malaria-related risk model; nodes which are not connected may then represent sampled hyperendemic transmission oriented variables which are conditionally independent of each other. Each node would then be associated with a probability function that would take as input a particular set of sampled district-level hyperendemic transmission oriented covariate coefficient values for the node's parent variables and thereafter render the probability of the geopredictive variable represented by the node. For instance, Jacob et al. (2011d) found that when the parents are m Boolean variables in a spectral endmember georeferenced aquatic larval habitat of *Anopheline arabiensis*, a major vector of malaria in a riceland agroecosystem in Mwea, Kenya, then the probability function could be represented by a table of 2^m entries, one entry for each of the 2^m possible combinations of its parents being true or false.

In programming languages that have a built-in Boolean data type, such as Pascal and Java, the comparison operators such as > and ≠ are usually defined to return a Boolean value. Conditional and iterative commands may be then defined to test Boolean-valued district-level time series geopredictive malaria-related risk model expressions. Languages without an explicit Boolean data type, like C90 and Lisp, may then geographically represent truth values in a robust time series geopredictive malaria-related district-level model by some other data type. Lisp uses an empty list for false, and any other value for true . C uses an integer type, where geopredictive relational expressions like i > j and logical expressions connected by && whereby || are defined to have a sampled district-level explanatory malaria-related hyperendemic transmission oriented covariate coefficient value 1 if true and 0 if false, whereas the test parts treat any non-zero value as true (see Kernighan and Ritchie 1978) Indeed, a robust Boolean malaria-related hyperendemic transmission oriented district-level geopredictor variable may be implemented as a numerical variable with a single binary digit (bit), which unfortunately can store only two values currently.

It is worth noting that the implementation of booleans in computers are most likely represented as a full word, rather than a bit; this is usually due to the ways computers transfer blocks of information. Most programming languages, even those that do not have an explicit Boolean type, have support for Boolean algebraic operations for robust geopredictive malaria-related district-level risk modeling such as conjunction (AND, &, *), disjunction (OR, |, +), equivalence (EQV, =, ==), exclusive or/non-equivalence (XOR, NEQV, ^, !=), and not (NOT, ~, !). In some

languages, the "true" and "false" values belong to separate classes(e.g. True and False, resp.)so there is no single Boolean "type." In SQL, which uses a three-valued logic for explicit comparisons because of its special treatment of Nulls, a Boolean district-level geopredictive malaria-related data type may also be defined to include more than two truth values, so that SQL "Booleans" can store all logical sampled hyperendemic transmission oriented covariate coefficient measurement values resulting from the evaluation of predicates in SQL. The column of the Boolean type seasonal malaria-related risk model outputs can then be restricted to just TRUE and FALSE. By so doing, reforecasting ideas may be applied to undirected, and possibly cyclic, graphs employing Markov networks for quantitating empirical sampled field/clinical/remote district-level hyperendemic transmission explanatory covariate coefficients. Efficient algorithms exist for seasonal vector arthropod-related risk mapping that perform inference and learning in Bayesian networks (Jacob et al. 2012b,Jacob et al. 2011c,Griffith 2005).

As such, generalizations of Bayesian networks may represent and solve decision problems in SPSS under uncertainty when constructing a robust geopredictive district-level malaria-related risk model. For instance, suppose that there are two events which could influence an empirical -sampled malaria-related hyperendemic transmission oriented covariate coefficient (e.g., total density count of *Anopheline gambiae s.l.* aquatic larval habitat) levels of statistical significance: either be it during periods of drought or conversely during periods of high rainfall. Also, suppose that the rain has a direct effect on the status of the district-sampled malaria-related mosquito aquatic habitat's total larval density count namely that when it rains, the habitat has higher immature count values). Then the situation can be modeled with a Bayesian network since all three district-sampled time series malaria-related geopredictive variables would then have two possible values, T (for true) and F (for false). The joint probability function then could be expressed as $\mathrm{P}(G, S, R) = \mathrm{P}(G|S, R)\mathrm{P}(S|R)\mathrm{P}(R)$where the names of the variables abbreviated in SPSS to *G = dry(yes/no), S =high habitat density count (yes/no),* and *R = Raining (yes/no)*. The model could then answer questions like "What is the probability that when it is raining, a georeferenced district-sampled *An. gambiae s.l.* aquatic habitat has high larval abundance count by using the conditional probability formula and summing over all nuisance variables:

$$\mathrm{P}(R = T \mid G = T) = \frac{\mathrm{P}(G = T, R = T)}{\mathrm{P}(G = T)} = \frac{\sum_{S \in \{T,F\}} \mathrm{P}(G = T, S, R = T)}{\sum_{S,R \in \{T,F\}} \mathrm{P}(G = T, S, R)}$$

$$= \frac{\mathrm{P}(G = T, S = T, R = T)_{TTT} + \mathrm{P}(G = T, S = F, R = T)_{TFT}}{\mathrm{P}(G = T, S = T, R = T)_{TTT} + \mathrm{P}(G = T, S = T, R = F)_{TTF} + \mathrm{P}(G = T, S = F, R = T)_{TFT} + \mathrm{P}(G = T, S = F, R = F)_{TFF}}$$

$$= \frac{(P(G=T|S=T,R=T)P(S=T|R=T)P(R=T))_{TTT} + (P(G=T|S=F,R=T)P(S=F|R=T)P(R=T))_{TFT}}{(P(G=T|S=T,R=T)P(S=T|R=T)P(R=T))_{TTT} + (P(G=T|S=T,R=F)P(S=T|R=F)P(R=F))_{TTF} + (P(G=T|S=F,R=T)P(S=F|R=T)P(R=T))_{TFT} + (P(G=T|S=F,R=F)P(S=F|R=F)P(R=F))_{TFF}}$$

As is pointed out explicitly in the example numerator, the joint probability function may be used to calculate each iteration of the summation function, marginalizing over Sin the numerator, and marginalizing over $S$and R in the denominator of the geopredictive malaria-related risk model.

Alternatively, if a malarialogist/experimenter desires to answer an interventional question: "What is the likelihood that a sampled district-level geopredictive habitat would be classified productive during a short rain period ?" . The answer would then be governed by the post-intervention joint distribution function $\mathrm{P}(S, R|do(G = T)) = P(S|R)P(R)$obtained by removing the factor $\mathrm{P}(G|S, R)$from the pre-intervention distribution. As expected, the likelihood of a sampled district-level larval habitat being classified productive would then be unaffected by the action: $\mathrm{P}(R|do(G = T)) = P(R)$.If, moreover, the malarialogist/experimenter wishes to forecast the impact of other geosampled district-level estimators on a sampled larval habitat (drought-related geopredictive variables), he or she may then simply employ $P(R, G|do(S = T)) = P(R)P(G|R, S = T)$with the term $P(S = T|R)$removed.

However, these district-level SPSS-derived time series geopredictive malaria-related hyperendemic transmission oriented data attributes may not be feasible when some of the variables are unobserved. The effect of the action

$do(x)$ in the sampled district-level hypendemic trasnmission oriented variables can still be forecasted however; whenever a criterion called "back-door" is satisfied. For instance, when constructing a Bayesian Network employing multiple geosampled district-level malaria-related geopredictive risk variables , a set *Z* of nodes can be observed that *d*-separates (or blocks) all back-door paths from *X* to *Y* then $P(Y, Z|do(x)) = P(Y, Z, X = x)/P(X = x|Z)$. A back-door path is one that ends with an arrow into *X* (see Cressie 1993). For instance, the set *Z* = *R* is admissible for forecasting the effect of *S* = *T* on *G* in a robust district-level geopredictive hyperendemic transmission oriented model may since *R* *d*-separate is the only back-door path $S \leftarrow R \rightarrow G$. However, if *S* is not observed in the district-level malaria-related risk model there would be no other set that *d*-separates this path and the effect of (*S* = *T*) and thus (*G*) will not be forecasted from passive sampled district-level regressed malaria-related hyperendemic transmission oriented observations. A malarialogist/experimenter then may remark that *P*(*G*|*do*(*S* = *T*)) is not "identifiable." Further, a malarialogist/experimenter may not be able to quantitate if the observed dependence between *S* and *G* is due to a causal connection or is spurious (e.g., apparent dependence arising from a common cause, *R*) (e.g.,Simpson's paradox).To determine whether a causal relation in the district-level geopredictive risk model is identified from an arbitrary Bayesian network with unobserved field/clinical/remote sampled variables, a malarialogist/experimenter may instead employ the three rules of "*do*-calculus"and test whether all *do* terms can be removed from the expression of that relation, thus confirming that the desired quantity is estimable from frequency data(http://reference.wolfram.com/mathematica/howto/DoCalculus.html)

Using a Bayesian network in SAS/GIS can save considerable amounts of memory, if the dependencies in the joint distribution are sparse. For example, a naive way of storing the conditional probabilities of 10 two-valued district-level malaria-related geopredictive hyperendemic transmission variables as a table requires storage space for $2^{10} = 1024$ sampled values. If the local distributions of no geopredictive malaria-related variable depends on more than 3 parent variables, the Bayesian district-level network representation in SAS/GIS would only need to store at most $10 \cdot 2^3 = 80$ values.One advantage of Bayesian networks for district-level malaria-related geopredictive risk modeling is that it is intuitively easier for a human to understand a sparse set of direct dependencies and local distributions than complete joint distribution.

The model nugget Model tab in SAS/GIS is split into two panels for generating Bayesian inferences from geopredictive district-level malaria-related risk models. The left panel view contains a network graph of nodes that would then display the relationship between the target (e.g., geosampled district-level statistically significant ecozonal hyperendemic transmission covariate) and its most important observational predictors, as well as the relationship between the predictors. The importance of each predictor could be displayed in SAS/GIS by the density of its color; (e.g., a strong color shows an important sampled district-level hyperendemic transmission oriented predictor, and vice versa). The bin values for nodes representing a range of the district-level empirical sampled hyperendemic transmission oriented estimators would then be displayed in a pop up ToolTip when a malarialogist/experimenter hovers the mouse pointer over the node. Thereafter, the malarialogist/experimenter can employ the Modeler's graph tools in SAS/GIS to interact, edit, and save an district-level malaria-related graph. This data may then be exported and processed in other statistical/cartographic software, for use in other applications such a MS Word.

Commonly SAS/GIS will display conditional probabilities in a malaria-related geopredictive district-level model for each node in the network as a mini graph. By hovering the mouse pointer over a district-level geopredictive time series malaria-related graph its hyperendemic transmission oriented values in a popup ToolTip will be displayed. In the right panel observational geopredictor importance may be quantitated. The output panel would then display a chart that indicates the relative importance of each district-level observational predictor in estimating the seasonal malaria-related risk model. Further, conditional probabilities in the residually forecasted uncertainty estimates would be quantitated**.** When a malariologist/experimenter selects a node or mini distribution graph in the left panel in SAS/GIS, an associated conditional probabilities table may be also displayed in the right panel. This table would contain the conditional probability explanatory hyperendemic transmission oriented covariate coefficient measurement value for each node value and each combination of

values in its parent nodes. In addition, the output will include the number of records observed in SAS/GIS for each record value and each combination of values in the parent nodes.

In addition to the usual ease-of-interpretation of hierarchical models, the Bayesian formulation in SAS/GIS could produce valid standard errors (which can be problematic for the frequentist lasso) which would be based on a geometrically ergodic Markov chain. From a homogeneous Markov chain $\xi(t)$with the following property: $p_j = \lim_{t\to\infty} p_{ij}(t), \quad \sum_j p_j = 1,$ may then be expressed employing a georefernced dataset of empirical sampled field/clinical/remote explanatory hyperendemic trasnmission oriented covariate coefficients where $p_{ij}(t) = \mathsf{P}\{\xi(t) = j \mid \xi(0) = i\}$ are the transition probabilities. The distribution $\{p_j\}$on the state space of the chain $\xi(t)$ would then be stationary distribution. If $\mathsf{P}\{\xi(0) = j\} = p_j$for all j in the malaria-related district-level dataset, then $\mathsf{P}\{\xi(t) = j\} = p_j$for all $j$and $t \geq 0$. Fortunately, a fundamental property of Markov chains, $$\mathsf{P}\{\xi(t) = j\} = \sum_i \mathsf{P}\{\xi(0) = i\}\, p_{ij}(t),$$ enables one to find the $\{p_j\}$without calculating the limits (Gilks 1996). Let $\tau_{jj} = \min\{t \geq 1 : \xi(t) = j \mid \xi(0) = j\}$be the moment of first return to the state j(for a discrete-time Markov chain), then $\mathsf{E}\,\tau_{jj} = p_j^{-1}$ (Gilks 1996). A similar but more complicated relation holds for a continuous-time Markov chain in a geopredictive time series malaria-related risk model. The trajectories of an ergodic Markov chain satisfy the ergodic theorem: If $f(.)$is a function on the state space of the chain $\xi(t)$, then, in the discrete-time case, $\mathsf{P}\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{t=0}^{n} f(\xi(t)) = \sum_i p_j f(j)\right\} = 1,$ while in the continuous-time case the sum on the left is replaced by an integral (Freedman 1975). A Markov chain for which there are $\rho < 1$and $C_{ij} < \infty$such that for all i, j, t, $|p_{ij}(t) - p_j| \leq C_{ij}\,\rho^t$, is called geometrically ergodic(Kemeny and Snell 1960).

Interestingly, a sufficient condition for geometric ergodicity of an ergodic Markov chain is the Doeblin condition (which for a discrete (finite or countable) Markov chain may be stated as follows: There are an $n < \infty$and a state $j$such that $\inf_i p_{ij}(n) = \delta > 0$. If the Doeblin condition is satisfied, then in a predictive malaria-related model then for the constants the relation $\sup_{i,j} C_{ij} = C < \infty$holds (Seneta,1981). As such, a malarialogist/experimenter could prove non explosiveness and a lower bound of the spectral gap via the strong Doeblin condition for a large class of stochastic processes in a robust geopredictive district-level time series malaria-related risk model by evolving in the interior of a region D μ Rd with boundary D according to an underlying Markov process with transition probabilities $p(t; x; dy)$ whereby , undergoing jumps to a random district-level hyperendemic transmission oriented sampled point x in D with distribution ^{o}»(dx) as soon as they reach a boundary point ». Besides usual regularity conditions on $p(t; x; dy)$, a malarialogist/experimenter would then simply require a tightness condition on the family of measures ^{o}»,
for preventing mass from escaping to the boundary. The setup can be applied to a multitude of geopredictive time series malaria-related risk models.

A malarialogist/experimenter then could compare the performance of the Bayesian lassos to their frequentist counterparts using simulated district-level regressed field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate coefficients in SPSS, if so desired.. Previous linear-based lasso papers employed for geopredictive malaria-related risk modeling have revealed problems for forecasting statistically significant seasonal hyperendemic geopredictive transmission explanatory oriented covariate coefficients. In terms of prediction mean squared error, the Bayesian lasso performance is similar to and, in some cases, better than, the frequentist malarial – related geopredictive district-level lasso (see Jacob et al. 2011c).

Additionally, in SPSS, a malariaologist/experimenter may point out that $\widehat{\sigma_e^2}$ is a biased estimator for the true variance, σ^2 in a geopredictive seasonal district-level malaria-related risk model. This estimator can be determined by letting $\widehat{\widehat{\sigma_e^2}}$ denote the unbiased form of the model parameter estimators in the Forecasting add-on module in SPSS (http://www.spss.com/worldwide) for approximating the error variance employing $\widehat{\widehat{\sigma_e^2}} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2.$ Deviance may be then defined as the log likelihood of the final model, multiplied in SPSS by calculating the deviance as $\sum_{i=1}^{n} 2(y_i \log \frac{y_i}{\hat{y}_i} - (y_i - \hat{y}_i))$, $\hat{y}_i$ is the predicted value of y_i. (www.spss.com). By so doing, under the assumption of normality then $\mathrm{BIC} = \chi^2 + k \cdot \ln(n).$ may be more found to be more tractable for district-level malaria-related risk modeling. Note that there is a constant added that must follow from the transition from log-likelihood to χ^2 in the SPSS dervived model; however, in using the BIC to determine the "best" model the constant becomes trivial. Given any two estimated models, the model with the lower value of BIC is the one to be preferred (Cressie 1993).

The BIC is an increasing function of σ_e^2 and an increasing function of *k*. That is, unexplained variation in the geopredictive malaria-related dependent variable (e.g., total seasonal district- level prevalence rates) and the number of explanatory hyperendemic transmission oriented variables would increase the value of BIC. Hence, lower BIC would imply either fewer explanatory district-level field/clinical/remote sampled malaria-related variables, better fit, or both in the residual forecasts. The BIC generally penalizes free geoparameter estimators more strongly than does the AIC in geopredictive autoregressive malarial risk modeling though it depends on the size of n and relative magnitude of *n* and *k*.(see Jacob et al. 2011d). The Forecasting add-on module employed with the SPSS Statistics Core system then could quantitate results for forecast values, fit values, observed values, upper and lower confidence limits, residual autocorrelations and partial autocorrelation.

Interestingly, partial autocorrelation plots are a commonly used tool for model identification in Box-Jenkins models (Box and Jenkins, 1970). The partial autocorrelation at lag *k* is the autocorrelation between X_t and X_{t-k} that is not accounted for by lags 1 through *k*-1 (Griffith 2003). There are multiple algorithms for computing the partial autocorrelation in a robust geopredictive time series district level malaria-related risk model based on a dataset of empirical sampled district-level explanatory field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients in SPSS. For instance, the residual partial autocorrelation function can display a table of latent autocorrelations coefficients by lag for each estimated seasonal predictive district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented regressor. Additionally, the table could include the confidence intervals for the autocorrelation in SPSS, if so desired. Specifically, partial autocorrelations are useful in identifying the order of an autoregressive model (see Brockwell, 1991). The partial autocorrelation of an AR(*p*) process is zero at lag *p*+1 (Griffith 2003). If the sample autocorrelation plot in SPSS indicates then that an AR-constructed district-level predictive seasonal forecasting field/clinical/remote sampled malaria-related hyperendmeic transmission oriented regression-based risk model may be appropriate, then the sample partial autocorrelation plot may be further examined to help identify the order of the estimates. In actuality, the malarialogist/experimenter would be searching for the geographic point on the plot where the partial autocorrelations in the risk model residual forecasts essentially become zero when quantitating the confidence interval for summarizing the statistical significance in the regressed district-level empirical parameter estimator dataset. Commonly, the approximate 95% confidence interval for the partial autocorrelations are at $\pm 2/\sqrt{N}$ when employing SPSS autocorrelation tests such as Ljung-Box *Q* tests .

The Ljung-Box Q-test is a quantitative way to test for autocorrelation at multiple lags jointly in SPSS. Commonly, the null hypothesis for the Ljung-Box Q-test for a predictive malaria-related district-level regression based risk model is that the first *m* autocorrelations are jointly zero in $H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0.$ If *N* is the length of the observed seasonal district-level field/clinical/remote sampled geopredictive malarial-related time series, for example, choosing $m = \ln(N)$ may reveal power for testing explanatory multiple hyperendemic transmission oriented covariate coefficient measurement values of *m*. The choice of *m* affects test performance (Box and Pierce 1970).If seasonal autocorrelation is possible, testing at larger values of *m* in the risk model, such as 10 or 15, may render robust unbiased dependent residual forecasts in a stochastic/deterministic interpolation algorithm (e.g., co-Kriged model). The Ljung-Box test statistic in SPSS is commonly given by $Q(m) = N(N+2)\sum_{h=1}^{m}\frac{\rho_h^2}{N-h}.$ (http://www.unt.edu/rss/class/Jon/SPSS_SC/Manuals/v18/PASW%20Forecasting%2018.pdf). Under the null hypothesis, Q (*m*), a seasonal predictive regression-based malaria-related district-level risk model χ^2_m distribution could then be robustly constructed in SPSS.This autocorrelation test is a modification of the Box-Pierce Portmanteau test statistic. The Ljung-Box Q-test test is an improved version of the Box–Pierce test, having been devised at essentially the same time; a seemingly trivial simplification (omitted in the improved test) was found to have a deleterious effect (see Ljung, and Box, 1978

On the other hand, a portmanteau test generated in SAS/GIS is a type of statistical hypothesis test in which the null hypothesis is well specified, but the alternative hypothesis is more loosely specified. Box and Pierce (1970) showed that a portmanteau statistic essentially is a multiple of a sum of squared residual autocorrelation coefficients that follows a chi-square distribution. This statistic is now generally known as the Box–Pierce statistic. Ljung and Box (1978) and McLeod (1978) suggested improved portmanteau statistics to rectify the conservative nature of the Box–Pierce statistic. McLeod (1978) then considered the distribution of residual autocorrelations using a martingale difference approach.

A basic definition of a discrete-time martingale in SAS/GIS is a discrete-time stochastic process (i.e., a sequence of randomized district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented variables) $X_1, X_2, X_3, \ldots$ that satisfies for any time $n, \mathbf{E}(|X_n|) < \infty \mathbf{E}(X_{n+1} \mid X_1, \ldots, X_n) = X_n$. That is, the conditional expected value of the next sampled hyperendemic transmission oriented field/clinical/remote observation, given all the past observations, is equal to the last observation. Due to the linearity of expectation, this second requirement would be equivalent to $\mathbf{E}(X_{n+1} - X_n \mid X_1, \ldots, X_n) = 0$ or $\mathbf{E}(X_{n+1} \mid X_1, \ldots, X_n) - X_n = 0$ which states that the average "winnings" from district-sampled malaria-related hyperendemic transmission oriented observation n to observation $n+1$ are 0, for instance. More generally, a sequence $Y_1, Y_2, Y_3 \ldots$ is said to be a martingale with respect to another sequence $X_1, X_2, X_3 \ldots$ if for all $n \mathbf{E}(|Y_n|) < \infty \mathbf{E}(Y_{n+1} \mid X_1, \ldots, X_n) = Y_n.$

Similarly, a continuous-time martingale with respect to the stochastic process X_t in a geopredictive seasonal malaria-related risk model would be a stochastic process Y_t such that for all $t \mathbf{E}(|Y_t|) < \infty \mathbf{E}(Y_t \mid \{X_\tau, \tau \leq s\}) = Y_s, \ \forall\, s \leq t.$ This model would express the property that the conditional expectation of a district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented observation at time *t*, given all the observations up to time s, is equal to the observation at time *s* provided of course that $s \leq t$. Further, in terms of full generality, a stochastic process $Y : T \times \Omega \to S$ in a geopredictive district-level malarial risk model would be a martingale with respect to a filtration Σ_* and probability measure P if Σ_* is a filtration of the underlying probability space $(\Omega, \Sigma, \mathrm{P})$; *Y* is adapted to the filtration Σ_*, i.e., for each *t* in the index set *T*, the random variable Y_t is a Σ_t-measurable function; for each *t*, Y_t lies in the L^p space $L^1(\Omega, \Sigma_t, \mathrm{P}; S)$, (i.e. $\mathbf{E}_\mathbf{P}(|Y_t|) < +\infty$; for all *s* and *t* with $s < t$ and all $F \in \Sigma_s$, $\mathbf{E}_\mathbf{P}([Y_t - Y_s]\chi_F) = 0$, where χ_F denotes the indicator function of the event *F*). In Grimmett and

Stirzaker's *Probability and Random Processes*, (He, Wang, Yan 1992), this last condition is denoted as $Y_s = \mathbf{E}_{\mathbf{P}}(Y_t|\Sigma_s)$, which is a general form of conditional expectation It is important to note that the property of being a martingale in a geopredictive malaria-related risk model constructed from an empirical dataset of district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented observation would involve both the filtration and the probability measure with respect to which the expectations were employed to construct the model. It is possible that *Y* could be a martingale with respect to one measure but not another one; the Girsanov theorem offers a way to find a measure with respect to which an Itō process is a martingale.

Girsanov's theorem is important in the general theory of stochastic processes since it enables the key result that if *Q* is a measure absolutely continuous with respect to *P* then every *P*-semimartingale is a *Q*-semimartingale. This theoreum underlying stochastic process is a Wiener process (see Cressie 1993). This special case is sufficient for risk-neutral pricing in the Black-Scholes model and in many other models (e.g. all continuous models) (see Jacob et al. 2013b). As such, employing this theorem $\{W_t\}$ when constructing a geopredictive malaria-related district-level model would be like qunatiating a Wiener process on the Wiener probability space $\{\Omega, \mathcal{F}, P\}$. Therefater , X_t would be a measurable process in the residually forecasted estimates adapted to the natural filtration of the Wiener process $\{\mathcal{F}_t^W\}$.]Given an adapted process $X_t$with $X_0 = 0$a malarialogist/experimenter could then define$Z_t = \mathcal{E}(X)_t$, in a a district-level malarial risk model where $\mathcal{E}(X)$is the stochastic exponential (i.e., Doléans exponential) of *X* with respect to *W*, i.e. $\mathcal{E}(X)_t = \exp\left(X_t - \frac{1}{2}[X]_t\right)$. Z_t in the risk model would then be a strictly positive local martingale, and as such a probability measure *Q* can be defined on $\{\Omega, \mathcal{F}\}$using a Radon–Nikodym derivative$\frac{dQ}{dP}|_{\mathcal{F}_t} = Z_t = \mathcal{E}(X)_t$

In mathematics, the Radon–Nikodym theorem is a result in measure theory that states that, given a measurable space (*X*,Σ), if a σ-finite measure νon (*X*,Σ) is absolutely continuous with respect to a σ-finite measure μon (*X*,Σ), then there is a measurable function *f* on *X* and taking values in [0,∞), such that $\nu(A) = \int_A f\, d\mu$ for any measurable set *A*. Then for each *t* the field measure *Q* restricted to the unaugmented sigma fields $\mathcal{F}_t^W$ would be equivalent to *P* restricted to $\mathcal{F}_t^W$ .Further, if *Y* is a local martingale in the empirical sampled field/clinical/remote sampled malaria-related hyperendmeic transmission oriented data attributes under *P* then the process$\tilde{Y}_t = Y_t - [Y, X]_t$is a *Q* local martingale on the filtered probability space $\{\Omega, F, Q, \{F_t^W\}\}$. Thus, if a malarialogist/experimenter lets ν, μ, and λ be σ-finite measures on the same measure space in a geopredictive malaria-related model as ν ≪ λ and μ ≪ λ where ν and μ are absolutely continuous in respect to λ, then $\frac{d(\nu+\mu)}{d\lambda} = \frac{d\nu}{d\lambda} + \frac{d\mu}{d\lambda}$ .If ν ≪ μ ≪ λ, then $\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$ .In particular, if μ ≪ ν and ν ≪ μ, then $\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}$ .If μ ≪ λ and *g* is a μ-integrable function, then $\int_X g\, d\mu = \int_X g\frac{d\mu}{d\lambda}\, d\lambda.$ If ν is a finite signed or complex measure, then $\frac{d|\nu|}{d\mu} = \left|\frac{d\nu}{d\mu}\right|$.

On the other hand, an Itō process may be defined to be an adapted stochastic process in a geopredictive time series malaria-related district-level risk model in SAS/GIS which can be expressed as the sum of an integral with respect to Brownian motion and an integral with respect to time, $X_t = X_0 + \int_0^t \sigma_s \, dB_s + \int_0^t \mu_s \, ds.$ Here, *B* is a Brownian motion and it is required that σ is a predictable *B*-integrable process, and μ is predictable and Lebesgue integrable. That is, $\int_0^t (\sigma_s^2 + |\mu_s|) \, ds < \infty$ for each *t* would then occur during the quantitation process of the field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients. For example. The stochastic integral can then be extended to an Itō processes, as $\int_0^t H \, dX = \int_0^t H_s \sigma_s \, dB_s + \int_0^t H_s \mu_s \, ds.$ This process may be defined in the district-level malarial risk model for all locally bounded and predictable integrands. More generally, it is required that *H*σ be *B*-integrable and *H*μ be Lebesgue integrable, so that $\int_0^t (H^2\sigma^2 + |H\mu|) ds < \infty.$ (see Cressie 1993). Such predictable field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented processes *H* would be considered *X*-integrable.

An important result for the study of Itō processes is Itō's lemma. In its simplest form, for any twice continuously differentiable function *f* on the reals and Itō process *X* as described above, it states that *f*(*X*) is itself an Itō process satisfying $df(X_t) = f'(X_t) \, dX_t + \frac{1}{2} f''(X_t) \sigma_t^2 \, dt.$ (Hagen 2004). This is the stochastic calculus version of the change of variables formula and chain rule. This mathematical output would differ from the standard results in a geopredictive seasonal malaria-related regression-based model due to the additional term involving the second derivative of *f*, which would be derived based on the property that Brownian motion has non-zero quadratic variation.

Another widely used technique for testing geopredictive seasonal district-level malaria-related model adequacy in SPSS is the score or Lagrange multiplier test procedure. Rao's score test, or the score test (often known as the Lagrange multiplier test in econometrics) is a statistical test of a simple null hypothesis that a parameter of interest θ is equal to some particular value θ_0. It is the most powerful test when the true value of θ is close to θ_0. The main advantage of the Score-test for predictive malaria-regression based modeling is that it does not require an estimate of the information under the alternative hypothesis or unconstrained maximum likelihood. This makes testing feasible when the unconstrained maximum likelihood estimate is a boundary point in the parameter space. It can be easily computed by letting L be the likelihood function in the malaria-related which would subsequently depend on a univariate parameter θa. Thereafter, by letting x be the seasonal district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented observational data. the score could be calculated as $U(\theta)$ where $U(\theta) = \frac{\partial \log L(\theta|x)}{\partial \theta}$. The observed information would be then quantitated employing $\mathcal{I}(\theta) = -\operatorname{E}\left[\frac{\partial^2}{\partial \theta^2} \log f(X;\theta) \middle| \theta\right]$. The statistic to test $\mathcal{H}_0 : \theta = \theta_0$ is $S(\theta_0) = \frac{U(\theta_0)^2}{I(\theta_0)}$ in the residually forecasted estimates would have an asymptotic distribution of χ_1^2, when $\mathcal{H}_0$ is true.

As pointed out by Box et al. (1994), the score test procedure is essentially a quadratic form in the residual autocorrelation quantitation procedures, but of a more complex form than the portmanteau. Ling and Li (1997) andWong and Li (2002) derived multivariate conditional heteroscedastic detection techniques using these models. As it is clear that the Ljung–Box statistic and McLeod–Li/Li–Mak statistics are sensitive to lack of fit in the first and second moments of the data structure respectively (Cressie 1993). As such, a malarialogist/experimenter can expect mixed statistics to be most powerful but only when the fitted predictive seasonal malaria-related risk model has disparity in both the first and second moments. In such circumstances the mixed statistics would simply be a sum of

the Ljung–Box and McLeod–Li statistic or Ljung–Box and Li–Mak statistic in the district-level malaria model output. The independence of the two statistics may then be shown then by a Monte Carlo simulation study of the size and power of the mixed statistic. In general, tests constructed in this context can have the property of being at least moderately powerful against a wide range of departures from the null hypothesis. Markov chain Monte Carlo (MCMC) methods (which include random walk Monte Carlo methods) are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution (Gilks 1996).

Monte Carlo methods (or Monte Carlo experiments) are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results by running simulations many times over in order to calculate those same probabilities heuristically. They are often used in physical and mathematical problems and are most suited to be applied when it is impossible to obtain a closed-form expression or infeasible to apply a deterministic algorithm. In seasonal malaria-related geopredictive risk modeling Monte Carlo methods are mainly used in three distinct problems for optimization, numerical integration and generation of samples from a probability distribution (see Jacob et al. 2011c). Monte Carlo simulation methods do not always require truly random numbers to be useful — while for some applications, such as primality testing, unpredictability is vital (Cressie 1993). Fortunately, most useful techniques in geopredictive district-level malaria risk modeling use deterministic, pseudorandom sequences, making it easy to test and re-run simulations. The only quality usually necessary to make good simulations is for the pseudo-random sequence to appear "random enough" in a certain sense. What this means depends on the application of the predictive risk model residual forecasts, but typically they should pass a series of statistical tests. Testing that the numbers are uniformly distributed or follow another desired distribution when a large enough number of elements of the sequence are considered is one of the simplest and most common ones.

Further, Monte Carlo simulation is the process of generating independent, random draws from a specified probabilistic model. When simulating time series models, one draw (or realization) is an entire sample path of specified length N, y_1, y_2,...,y_N. When a malarialogist/experimenter generates a large number of draws from a regressed empirical dataset of district-level geopredictive malaria-related explanatory hyperendemic transmission oriented covariate coefficinets for instance, say M, then M sample paths will be generated each of length N. Commonly applications of Monte Carlo simulation are used for demonstrating theoretical results, forecasting future events and estimating the probability of future events (Gilks 1996) The time series portion of a predictive district-level malaria-related risk model would then specify the dynamic evolution of the unconditional disturbance process over time through a conditional mean structure. To perform Monte Carlo simulation of a robust geopredictive malaria-related regression based risk model, residually forecasted estimates with ARIMA errors therefater, SPSS would simply specify presample innovations or unconditional disturbances or use default presample data.By so doing, uncorrelated innovation series from a probability distribution would be generated from the regressed empirical explanatory hyperendemic transmission oriented covariates. Thereafter, by filtering the innovations through the ARIMA error model simulated unconditional disturbances can be parsimoniously derived.

Thus, in applied statistics for seasonal geopredictive malaria-related district-level risk models, a portmanteau test could provide a reasonable way of proceeding as a general check of a model's match to an empirical sampled dataset of explanatory geopredictive field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients to determine uncertain departures from the underlying district-level data generating process. Fortunately, in SPSS use of such tests would avoid having to be very specific about the particular type of departure being tested. This statistical test can then determine whether any of a group of autocorrelations in a regressed empirical sampled district-level malarial-related ecological dataset in SPSS is different from zero. Instead of testing randomness at each distinct lag. The portmanteau test would evaluate the "overall" randomness in the district-level geopredictive field/clinical/remote sampled malaria-related hyperendemic transmission oriented estimators based on a number of lags. Tests constructed in this context can have the property of being at least moderately powerful against a wide range of departures from the null hypothesis in a malarial district-level model output. For instance, in applied statistics for district-level predictive malaria-related risk modeling, a portmanteau test can provide a reasonable way of proceeding as a general check of a model's match to an empirical sampled dataset of district-level field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients to summarize different ways in which the model may depart from the underlying district-level data generating

process. Fortunately, use of such tests can avoid having to be very specific in SPSS about the particular type of departure being tested in a robust predictive district-level malaria-related time series analysis.

The portmanteau test is also useful in working with ARIMA models (Griffith 2003). These statistics could then be employed to test for significant correlation up to lag in a robust geopredictive district-level malaria-related hyperendemic transmission oriented risk model. It is well known that for i.d.d data, the autocorrelations behave as independent normally distributed random variables, and therefore under the null hypothesis (e.g., correctly fitted predictive time series malaria-related risk model) both these data attributes may be shown to be asymptotically distributed chi-squared random variables with degrees of freedom, when the order of AR and MA terms are estimated accurately in a fitted model. Further, in the context of regression analysis, including an analysis with time series structures, a portmanteau test can be devised in SPSS which allows for a general test to made for quantitating a range of seasonal nonlinear field/clinical/remote sampled malaria-related hyperendemic transmission oriented data transformations based on various combinations of the explanatory district-level malaria-related covariate coefficients. By so doing, chi-squared critical values of a particular district-level parameter estimator significance level can be quantitated to determine if there is evidence to suggest the fitted ARIMA process does not adequately model the correlation in the sampled empirical dataset.

A district-level ARIMA predictive autoregressive malaria-related empirical-sampled dataset employing Φ and Θ in SPSS can then be selected, if so desired, so that the zeros of both polynomials lie outside the unit circle in order to avoid generating unbounded processes in the residual forecasts. By so doing, unbiased optimal district-level explanatory hyperendemic transmission-oriented uncertainty covariate coefficients would be rendered with the regressed residuals and their estimated significance levels very parsimoniously. In the mathematical field of numerical analysis, interpolation is a method of constructing new data points within the range of a discrete set of known data points(Cressie 1993) As such, the regressed district-level explanatory district-level field/clinical/remote malaria-related hyperendemic transmission oriented residual forecasts would be based on rigorously quantitated difference operator ``unit root" (1-*B*) behavior in the time series for *d*>0.5, for example. A robust ARIMA seasonal geopredictive malaria-related district-level risk model can then be provided in SPSS by the ARIMA (1, 0, and 0) first order autoregressive model [i.e., $x_y = \phi_1 x_{y-1} + a_y$].

Additionally, the Time Series Modeler procedure estimates in SPSS can also be employed to construct exponential smoothing, univariate ARIMA, and multivariate ARIMA and/or transfer function models from empirical ecological datasets of time series district-level hyperendemic transmission oriented predictive malaria-related field/clinical/ remote sampled data attributes. The procedure would include an Expert Modeler that would automatically identify and estimate the best-fitting ARIMA or exponential smoothing model for one or more dependent variable series, thereby eliminating the need to identify an appropriate model fit through trial and error. The Time Series Modeler allows building custom non-seasonal or seasonal ARIMA models with or without a fixed set of data variable(http://www-01.ibm.com/software/analytics/spss/). Thereafter, transfer functions can be defined for any or all of the regressed hyperendemic transmission-oriented seasonal-sampled covariates. By so doing, goodness-of-fit measures in SPSS for the interpolated regression-based model residually forecasted hyperendemic transmission oriented data attributes can then include: stationary R-square, R^2 MAE, MAPE, MaxAE, MaxAPE, and normalized BIC estimators along with the standard RSME.

Thereafter, if so desired, the Expert Modeler in SPSS could statistically locate the optimal *p,d,q* district-level optimal field/clinical/remote sampled malaria-related hyperendemic transmission oriented estimators rendered from the predictive district-level malarial-related risk model. The Expert Modeler can automatically find the best-fitting model for each dependent series (www.spss.com). If the seasonal sampled empirical dataset of the malaria-related independent predictor variables are specified, for instance, the Expert Modeler would then select, for inclusion in the ARIMA model, those that have a statistically significant relationship with the dependent series. Routinely, district-level time series malaria-related hyperendemic transmission oriented model variables would then be transformed where appropriate using differencing and/or a square root or natural log transformation. By default, the Expert Modeler would consider both exponential smoothing and ARIMA models. SPSS offers a variety of exponential smoothing models that differ in their treatment of trend and seasonality (www.spss.com). For instance, exponential

smoothing district-level predictive hyperendemeic transmission oriented risk models can be classified as either seasonal or nonseasonal. Thereafter, outlier estimates can be determined in a robust model. An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs (Grubbs 1969).

Outliers can occur by chance in any regressed empirical dataset of seasonally forecasted district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive explanatory covariate coefficients which are often indicative either of measurement error or, that the sampled population has a heavy-tailed distribution (see Jacob et al. 2012b, Jacob et al. 2009d). Fortunately, in an SPSS derived district-level malarial-related model, heavy-tailed distributions (i.e., district-level probability distributions) can be delineated whereby, the tails are not exponentially bounded: that is, they have heavier tails than the exponential distribution. In many applications, it is the right tail of the distribution that is of interest to malariologists/experimenters but, a district-level forecasted risk distribution may have a heavy left tail, or both tails may be heavy. A right heavy tailed distribution is one with infinite moment generating function on $(0,\infty)$, that is, X has right heavy tail if, $E(e^{tX})= \infty$ $t>0$ (Hosmer and Lemeshew 2000). There are three important subclasses of heavy-tailed distributions in seasonal geopredictive time series malarial risk model construction; the fat-tailed distributions, the long-tailed distributions and the sub-exponential distributions (Jacob et al. 2011b, Jacob et al. 2009d). In SPSS all commonly used heavy-tailed distributions are classified as the sub-exponential (www.spss.com). Although there is still some discrepancy over the classification of the term heavy-tailed, seasonal forecasted malaria-related risk model residual forecast error distributions, SPSS can accurately target explanatory hyperendemic transmission covariate coefficients based on regressed seasonally-sampled district –level sampled field/clinical/remote sampled measurement values .

Importantly, in SPSS autoregressive algorithms are often employed when regressing the conditional mean of a geopredictive malarial risk based model since these processes are considered causal at each georefernced district-level sample point in time. In these algorithms the error terms ε_t are generally assumed to be i.i.d. sampled from a normal distribution with zero mean: $\varepsilon_t \sim N(0,\sigma^2)$ where σ^2 is the variance. These assumptions may be weakened but, doing so will change the properties in any geopredictive district-level malarial risk model residual forecasts.

In some texts predictive district-level malarial risk models can be specified in terms of the lag operator *L*. In time series analysis, the lag operator or backshift operator then would run on an element of a time series to produce the previous element. For instance, given some time series empirical sampled dataset of field and remote sampled district level malarial-related hyperendemic transmission oriented parameter estimators in SPSS [e.g. $x=\{x^1,x^2 \ldots x^n\}$], then $LX_t = X_{t-1}$ for all t>1 or, equivalently $X_t = LX_{t+1}$ for all $t \geq 1$ where *L* is the lag operator. Note, that the lag operator can be raised to arbitrary integer powers so that $L^{-1}X_t = X_{t+1}$. As such, then in the SPSS derived predictive malaria-related AR(*p*) risk model would be given by $\varepsilon_t = \left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \varphi(L) X_t$ where φ represents the polynomial $\varphi(L) = 1 - \sum_{i=1}^{p} \varphi_i L^i.$ The MA(*q*) portion of the model would then be given by $X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t = \theta(L)\varepsilon_t,$ where θ represents the polynomial $\theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i.$ Finally, the combined ARMA(*p*, *q*) time series predictive district-level malaria-related regression-based risk model would be given by $\left(1 - \sum_{i=1}^{p} \varphi_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$, or more concisely, $\varphi(L)X_t = \theta(L)\varepsilon_t$ or $\frac{\varphi(L)}{\theta(L)} X_t = \varepsilon_t$.

Additionally, tests constructed in a seasonal SPSS-derived ARMA (*p*, *q*) time series geopredictive district-level malaria-related risk model context can reveal specific properties of the sampled field/clinical/remote sampled hyperendemic transmission oriented parameter uncertainty estimators. For example, in a spatiotemporal district-

level geopredictive malaria-related model regression-based framework, Q_k could be approximately distributed as a chi-square distribution with *k-m* degrees of freedom, where *m* is the number of sampled hyperendemic transmission oriented parameter estimators employed in fitting the model, excluding any constant term or other variables (i.e. including just the *p,d,q* triples). The district-level geopredictive malarial-related Ljung–Box test, for example, then could be defined as follows: H_0: the seasonal-sampled field/clinical/remote sampled malaria-related hyperendemic transmission oriented data attributes are independently distributed (i.e. the correlations in the sampled district-level population from which the sample is taken are 0), so that any observed correlations in the sampled data result from randomness of the sampling process; and, H_a: the data is not independently distributed. In SPSS the test statistic for testing such a hypotheses would be $Q = n\,(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$ where *n* is the district-level sample size, $\hat{\rho}_k$ is the sample autocorrelation at lag *k*, and *h* is the number of lags being tested. For significance level α, the critical region for rejection of the hypothesis of randomness in the geopredictive time series seasonal district-level SPSS derived malarial-related uncertainty model would then be $Q > \chi^2_{1-\alpha,h}$when $\chi^2_{1-\alpha,h}$is the α-quintile of the chi-squared distribution with *h* degrees of freedom.

In probability and statistics, the quantile function, (also called percent point function or inverse cumulative distribution function) of the probability distribution of a random variable specifies, for a given probability, the value which the random variable will be at, or below, with that probability (Hosmer and Lemeshew 2000). The quantile function is one way of prescribing a probability distribution in a robust geopredictive district-level regression-based malaria-related risk hyperendemic transmission oriented model and it is an alternative to the probability density function (pdf) or probability mass function (pmf), the cumulative distribution function (cdf) and the characteristic function (see Jacob et al. 2009d). The quantile function, *Q*, of a probability distribution is the inverse of its cumulative distribution function *F*.(Cressie 1993).

Interestingly, the derivative of the quantile function in a geopredictive seasonal district-level malarial risk model in SPSS would be the quantile density function for prescribing a probability distribution. This function in the risk model would then be the reciprocal of the pdf composed with the quantile function. Assuming a continuous and strictly monotonic distribution function in a time series predictive district-level malaria-related model,$F{:}R\to(0,1)$, the quantile function then in the derivative would return the value below which the random draws from the given district-level distribution which would fall $p\times100$ percent of the time. That is, the derivative of the quantile funtion in SPSS would return the sampled hyperendemic transmission oriented covariate coefficient measurement value of *x* such that$F(x) = \Pr(X \le x) = p.$ If the probability distribution is discrete rather than continuous in the risk model derivatives then there may be gaps between the sampled empirical dataset of malaria-related explanatory district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficient measurement values in the domain of its cumulative distribution function (cdf). The cdf of a real-valued random variable *X* is the function givenby$F_X(x) = \mathrm{P}(X \le x)$, where the right-hand side represents the probability that the random variable *X* takes on a value less than or equal to *x*. The probability that *X* lies in the semi-closed interval (*a*, *b*], where $a < b$, is therefore$\mathrm{P}(a < X \le b) = F_X(b) - F_X(a)$.(Hosmer and Lemeshew 2000.

However if the cdf is only weakly monotonic there may be "flat spots" in its range. In either case, the quantile function would be $Q(p) = \inf\{x \in R : p \le F(x)\}$ for a probability $0 < p < 1$ and the quantile function would return the minimum value of *x* for which the previous probability statement holds in the risk model geopredictive residual forecasts. For instance, the quantile function for Exponential (*λ*) (i.e. intensity *λ* and expected value 1/*λ*) in a predictive district-level malaria-related risk model may be $Q(p;\lambda) = \frac{-\ln(1-p)}{\lambda}$, for $0 \le p < 1$. The quartiles would therefore be represented as first quartile $\ln(4/3)/\lambda$median $\ln(2)/\lambda$ third quartile $\ln(4)/\lambda$. Subsequently, a non-linear ordinary differential time series district level malarial-related geopredictive regression-based equation for the normal quantile, *w*(*p*), may be then given as

$\frac{d^2w}{dp^2} = w\left(\frac{dw}{dp}\right)^2$ with the center (boundary) conditions $w(1/2) = 0, w'(1/2) = \sqrt{2\pi}$] By so doing, the non-linear ordinary differential equation could also be given for normal distribution in a robust geopredictive seasonal district-level malaria-related risk model employing any quantile function whose second derivative exists. In general the equation for a quantile, $Q(p)$, may be given as $\frac{d^2Q}{dp^2} = H(Q)\left(\frac{dQ}{dp}\right)^2$ (see Cressie 1993)which may be augmented by suitable boundary conditions, in some district-level malarial risk modeling circumstances where $H(x) = -\frac{d\log[f(x)]}{dx}$ and $f(x)$ is the pdf. The forms of this time series equation, and its classical analysis by series and asymptotic solutions, for the cases of the normal, Student, gamma and beta distributions has been elucidated by Steinbrecher and Shaw (2008). Such solutions may provide accurate benchmarks, and in the case of the Student, suitable series for live Monte Carlo use for regressing district-level seasonal explanatory district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients.

Additionally, employing the ARIMA geopredictive district-level malaria-related uncertainty regression model output, simulated unconditional disturbances for the field/clinical/remote sampled hyperendemic transmission oriented geopredictive autoregressive regressors can be obtained. For instance, suppose a malarialogist/experimenter considers simulating N responses from a regression –based geopredictive malarial-related district-level risk model with ARMA(2,1) errors: $y_t = X_t\beta + u_t\ u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + \varepsilon_t + \theta_1\varepsilon_{t-1}$, where ε_t is Gaussian with mean 0 and variance σ^2. Given presample unconditional disturbances (u_0 and u_{-1}) and innovations (ε_0), N independent innovations from the Gaussian distribution: $\{\varepsilon_1, \varepsilon_2, ..., \varepsilon_N\}$. will be generated in SPSS. Thereafter, by filtering the innovations recursively the unconditional disturbances [e.g. $\hat{u}_1 = \phi_1 u_0 + \phi_2 u_{-1} + \hat{\varepsilon}_1 + \varepsilon_0$ b) $\hat{u}_2 = \phi_1\hat{u}_1 + \phi_2 u_0 + \hat{\varepsilon}_2 + \hat{\varepsilon}_1$ c) $\hat{u}_3 = \phi_1\hat{u}_2 + \phi_2\hat{u}_1 + \hat{\varepsilon}_3 + \hat{\varepsilon}_2$ d) $\hat{u}_N = \phi_1\hat{u}_{N-1} + \phi_2\hat{u}_{N-2} + \hat{\varepsilon}_N + \hat{\varepsilon}_{N-1}$.] can be derived from any empirical sampled dataset of predictive district-level malaria-related explantory hyperendemic trasnmission oriented covariate coefficients. Therefater, a malarialogist/experimenter could obtain simulated responses employing the unconditional disturbances, regression model, and the other sampled malaria-related exploratory district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive predictors: $\hat{y}_t = X_t\beta + \hat{u}_t$. ,if so desired.

Interestingly, in SPSS the simulation test would be applied to the residuals of a fitted district-level geopredictive malaria-related explanatory hyperendemic transmission oriented ARIMA model, not the original series. As such, the hypothesis actually being tested when constructing a robust SPSS time series risk model would be the level of uncertainty error coefficients in the residually forecasted district-level field/clinical/remote sampled hyperendemic transmission oriented predictive autoregressive estimates. The model components would then be actually be based on whether the residuals from the ARIMA model contain uncertainty coefficients (e.g.,latent autocorrelation). Fortunately, when testing ARIMA models, no adjustment to the test statistic or to the critical region of the test are made in relation to the structure of the ARIMA error model.

By so doing, at lag k in the ARIMA SPSS derived district-level hyperendemic transmission oriented robust geopredictive malaria-related risk model, the Box-Ljung statistic could be defined by Qk=n (n+2) $k\Sigma l$=1r2ln−l. As such, when n is large, Qi would have a chi-square distribution with degrees of freedom $k−p−q$, in the residual forecasts for precision targeting the significant district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive explanatory covariate coefficients when p and q are autoregressive and moving average orders, respectively. The significance level of Qi could then be calculated from the chi-square distribution with $k−p−q$ degrees of freedom. If the measure is statistically significant in the regressed georeferenced empirical district-level empirical sampled dataset, it would then indicate that the residuals forecasts targeting the highly significant hyperendemic transmission oriented explanatory covariate coefficients still contain significant latent unobserved autocorrelation uncertainty coefficients after the model has been fitted. This then would suggest that an improved model should be sought.

In statistics, OLS or linear least squares is a method for estimating the unknown parameters in a linear regression model (Rao 1973). In SPSS the estimated district-level predictive malaria-related time series regression-based equation would be $Y = ß_0 + ß_1X_1 + ß_2X_2 + ß_3D + ê$ where the ßs would be the OLS estimates of the Bs(www-01.ibm.com/software/analytics/spss/). The residual, ê, would then be the difference between the actual Y and the predicted Y and would thus possess a zero mean in the risk model. In other words, OLS would be able to calculate the slope coefficients in an empirical regressed dataset of sampled district-level explanatory hyperendemic transmission oriented covariate coefficients so that the difference between the forecasted Y and the actual Y could be minimized. OLS minimizes the sum of the squared residuals OLS minimizes SUM $ê^2$(Hosmer and Lemeshew 2000).

Routinely, the residual forecasts rendered from the seasonal geopredictive malarial-related district-level regression-based risk model would then be squared in order to compare negative errors to positive errors more efficiently in the regressed hyperendemic transmission oriented derivatives. The OLS estimates of the ßs in the seasonal district-level SPSS model would be unbiased –(e.g., the ßs are centered around the true sample population values of the Bs). Further, the residual forecasts would display minimum variance whereby, the distributions of the ß malarial-related district-level hyperendemic transmission oriented sampled estimates around the true Bs would be spatially configured as tight as possible. This model would be consistent - as the sample size (n) approaches infinity and the estimated ßs would converge on the true Bs. As such, the rendered residual district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive forecasts would be normally distributed since the statistical tests would be based on a district-level simulated normal distribution.

Interestingly, statistical computing packages in SPSS routinely print out the estimated ßs when estimating a regression equation (e.g. ols1.t). As such, OLS could be employed to minimize the sum of the squared residuals in a robust geopredictive district-level risk uncertainty specified malarial risk model. By so doing, OLS could calculate the slope coefficients so that the difference between the predicted Y and the actual Y being minimized in the geopredictive district-level risk model error specification would subsequently provide a "best" fit for the sampled data points. Here the "best" may be understood, as the optimal risk model for minimizing the sum of squared residuals of the linear regression model. By so doing, α (the y-intercept) and β (the slope) would be efficiently quantitated in the empirical regressed geosampled district-level malaria- related explanatory hyperendemic transmission oriented covariate coefficient measurement values .

Thereafter, by employing calculus, the geometry of inner product spaces in a geopredictive time series district-level malarial risk model could even be further quantitated, if so desired, by simply expanding the model to attain a quadratic in α and β, as in Jacob et al. (2013c). By so doing, the sampled values of α and β in the risk model would then be able to minimize the objective function when r_{xy} is the sample correlation coefficient between x and y, s_x is the standard deviation of x, and s_y is correspondingly the standard deviation of y. For instance, if the seasonal SPSS constructed geopredictive malaria-related district-level risk model is described using $\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$ then substituting the above expressions for $\hat{\alpha}$ and $\hat{\beta}$ into $y = \hat{\alpha} + \hat{\beta}x$, would yield $\frac{y - \bar{y}}{s_y} = r_{xy}\frac{x - \bar{x}}{s_x}$ Fortunately, since ARIMA uncertainty models include only AR terms they can be fitted by OLS.

Further, the quantitatively regressed seasonal district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive data attributes may be squared in in SPSS in order to compare the negative errors to positive errors more proficiently. Optimally, as previously mentioned, the OLS estimates of the ßs in the residual forecasts would be unbiased – (e.g., the ßs centered around the true sampled district-level malarial-related population values) and would have minimum variance. By so doing, the SPSS OLS malarial model derivatives would minimize the sum of squared vertical distances between the georeferenced district-level predictive observed responses in any regressed empirical-sampled dataset and their responses for seasonal approximation of unbiased exploratory hyperendemic transmission oriented predictors (e.g., prolific larval habitat based on field sampled density count data) and their geolocations.

Importantly, the SPSS OLS estimator is consistent when 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 would provide minimum-variance mean-unbiased residual forecast error decomposition and estimation for any regressed empirical sampled dataset of geopredictive malarial-related district-level explanatory hyperendemic transmission –transmission-oriented covariate as the errors would have finite variances in the risk model outputs. Additionally, under the additional assumption that the error terms be normally distributed, the OLS would be the maximum likelihood estimator (MLE) in the residual forecasts targeting the -level field/clinical/remote sampled geopredictive covariate coefficients. In statistics, MLE is a method of estimating the uncertainty-related parameters of a statistical model assuming that the heights are normally (i.e., Gaussian) distributed with some unknown mean and variance(see Hosmer and Lemeshew 2000).

As such, SPSS can be used to regress the number of malaria cases in neighboring districts at an epidemiological study site using specific explanatory district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive explanatory covariates (e.g. weekly rainfall) for forecasting selected seasonal predictive ARIMA model outputs. For example, Zhu et al. (2007) constructed a geopredictive ARIMA model based on the monthly malaria incidence of Huaiyuan and Tongbai counties in Huaihe River Valley, China from January 1998 to December 2005 in SPSS 13.0 software. Akaike's information criterion and BIC were then employed to confirm the fitness of the model. Thereafter, the SPSS derived ARIMA-related geopredictive district-level malarial-related risk model was employed to predict the monthly malaria incidence in 2006 using the linear residual forecasts. The data was then compared with the actual incidence so as to evaluate the model's geopredictive power. Malaria incidence of 2007 was correctly forecasted by the ARIMA-related model based on malaria incidence from 1998 to 2006. The results indicated that statistics assisted estimation of the significance of the fitted autoregressive and seasonal moving average coefficients were unbiased (i.e., AR1=0.512, SMA1=0.609, $P<0.01$). An ARIMA model, with AIC=67.01, BIC= 71.87 and white noise exactly fitted the incidence of the previous monthly incidence rates from January 1998 to December 2005. Additionally, the predicted residual variance revealed that district-level malarial monthly incidence rates in 2006 were consistent with the actual incidence rates. Malaria incidence of 2007 was 106.50/100, 000, with a peak incidence during July and October. Thus, the spatiotemporal SPSS constructed ARIMA model was an appropriate model to fit exactly the changes of seasonal malaria incidence rates and to forecast incidence trends with a high precision based on the short term time-series.

Unfortunately, Expert Modeler from ARIMA in SPSS 13, 14 and 15 have found that Trends Version 14 render different results for point predictions and upper and lower confidence limits for output values when employing the ARIMA stand-alone command when compared to employing the ARMIA option under the Expert Modeler (i.e., TSMODEL). Thus, if the exact same ARIMA-related geopredictive district-level malarial specification found by the Expert Modeler were repeatedly run in ARIMA employing the sampled district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive data from the Huaihe River Valley epidemiological study sites, the point estimates and the upper and lower confidence limits would differ. In particular, under the Expert Modeler, the geopredictive intervals would be substantially narrower and the differences with the ARIMA stand alone command would increase as the forecast horizon lengthens. Even though over the estimation period, the two commands may yield results from the geopredictive autoregressive seasonal trend district–level malarial model residual forecasts, they would not be identical. Disastrously, since the Expert Modeler and ARIMA uses different algorithms for choosing the starting values for the geopredictions in time series, the forecasting context may begin to diverge rendering misspecifications in the district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive outputs. Although , the narrower predictive intervals found may be evidence of the superiority of its implementation over the older ARIMA procedure, the newer SPSS TSMODEL commands, such EXSMOOTH, may not provide sufficient functionality for generating robust residual forecasts from a predictive seasonal district-level malaria-related hyperendemic transmission oriented regression based risk model.

Further, although SPSS has a nice routine in their regression models (e.g., logistic) for estimating and validating district-level geopredictive seasonal malaria-related time series explanatory hyperendemic transmission oriented covariate coefficient interactions, it is only a trivial advantage since the software's useful multivariate error analysis procedures are limited to logit models. The logit function is the inverse of the sigmoidal "logistic" function used in

mathematics, especially in regression-based statistics(Rao 1973). The sigmoid function, also called the sigmoidal curve or logistic function, is the function $y = \frac{1}{1+e^{-x}}$. which has as a derivative $\frac{dy}{dx} = [1 - y(x)]\, y(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{e^{x}}{(1+e^{x})^2}$ with an indefinite integral $\int y\, dx = x + \ln(1+e^{-x}) = \ln(1+e^{x})$. (see von Seggern 2007). Log-odds and logit are synonyms (Cressie 1993). Although the logit of a number *p* between 0 and 1 can be given by the formula $\operatorname{logit}(p) = \log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p)$, it would be the base of the logarithm function employed in a district-level geopredictive seasonal malaria-related regression based risk model would be of little importance since the forecast coefficients would be greater than 1.

Currently the natural logarithm with base *e* is the one most often employed in geopredictive district-level malarial risk modeling. The "logistic" function of any seasonal-sampled empirical dataset of district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive explanatory covariate coefficient measurement values then in an SPSS constructed model would be α which could be given by the inverse-logit: $\operatorname{logit}^{-1}(\alpha) = \frac{1}{1+\exp(-\alpha)} = \frac{\exp(\alpha)}{1+\exp(\alpha)}$(see Cramer 2003). Thus, if *p* is a probability in a robust predictive district-level malaria-related risk model, then *p*/ (1 − *p*) would be the corresponding odds in the residual forecasts targeting the seasonal explanatory hyperendemic transmission-oriented covariate coefficients by their respected significance levels where the logit of the probability would be the logarithm of the odds in the model. Similarly, the difference between the logits of two probabilities in the residual forecasts in the SPSS constructed model would be the logarithm of the odds ratio (*R*);thereby, providing a shorthand for writing the correct combination of odds ratios simply by adding and/or subtracting $\log(R) = \log\left(\frac{p_1/(1-p_1)}{p_2/(1-p_2)}\right) = \log\left(\frac{p_1}{1-p_1}\right) - \log\left(\frac{p_2}{1-p_2}\right) = \operatorname{logit}(p_1) - \operatorname{logit}(p_2).$ (see Cressie 1993). The logit in logistic regression –based seasonal district-level predictive malaria-related district-level uncertainty risk model then would be a special case of a link function in a generalized linear model (GLM) which, in turn, would be the canonical link function for a binomial distribution. The logit function is the negative of the derivative of the binary entropy function (Hosmer and Lemeshew 2000). In information theory, the binary entropy function, denoted $H(p)$or $H_{\mathrm{b}}(p)$, is defined as the entropy of a Bernoulli process with probability of success *p* (see Hosmer and Lemeshew 2000).

Theoretically, in SPSS, the Bernoulli trial may be modeled geomathematically, in any malaria-related regression-based analyses as a seasonal sampled district-level field/clinical/remote sampled explanatory hyperendemic transmission oriented autoregressive random variable *X* Geomathematics or mathematical geosciences is the application of mathematics to the geosciences which commonly employs computer based technology and which is a form of geophysics (Freeden,et al. 2010). This variable, however, can take on only two sampled explanatory district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive estimator values: 0 and 1. In such circumstances when $X = 1$ the malaria-related district-level risk model residual forecasts would be considered a success and the event $X = 0$ would be considered a failure. These two events would then be mutually exclusive and exhaustive however, if $\Pr(X = 1) = p$, then $\Pr(X = 0) = 1 - p$and the entropy of *X* is provided by$H(X) = H_{\mathrm{b}}(p) = -p\log_2 p - (1-p)\log_2(1-p).$ where $0log_20$ would be taken to be 0 (see Cramer 2003). The logarithms in this formula in SPSS would thereafter be taken to the base 2 in the risk model residual forecasts (i.e., binary logarithm).When $p = \frac{1}{2}$,the binary entropy function in the malaria-related hyperendemic transmission oriented geopredictive autoregressive malaria-related risk model would then attain its maximum value in the residual error uncertainty estimation. Unfortunately, SPSS cannot distinguish the appropriate function in time series by its argument and, as such, the statistical package will confuse functions in the latent geopredictive district-level malaria-related risk model output with the analogous

function related to Rényi entropy, so $H_{\mathrm{b}}(p)$would not be able to dispel error-prone ambiguous terms in the forecasted explanatory hyperendemic transmission oriented uncertainty covariate coefficients.

The Rényi entropy is important in ecology and geostatistics as indices of diversity. The Rényi entropy can provide important quantum information, which can be used as a measure of uncertainty entanglement for parsimoniously constructing a robust seasonal geopredictive time series district-level malaria-related model with robust interpolated residual forecasts (see Jacob et al. 2013b). For example, in the Heisenberg XY spin chain model commonly generated in SPSS, the Rényi entropy is a function of α which can be calculated explicitly by virtue of the fact that it is an automorphic function with respect to a particular subgroup of the modular group. In theoretical computer science, the min-entropy is used in the context of randomness extractors (Cressie 1993).

Unfortunately, an SPSS logit model employed for quantitating district-level time series geosampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients and its entropy would generate mispecified residual uncertainty forecast estimates. However, if a malarialogist/experimenter regresses an empirical dataset in SPSS using multiple district-level hyperendemic transmission-oriented georeferenced explanatory covariate coefficients, a linear geopredictive seasonal regression-based malarial model(i.e., $y = X\beta + u,$) can be constructed instead where X is the design matrix and β is a $k \times 1$ column vector of the seasonal-sampled parameters to be estimated. By so doing, a model estimator for Renyi's quadratic entropy can be developed using kernel density estimation (KDE).

The KDE s a non-parametric way to estimate the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. By letting $(x_1, x_2, \ldots, x_n)$ be an i.i.d. sample drawn from some distribution (e.g., regressed empirical dataset of district-level malaria-related hyperendemic transmission oriented explanatory covariate coefficients) with an unknown density f, the shape of this function f can be easily estimated employing $\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) \quad = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right),$ where $K(\bullet)$ is the kernel — a symmetric but not necessarily positive function that integrates to one — and $h > 0$ is a smoothing parameter called the bandwidth. A kernel with subscript h is called the scaled kernel and defined as $K_h(x) = 1/h\ K(x/h)$(Cressie 1993). Intuitively if a malarialogist/experimenter wants to choose h as small as the sampled district-level field/clinical/remote sampled hyperendemic transmission oriented geopredictive autoregressive data allows, there will be a trade-off between the bias of the estimator and its variance in the risk model. A range of kernel functions are commonly used for geopredictive malarial risk modeling including: uniform, triangular, biweight, triweight, Epanechnikov, normal, and others (see Jacob et al. 2009d). The Epanechnikov kernel derived commonly from $K(u) = \frac{3}{4}(1 - u^2)\,\mathbf{1}_{\{|u| \le 1\}}$,for instance, is optimal in a minimum variance sense for district-level risk modeling although the normal kernel is often used [i.e., $K(x) = \phi(x)$], where ϕ is the standard normal density function. The construction of a kernel density estimate finds interpretations in fields outside of density estimation (Cressie 1993). For instance, in thermodynamics, this is equivalent to the amount of heat generated when heat kernels (e.g., 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.

With cost functions for adaptation in mind, the properties of this Information Potential (IP) estimator can then be carefully presented in a robust geopredictive time series district-level malaria-related hyperendemic transmission oriented model by including its bias and variance in the residual forecasts. The Rényi entropy of order α, where $\alpha \ge 0$and $\alpha \ne 1$, can then be defined as $H_\alpha(X) = \frac{1}{1-\alpha}\log\left(\sum_{i=1}^{n} p_i^\alpha\right)$ (see Hosmer and Lemeshow 2000). Here, $X$is a discrete random variable with possible outcomes $1, 2, ..., n$and corresponding probabilities $p_i \doteq \Pr(X = i)$for $i = 1, \ldots, n$, and the logarithm is base 2. If the probabilities are $p_i = 1/n$for all $i = 1, \ldots, n$, then all the Rényi entropies of the distribution in a robust geopredictive district-level malaria-

related risk model would be equal: $H_\alpha(X) = \log n$. In general, for all discrete random variables X, $H_\alpha(X)$ is a non-increasing function in α.(Cressie 1993) Applications often exploit the following relation between the Rényi entropy and the p-norm: $H_\alpha(X) = \frac{\alpha}{1-\alpha} \log \left(\|X\|_\alpha \right)$ (Griffth 2003). In such circumstances, the discrete probability distribution X in a robust geopredictive malaria-related hyperendemic trasnmission oriented risk model may be interpreted as a vector in $\mathbb{R}^n$ with $X_i = p_i \geq 0$ and $\sum_{i=1}^{n} X_i = 1$.

Further, based on the Nystrom approximation and the primal-dual formulation of Least Squares Support Vector Machines (LS-SVM), it becomes possible to apply a nonlinear geopredictive district-level malaria-related hyperendemic transmission oriented risk model to a large scale regression problem. Least squares support vector machines are least squares versions of support vector machines (SVM), which are a set of related supervised learning methods that analyze data and recognize patterns, and which are used for classification and regression analysis. The Nystrom method is an efficient technique for the eigenvalue decomposition of large kernel matrices (Rao 1973) In machine learning for example, eigenvalue decomposition is used in kernel principal component analysis and kernel Fisher discriminant analysis for the extraction of nonlinear structures and decision boundaries from the kernel matrix. The eigenvectors of the kernel or affinity matrix are also used in many spectral clustering (von Luxburg, 2007) and manifold learning algorithms (Belkin and Niyogi, 2002; Tenenbaum et al., 2000) for the discovery of the intrinsic clustering structure or low-dimensional manifolds. In this version a malarialogist/experimenter would devise the a solution for constructing an autoregressive geopredictive district-level malaria-related hyperendemic transmission oriented risk model by solving a set of linear equations instead of a convex quadratic programming (QP) problem for classical SVMs. Least squares SVM classifiers, were proposed by Suykens and Vandewalle (1999). LS-SVMs are a class of kernel-based learning methods. In numerical analysis, the Nyström method of discretizing an integral equation uses a quadrature rule; (i.e. applying the quadrature rule $\int_a^b h(x)\,\mathrm{d}x \approx \sum_{k=1}^{n} w_k h(x_k)$ to, for instance , the inhomogeneous Fredholm equation of the second kind $f(x) = \lambda u(x) - \int_a^b K(x,x') f(x')\,\mathrm{d}x'$ results in $f(x) \approx \lambda u(x) - \sum_{k=1}^{n} w_k K(x,x_k) f(x_k)$ (see Leonard and Walsh 1974).

An inhomogeneous Fredholm equation of the first kind is written as: $g(t) = \int_a^b K(t,s) f(s)\,\mathrm{d}s$ in SPSS and the problem is, given the continuous kernel function *K(t,s)*, and the function *g(t)*, to find the function *f(s)*.If the kernel is a function only of the difference of its arguments, in a geopredictive district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented autoregressive model namely $K(t,s) = K(t-s)$, the limits of integration would be $\pm\infty$, and, as such, the regression based equation can be rewritten as a convolution of the functions *K* and *f* . By so doing, a solution can be derived by $f(t) = \mathcal{F}_\omega^{-1}\left[\frac{\mathcal{F}_t[g(t)](\omega)}{\mathcal{F}_t[K(t)](\omega)}\right] = \int_{-\infty}^{\infty} \frac{\mathcal{F}_t[g(t)](\omega)}{\mathcal{F}_t[K(t)](\omega)} e^{2\pi i \omega t} \mathrm{d}\omega$ where $\mathcal{F}_t$ and $\mathcal{F}_\omega^{-1}$ are the direct and inverse Fourier transforms respectively. An inhomogeneous Fredholm equation of the second kind can also be given for an SPSS derived risk model as $\phi(t) = f(t) + \lambda \int_a^b K(t,s)\phi(s)\,\mathrm{d}s.$ Given the kernel *K(t,s)*, and the function $f(t)$, the problem is typically to find the function $\phi(t)$(Cressie 1993).

Interestingly, a standard approach to solving this in a time series dataset of district-level SPSS derived malaria-related explanatory hyperendemic transmission oriented covariate coefficients may be to use the resolvent

formalism. Written as a series, the solution in these types of foramlisms are known as the Liouville-Neumann. The Liouville–Neumann series is defined as $\phi(x) = \sum_{n=0}^{\infty} \lambda^n \phi_n(x)$ which is a unique, continuous solution of a Fredholm integral equation of the second kind: $f(t) = \phi(t) - \lambda \int_a^b K(t,s)\phi(s)\,ds.$ (Hosmer and Lemeshew 2000). If the *n*th iterated kernel is then defined in a predictive malaria-related district-level hyperendemic transmission oriented risk model as $K_n(x,z) = \int\int \cdots \int K(x,y_1)\,K(y_1,y_2)\cdots K(y_{n-1},z)\,dy_1 dy_2 \cdots dy_{n-1}$ where the $\phi_n(x) = \int K_n(x,z)\,f(z)\,dz$ and $\phi_0(x) = f(x)$ .The resolvent or solving kernel may then be given by $K(x,z;\lambda) = \sum_{n=0}^{\infty} \lambda^n K_{n+1}(x,z)\,.$ The solution of the integral equation thereafter will be $\phi(x) = \int K(x,z;\lambda)\,f(z)\,dz.$ This computation can be performed in SPSS using a sparse approximation of the non-linear oriented mapping data attributes induced by the kernel matrix, within an active selection of support vectors based on quadratic Renyi entropy criteria. SPSS-Macros supports kernel density estimation (http://www01.ibm.com/software/analytics/spss).

By so doing, the OLS estimator $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$.in SPSS could be employed for determining optimal the statistically significant hyperendemic oriented district-level geopredictive malaria-related residual error estimators. For instance, if sample errors from the time series malarial risk model have equal variance σ^2 and are uncorrelated, the least-squares estimate of β would be the best linear unbiased district-level field/clinical/remote sampled geopredictive error estimates and its variance could be quantitated with $v_{OLS}[\hat{\beta}_{OLS}] = s^2(X'X)^{-1}, s^2 = \frac{\sum_i \hat{u}_i^2}{n-k}$ where $\hat{u}_i$ would be the regression residuals regardless of the Rényi entropy. Unfortunately, assumptions of $E[uu'] = \sigma^2 I_n$are easily violated in SPSS; thus, the OLS estimator would lose its desirable properties for targeting the district-level georeferenced seasonal time series explanatory hyperendemic transmission oriented uncertainty covariate coefficients. Indeed, $V[\hat{\beta}] = V[(X'X)^{-1}X'Y] = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$where $\Sigma = V[u]$ in an autoregressive geopredictive malaria-related district-level risk model (see Jacob et al. 2009d). Therefore, while the SPSS derived OLS point forecasted district-level seasonal sampled malaria-related hyperendemic transmission oriented parameter estimators would remain unbiased; it would not be optimal in the sense of having minimum mean square error. Further, the SPSS derived OLS variance estimator $v_{OLS}[\hat{\beta}_{OLS}]$would not provide a consistent estimate of the variance of the residual uncertainty based on of the OLS estimates in the district-level risk model outputs.

Importantly, SPSS lags are more geomathematical rather than statistical for modern data district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive uncertainty risk based analysis. For example, bootstrapping approaches are non-existent in SPSS for seasonal district-level malaria-related risk forecasting. Basic tests of analytical assumptions for efficient geopredictive risk model construction (e.g., assumptions of independence of malarial indicators) are often not available. SPSS menu offerings for seasonal geopredictive risk modeling are typically the most basic of any standard regression analysis. Additionally, the default graphics are poor and not easily customizable for autoregressive parsimonious district-level malarial risk model construction. However, the major disadvantage for generating unbiased residual forecasts from regressed empirical datasets of district sampled malaria-related time series georeferenced explanatory hyperendemic transmission oriented covariates in SPSS, is that it does not have an undo option when a variable is deleted in the regression weighted error matrix. Therefore, suppose the empirical-sampled data consists of *n* sampled quantitated district-level malaria-related hyperendemic transmission oriented observations using yi, *xi*} Ni=1, where each

sampled field/clinical/remote observation in SPSS would include a scalar response y_i and a vector of p (i.e.predictors) or repressors (i.e.x_i.) In a linear-based robust predictive time series regression based risk model the response variable would then be a linear function of the regressors: $y_i = x_i'\beta + \varepsilon_i$, where β is a $p\times1$ vector of unknown hyperendemic transmission oriented parameter estimators and ε_i's would be the unobserved scalar random variables (i.e., errors). This SPSS derived seasonal district-level malaria-related uncertainty risk model would not be able to account for the discrepancy between the actually observed responses y_i and the "geopredicted risk outcomes" $x'_i\beta$. This procedure would also not be able to denote the matrix transpose in SPSS so that $x'\beta$ is the dot product between the vectors x and β.

In linear algebra, the transpose of a matrix A is another matrix A^T (also written A′, A^{tr},^tA or A^t) (Cressie 1993).Thus, formally, the i th row, j th column element of A^T for any robust seasonal predictive SPSS derived district-level malaria-related risk model would be the j th row, i th column element of A when $[\mathbf{A}^{\mathbf{T}}]_{ij} = [\mathbf{A}]_{ji}$ and A is an $m \times n$ matrix and when A^T is an $n \times m$ matrix. Although this model can also be written in matrix notation in SPSS as $y = X\beta + \varepsilon$, where y and ε are $n\times1$ vectors, and X is an $n\times p$ matrix of regressors, (i.e., design matrix) which may provide statistical significance of any regressed explanatory hyperendemic transmission oriented covariate coefficient in a geopredictive autoregressive district-level malaria-related risk model, the specifications would be erroneous .

It is important to remember, in statistics, a design matrix is a matrix of variables, often denoted by X, that is used in certain statistical models [e.g., the general linearized model] which can contain indicator variables (i.e, ones and zeros) that indicate group membership in an ANOVA, or it can contain values of continuous variables. Thus, as a rule, the constant term is always included in an empirical georeferenced datasets of malarial-related hyperendemic transmission oriented regressors X, by taking $x_{i1} = 1$ for all $i = 1, \ldots, n$(see Jacob et al. 2009d). By so doing, the coefficient β_1 corresponding to this regressor in a SPSS constructed model then would be the intercept identifying and quantitating the explanatory hyperendemic transmission-oriented covariate coefficients for accurately determining statistical significance levels in the residually forecasted district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive estimators. But there may be some inconspicuous latent unaccounted relationships between the regressors in the forecasts that may still remain residually unquantitated in the risk model in SPSS. For instance, the third district-level malarial regressor may be the square of the second regressor in the risk model outputs. In this case SPSS would render a quadratic model in the second regressor assuming that the first regressor is constant in the model. Disadvantages of quadratic and higher-order polynomials are: 1) they may require more reference standards to capture the region of curvature, 2) the correction for bias is more complicated than for the linear model; and, 3) the uncertainty analysis is difficult (Homer and Lemeshew 2000).

Additionally, SPSS does not have any multiple pooled cross-sectional time series routines for adequately categorizing regressed seasonal-sampled district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive estimates. A database that provides a multivariate statistical history for each of a number of individual entities is called a pooled cross-sectional and time series data base in literature (Rao 1973).Further, since there are no count procedures (e.g., Poisson, negative binomial and the zero routines) in SPSS, other MLEs such as Tobit, multinomial logit, ordinal or probit, and complementary log-log models would not be readily available for district-level predictive malarial related risk modeling for quantitating seasonal sampled field/clinical/remote sampled geopredictive autoregressive explanatory hyperendemic transmission oriented covariate coefficient significance levels. Commonly specifications for determining conspicuous latent autocorrelation error coefficients in empirical datasets of seasonal sampled malarial-related explanatory hyperendemic transmission oriented covariate coefficients variables, a Poisson specification with a non-homogenous gamma distributed mean is required (see Jacob et al. 2005b), which unfortunately is not currently available in SPSS.

Also in SPSS no correction for heteroskedascity (e.g., Huber-White) are included. Thus, for example, suppose there is a sequence of seasonal-sampled district-level malaria-related hyperendemic transmission oriented random variables [e.g., $\{Y_T\}_{t=1}^{n}$] in an empirical dataset with a sequence of vectors of random variables, $\{X_t\}_{t=1}^{n}$. In dealing with conditional expectations of Y_t given X_t, the sequence $\{Y_t\}_{t=1}^{n}$ in the residual forecasted parameter estimators, identifying the statistically important seasonal district-level field/clinical/remote sampled malaria-related autoregressive hyperendemic transmission-oriented covariate coefficients would then lead to heteroskedastic

parameters as the conditional variance of Y_t given X_t would change with *t*. Some authors refer to this as conditional heteroscedasticity to emphasize the fact that it is the sequence of conditional variances that changes and not the unconditional variance. In fact it is possible to observe conditional heteroscedasticity in an ecological empirical regressed dataset of seasonal district-level geopredictive malaria-related hyperendemic transmission oriented explanatory covariate coefficients even when dealing with a sequence of unconditional homoscedastic random variables, however, the opposite does not hold. Heteroskedastic parameters would not cause the seasonal malaria-related district-level OLS coefficient estimates to be biased, although they would cause OLS estimates of the variance and, thus, standard errors of the coefficients to be error prone, possibly above or below the true or sampled population variance. Thus, seasonal regression analysis of empirical sampled hyperendemic transmission-oriented observational geopredictive variables would render unbiased estimates when quantitating the relationship between the seasonal-sampled variables and the outcome in the district-level risk model. Biased standard errors lead to biased inference, so results of hypothesis tests are possibly wrong (Hosmer and Lemeshew 2000).

As a consequence of biased standard error estimation, if heteroscedasticity is present in a SPSS constructed predictive district-level ARIMA risk model output, a malarialogist/experimenter might find compelling results against the rejection of a null hypothesis at a given field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive covariate coefficient significance level which may be interpreted as significant, when that null hypothesis is actually uncharacteristic of the actual seasonal sampled regressed covariate coefficients (i.e., make a type II error). In ARIMA modeling in SPSS, as in any statistical software packages, the selection of a best model fit as associated to historical data is directly related to whether residual analysis is performed well. Diagnostic checks including the independence, normality and homoscedasticity of residuals is the most important stage of a time series district-level malarial-related autoregressive geopredictive model building process (Jacob et al. 2005b).

Under certain assumptions, however, the SPSS derived OLS estimator in a geopredictive district-level hyperendemic transmission oriented malaria-related risk model would have a normal asymptotic distribution when properly normalized and centered even when the time series data does not come from a normal distribution. This result would then justify using a normal distribution, or a chi square distribution depending on how the test statistic in the geopredictive district-level malarial risk model is calculated for conducting the hypothesis test. This would hold even under heteroscedasticity in the residually forecasted district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive estimates. More precisely, the SPSS OLS estimator in the presence of heteroscedasticity would be asymptotically normal, when properly centered with a variance-covariance matrix that differs from the case of homoscedasticity in a seasonal district-level malaria-related geopredictive risk model. Unfortunately, since SPSS does not offer any resolution for heteroskedastic parameters specified in residually forecasted uncertainty estimators for accurately identifying optimal unbiased hyperendemic transmission oriented explanatory covariate coefficients. As such unqunatitated heteroskedastic parameters would be a major practical issue encountered in ANOVA inferences generated from regressed seasonal district-level malarial based risk model field and remote sampled data attributes.

ANOVA is a particular form of statistical hypothesis testing heavily used in the analysis of experimental district-level malarial risk-based data analyses. A geopredictive malarial related regression test result, similar to any other statistical test (e.g., calculated from the null hypothesis and a district-level sample) is significant if it is deemed unlikely to have occurred by chance, assuming the truth of the null hypothesis (see Jacob et al. 2005b). Thus, a statistically significant result (e.g., probability (i.e.,*p*-value) rendered from a district-level seasonal sampled risk based malaria-related geopredictive regression-based risk analyses is less than a threshold (i.e., significance level) that would subsequently justify the rejection of the null hypothesis. In the typical application of ANOVA for seasonal district-level malaria-related autoregressive geopredictive risk modeling, the null hypothesis is that all groups are simply random samples of the same seasonal-sampled population (see Hosmer and Lemeshew 2000).This implies that all explanatory district-level field/clinical/remote sampled hyperendemic transmission oriented geopredictive autoregressive covariate coefficient interaction effects in an empirical sampled district-level malaria-related dataset have the same effect. Rejecting the null hypothesis then implies that different regressed covariates results have altered sampled effects in the predictive malarial risk model output.

By construction, hypothesis testing limits the rate of Type I errors (i.e., false positives leading to false claims) to a significance level in a robust malaria-related geopredictive district-level hyperendemic transmission oriented risk model. Commonly, malarialogists and other experimenters wish to limit Type II errors (i.e., false negatives resulting in missed discoveries). The Type II error rate is a function of several things in a district-level malarial geopredictive risk model including sample size (i.e., positively correlated with experiment cost), significance level (e.g., when the standard of proof is high, the chances of overlooking a discovery are also high) and effect size (e.g., when the effect is obvious to the casual observer, Type II error rates are low) (see Jacob et al. 2012b) The terminology of ANOVA is largely from the statistical design of experiments (Cressie 1993). The malarialogist/ experimenter thus would adjust factors and measures responses in the predictive risk model in an attempt to determine an effect (e.g.district-level rainfall on prevalence rates). Routinely, factors in seasonal malaria related district-level geopredictive risk modeling are assigned to experimental units by a combination of randomization and blocking to ensure the validity of the results. Blinding keeps the weighing impartial (Cressie 1993). Responses in these models can show a variability that is partially the result of the effect and is partially random error.

ANOVA is the synthesis of several ideas which is used for multiple purposes in district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive modeling. As a consequence, it is difficult to define concisely or precisely the role of ANOVA in geopredictive malarial risk modeling. For instance, in exploratory ANOVA data analysis, in malarial risk modeling, an organization of additive data decomposition can be conducted by the sums of squares which could indicate the variance of each component of the decomposition (or, equivalently, each set of terms of a linear model). By so doing, comparisons of mean squares, along with F-tests can allow testing of a nested sequence of geopredictive autoregressive district-level malarial model residual forecasts estimates. ANOVA is a linear model fit with coefficient estimates and standard errors (Cressie 1993). Thus, ANOVA can be a statistical tool used in several ways to develop and confirm an explanation for time series observed malarial-related district level hyperendemic transmission oriented data attributes.

Additionally, the ANOVA model is computationally elegant and relatively robust against violations of its assumptions commonly observed in geopredictive district-level malarial risk models [e.g., linear correlated covariate coefficients]. ANOVA provides industrial strength (i.e., multiple sample comparison) statistical analysis (see Hosmer and Lemeshew 2000). ANOVA, however, is difficult to interpret particularly for complex seasonal geopredictive malarial-related field experiments, with split-plot designs being notorious. In statistics, restricted randomization occurs in the design of experiments and in particular in the context of randomized experiments and randomized controlled trials (Bailey 1987). Restricted randomization in geopredictive district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented autoregressive modeling would thus allow intuitively poor allocations of treatments to experimental units to be avoided, while retaining the theoretical benefits of randomization. In some cases the proper application of the method in a robust geopredictive malarial district-level risk model is best determined by problem pattern recognition followed by the consultation of a classic authoritative test. Fortunately, it has been adapted to the analysis of a variety of experimental designs in SPSS.

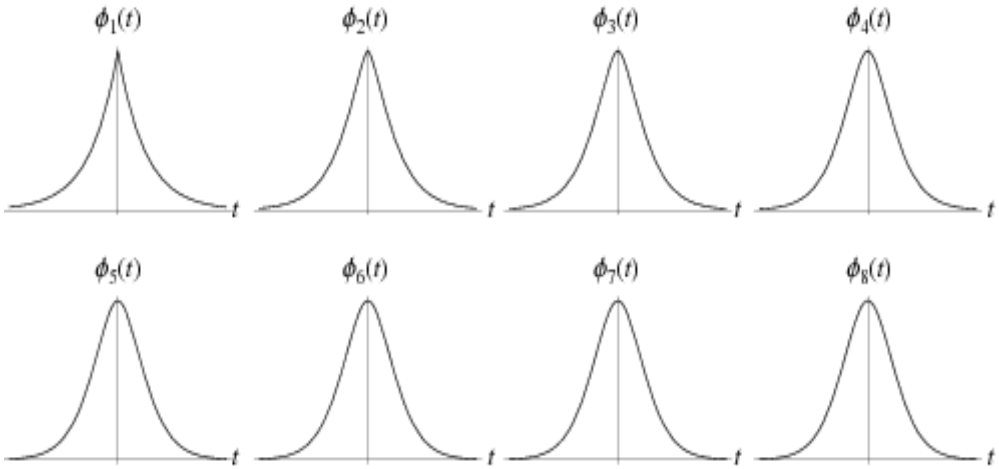
For example, the One-Way ANOVA in SPSS procedure employs a one-way analysis of variance for vigorously regressing a quantitative seasonal sampled malarial –related district level dependent variable (total malarial mosquito *Anopheles gambiae s.l.* aquatic larval habitat field sampled spatiotemporal density count data) by a single factor (independent) variable (e.g. daily humidity). Analysis of variance can then be used to test the hypothesis that several means are equal in the district-level geopredictive risk model. This technique would be an extension of the two-sample *t* test. In addition to determining that differences exists among the means in the sampled regressed hyperendemic transmission oriented data attributes, a malariologist /experimenter can determine which means differ in the empirical sampled dataset. There are two types of tests for comparing means in a geopredictive malarial risk modeling: a priori contrasts and post hoc tests (see Jacob et al.2000d). Contrasts are tests that can be set up before running the geopredictive malarial risk model experiment or, while post hoc tests are being conducted. By so doing, a malariologist /experimenter may additionally test for trends in the empirical sampled district-level malarial-related hyperendemic transmission oriented model parameter estimators across multiple field-sampled categories. When only two groups need to be compared, the studentized range distribution is similar to the Student's t distribution, differing only in that it takes into account the number of means under consideration(Cressie 1993).

International Journal of Geographic Information System
Vol. 1, No. 1, November 2013, PP: 01 - 143
Available Online at http://acascipub.com/Journals.php

A statistical distribution published by Gosset in 1908., stated that given N independent measurements x_i, and then letting $t \equiv \frac{\bar{x} - \mu}{s/\sqrt{N}}$, where $\mu$is the population mean, $\bar{x}$ is the sample mean, and s is the estimator for population standard deviation (i.e., the sample variance) as defined by $s^2 \equiv \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$. then a Student's t-distribution may be defined as the distribution of the random variable t which is the "best" that a malarialogist/experimenter can do not knowing σ. The Student's t-distribution with $n$degrees of freedom is implemented in SPSS as StudentTDistribution[n]. If $\sigma = s$, $t = z$and the distribution becomes the normal distribution the Student's t-distribution will also approach the normal distribution. The Student's t-distribution can then be derived by transforming Student's z-distribution using $z \equiv \frac{\bar{x} - \mu}{s}$, and then defining$t \equiv z\sqrt{n-1}$ .The resulting probability and cumulative distribution functions would then be $f_r(t) = \frac{\Gamma\left[\frac{1}{2}(r+1)\right]}{\sqrt{r\pi}\,\Gamma\left(\frac{1}{2}r\right)\left(1+\frac{t^2}{r}\right)^{(r+1)/2}} = \frac{\left(\frac{r}{r+t^2}\right)^{(1+r)/2}}{\sqrt{r}\,B\left(\frac{1}{2}r, \frac{1}{2}\right)}$,

$F_r(t) = \frac{1}{2} + \frac{1}{2}\left[I\left(1; \frac{1}{2}r, \frac{1}{2}\right) - I\left(\frac{r}{r+t^2}, \frac{1}{2}r, \frac{1}{2}\right)\right]\operatorname{sgn}(t) = \frac{1}{2} - \frac{i\,t\,B\left(-\frac{t^2}{r}; \frac{1}{2}, \frac{1}{2}(1-r)\right)\Gamma\left(\frac{1}{2}(r+1)\right)}{2\sqrt{\pi}\,|t|\,\Gamma\left(\frac{1}{2}r\right)} =$

$\frac{1}{2} + \frac{t\,\Gamma\left(\frac{1}{2}(r+1)\right){}_2F_1\left(\frac{1}{2}, \frac{1}{2}(r+1); \frac{3}{2}; -\frac{t^2}{r}\right)}{\sqrt{\pi r}\,\Gamma\left(\frac{1}{2}r\right)}$, where $r \equiv n-1$ is the number of degrees of freedom, $-\infty < t < \infty$, $\Gamma(z)$is the gamma function, $B(a,b)$is the beta function, ${}_2F_1(a,b;c;z)$is a hypergeometric function, and $I(z;a,b)$is the regularized beta function as defined by $I(z;a,b) = \frac{B(z;a,b)}{B(a,b)}$. The mean, variance, skewness, and kurtosis of Student's t-distribution of a district-level predictive seasonal malarial hyperendemic transmission oriented risk model then would be μ=0 , $\sigma^2 = \frac{r}{r-2}$.,γ_1 =0 and $\gamma_2 = \frac{6}{r-4}$. where



By do doing, the characteristic functions $\phi_n(t)$for the first few sampled district-level hyperendemic transmission oriented covaraite coefficients values would then be of n then would

be $\phi_1(t)=e^{-|t|}$, $\phi_2(t)=\sqrt{2}\,|t|\,K_1\left(\sqrt{2}\,|t|\right)$, $\phi_3(t)=e^{-\sqrt{3}\,|t|}\left(1+\sqrt{3}\,|t|\right)$, $\phi_4(t)=2\,t^2\,K_2(2\,|t|)$, $\phi_5(t)=\frac{1}{3}\,e^{-\sqrt{5}\,|t|}\left(3+3\sqrt{5}\,|t|+5\,t^2\right)$, and so on, where $K_n(x)$ is a modified Bessel function of the second kind.

The multivariate form of the Student's *t*-distribution with correlation matrix r and *m* degrees of freedom for an empirical sampled dataset of malaria-related district-level geopredictive field and remote hyperendemic transmission oriented covariate coefficient may be implemented as MultivariateTDistribution[*r*, *m*] using SPSS multivariate statistics`. By so doing, the so-called $A(t\,|\,n)$ distribution may be useful for testing if two observed distributions have the same mean in the dataset. $A(t\,|\,n)$ gives the probability that the difference in two observed means for a certain statistic *t* with *n* degrees of freedom would be smaller than the observed value purely by chance employing

$$A(t\,|\,n)=\frac{1}{\sqrt{n}\;B\left(\frac{1}{2},\frac{1}{2}n\right)}\int_{-t}^{t}\left(1+\frac{x^2}{n}\right)^{-(1+n)/2}dx.$$

(Cressie 1993). Thus, if a malarialogist/experimenter lets X be a normally distributed randomized district-level seasonal-sampled geopredictive hyperendemic transmission oriented variable with mean 0 and variance σ^2, and also lets Y^2/σ^2 have a chi-squared distribution with *n* degrees of freedom where X and Y are independent then $t\equiv\frac{X\sqrt{n}}{Y}$ would be distributed as Student's *t* with *n* degrees of freedom in the risk model. The means in a geopredictive district-level regression-based malarial risk model that the larger the critical value the more precise the residually forecasted estimates. This makes sense since the more means there are; the greater the likelihood that at least some differences between pairs of hyperendemic transmission oriented variable means will be large due to chance alone in the residually forecasted estimates. Typically, the empirical seasonal sampled district-level malaria-related time series data geopredictive variables is either the sample derivative from fitting the model to an observed time series, or the standardized residuals obtained by dividing the sample data by the conditional standard deviations (see Jacob et al. 2009d).

Alternatively, a malarialogist/experimenter can test for conditional heteroscedasticity in a geopredictive district-level malaria-related risk model by conducting an Engle's ARCH test (1982) (archtest)ARCH test in SAS on any squared residual series. SAS /GIS software provides an interactive Geographic Information System (GIS) within the SAS System. Many types of data have a spatial aspect, including demographics, marketing surveys, and epidemiological studies. This software also enables you to do more than simply view data in its spatial context. The statistical software package allows you to interact with data by selecting features and performing actions based on those selections. SAS/GIS software draws on computing capabilities of the SAS System and enables you to access, manage, analyze, and present your data easily. SAS/GIS software uses two basic types of data: Spatial data - containing the coordinates and identifying information describing the map itself; and, Attribute data - containing information that can be linked to the spatial data--for example, matching addresses or coordinates in the spatial data For example, the U.S. Census Bureau distributes both types of data: *TIGER line files* - contain spatial information that you can use to build maps and summary tape files - contain population and other demographic information that you can link to the maps (www.esri.com).

Among the most important methodology in which a malarialogist/experimenter can use sampled district-level seasonal hyperendemic transmission oriented attribute data in SAS/GIS include using variables from the attribute data as themes for layers. For example, an attribute dataset containing district-level malaria population data attributes could provide a theme for a map of census tracts by creating actions that display or manipulate the attribute data when features are selected in the map. The actions can range from simple, such as displaying sampled hyperendemic transmission oriented observations from an attribute dataset that relate to features in the map, to complex analyses, such as submitting procedures from SAS/STAT software to perform spatial statistical analyses. One of the key concepts with SAS/GIS software is selecting features from a district-level malarial map and then performing actions on the attribute data associated with those features (www.sas.com). Actions defined for robustly quantitating seasonal-sampled district-level geopredictive malarial-related hyperendemic transmission oriented explanatory covariate coefficients then include: 1) displaying field/clinical/remote sampled observations from the attribute data sets that relate to the selected map features 2) opening additional maps that relate to selected map

features 3) displaying graphic images that relate to the selected map features 4) interactively sub-setting the attribute datasets according to the subset of selected map features, and 5) submitting SAS programs for processing subsets of the attribute data that relate to the selected map features (see Jacob et al. 2008c).

The simple ARCH(2) model, in SAS/GIS for instance, can be estimated using the AUTOREG procedure in the statistical software package. The MODEL statement option GARCH=(Q=2) specifies the ARCH(2) model(www.sas.edu). The OUTPUT statement with the CEV= option would then produce the conditional variances V from the ecological empirical dataset of field and remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients. The conditional variance and its residual forecasts can then be calculated using the sampled parameter estimates in $h_t = \hat{\omega} + \hat{\alpha}_1 \varepsilon_{t-1}^2 + \hat{\alpha}_2 \varepsilon_{t-2}^2$ $\mathbf{E}(\varepsilon_{t+d}^2 | \Psi_t) = \hat{\omega} + \sum_{i=1}^{2} \hat{\alpha}_i \mathbf{E}(\varepsilon_{t+d-i}^2 | \Psi_t)$ where $d > 1$.This SAS/GIS derived malaria-related risk model can be estimated for example, as follows:

```
proc autoreg data=ibm maxit=50;

  model r = / noint garch=(q=2);
  output out=a cev=v;
run;
```

While conventional time series SAS/GIS derived malaria-related geopredictive district-level models operate under an assumption of constant variance, the ARCH process introduced allows the conditional variance to change over time as a function of past errors leaving the unconditional variance constant. This type of model behavior has already proven useful in modeling several different economic phenomena. In Engle (1982), Engle (1983) and Engle and Kraft (1983), models for the inflation rate were constructed for recognizing that the uncertainty of inflation tended to change over time. In Coulson and Robins (1985) the estimated inflation volatility was related to some key macroeconomic variables. Models for the term structure using an estimate of the conditional variance as a proxy for the risk premium were also given in Engle et al. (1985). The same idea was applied to the foreign exchange market in Domowitz and Hakkio (1985). In Weiss (1984) ARMA models with ARCH errors were found to be successful in modeling thirteen different U.S. macroeconomic time series. Common to most of the above applications however, is the introduction of a rather arbitrary linear declining lag structure in the conditional variance equation to take account of the long memory typically found in empirical dataset, since estimating a totally free lag distribution often leads to violation of the non-negativity constraints. Thus, SAS/GIS-derived ARCH errors may not be by themselves relevant for geopredictive district-level seasonal malarial risk modeling exercises.

A new, more general class of processes, GARCH (Generalized Autoregressive Conditional Heteroskedastic) in SAS/GIS may instead be introduced for allowing a much more flexible lag structure in a geopredictive district-level malarial risk model. The extension of the ARCH process to the GARCH process bears much resemblance to the extension of the standard time series AR process to a more generalized ARMA process (Cressie 1993) and, as such, may permit a more robust description in many predictive seasonal malarial uncertainty risk modeling situations. By so doing, a new class of processes may be formally also presented and conditions for their wide-sense stationary quantitated. The simple GARCH(1, 1) process, for instance, may be considered for regressing and then determining an empirical sampled dataset of district-level seasonal sampled field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive covariate coefficients statistical significance. As previously mentioned, it is well established, that the autocorrelation and partial autocorrelation functions are useful tools in identifying and checking time series behavior of the ARMA form in the conditional mean.

Similarly the autocorrelations and partial autocorrelations for the squared process may prove helpful in identifying and checking GARCH behavior in the conditional variance equation of a geopredictive district-level malarial regression based risk model. By so doing, the MLE of the linear regression model with GARCH errors may be put forward to effectively quantitate the asymptotic independence between the estimates of the mean employing the variance parameters carried over from the ARCH regression model. It may be argued that a simple GARCH related geopredictive autoregressive seasonal district-level malarial –related uncertainty risk model provides a marginally

better fit and a more plausible learning mechanism than the ARCH model with an eight order linear declining lag structure as in Engle and Kraft (1983).

Alternatively, a new, more general class of processes, GARCH (Generalized Autoregressive Conditional Heteroskedastic) in SAS/GIS may be introduced, allowing for a much more flexible lag structure in a geopredictive malarial model. For example, AutoRegressive Conditional Heteroskedasticity (ARCH) models(Engle, 1982) may be then used to characterize and model observed time series in an empirical ecological dataset of malaria-related explanatory hyperendemic transmission oriented covariate coefficients . Based on skewness and kurtosis, Jarque and Bera (1980) calculated the test statistic $T_N = \left[\frac{N}{6}b_1^2 + \frac{N}{24}(b_2-3)^2\right]$ where $b_1 = \frac{\sqrt{N}\sum_{t=1}^{N}\hat{u}_t^3}{\left(\sum_{t=1}^{N}\hat{u}_t^2\right)^{\frac{3}{2}}} b_2 = \frac{N\sum_{t=1}^{N}\hat{u}_t^4}{\left(\sum_{t=1}^{N}\hat{u}_t^2\right)^2}$. The $\chi_2(2)$ distribution rendered an approximation to the normality test T_N. When the GARCH model was estimated, the normality test was obtained using the standardized residuals $\hat{u}_t = \hat{\epsilon}_t/\sqrt{h_t}$. As such, this normality test may be used to detect misspecification in an empirical dataset of district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive covariate coefficients constructed from a family of ARCH models.

SAS/GIS derived ARCH models are employed whenever there is reason to believe that, at any point in a series, the terms will have a characteristic size, or variance. In particular ARCH models assume the variance of the current error term or innovation to be a function of the actual sizes of the previous time periods' error terms: often the variance is related to the squares of the previous innovations. ARCH models are also employed in modeling financial time series, for example, that exhibit time-varying volatility clustering, (i.e. periods of swings) followed by periods of relative calm Thus, suppose a malarialogist/experimenter wants to model a time series dataset of district-level malaria-related geopredictive time series explanatory hyperendemic transmission oriented covariate coefficients using an ARCH process. In these models , ϵ_t would denote the error terms (i.e., return residuals, with respect to a mean process) in the series terms. These ϵ_t would then be split into a stochastic piece z_t and a time-dependent standard deviation σ_t for characterizing the typical size of the terms so that $\epsilon_t = \sigma_t z_t$ could be efficitively qunatiated in the risk model. The random variable z_t is a strong white noise process (Cressie 1993). The series σ_t^2 would then be modeled by $\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_q\epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2$ where $\alpha_0 > 0$ and $\alpha_i \geq 0,\ i > 0$. An ARCH(q) district-level geopredictive malarial related risk model can then be estimated using OLS.

A methodology to test for the lag length of SAS ARCH errors using the Lagrange multiplier test can also then be proposed. This procedure could include :1) estimating the best fitting geopredictive autoregressive model AR(q) $y_t = a_0 + a_1 y_{t-1} + \cdots + a_q y_{t-q} + \epsilon_t = a_0 + \sum_{i=1}^{q} a_i y_{t-i} + \epsilon_t$. By so doing, the squares of the error $\hat{\epsilon}^2$ in the model can be obtained and thereafter regressed employing a constant and *q* lagged values in $\hat{\epsilon}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{q}\hat{\alpha}_i\hat{\epsilon}_{t-i}^2$ where *q* is the length of ARCH lags. The null hypothesis for the geopredictive autoregressive district-level malarial risk model would be that, in the absence of ARCH components, $\alpha_i = 0$ would be rendered for all $i = 1, \cdots, q$. The alternative hypothesis would be that, in the presence of ARCH components, at least one of the estimated α_i district sampled explanatory hyperendemic trasnmission oriented coavariate coefficients must be significant. In a sample of the residuals under the null hypothesis of no ARCH errors, the test statistic TR^2 would then follow χ^2 distribution with *q* degrees of freedom. If TR^2 is greater

than the chi-square table value in the district-level predictive risk model residual forecasts, the malarialogist/experimenter can reject the null hypothesis and conclude there is an ARCH effect in the ARMA model. If TR^2 is smaller than the chi-square table value the null hypothesis may not be rejected. Further, if an SAS derived ARMA-related residual predictive district-level malarial risk model is assumed for the error variance, the model could be delineated as a generalized autoregressive conditional heteroskedasticity [GARCH, Bollerslev(1986)] model. In such circumstances, the GARCH(p, q) model (where p is the order of the GARCH terms σ^2 and q) would be based on the order of the ARCH terms ϵ^2 as given by $\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \cdots + \alpha_q\epsilon_{t-q}^2 + \beta_1\sigma_{t-1}^2 + \cdots + \beta_p\sigma_{t-p}^2 = \alpha_0 + \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2 + \sum_{i=1}^{p}\beta_i\sigma_{t-i}^2$

Generally, when testing for heteroskedasticity in econometric models, the best test has been the White test. However, when dealing with time series sampled geopredictive district-level regression-based malarial data, this means testing for ARCH errors and GARCH errors in the residually forecasted field/clinical/remote sampled malaria-related hyperendemic transmission oriented autoregressive estimates. The lag length *p* of a GARCH(p, q) process in the time series geopredictive seasonal district-level malarial risk model can then be established by estimating the best fitting AR(q) model using $y_t = a_0 + a_1 y_{t-1} + \cdots + a_q y_{t-q} + \epsilon_t = a_0 + \sum_{i=1}^{q} a_i y_{t-i} + \epsilon_t$ and by computing and plotting the autocorrelations of ϵ^2 by $\rho = \frac{\sum_{t=i+1}^{T}(\hat{\epsilon}_t^2 - \hat{\sigma}_t^2)(\hat{\epsilon}_{t-1}^2 - \hat{\sigma}_{t-1}^2)}{\sum_{t=1}^{T}(\hat{\epsilon}_t^2 - \hat{\sigma}_t^2)^2}$ The asymptotic, for large samples, standard deviation of $\rho(i)$ would then be $1/\sqrt{T}$.(see Engle 1972) Individual time series explanatory hyperendemic transmission oriented covariate coefficient measurement values that are larger than $1/\sqrt{T}$ would then indicate seasonal district-level geopredictive GARCH errors. Further, to estimate the total number of lags, the Ljung-Box test may be employed for determining the statistical significance levels of the sampled explanatory hyperendemic transmission oriented covariate coefficients. The Ljung-Box Q-statistic follows χ^2 distribution with *n* degrees of freedom if the squared residuals ϵ_t^2 are uncorrelated(Cressie 1993). It is recommended to consider up to T/4 values of *n*. (see Engle 1972). The null hypothesis then for a seasonal geopredictive district-level malarial risk models would state that there are no ARCH or GARCH errors in the sampled hyperendemic transmission oriented residual forecasts. Rejecting the null in the risk model would thus means that such errors exist only in the conditional variance.

Nonlinear GARCH (NGARCH) also known as Nonlinear Asymmetric GARCH(1,1) (NAGARCH) as introduced by Engle and Ng in 1993 [i.e $\sigma_t^2 = \omega + \alpha(\epsilon_{t-1} - \theta\sigma_{t-1})^2 + \beta\sigma_{t-1}^2\alpha,\ \beta \geq 0;\ \omega > 0$] may be also useful for regressing explanatory district-level predicting malarial-related hyperendemic transmission oriented covariate coefficients. By so doing, then if the residual forecasts would reflect the leverage effect which in turn would increase future volatility in the coefficients by a larger amount than positive uncorrelated effects of the same magnitude. As such, this would signify that negative biased effects in the regressed parameter estimators. This model should not be confused, however, with the NARCH model, together with the NGARCH extension as introduced by Higgins and Bera in 1992.

SAS/GIS can also generate Integrated Generalized Autoregressive Conditional Heteroskedasticity IGARCH which is a restricted version of the GARCH model, where the persistent explanatory hyperendemic transmission oriented district-level parameter estimators can be made to sum up to one. By so doing, there would a unit root in the GARCH process. The condition for this would be $\sum_{i=1}^{p}\beta_i + \sum_{i=1}^{q}\alpha_i = 1$ (see Engel 1972).The exponential

generalized autoregressive conditional heteroskedastic (EGARCH) model by Nelson (1991) is another form of the GARCH model that may be employed for uncertainty quantitation of regressed residuals rendered from autoregressive time series geopredictive district-level malarial risk model. Formally, an EGARCH(p,q): $\log \sigma_t^2 = \omega + \sum_{k=1}^{q} \beta_k g(Z_{t-k}) + \sum_{k=1}^{p} \alpha_k \log \sigma_{t-k}^2$ where $g(Z_t) = \theta Z_t + \lambda(|Z_t| - E(|Z_t|))$, σ_t^2 is the conditional variance, ω, β, α, $\theta$and $\lambda$are coefficients, and $Z_t$may be a standard normal variable or, come from a generalized risk model related error distribution. The formulation for $g(Z_t)$ thus would allow the sign and the magnitude of $Z_t$to in the geopredictive district-level malaria-related risk model to have separate effects on the volatility. This may be particularly useful to assess the significance of the sampled explanatory hyperendemic transmission oriented covariate coefficients. Fortunately, since $\log \sigma_t^2$ may be negative there would be no restrictions on the malarial-related district -level parameter estimator dataset.

Further, the GARCH-in-mean (GARCH-M) model adds a heteroskedasticity term into the mean equation. It has the specification: $y_t = \beta x_t + \lambda \sigma_t + \epsilon_t$, thus the residual ϵ_t in a geopredictive seasonal district-level malarial risk model would be defined as $\epsilon_t = \sigma_t \times z_t$. The Quadratic GARCH (QGARCH) model by Sentana (1995) may also be employed to model symmetric effects of positive and negative shocks in a geopredictive district-level malarial risk model. In the example of a GARCH(1,1) model, the residual process σ_t then would be $\epsilon_t = \sigma_t z_t$where $z_t$is i.i.d. and$\sigma_t^2 = K + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \phi \epsilon_{t-1}$. Similar to QGARCH, the Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model by Glosten et al. (1993) may also model seasonal sampled geopredictive explanatory hyperendemic transmission oriented covariate coefficients asymmetry in a risk empirical sampled parameter estimator model ARCH derived process. The idea here would be to model $\epsilon_t = \sigma_t z_t$where $z_t$is i.i.d., and$\sigma_t^2 = K + \delta \sigma_{t-1}^2 + \alpha \epsilon_{t-1}^2 + \phi \epsilon_{t-1}^2 I_{t-1}$where $I_{t-1} = 0$if,$\epsilon_{t-1} \geq 0$, and $I_{t-1} = 1$if $\epsilon_{t-1} < 0$.The Threshold GARCH (TGARCH) model by Zakoian (1994) may also be employed for robust district-level modeling using field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive as it is similar to GJR GARCH, whereby the specification would be one based on conditional standard deviations in the empirical sampled datasets. The only difference between the models is that instead of conditional variance:$\sigma_t = K + \delta \sigma_{t-1} + \alpha_1^+ \epsilon_{t-1}^+ + \alpha_1^- \epsilon_{t-1}^-$where $\epsilon_{t-1}^+ = \epsilon_{t-1}$if $\epsilon_{t-1} > 0$, and $\epsilon_{t-1}^+ = 0$if $\epsilon_{t-1} \leq 0$.$\epsilon_{t-1}^- = \epsilon_{t-1}$if $\epsilon_{t-1} \leq 0$, and $\epsilon_{t-1}^- = 0$if $\epsilon_{t-1} > 0$would be used.

The extension of the ARCH process to the GARCH process may reveal as much resemblance to the extension of the standard time series AR process to the general ARMA geopredictive seasonal district-level malarial data processing. As such a more parsimonious description in many risk modeling situations may be permitted in SAS/GIS. By so doing, a new class of processes may be also formally presented and conditions for their wide-sense stationarity. The simple GARCH (1, 1) process, for instance, may be considered for quantitating a correlated empirical sampled dataset of malarial-related explanatory district level hyperenedmic transmission oriented covariate coefficients. As mentioned it is well established, that the autocorrelation and partial autocorrelation functions are useful tools in identifying and checking time series behavior of the ARMA form in the conditional mean.

The autocorrelations and partial autocorrelations for the squared process may prove helpful in identifying and checking GARCH behavior in the conditional variance equation of a geopredictive malarial risk model residual forecasts. As such, the MLE of the linear regression model with GARCH errors may be put forward, to quantitate the asymptotic independence between the estimates of the mean employing the variance parameters carried over from the ARCH regression model. Additionally, a unified approach to generating standardized-residuals-based correlation tests for checking GARCH-type models may be pursued. This approach may be valid in the presence of geopredictive seasonal malarial-related hyperendemic transmission oriented uncertainty model estimation, using

various standardized error distributions which may in turn be applicable to testing various types of misspecifications in residually forecasted estimates. By using this approach, a malarialogist/experimenter could theoretically also propose a class of power-transformed-series (PTS) correlation tests for providing certain robustifications and power extensions to the Box–Pierce, McLeod–Li, Li–Mak, and Berkes–Horv´ath–Kokoszka tests for diagnosing GARCH-type malarial district-level geopredictive risk models It may be then argued that a simple GARCH related district-level predictive malarial risk model provides a marginally better fit and a more plausible learning mechanism than the ARCH model with an eight order linear declining lag structure as in Engle and Kraft (1983).

The Ljung-Box statistic any also be provided in the SAS/GIS procedure ARIMA for an assortment of lags . For large, the Box- Pierce and Ljung-Box statistics are essentially equivalent. The Ljung-Box (1978) statistic is typically used since it better approximates a chi-squared random variable for smaller. Interestingly, a similar statistic to the Ljung-Box statistic was introduced by Monti (1994) which uses the standardized partial autocorrelation function up to lag where the residual partial autocorrelation is at lag. Recently, Peña and Rodríguez (2002) proposed a statistic based on the determinant of the residual autocorrelation matrix. Under the null hypothesis the authors suggested a fitted adequate model for the ARMA process. By so doing, the authors where also able to generate a matrix in SAS/GIS that approximated the identity matrix.

The identity matrix is the simplest nontrivial diagonal matrix, defined such that $\mathbf{I}(\mathbf{X}) \equiv \mathbf{X}$ for all vectors $\mathbf{X}$. An identity matrix may be denoted I or E (the latter being an abbreviation for the German term "Einheitsmatrix"; Courant and Hilbert 1989, p. 7). Identity matrices are sometimes also known as unit matrices (Akivis and Goldberg 1972,). The $n \times n$ identity matrix is given explicitly by $I_{ij} = \delta_{ij}$ for $i, j = 1, 2, \ldots, n$, where δ_{ij} is the Kronecker delta. Written explicitly, $\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$. The $n \times n$ identity matrix is implemented in SAS as IdentityMatrix[n]. "Square root of identity" matrices can be defined for $\mathbf{I}_n$ by solving $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix}$. For $n = 2$, the most general form of the resulting square root matrix is $\mathbf{I}_2^{1/2} = \begin{bmatrix} \pm d & \frac{1-d^2}{c} \\ c & \mp d \end{bmatrix}, \begin{bmatrix} \pm d & c \\ \frac{1-d^2}{c} & \mp d \end{bmatrix}$ giving $\begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}, \begin{bmatrix} \pm 1 & 0 \\ c & \mp 1 \end{bmatrix}, \begin{bmatrix} \pm 1 & c \\ 0 & \mp 1 \end{bmatrix}$ as limiting cases. Testing for model adequacy is equivalent to testing if is approximately the identity matrix (Cressie 1993). Thus, by employing an identity matrix district-level regressed field/clinical/remote hyperendemic transmission oriented residual forecasts rendered may be asymptotically distributed as a linear combination of chi-squared random variables while simultaneously approximately a Gamma distributed for larger covariate coefficient measurement values. In practice, they however recommended that the matrix be constructed using the standardized residual forecasts as this would, according to the authors improve the Gamma distribution approximation.

Further, in Peña and Rodríguez (2006) they showed that the log of the determinant follows the same asymptotic distribution as and the residual forecasts according to them can be better in small sample time series. The statistic then may determines whether the matrix of a geopredictive seasonal malaria-related hyperendemic transmission oriented risk model is an identity matrix, or equivalent (i.e., if the fitted model is adequate). It has been demonstrated that both and improve over the Ljung-Box and Box-Pierce statistics; see Monti (1994) or Peña and Rodríguez (2002, 2006). However, neither appears to be frequently implemented in applications of time series district-level malarial-related data attributes. Particularly, the Peña and Rodríguez statistic may be difficult to implement since it involves calculating the determinant of a matrix. As pointed out in Lin and McLeod (2006), the statistic constructed using the standardized residuals may be degenerate in practice since the matrix could be ill-conditioned or singular.

The Ljung-Box Q-test in SAS may be also employed to assess autocorrelation in any empirical sampled district level malarial –related series with a constant mean. This includes residually forecasted series, which can be tested for autocorrelation during residual model diagnostic checks. Additionally, if the residuals result from fitting a model with *g* geoparameter estimators, a malarialogist and or experimenter could compare the test statistic to a χ^2 distribution with *m* – *g* degrees of freedom, if so desired. Optional input arguments to lbqtest in SPSS could then modify the degrees of freedom of the null distribution in the dataset(ww.sas.edu). lbqtest computes the sample Q-statistic whereby the last row of the series contains the most recent observation of a stochastic sequence (www-01.ibm.com/software/analytics/spss/).

The Q-statistic is a test statistic output by either the Box-Pierce test or, in a modified version which provides better small sample properties, by the Ljung-Box test. The q statistic or studentized range statistic is a statistic used for multiple significance testing across a number of means. The formula for Tukey's test is: $q_s = \frac{Y_A - Y_B}{SE}$, where Y_A is the larger of the two means being compared, Y_B is the smaller of the two means being compared, and SE is the standard error of the data in question. This q_s value can then be compared to a q value in SAS constructed geopredictive district-level seasonal malaria-related risk model employing a studentized range distribution. If the q_s value is larger than the $q_{critical}$ value obtained from the distribution, the two means would then be deemed significantly different. Tukey's test compares the means of every treatment to the means of every other treatment; that is, it applies simultaneously to the set of all pairwise comparisons $\mu_i - \mu_j$ and identifies any difference between two means that is greater than the expected standard error (Hosmer and Lemeshew 2000). The confidence coefficient for the district-level malaria-related empirical sampled dataset for all sample sizes would then be equal and exactly $1 - \alpha$. For unequal sample sizes, the confidence coefficient is greater than $1 – \alpha$ (Cressie 1993) In other words, the Tukey method would be conservative when there are unequal sample sizes in the malarial risk model. Since the null hypothesis for Tukey's test in SAS/GIS routinely compares all means from the same population (i.e. $\mu_1 = \mu_2 = \mu_3 = ... = \mu_n$) (http://www-01.ibm.com/software/analytics/spss/), the means in the geopredictive district-level malarial-related regression-based risk model would be as normally distributed according to the central limit theorem (CLT).

In probability theory, the CLT states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, is approximately normally distributed. That is, suppose that a sample in a district-level predictive seasonal malarial –related empirical ecological dataset is obtained containing a large number of field/clinical/remote hyperendemic transmission oriented observations. Further, suppose each malaria-related geopredictor is randomly generated in such a way that it does not depend on the values of the other hyperendemic transmission oriented observations in the dataset, and that the arithmetic average of the observed values is computed. If this procedure is performed many times, regardless of statistical software package employed for the analyses, the computed average will not always be the same each time; the CLT states that the computed values of the average will be distributed according to the normal distribution (commonly known as a "bell curve") (see Rao 1973).

The CLT 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, given that they comply with certain conditions. In more general probability theory, a CLT is any of a set of weak-convergence theorems. They all express the fact that a sum of many i.i.d. random 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. variables is finite in a robust geopredictive district-level malarial risk model, the attractor distribution is the normal distribution (see Jacob et al. 2012b). In contrast, the sum of a number of i.i.d. random variables with power law tail distributions will 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 variables grows in the district-level autoregressive malarial geopredictive risk model.

For instance , if a malarialogist/experimenter lets$\{X_1, ..., X_n\}$ be a random sample of size n—that is, a sequence of i.d.d. random variables drawn from empirical distributions of seasonal sampled georeferenced explanatory field/clinical/remote hyperendemic transmission oriented covariate coefficient expected values given by μ then the finite variances given by σ^2. By the law of large numbers, the sample averages in the geopredictive district-level malarial risk model would converge in probability and almost surely to the expected value μ as $n \to \infty$. However, suppose, the malarialogist/experimenter is interested in the sample average $S_n := \frac{X_1 + \cdots + X_n}{n}$ random variables. The classical CLT would then describe the size and the distributional form of the stochastic fluctuations around the deterministic number μ during convergence. More precisely, the theorem states that as n gets larger in the geopredictive seasonal malarial-related risk model, the distribution of the difference between the sample average S_n and its limit μ, must be multiplied by the factor $\sqrt{n}$ (that is $\sqrt{n}(S_n - \mu)$), so as to approximate the normal distribution with mean 0 and variance σ^2. For large enough n, the distribution of S_n in the risk model would then be close to the normal distribution with mean μ and variance σ^2/n. The usefulness of the theorem is that the distribution of $\sqrt{n}$ $(S_n - \mu)$ in the geopredictive malarial-related georefernced hyperendemic transmission oriented data points would approach normality regardless of the shape of the distribution of the individual X_i's.in the regressed dataset.

For example, suppose $\{X_1, X_2,\}$ is a sequence of i.i.d. random variables in a district-level ecological empirical dataset of hyperendemic malarial transmission oriented observational predictors regressed with with $E[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$. Then as n approaches infinity, in SAS/GIS, the random sampled district-level variables $\sqrt{n}(S_n - \mu)$ will converge in distribution to a normal $N(0, \sigma^2)$[e.g. $\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \mu\right) \xrightarrow{d} N(0, \sigma^2).$]. In the case $\sigma > 0$, in the residually forecasted estimates, the convergence in distribution would signify that the cumulative distribution functions of $\sqrt{n}(S_n - \mu)$ and, as such, would converge point wise to the cdf of the $N(0, \sigma^2)$ distribution. As such, for every sampled district-level geopredictive malarial-related field/clinical/remote hyperendemic transmission oriented covariate coefficient measurement values z, $\lim_{n\to\infty} \Pr[\sqrt{n}(S_n - \mu) \le z] = \Phi(z/\sigma),$ where $\Phi(x)$ would be the standard normal and as such cdf could be evaluated at x. Note that the convergence would be uniform in z the geopredictive risk model in the sense that $\lim_{n\to\infty} \sup_{z\in\mathbf{R}} \left| \Pr[\sqrt{n}(S_n - \mu) \le z] - \Phi(z/\sigma) \right| = 0,$ would denote the least upper bound or supremum of the regressed district-level data attributes. This then would then give rise to the normality assumption of Tukey's test. The assumptions in the geopredictive district-level seasonal malarial-related risk model then would be that the sampled observational predictors being tested are independent and that there is equal within-group variance across the groups associated with each mean in the test (i.e., homogeneity of variance).

Importantly, when testing the residuals of an estimated ARIMA malarial-related predictive risk models in SAS/GIS the degrees of freedom would need to be adjusted to reflect the geoparameter estimation. For instance, for an ARIMA (p,0,q) SAS derived malarial-related geopredictive risk model, the degrees of freedom should be set to $m - p - q$. The studentized range computed from a list $x_1, ..., x_n$ of the seasonal sampled field/clinical/remote hyperendemic transmission oriented explanatory covariate coefficients then would be given by the formulas $q_{n,\nu} = \frac{\max\{x_1, \ldots x_n\} - \min\{x_1, \ldots x_n\}}{s} = \max_{i,j=1,\ldots,n} \left\{ \frac{x_i - x_j}{s} \right\}$

where $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2,$ where the square of the sample standard deviation s, could be computed as $\overline{x} = \frac{x_1 + \cdots + x_n}{n}$ (i.e.,the sample mean).The critical value of q would then be based on three factors:α

(i.e., the probability of rejecting a true null hypothesis) *n* (i.e, the number of district-level hyperendemic transmission oriented observations or groups) :and.*v* (i.e., degrees of freedom in the second sample). Thus, if X_1, ..., X_n are i.d.d. predictive district-level malarial-related random variables that are normally distributed in an empirical sampled dataset, the probability distribution of their studentized range would be the studentized range distribution. This probability distribution in the risk model would then be the same regardless of the expected value and standard deviation of the normal distribution from which the sample is drawn. This probability distribution has applications to hypothesis testing in malaria research. For example, Tukey's range test and Duncan's new multiple range test (MRT), which uses q statistics, can be used as post-hoc analysis to test between two groups especially if there is a significant difference after rejecting null hypothesis by ANOVA.

Conversely, for a pure geopredictive malarial district-level malaria-related risk model the Yule-Walker (YW) equations may be used to provide a fit 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 MacOS (http://www.r-project.org) comprehensive. R incorporates all of the standard statistical tests, models, and analyses, as well as providing a comprehensive language for managing and manipulating data Presently R is maintained by a core team of 19 developers, including some very senior statisticians. The graphical capabilities of R are outstanding, providing a fully programmable graphics language that surpasses most other statistical and graphical packages. The validity of the R software is ensured through openly validated and comprehensive governance as documented for the US Food and Drug Administration (R Foundation for Statistical Computing, 2008). Because R is open source, unlike closed source software, it has been reviewed by many internationally renowned statisticians and computational scientists. R is licensed under the GNU General Public License, with copyright held by The R Foundation for Statistical Computing. R has no license restrictions (other than ensuring our freedom to use it at our own discretion), and so a malarialogist/experimeneter may run it anywhere and at any time, and even sell it under the conditions of the license. R has over 4800 packages available from multiple repositories specializing in topics like econometrics, data mining, spatial analysis, and bio-informatics. (http://www.r-project.org). R is cross-platform. R runs on many operating systems and different hardware. It is popularly used on GNU/Linux, Macintosh, and Microsoft Windows, running on both 32 and 64 bit processors. R plays well with many other tools, importing data, for example, from CSV les, SAS, and SPSS, or directly from Microsoft Excel, Microsoft Access, Oracle, MySQL, and SQLite. It can also produce graphics output in PDF, JPG, PNG, and SVG formats, and table output for LATEX and HTML. R has active user groups where questions can be asked and are often quickly responded to, often by the very people who developed the environment. This support is second to none. ˆ New books for R (the Springer Use R! series) are emerging, and there is now a very good library of books for using R.

Jacob et al. (2010b) constructed multiple Eastern Equine Encephalitis Virus (EEEV) mosquito (e.g.*Culex erracticus*)-related geopredictive risk model to seasonally quantitate environmental estimators of arboviral disease transmission in Central Alabama in R. The YW equations the authors used were based on following set of equations

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

where $m = 0, ..., p$, yielding $p + 1$ equations. In this model $\gamma_m$was the autocovariance function of X_t, which was the standard deviation of the input noise process which also was the Kronecker delta function. In mathematics, the Kronecker delta or Kronecker's delta is a function of two variables, usually integers (Cressie 1993). The function was 1, if the sampled *Cx. erracticus*) geopredictive variables were equal, and 0 otherwise:

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

where Kronecker delta δ_{ij} was a piecewise function of variables $i$and j. For example, $\delta_{1\,2} = 0$, whereas $\delta_{3\,3} = 1$ in the model. 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 $\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{ij} a_i \delta_{ij} b_j$ (Spiegel et al.1997). Because the last part of the *Cx erracticus* geopredictive risk model, the equation was non-zero only if $m = 0$. A set of equations were then solved by representing the equations for $m > 0$ in matrix form, thus rendering the

equation $\begin{bmatrix}\gamma_1\\ \gamma_2\\ \gamma_3\\ \vdots\\ \gamma_p\end{bmatrix} = \begin{bmatrix}\gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots\\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots\\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots\\ \vdots & \vdots & \vdots & \ddots\\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots\end{bmatrix}\begin{bmatrix}\varphi_1\\ \varphi_2\\ \varphi_3\\ \vdots\\ \varphi_p\end{bmatrix}$ which was then solved for all $\{\varphi_m; m = 1, 2, \cdots, p\}$. The remaining equation for *m* = 0 was $\gamma_0 = \sum_{k=1}^{p} \varphi_k \gamma_{-k} + \sigma_\varepsilon^2,$ and as such $\{\varphi_m; m = 1, 2, \cdots, p\}$ was solved for σ_ε^2.

An alternative formulation was then constructed in Jacob et al. (2010b) in terms of the autocorrelation function. The autocorrelation function measures the correlation of a signal *x(t)* with itself shifted by some time delay τ which can then be used to detect repeats or periodicity in a signal (Griffith 2003). The authors then used the autocorrelation to assess the effect of fluctuations (i.e., noise) on a periodic signal.The *Cx erracticus* habitat model AR parameters were then determined by the first p+1 elements $\rho(\tau)$ of the autocorrelation function. The full autocorrelation function was then derived by recursively calculating $\rho(\tau) = \sum_{k=1}^{p} \varphi_k \rho(k - \tau)$ The YW equations for an AR(2) process was then $\gamma_1 = \varphi_1\gamma_0 + \varphi_2\gamma_{-1}$ and $\gamma_2 = \varphi_1\gamma_1 + \varphi_2\gamma_0$ Using the first equation then yielded $\rho_1 = \gamma_1/\gamma_0 = \frac{\varphi_1}{1 - \varphi_2}$ while the recursion formula yielded $\rho_2 = \gamma_2/\gamma_0 = \frac{\varphi_1^2 - \varphi_2^2 + \varphi_2}{1 - \varphi_2}$ The solution set also included one solution, the minimal norm solution, which defined the autoregressive system in the geopredictive *Cx. erraticus* habitat risk model whose characteristic polynomial had either only stable zeros implying that only one stationary output existed for this system . Intermittently, the set was linearly regular or, had stable zeros as well as zeros of unit modulus, implying that stationary solutions of the system were a sum of a linearly regular process and a linearly singular process. The numbers of stable and unit circle zeros of the characteristic polynomial of the defined *Cx. erraticus* habitat autoregressive forecasting model system which was characterized in terms of the ranks of certain error matrices, and the characteristic polynomial of the autoregressive model defined by the minimal norm solution. By so doing, the residual forecasts had the least number of unit circle zeros and the most number of stable zeros over all possible solutions. Autoregressive statistics were then generated using the AUTOREG procedure in R to estimate whether the OLS regression estimates indicated significant serial correlation with an estimated order of a lagged covariance of 1. The AUTOREG procedure corrected for serial correlation using the YW method. The statistic indicated that serial correlation was not significant in the YW corrected *Cx. erraticus* predictive risk model. The YW estimates for the model indicated a $R^2 = 0.632$, F statistics of 39.177, and Durbin-Watson score of 1.935.

In statistics, the Durbin–Watson statistic is a test statistic used to detect the presence of autocorrelation (i.e., a relationship between values separated from each other by a given time lag) in the residuals (prediction errors) from a regression analysis. Under the assumption of normally distributed disturbances, the null distribution of the Durbin-Watson statistic is the distribution of a linear combination of chi-squared variables (see Griffith 2003). The *p*-value is computed using the Fortran version of Applied Statistics Algorithm as in Farebrother (1980, 1984). This algorithm is called "pan" or "gradsol". This p value is computed using a normal approximation with mean and variance of the Durbin-Watson test statistic If e_t is the residual associated with a seasonal sampled malarial-related district-level field/clinical/remote sampled hyperendemic transmission oriented geopredictive autoregressive explanatory hyperendemic transmission oriented observational predictors at time *t*, then the test statistic is $d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2},$ where *T* is the number of observations. Since *d* is approximately equal to $2(1 - r)$, where *r* is the sample autocorrelation of the residuals, $d = 2$ would indicate no autocorrelation in the malarial model residuals. The value of *d* always lies between 0 and 4 (Cressie 1993). If the Durbin–Watson statistic is substantially

less than 2, there would be evidence of positive serial correlation in the malaria-related hyperendemic transmission oriented residual forecasts.

For instance, a seasonal-sampled malaria-related district-level explanatory hyperendemic trasnmission generalized Durbin-Watson Test may consider the following linear regression model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\nu}$ using mupliple covariate coeffcients where $\mathbf{X}$is a $N \times k$data matrix, $\beta$is a $k \times 1$coefficient vector, and $\boldsymbol{\nu}$is a $N \times 1$ disturbance vector. The error term $\boldsymbol{\nu}$ would then be assumed to be generated by the jth-order autoregressive process $\nu_t = \varepsilon_t - \varphi_j \nu_{t-j}$where$|\varphi_j| < 1$, $\varepsilon_t$is a sequence of independent normal error terms with mean 0 and varianceσ^2. Usually, the Durbin-Watson statistic is used to test the null hypothesis $H_0 : \varphi_1 = 0$against$H_1 : -\varphi_1 > 0$. (Griffith 2003). Vinod (1973) generalized the Durbin-Watson statistic: $d_j = \frac{\sum_{t=j+1}^{N} (\hat{\nu}_t - \hat{\nu}_{t-j})^2}{\sum_{t=1}^{N} \hat{\nu}_t^2}$ where $\hat{\boldsymbol{\nu}}$are OLS residuals. Using the matrix notation, $d_j = \frac{\mathbf{Y}'\mathbf{M}\mathbf{A}_j'\mathbf{A}_j\mathbf{M}\mathbf{Y}}{\mathbf{Y}'\mathbf{M}\mathbf{Y}}$ where $\mathbf{M} = \mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$and $\mathbf{A}_j$ was then a $(N-j) \times N$matrix: $\mathbf{A}_j = \begin{bmatrix} -1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & -1 & 0 & \cdots & 0 & 1 \end{bmatrix}$ and there were $j-1$zeros between $-1$and 1 in each row of matrix$\mathbf{A}_j$. The QR factorization of the design matrix $\mathbf{X}$ yielded an $N \times N$orthogonal matrix$\mathbf{Q}$: $\mathbf{X} = \mathbf{Q}\mathbf{R}$where R was an $N \times k$upper triangular matrix. There then existed a $N \times (N-k)$submatrix of $\mathbf{Q}$ such that $\mathbf{Q}_1\mathbf{Q}_1' = \mathbf{M}and\mathbf{Q}_1'\mathbf{Q}_1 = \mathbf{I}_{N-k}$. Consequently, the generalized Durbin-Watson statistic for constructing a robust seasonal district-level malaria-related risk model would be stated as a ratio of two quadratic forms: $d_j = \frac{\sum_{l=1}^{n} \lambda_{jl} \xi_l^2}{\sum_{l=1}^{n} \xi_l^2}$ where $\lambda_{j1} \ldots \lambda_{jn}$ was upper n eigenvalues of $\mathbf{M}\mathbf{A}'_j\mathbf{A}_j\mathbf{M}$and ξ_l is a standard normal variate, and$n = \min(N-k, N-j)$. These eigenvalues may be obtained by a singular value decomposition of $\mathbf{Q}'_1\mathbf{A}'_j$(see Golub and Van Loan; 1989; Savin and White; 1978). By so doing, the marginal probability (or p-value) for $d_j$given c_0 in the risk model then would be $\text{Prob}\left(\frac{\sum_{l=1}^{n} \lambda_{jl}\xi_l^2}{\sum_{l=1}^{n} \xi_l^2} < c_0\right) = \text{Prob}(q_j < 0)$ where $q_j = \sum_{l=1}^{n} (\lambda_{jl} - c_0)\xi_l^2$ Further, when the null hypothesis $H_0 : \varphi_j = 0$holds, the quadratic form q_j in the risk model residual forecasts targeting the stasictically significant field/clinical/remote explanatory hyperendemic transmission oriented covariate coefficients would have the characteristic function $\phi_j(t) = \prod_{l=1}^{n} (1 - 2(\lambda_{jl} - c_0)it)^{-1/2}$ The distribution function would then be uniquely determined by the characteristic function $F(x) = \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{e^{itx}\phi_j(-t) - e^{-itx}\phi_j(t)}{it} dt$.

As a rough rule of thumb, if Durbin–Watson is less than 1.0 in a geopredictive district-level malarial related hyperendemic risk model, there may be cause for alarm. Small values of *d* indicate successive error terms are, on average, close in value to one another, or positively correlated. If *d* > 2, successive error terms are, on average, much different in value from one another, (i.e., negatively correlated) (Cressie 1993). In malarial-related regressions, this can imply an underestimation of the level of statistical significance in R

Unfortunately Durbin Watson statistic can be biased towards 2, thus falsely showing that there is no autocorrelation when lagged values of the dependent variable are used as independent variables which is common in district-level malaria-related autoregressive geopredictive risk modeling explanatory seasonal sampled field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients. As such, sampled explanatory hyperendemic transmission oriented covariate coefficient estimates on lagged independent variables may merely reflect the presence of an omitted variable or measurement error bias. Thus, adding up the contemporaneous and

lagged coefficients in a predictive district-level seasonal malaria-related model can actually increase, rather than reduce, bias in the residual forecasts. Also, specification tests based on the data at hand (as opposed to external data, such as external instruments), will generally not allow a malarialogist/experimenter to distinguish between true lagged effects, measurement error and/or omitted variable bias in the risk model. Although lagged effects cannot reflect omitted variable bias or bias due to a mismeasured independent variable in a malaria-related geopredictive risk model the differential effect of short and long-term changes in conditions will fall into the "indeterminate" range(i.e., it would render an ambiguous result). The statistic tests only for correlation between the current error and the immediately preceding error (i.e. first order autocorrelation).

Conversely, in R a vector autoregression (VAR) time series geopredictive district level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive malarial risk model may be constructed. A VAR is a statistical model which may be used to capture the linear interdependencies among multiple time series in a geopredictive district-level malarial risk model. VAR models would generalize the univariate AR by allowing for more than one evolving explanatory hyperendemic transmission oriented predictive variable in the model. All the variables in a VAR are treated symmetrically in a structural sense (although the estimated quantitative response coefficients will not in general be the same); each variable has an equation explaining its evolution based on its own lags and the lags of the other model variables (Griffith 2003). VAR modeling would not require as much knowledge about the forces influencing a hyperendemic transmission oriented geopredictive variable as do time series malarial-related structural models with simultaneous equations. The only prior knowledge required would be a list of the variables which can be hypothesized to affect each other intertemporally in the predictive district-level risk model. VARs are most successful, flexible, and easy to use models for the analysis of multivariate time series. It is a natural extension of the univariate autoregressive model to dynamic multivariate time series (Cressie 1993).

VAR –related geopredictive R derived risk models could describe the evolution of an empirical dataset of k explanatory hyperendemic transmission oriented variables (i.e., seasonal malarial-related endogenous variables) over the same sample period (t = 1, T) as a linear function of their past values. The variables would be collected in a k × 1 vector i_t, which would have the I ^{the} element, $y_{en,\ t}$ and the time (t) hyperendemic transmission oriented predictive variable observation of the i th variable. For example, if the *i* Th seasonal sampled hyperendemic transmission oriented variable is represented by y, then $y_{i,\ t}$ is the value of the sampled explanatory covariate coefficient at time t. Thereafter, a p-th order VAR can be denoted in R as VAR(p), is $y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t$, where the l-periods back observation y_{t-l} would be the l-th lag of y when c is a k × 1 vector of constants (i.e., intercepts), A_i is a time-invariant k × k matrix and e_t is a k × 1 vector of error terms satisfying the following $\mathrm{E}(e_t) = 0$— where every error term has mean zero. As such, $\mathrm{E}(e_t e_t') = \Omega$ would be the contemporaneous covariance matrix of error terms when Ω is a k × k positive-semidefinite matrix; and, the quantitation of the residual forecasts would render $\mathrm{E}(e_t e_{t-k}') = 0$ for any non-zero k (i.e., there is no correlation across time; in particular, no serial correlation in individual error terms). A pth-order VAR is also called a VAR with p lags (Cressie 1993). The process of choosing the maximum lag p in the VAR predictive seasonal district-level malarial risk model would then render inferences dependent on the correctness of the selected lag order.

Thereafter, a general example of a VAR(p) in R with k hyperendemic transmission oriented predictor variables can be employed to construct a seasonal district level malarial-related empirical datasets of field/clinical/remote autoregressive explanatory covariate coefficients using a General matrix notation of a VAR(p).This would require using the equation $y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t$, where each y_i is a k × 1 vector and each A_i is a k × k matrix. A dataset of VAR (1) predictive district-level malarial related risk model variables can then be written in matrix form employing $\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$, in which only a single A matrix appears. This is because in a VAR(1) matric a maximum lag p equal to 1. Equivalently, a malarial-related risk model may be

constructed in R using the following system of two equations $y_{1,t} = c_1 + A_{1,1}y_{1,t-1} + A_{1,2}y_{2,t-1} + e_{1,t}$ and $y_{2,t} = c_2 + A_{2,1}y_{1,t-1} + A_{2,2}y_{2,t-1} + e_{2,t}$.A matrix notation can then be constructed as :

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} + \begin{bmatrix} a^1_{1,1} & a^1_{1,2} & \cdots & a^1_{1,k} \\ a^1_{2,1} & a^1_{2,2} & \cdots & a^1_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a^1_{k,1} & a^1_{k,2} & \cdots & a^1_{k,k} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{k,t-1} \end{bmatrix} + \cdots + \begin{bmatrix} a^p_{1,1} & a^p_{1,2} & \cdots & a^p_{1,k} \\ a^p_{2,1} & a^p_{2,2} & \cdots & a^p_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a^p_{k,1} & a^p_{k,2} & \cdots & a^p_{k,k} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{k,t-p} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \\ \vdots \\ e_{k,t} \end{bmatrix}$$

Rewriting the y variables in the seasonal malarial predictive risk model would then render:

$$y_{1,t} = c_1 + a^1_{1,1}y_{1,t-1} + a^1_{1,2}y_{2,t-1} + \cdots + a^1_{1,k}y_{k,t-1} + \cdots + a^p_{1,1}y_{1,t-p} + a^p_{1,2}y_{2,t-p} + \cdots + a^p_{1,k}y_{k,t-p} + e_{1,t}$$

,

$$y_{2,t} = c_2 + a^1_{2,1}y_{1,t-1} + a^1_{2,2}y_{2,t-1} + \cdots + a^1_{2,k}y_{k,t-1} + \cdots + a^p_{2,1}y_{1,t-p} + a^p_{2,2}y_{2,t-p} + \cdots + a^p_{2,k}y_{k,t-p} + e_{2,t}$$

and

$$y_{k,t} = c_k + a^1_{k,1}y_{1,t-1} + a^1_{k,2}y_{2,t-1} + \cdots + a^1_{k,k}y_{k,t-1} + \cdots + a^p_{k,1}y_{1,t-p} + a^p_{k,2}y_{2,t-p} + \cdots + a^p_{k,k}y_{k,t-p} + e_{k,t}$$

By so doing, A malariologist/or experimenter could, if so desired, rewrite a VAR(p) with k variables in a general way which could then include T+1 seasonal sampled hyperendemic transmission oriented observations $y_0$through $y_T Y = BZ + U$ where:

$$Y = \begin{bmatrix} y_p & y_{p+1} & \cdots & y_T \end{bmatrix} = \begin{bmatrix} y_{1,p} & y_{1,p+1} & \cdots & y_{1,T} \\ y_{2,p} & y_{2,p+1} & \cdots & y_{2,T} \\ \vdots & \vdots & \vdots & \vdots \\ y_{k,p} & y_{k,p+1} & \cdots & y_{k,T} \end{bmatrix}$$

$$B = \begin{bmatrix} c & A_1 & A_2 & \cdots & A_p \end{bmatrix} = \begin{bmatrix} c_1 & a^1_{1,1} & a^1_{1,2} & \cdots & a^1_{1,k} & \cdots & a^p_{1,1} & a^p_{1,2} & \cdots & a^p_{1,k} \\ c_2 & a^1_{2,1} & a^1_{2,2} & \cdots & a^1_{2,k} & \cdots & a^p_{2,1} & a^p_{2,2} & \cdots & a^p_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ c_k & a^1_{k,1} & a^1_{k,2} & \cdots & a^1_{k,k} & \cdots & a^p_{k,1} & a^p_{k,2} & \cdots & a^p_{k,k} \end{bmatrix}$$

$$Z = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_{p-1} & y_p & \cdots & y_{T-1} \\ y_{p-2} & y_{p-1} & \cdots & y_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_0 & y_1 & \cdots & y_{T-p} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_{1,p-1} & y_{1,p} & \cdots & y_{1,T-1} \\ y_{2,p-1} & y_{2,p} & \cdots & y_{2,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k,p-1} & y_{k,p} & \cdots & y_{k,T-1} \\ y_{1,p-2} & y_{1,p-1} & \cdots & y_{1,T-2} \\ y_{2,p-2} & y_{2,p-1} & \cdots & y_{2,T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k,p-2} & y_{k,p-1} & \cdots & y_{k,T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1,0} & y_{1,1} & \cdots & y_{1,T-p} \\ y_{2,0} & y_{2,1} & \cdots & y_{2,T-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k,0} & y_{k,1} & \cdots & y_{k,T-p} \end{bmatrix}$$

and

$$U = \begin{bmatrix} e_p & e_{p+1} & \cdots & e_T \end{bmatrix} = \begin{bmatrix} e_{1,p} & e_{1,p+1} & \cdots & e_{1,T} \\ e_{2,p} & e_{2,p+1} & \cdots & e_{2,T} \\ \vdots & \vdots & \ddots & \vdots \\ e_{k,p} & e_{k,p+1} & \cdots & e_{k,T} \end{bmatrix}.$$

These matrices would then solve for the coefficient matrix B using an OLS estimation of $Y \approx BZ$. Each hyperendemic transmission oriented geopredictive variable in the model would then have one equation. The current (time t) district-level field/clinical/remote sampled malaria-related predictive autoregressive hyperendemic transmission oriented observation of each of the district level seasonal sampled variable would then simply depend on its own lagged values as well as on the lagged values of each other variable in the VAR.

A VAR with p lags in a predictive seasonal malarial-related risk model can also be rewritten in R as a VAR with only one lag by appropriately redefining the dependent variable (e.g., district level malarial prevalence rates). The transformation amounts to stacking the lags of the VAR(p) variable in the new VAR(1) dependent variable in R and appending identities to complete the number of equations. Thereafter, a VAR (2) derived predictive district level seasonal malarial-related risk model could be generated from l where $y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + e_t$ which then could be recast as the VAR (1) model suing

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} e_t \\ 0 \end{bmatrix},$$

where I is the identity matrix.

In the class of multivariate linear models, pure VARs dominate in macroeconomic applications. However, VAR models may require a rather large lag length in order to accurately describe a time series of district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive explanatory covariate coefficients adequately. This means a loss of precision because many parameter estimators in the risk model have to be estimated. The problem could be avoided by using VARMA models that may provide a

more parsimonious description of the data generating process (DGP). In contrast to the class of VARMA models, the class of VAR models is not closed under linear transformations. For example, a subset of variables generated by a VAR process is typically generated by a VARMA, not by VAR process (Lutkepohl 1984*a*,*b*). The VARMA class includes many models of interest such as unobserved component models. It is well known that linearized dynamic stochastic general equilibrium (DSGE) models imply that the variables of interest are generated by a finite-order VARMA process. Ferndandez-Villaverde et al.(2007) show formally how DSGE models and VARMA processes are linked. Also Cooley and Dwyer (1998) claim that modeling macroeconomic time series systematically as pure VARs is not justified by any underlying economic theory. The recent debate between Chari, Kehoe and McGrattan (2008) and Christiano, Eichenbaum and Vigfusson (2006) on the ability of structural VARs to uncover fundamental shocks also questions implicitly the ability of pure VARs to capture the dynamics of any malarial-related district-level predictive epidemiological study site.

Further, there are also some complications that make VARMA modeling more difficult for district level geopredictive hyperendemic transmission oriented malarial risk modeling First, VARMA representations are not unique. That is, there are typically many parameterizations that can describe the same DGP (see Lutkepohl 2005). Therefore, a malarialogist/experimenter has to choose first an identified representation. In any case, an identified VARMA representation has to be specified by more integer-valued parameter estimators than a VAR representation that is determined just by one integer estimator, the lag length. This aspect introduces additional uncertainty at the specification stage of the modeling process, although procedures for VARMA models do exist which could be used in a completely automatic way. An identified representation, however, may be then for consistent estimation in a predictive malarial risk model. Apart from a more involved specification stage, the estimation stage may be affected by an identity matrix problem because it would have to examine many different models in order to properly quantitate district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive explanatory covariate coefficients.

Fortunately, where $n \times n$ matrices in a robust geopredictive malarial risk model are used to represent linear transformations from an n-dimensional vector space to itself, I_n would represent the identity function, regardless of the basis. The ith column of an identity matrix is the unit vector e_i. (Cressie 1993). It follows that the determinant of the identity matrix is 1 and the trace is n. Using the notation that is sometimes used to concisely describe diagonal matrices, a malarialogist/experimenter could then write: $I_n = \operatorname{diag}(1, 1, \ldots, 1)$. It can also be written using the Kronecker delta notation: $(I_n)_{ij} = \delta_{ij}$. Identity matrices are sometimes also known as unit matrices (Akivis and Goldberg 1972). The $n \times n$ identity matrix would then be given explicitly by $I_{ij} = \delta_{ij}$ for $i, j = 1, 2, \ldots, n$, where δ_{ij} is the Kronecker delta which in turn could be written explicitly as

$$\mathrm{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

In mathematics, the Kronecker delta is a function of two variables, usually integers. The function is 1 if the empirical dataset of geopredictive malarial –related district level risk model hyerendemic transmission oriented variables, for example, are equal, and 0 otherwise:

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

where Kronecker delta δ_{ij} is a piecewise function of variables i and j. For example, $\delta_{1\,2} = 0$, whereas $\delta_{3\,3} = 1$.For a robust seasonal predictive district level seasonal malarial related risk model, the identity matrix can then be written as $(\delta_{ij})_{i,j=1}^{n}$ and the inner product of vectors can be written as $\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{ij} a_i \delta_{ij} b_j$. The identity matrix for the risk model would then have the property that, when it is the product of two square matrices, the matrices would be the inverse of one another. The identity matrix of a given size is the only idempotent matrix of that size having full rank (Cressie 1993). In algebra, an idempotent matrix is a matrix which, when multiplied by itself, yields itself

(http://mathworld.wolfram.com/IdentityMatrix.html). As such, the matrix *M* in a robust geopredictive seasonal district-level malarial model is idempotent if, and only if ,$MM = M$. For this product *MM* to be defined in the risk however, *M* must be a square matrix. Viewed this way, idempotent matrices for geopredictive seasonal district-level malarial risk modeling are idempotent elements of matrix rings. With the exception of the identity matrix, an idempotent matrix is singular; that is, its number of independent rows (and columns) is less than its number of rows (and columns). This can be seen from writing $MM = M$ for the predictive seasonal district-level malarial model, assuming that *M* has full rank (i.e., non-singular), by pre-multiplying by M^{-1} to obtain $M = M^{-1}M = I$. When an idempotent matrix is subtracted from the identity matrix, the result is also idempotent. This holds since $[I - M]$ $[I - M] = I - M - M + M^2 = I - M - M + M = I - M$. An idempotent matrix is always diagonalizable and its eigenvalues are either 0 or 1(Cressie 1993). The trace of an idempotent matrix — the sum of the elements on its main diagonal — equals the rank of the matrix and thus is always an integer (Meyer 2000). This provides an easy way of computing the rank, or alternatively an easy way of determining the trace of a matrix whose elements are not specifically know. This feature may be useful in a robust predictive seasonal district-level malarial risk model for example, in establishing the degree of bias in using a sample variance as an estimate of a population variance.

Idempotent matrices arise frequently in regression analysis. For example, in ordinary least squares, the regression problem is to choose a vector β of coefficient estimates so as to minimize the sum of squared residuals (i.e., mispredictions) using e_i: in matrix form and minimizing $(y - X\beta)^T(y - X\beta)$ where *y* is a vector of dependent variable hyperendemic transmission oriented observations, when *X* is a matrix each of whose columns is a column of observations on one of the independent variables. The resulting estimator then would be $\beta = (X^T X)^{-1} X^T y$ where superscript *T* indicates a transpose, and the vector of residuals is $e = y - X\beta = y - X(X^T X)^{-1} X^T y = [I - X(X^T X)^{-1} X^T]y = My.$ By so doing, both *M* and $X(X^T X)^{-1} X^T$ (the latter being known as the hat matrix) would be idempotent matrices in the geopredictive seasonal district-level malarial risk model, a fact which would then allow simplification when the sum of squared residuals is computed using $e^T e = (My)^T(My) = y^T M^T My = y^T MMy = y^T My.$ The idempotency of *M* can also then plays a role in other calculations as well, such as in determining the variance of the estimator β in the malarial risk model.

Further, since an idempotent linear operator *P* is a projection operator on the range space *R(P)* along its null space *N(P)*(Griffith 2003) *P* would an orthogonal projection operator if, and only if ,it is idempotent and symmetrical in the geopredictive seasonal district-level malarial risk model. In linear algebra and functional analysis, a projection is a linear transformation *P* from a vector space to itself such that $P^2 = P$ (Cressie 1993) That is, whenever *P* is applied twice to any seasonal sampled geopredictive seasonal district-level malarial risk model hyperendemic transmission oriented covariate coefficient measurement values, it would render the same result as if it were applied once (i.e., idempotence). Though abstract, this definition of "projection would formalize and generalizes the idea of graphical projection in a robust geopredictive seasonal district-level malarial risk model. A malarialogist/experimenter can also consider the effect of a projection on a district-level related geometrical object (e.g., malarial mosquito larval habitat) by examining the effect of the projection on the georeferenced points in the object. For example, a function may be employed to map the district level seasonal sampled habitat point (x, y, z) in three-dimensional space R^3 to the point $(x, y, 0)$ is a projection onto the *x*–*y* plane. This function may be represented by the matrix $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. the action of this matrix on an arbitrary vector then would be

$$P\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

(see Meyer 2000). To see that P is indeed a projection, i.e., $P = P^2$, the malariologist/experimenter could then compute: $$P^2\begin{pmatrix} x \\ y \\ z \end{pmatrix} = P\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = P\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Thereafter, by letting W be a finite dimensional vector space in the geopredictive seasonal district-level malarial risk model the subspaces U and V would be the range and kernel of P respectively. Then P in the risk model would have the following basic properties: P would be the identity operator I on U: $\forall x \in U : Px = x.$, a direct sum $W = U \oplus V$ could be achieved, every vector x in W may be decomposed uniquely as $x = u + v$ with $u = Px$ and $v = x - Px = (I - P)x$, where u is in U and v is in V and P is idempotent and satisfies $P^2 = P$. The range and kernel of a projection would be complementary in the predictive seasonal district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive malarial risk model residual forecasts as P and Q would be equal to $I - P$. The operator Q is a projection and the range and kernel of P become the kernel and range of Q and vice-versa (Cressie 1993). Thus, P would be a projection along V onto U (i.e., kernel/range) and Q would be a projection along U onto V in the predictive seasonal district-level malaria-related risk model In infinite dimensional vector spaces spectrum of a projection in the risk model then would be contained in {0, 1}, as $(\lambda I - P)^{-1} = \frac{1}{\lambda} I + \frac{1}{\lambda(\lambda - 1)} P$. Only 0 and 1 can be an eigenvalue of a projection (Griffith 2003). The corresponding eigenspaces in the predictive seasonal district-level malarial risk model would then respectively be the kernel and range of the projection.

Decomposition of a vector space geopredictive seasonal district-level malarial risk model into direct sums is not unique in general. Therefore, given a subspace V, in general there are many projections whose range (or kernel) is V.If a projection is nontrivial in the risk model it would then have a minimal polynomial [e.g. $X^2 - X = X(X - I)$], which would factor into the roots, and thus P would be diagonalizable in the residual forecasts targeting the statistically significant explanatory field/clinical/ geopredictive autoregressive hyperendemic transmission oriented covariate coefficients. When the vector space W has an inner product (e.g., Hilbert space) the concept of orthogonality can be used in the district-level geopredictive malarial risk model

A Hilbert space is a vector space $H$with an inner product $\langle f, g\rangle$such that the norm defined by $|f| = \sqrt{\langle f, f\rangle}$ turns Hinto a complete metric space(Griffith 2003). If the metric defined by the norm is not complete, then $H$is instead known as an inner product space. Examples of finite-dimensional Hilbert spaces include :. The real numbers $\mathbb{R}^n$with $\langle v, u\rangle$the vector dot product of $v$and u the complex numbers $\mathbb{C}^n$with $\langle v, u\rangle$the vector dot product of $v$and the complex conjugate of u. An example of an infinite-dimensional Hilbert space is L^2, whereby the set of all functions $f : \mathbb{R} \to \mathbb{R}$ issuch that the integral of f^2over the whole real line is finite. In this case, the inner product is $\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\, g(x)\, dx.$

Importantly, an orthogonal projection is a projection for which the range U and the null space V are orthogonal subspaces (Cressie 1993). A projection is orthogonal in a robust geopredictive seasonal district-level malarial risk model if ,and only if, it is self-adjoint(Jacob et al. 2009d). Using the self-adjoint and idempotent properties of P, for any x and y in W in the risk model then would reflect $Px \in U$, $y - Py \in V$, and$\langle Px, y - Py\rangle = \langle P^2 x, y - Py\rangle = \langle Px, P(I - P)y\rangle = \langle Px, (P - P^2)y\rangle = 0$ where

$\langle \cdot, \cdot \rangle$is the inner product associated with *W*. Therefore, *Px* and *y* − *Py* would be orthogonal in the residual forecasts statistically targeting the important field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented predictive autoregressive covariate coefficients. Further, for quantitating finite dimensional complex or real vector spaces in a predictive seasonal district-level malarial risk model, the standard inner product can be substituted for $\langle \cdot, \cdot \rangle$.A simple case occurs when the orthogonal projection is onto a line (Meyers 2000). If *u* is a unit vector on the line in the risk model , then the projection could be given by$P_u = uu^{\mathrm{T}}$ .This operator would leave *u* invariant in the model and it would annihilate all vectors orthogonal to *u*, proving that it is indeed the orthogonal projection onto the line containing *u* in the residual forecasts. A simple way to see this is to consider an arbitrary vector $x$as the sum of a component on the line (i.e. the projected vector) and another perpendicular to it, $x = x_{\parallel} + x_{\perp}$. Applying projection to the predictive seasonal district-level malarial risk model in R would then render $P_u x = uu^{\mathrm{T}} x_{\parallel} + uu^{\mathrm{T}} x_{\perp} = u|x_{\parallel}| + u0 = x_{\parallel}$by the properties of the dot product of parallel and perpendicular vectors.

In R the dot product, or scalar product (or sometimes inner product in the context of Euclidean space), is an algebraic operation that takes two equal-length sequences of numbers (usually coordinate vectors) and returns a single number. This operation can be defined either algebraically or geometrically. Algebraically, it is the sum of the products of the corresponding entries of the two sequences of numbers. Geometrically, it is the product of the magnitudes of the two vectors and the cosine of the angle between them. The name "dot product" is derived from the centered dot " · " that is often used to designate this operation; the alternative name "scalar product" emphasizes the scalar (rather than vectorial) nature of the result. In three-dimensional space, the dot product contrasts with the cross product of two vectors, which produces a pseudovector as the result.

The dot product in a robust district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive malarial risk model thus would be directly related to the cosine of the angle between two vectors in Euclidean space of any number of dimensions. The dot product of two vectors a = [a_1, a_2, ..., a_n] and b = [b_1, b_2, ..., b_n] would then be defined as:

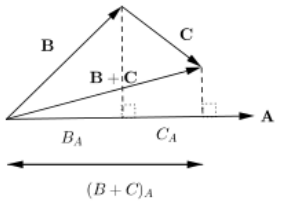
$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

where Σ denotes summation notation and *n* is the dimension of the vector space. For instance, in three-dimensional space, the dot product of vectors [1, 3, −5] and [4, −2, −1] in a robust predictive district level time series model would be

$$[1, 3, -5] \cdot [4, -2, -1] = (1)(4) + (3)(-2) + (-5)(-1) = 4 - 6 + 5 = 3.$$

In Euclidean space, a Euclidean vector is a geometrical object that possesses both a magnitude and a direction. A vector can be pictured as an arrow. Its magnitude is its length, and its direction is the direction the arrow points. As such, the magnitude of a vector A can be denoted by$\|\mathbf{A}\|$. The dot product of two Euclidean vectors A and B would then be b defined by$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \, \|\mathbf{B}\| \cos\theta$,where θ is the angle between A and B. In particular, if A and B are orthogonal in the risk model, then the angle between them is 90° and$\mathbf{A} \cdot \mathbf{B} = 0$.At the other extreme, if they are codirectional, then the angle between them is 0° and$\mathbf{A} \cdot \mathbf{B} = \|\mathbf{A}\| \, \|\mathbf{B}\|$This implies that the dot product of a vector A by itself is$\mathbf{A} \cdot \mathbf{A} = \|\mathbf{A}\|^2$,which gives$\|\mathbf{A}\| = \sqrt{\mathbf{A} \cdot \mathbf{A}}$,the formula for the Euclidean length of the vector.

The scalar projection or scalar component of a Euclidean vector in a geopredictive district-level malarial risk model could then be quantitated as A in the direction of a Euclidean vector B is given by$A_B = \|\mathbf{A}\| \cos\theta$where θ is the angle between A and B between the estimators in geospace. In terms of the geometric $A_B = \mathbf{A} \cdot \widehat{\mathbf{B}}$definition of the dot product, can then be rewritten where$\widehat{\mathbf{B}} = \mathbf{B}/\|\mathbf{B}\|$is the unit



vector in the direction of B. The dot product would thus would be characterized geometrically by $\mathbf{A} \cdot \mathbf{B} = A_B \|\mathbf{B}\| = B_A \|\mathbf{A}\|$. The dot product, defined in this manner would be homogeneous under scaling therefatre for each sampled explanatory hyperendemic transmission oriented district-level field/clinical/remote sampled malaria-related predictive autoregressive predictor variable, meaning that for any scalar α, $(\alpha\mathbf{A}) \cdot \mathbf{B} = \alpha(\mathbf{A} \cdot \mathbf{B}) = \mathbf{A} \cdot (\alpha\mathbf{B})$. The dot product would also then satisfy the distributive law, meaning that $\mathbf{A} \cdot (\mathbf{B} + \mathbf{C}) = \mathbf{A} \cdot \mathbf{B} + \mathbf{A} \cdot \mathbf{C}$. As a consequence, if $\mathbf{e}_1, \ldots, \mathbf{e}_n$ are the standard basis vectors in $\mathbb{R}^n$, in a geopredictive district-level malarial risk model output then

$$\mathbf{A} = [A_1, \ldots, A_n] = \sum_i A_i \mathbf{e}_i \quad \mathbf{B} = [B_1, \ldots, B_n] = \sum_i B_i \mathbf{e}_i$$

writing and would render

$$\mathbf{A} \cdot \mathbf{B} = \sum_i B_i (\mathbf{A} \cdot \mathbf{e}_i) = \sum_i B_i A_i$$

which is precisely the algebraic definition of the dot product. More generally, the same identity would holds when e_i is replaced by any orthonormal basis in the risk model residual forecasts.

The dot product would thus fulfill the following properties if a, b, and c are real vectors in the predictive district level malarial risk model and *r* is a scalar. The commutative (i.e., $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$.)would follow from the definition $\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos\theta = \|\mathbf{b}\| \|\mathbf{a}\| \cos\theta = \mathbf{b} \cdot \mathbf{a}$ while the distributive features would be illustrated over vector addition: $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$. Further, the Bilinear and , Scalar multiplication[i.e., $(c_1\mathbf{a}) \cdot (c_2\mathbf{b}) = c_1 c_2 (\mathbf{a} \cdot \mathbf{b})$] and Orthogonal data attributes could be quantitated. Two non-zero vectors a and b are orthogonal if and only if a · b = 0 (Griffith 2003). Unlike multiplication of ordinary numbers, where if *ab* = *ac*, then *b* always equals *c* unless *a* is zero, the dot product does not obey the cancellation law: If a · b = a · c and a ≠ 0, then a malarialogist/experimenter could write: a · (b − c) = 0 by the distributive law whereby the result would signify whether a is perpendicular to (b − c), which would allow (b − c) ≠ 0, and therefore b ≠ c. If a and b are functions, then the derivative (denoted by a prime ′) of a · b is a′ · b + a · b′ (Cressie 1993) Triangle with vector edges a and b, separated by angle *θ*. Thus, given two vectors a and b separated by angle *θ* (see image right), they form a triangle with a third side c = a − b. The dot product of this with itself is:

$$\begin{aligned}
\mathbf{c} \cdot \mathbf{c} &= (\mathbf{a} - \mathbf{b}) \cdot (\mathbf{a} - \mathbf{b}) \\
&= \mathbf{a} \cdot \mathbf{a} - \mathbf{a} \cdot \mathbf{b} - \mathbf{b} \cdot \mathbf{a} + \mathbf{b} \cdot \mathbf{b} \\
&= a^2 - \mathbf{a} \cdot \mathbf{b} - \mathbf{a} \cdot \mathbf{b} + b^2 \\
&= a^2 - 2\mathbf{a} \cdot \mathbf{b} + b^2 \\
c^2 &= a^2 + b^2 - 2ab\cos\theta
\end{aligned}$$

This formula can be generalized to orthogonal projections on a subspace of arbitrary dimension in a geopredictive district-level malarial related risk model. Let u_1, ..., u_k be an orthonormal basis of the subspace *U*, and let *A* denote the *n*-by-*k* matrix whose columns are u_1, ..., u_k.then the projection is given by $P_A = AA^{\mathrm{T}}$ (Griffith 2003) which

$$P_A = \sum_i \langle u_i, \cdot \rangle u_i.$$

for district level malarial predictive modeling can be rewritten as The matrix A^{T} is the partial isometry that vanishes on the orthogonal complement of U and A is the isometry that embeds U into the underlying vector space (Cressie 1993). The range of P_A would therefore be the final space of A in a robust predictive malarial model It is also clear that $A^{\mathrm{T}}A$ would then be the identity operator on U in the residual forecasts targeting the statistically important explanatory hyperendemic transmission oriented covariate coefficients The orthonormality condition can also be dropped in the forecasts. If u_1, ..., u_k is a (not necessarily orthonormal) basis, and A is the matrix with these vectors as columns, then the projection is $P_A = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. (Griffith 2003)

Interestingly, the matrix A would still embed U into the underlying vector space in the geopredictive seasonal malarial-related risk model but it would no longer be isometric. The matrix $(A^{\mathrm{T}}A)^{-1}$ is a "normalizing factor" that recovers the norm. For example, the rank-1 operator uu^{T} would not be a projection if $\|u\| \neq 1$ in a robust geopredictive malarial risk model residual forecast. After dividing by $u^{\mathrm{T}}u = \|u\|^2$, a malarialogist/experimenter would then obtain the projection $u(u^{\mathrm{T}}u)^{-1}u^{\mathrm{T}}$ onto the subspace spanned by u. When the range space of the projection is generated by a frame (i.e. the number of generators is greater than its dimension), the formula for the projection would take the form $P_A = A(A^{\mathrm{T}}A)^{+}A^{\mathrm{T}}$. Here A^{+} stands for the Moore–Penrose pseudoinverse. This is just one of many ways to construct the projection operator for a geopredictive district level malaria-related risk model. Further, if a matrix $[A\ B]$ is non-singular and $A^{\mathrm{T}}B = 0$ (i.e., B is the null space matrix of A) in the residual forecast then the following holds: $I = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} + B(B^{\mathrm{T}}B)^{-1}B^{\mathrm{T}}$. On the other hand, if the orthogonal condition is enhanced to $A^{\mathrm{T}}WB = A^{\mathrm{T}}W^{\mathrm{T}}B = 0$ with W being non-singular in the residual forecasts, then the following holds:

$$I = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} (A^{\mathrm{T}}WA)^{-1}A^{\mathrm{T}} \\ (B^{\mathrm{T}}WB)^{-1}B^{\mathrm{T}} \end{bmatrix} W.$$

All these formulas would hold for complex inner product spaces in the geopredictive malarial-related model provided that the conjugate transpose is used instead of the transpose.

The term oblique projections is sometimes used to refer to non-orthogonal projections. These projections are also used to represent spatial figures in two-dimensional drawings (see Griffith 2003), though not as frequently as orthogonal projections. Oblique projections are defined by their range and null space. A formula for the matrix representing the projection with a given range and null space can be found as follows. Thus if a malarialogist/experimenter lets the vectors u_1, ..., u_k form a basis for the range of the projection in the risk model and assemble these vectors in the n-by-k matrix A the range and the null space would be complementary spaces, so the null space has a dimension $n - k$. It follows then that the orthogonal complement of the null space in the model would then have dimension k. There are by letting v_1, ..., v_k form a basis for the orthogonal complement of the null space in the residual forecast of the projection, and assembling these vectors in the matrix B. the projection could be defined by $P = A(B^{\mathrm{T}}A)^{-1}B^{\mathrm{T}}$. This expression would generalize the formula for orthogonal projections then for a robust geopredictive malaria-related risk model.

It is important to remember that any projection $P = P^2$ on a vector space of dimension d over a field is a diagonalizable matrix, since its minimal polynomial is $x^2 - x$ normally splits into distinct linear factors. Thus, there exists a basis in which P has the form $P = I_r \oplus 0_{d-r}$ where r is the rank of P. Here I_r is the identity matrix of size r, and 0_{d-r} is the zero matrix of size $d - r$. If the vector space is complex and equipped with an inner product, then there is an orthonormal basis in which the matrix of P is

$$P = \begin{bmatrix} 1 & \sigma_1 \\ 0 & 0 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 1 & \sigma_k \\ 0 & 0 \end{bmatrix} \oplus I_m \oplus 0_s$$

where $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k > 0$. The integers k, s, m and the district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive explanatory covariate coefficient measurement values σ_i then would be uniquely determined. Note that since $2k + s + m = d$ (Cressie 1993) the factor $I_m \oplus 0_s$ would correspond to the maximal invariant subspace in

the risk model on which P acts as an orthogonal projection (so that P itself is orthogonal if and only if $k = 0$) and the σ_i-blocks correspond to the oblique components.

Interestingly, when the underlying vector space X is a (not necessarily finite-dimensional) normed vector space in R analytic questions, irrelevant in the finite-dimensional case, need to be considered. Assume now X is a Banach space. A given direct sum decomposition of X into complementary subspaces in a geopredictive malaria-related model would then specify a projection, and vice versa. If X is the direct sum $X = U \oplus V$, then the operator defined by $P(u + v) = u$ is still a projection with range U and kernel V. It is also clear that $P^2 = P$. Conversely, if P is projection on X, i.e. $P^2 = P$ in the risk model then it would be easily verified that $(I - P)^2 = (I - P)$. In other words, $(I - P)$ would also be a projection in a robust predictive malaria-related risk model. The relation $I = P + (I - P)$ implies X is the direct sum Ran $(P) \oplus$ Ran $(I - P)$ (Griffith 2003).

However, in contrast to the finite-dimensional case, projections need not be continuous in general in a robust geopredictive district level malarial risk model. If a subspace U of X is not closed in the norm topology, then projection onto U would not be continuous in the derivatives rendered from the model output. In other words, the range of a continuous projection P must be a closed subspace in order to attain statistical significance of the sampled explanatory hyperendemic transmission oriented covariate coefficients. Further, the kernel of a continuous projection would be closed. Thus a continuous projection in a robust predictive malarial district-level model P would render a decomposition of X into two complementary closed subspaces: $X = \text{Ran}(P) \oplus \text{Ker}(P) = \text{Ran}(P) \oplus \text{Ran}(I - P)$ in the resdiual forecasts.

The converse will also hold in a geopredictive malaria-related district level risk model if an additional assumption is employed. For example, suppose U is a closed subspace of X in the district level risk model *If* there exists a closed subspace V such that $X = U \oplus V$, then the projection P with range U and kernel V would then be continuous. This follows from the closed graph theorem. Further, suppose $x_n \to x$ and $Px_n \to y$ in the model residual forecasts. The malarialogist/experimenter then would have to show $Px = y$. Since U is closed and $\{Px_n\} \subset U$, y lies in U, i.e. $Py = y$. then, $x_n - Px_n = (I - P)x_n \to x - y$. Because V would then be closed and $\{(I - P)x_n\} \subset V$, h $x - y \in V$, (i.e. $P(x - y) = Px - Py = Px - y = 0$) would then be rendered by the risk model.

The above argument makes use of the assumption that both U and V would be closed in the geopredictive malaria-related. district-level model. In general, given a closed subspace U, there need not exist a complementary closed subspace V, although for Hilbert spaces this can always be done by taking the orthogonal complement. For Banach spaces, a one-dimensional subspace always has a closed complementary subspace. This is an immediate consequence of Hahn–Banach theorem whereby if a malarialogist/experimenter lets U be the linear span of u then there would exist a bounded linear functional Φ such that $\varphi(u) = 1$. The operator $P(x) = \varphi(x)u$ would then satisfy $P^2 = P$ in the residual forecasts. Boundedness of φ implies continuity of P and therefore $\text{Ker}(P) = \text{Ran}(I - P)$ is a closed complementary subspace of U.(Cressie 1993). However, every continuous projection on a Banach space in the risk model would be an open mapping source. That is, the district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive risk model would have the only matrix such that (a) when multiplied by itself the result is itself, and (b) all of its rows, and all of its columns would be linearly independent. The principal square root of an identity matrix is itself, and this is its only positive definite square root(Griffith 2003).Further, every identity matrix in a malarial predictive risk model with at least two rows and columns has an infinitude of symmetric square r(see Jacob et al. 2009d),

The equivalent VAR(1) geopredictive seasonal malarial-related risk model would then form a more convenient for analytical derivations and allows more compact residual forecasted statements. Meanwhile a A structural VAR with p lags predictive seasonal malarial-related risk model could be described by $B_0 y_t = c_0 + B_1 y_{t-1} + B_2 y_{t-2} + \cdots + B_p y_{t-p} + \epsilon_t$, where c_0 is a k × 1 vector of constants, B_i is a k × k matrix (for every i = 0, ..., p) and ε_t is a k × 1 vector of error terms. The main diagonal terms of the B_0 matrix (i.e., the coefficients on the i^{th} variable in the i^{th} equation) would then be scaled to 1. The error terms ε_t (i.e., structural shocks) would then be satisfied in the model with the particularity that all the elements off the main diagonal of the covariance matrix $\mathrm{E}(\epsilon_t \epsilon_t') = \Sigma$ would be zero. That is, the structural shocks in the predictive

district-level malarial risk model would be uncorrelated. For example, a two sampled district level hyperendemic transmission oriented variable structural VAR(1) could be constructed using:

$$\begin{bmatrix} 1 & B_{0;1,2} \\ B_{0;2,1} & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_{0;1} \\ c_{0;2} \end{bmatrix} + \begin{bmatrix} B_{1;1,1} & B_{1;1,2} \\ B_{1;2,1} & B_{1;2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix},$$ where

$\Sigma = \mathrm{E}(\epsilon_t \epsilon_t') = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$; that is, the variances of the structural shocks denoted are $\operatorname{var}(\epsilon_i) = \sigma_i^2$ (i = 1, 2) and the covariance is $\operatorname{cov}(\epsilon_1, \epsilon_2) = 0$. Writing the first equation explicitly and passing $y_{2,t}$ to the right hand side a malarialogist/experimenter would obtain $y_{1,t} = c_{0;1} - B_{0;1,2} y_{2,t} + B_{1;1,1} y_{1,t-1} + B_{1;1,2} y_{2,t-1} + \epsilon_{1,t}$ from the regressed emperical sampled datset.

Note that $y_{2,t}$ can have a contemporaneous effect on $y_{1,t}$ if $B_{0;1,2}$ is not zero. This is different from the case when B_0 is the identity matrix (all off-diagonal elements are zero — the case in the initial definition), when $y_{2,t}$ can impact directly $y_{1,t+1}$ and subsequent future values, but not $y_{1,t}$. Because of the parameter identification problem, OLS estimation of the structural VAR in a predictive malaria-related district model would yield inconsistent geoparameter estimates. This problem can be overcome by rewriting the VAR in reduced form. As such, if the joint dynamics of a set of hyperendemic transmission oriented variables are represented by a VAR model, then the structural form would be a depiction of the underlying, "structural", sampled estimators relationships. A key feature of the structural form which may make it the preferred candidate to represent the underlying relations in a robust predictive malaria-related district level risk model is that the error terms would not be correlated. The structural shocks which drive the dynamics of the variables would then be assumed to be independent, which implies zero correlation between error terms as a desired property. This is helpful for separating out the effects of unrelated influences in the VAR derived malaria-related risk model.

Further by premultiplying the structural VAR in R with the inverse of B_0 $y_t = B_0^{-1} c_0 + B_0^{-1} B_1 y_{t-1} + B_0^{-1} B_2 y_{t-2} + \cdots + B_0^{-1} B_p y_{t-p} + B_0^{-1} \epsilon_t,$ and denoting $B_0^{-1} c_0 = c, \quad B_0^{-1} B_i = A_i$ for $i = 1, \ldots, p$ and $B_0^{-1} \epsilon_t = e_t$ a malarialogist/experimenter would obtain the pth order reduced VAR $y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + e_t$ Note, that in the reduced form all right hand side district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive oriented variables would be predetermined at time t in the model. As there are no endogenous variables on the right hand side, no hyperendemic transmission oriented variable would have a direct contemporaneous effect on other sampled variables in the risk model. However, the error terms in the reduced VAR would be composites of the structural shocks $e_t = B_0^{-1} \varepsilon_t$. Thus, the occurrence of one structural shock $\varepsilon_{i,t}$ can potentially lead to the occurrence of shocks in all error terms $e_{j,t}$, in the risk model thus creating contemporaneous movement in all endogenous variables. Consequently, the covariance matrix of the reduced VAR $\Omega = \mathrm{E}(e_t e_t') = \mathrm{E}(B_0^{-1} \epsilon_t \epsilon_t' (B_0^{-1})') = B_0^{-1} \Sigma (B_0^{-1})'$ can have non-zero off-diagonal elements, thus allowing non-zero correlation between error terms.

The VAR model in R has proven to be especially useful for describing the dynamic behavior of economic and financial time series and for forecasting. It often provides superior forecasts to those from univariate time series models and elaborate theory-based simultaneous equations models. Forecasts from VAR models are quite flexible because they can be made conditional on the potential future paths of specified variables in the model (Cressie 1993).In addition to data description and forecasting, the ARIMA model may also be used for structural inference and policy analysis for laying down the foundation of an integrated vector management program. In structural analysis, certain assumptions about the causal structure of the data under investigation can then be imposed, and the resulting causal impacts of unexpected shocks or innovations to specified variables on the variables in the model can be summarized. These causal impacts are usually summarized with impulse response functions and forecast error variance decompositions.

The structure of the package vars and its implementation of vector autoregressive-structural vector autoregressive- and structural vector error correction models can also be generated in R. In addition to the three cornerstone functions VAR(), SVAR() and SVEC() for estimating such models, functions for diagnostic testing, estimation of a restricted models, prediction, causality analysis, impulse response analysis and forecast error variance decomposition can be provided too. It is further possible to convert vector error correction models in R. into their level VAR representation for the seasonal predictive district -level malarial model residual forecasts

However, R has a steep learning curve |. R is not so easy to use for the novice. There are several simple-to use graphical user interfaces (GUIs) for R that encompass point and-click interactions, but they generally do not have the polish of the commercial offerings. Further, documentation is sometimes patchy and terse, and impenetrable to the non-statisticians. However, some very high-standard books are increasingly plugging the documentation gaps. The quality of some packages is less than perfect, although if a package is useful to many malarialogists/experimenters, it will quickly evolve into a very robust product through collaborative efforts. Further, many R commands give little thought to memory management, and so R can very quickly consume all available memory. This can be a restriction when doing data mining for constructing a robust predictive malarial related risk model.

Conversely, the SAS/GIS procedure PROC REG can be used instead to obtain generalized least squares (GLS) estimates to regress log- transformed seasonal district-level malarial data. Within PROC REG the MODEL statement option DW produces the Durbin-Watson statistic. SAS also contains a powerful procedure, PROC AUTOREG, documented in the SAS/ETS User's Guide for estimating linear regression models with autocorrelation. A PROC IML appendix can then be provided to illustrate any iterative estimation procedures, (e.g., the CochraneOrcutt technique).

Cochrane–Orcutt estimation is a procedure which adjusts a linear model for serial correlation in the error term Consider the geopredictive seasonal district-level malarial risk model $y_t = \alpha + X_t\beta + \varepsilon_t$, where y_t is the value of the dependent variable of interest at time *t*(e.g., total district larval density count), β is a column vector of coefficients to be estimated, X_t is a row vector of explanatory variables at time *t*, and ε_t is the error term at time *t*. If it is found via the Durbin–Watson statistic that the error term is serially correlated over time in the risk model, then standard statistical inference as normally applied to regressions would be invalid because standard errors are estimated with bias (see Homer and Lemeshew 2000). To avoid this problem, the residual forecasts rendered from the geopredictive district-level risk model must be adequately modeled. If the process generating the residual forecasts is found to be a stationary first-order autoregressive structure, $\varepsilon_t = \rho\varepsilon_{t-1} + e_t, \ |\rho| < 1$, with the errors $\{e_t\}$ being white noise, then the Cochrane–Orcutt procedure can be used to transform the predictive seasonal district-level malarial risk model by taking a quasi-difference: $y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + e_t$. In this specification the error terms would be white noise, so statistical inference would be valid. Then the sum of squared residuals (i.e.,the sum of squared estimates of e_t^2) would be minimized in the risk model residual forecasts with respect to (α, β), conditional on ρ. If ρ is not known, then it could be estimated by first regressing the untransformed model and obtaining the residuals $\{\hat{\varepsilon}_t\}$, and regressing $\hat{\varepsilon}_t$ on $\hat{\varepsilon}_{t-1}$, leading to an estimate of ρ for making the transformed regression above feasible. (Note that one data point, the first, is lost in this regression.) This procedure of autoregressing estimated predictive seasonal district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented autoregressive model residuals can then be conducted parsimoniously once the resulting value of ρ is used in the transformed *y* regression, or the residuals of the residual autoregression which can themselves be autoregressed in consecutive steps until no substantial change in the estimated value of ρ is observed.

Alternatively, White (1980) proposed a consistent estimator for the variance-covariance matrix of the asymptotic distribution of the OLS estimator, which would validate the use of hypothesis testing employing OLS estimators under heteroscedasticity in STATA. The test is implemented in STATA which is a general-purpose statistical software package which includes data management, statistical analysis, graphics, simulations, and custom

programming. There are three major builds of each version of STATA: 1) STATA/MP for multiprocessor computers (including dual-core and multicore processors), 2) STATA/SE for large databases; and 3) STATA/IC, which is the standard version (www.stata.com). STATA emphasizes a command-line interface which also facilitates replicable analyses. Recently, STATA has included a graphical user interface which uses menus and dialog boxes to give access to nearly all built-in commands. This could then generate codes in a seasonal geopredictive malarial-related risk model which can then be displayed, easing the transition to the command line interface and more flexible scripting language for robust residually forecasted district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive uncertainty modeling. The dataset could thereafter be viewed or edited in a spreadsheet format. STATA can then import error prone district-level malarial data attributes in a variety of formats including ASCII data formats such as CSV or databank formats and spreadsheet formats including various Excel formats, if so desired.

Further, STATA proprietary file formats are platform independent, so malarialogists/experimenters employing different operating systems can easily exchange residually forecasted district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive uncertainty estimators. Additionally, every version of STATA can read all older sampled district-level dataset formats, and can even write both the current and most recent previous dataset formats, using the *saveold* command. STATA can also read and write SAS XPORT format datasets natively, using the *fdause* and *fdasave* commands.

filename.xpt, which contains the Since White's test would involve regressing the squared error term from the OLS regression on the independent variables in the malarial regression equation in STATA, the R squared values renderd from that regression would be multiplied by *n* (i.e. sampled district-level hyperendemic transmission-oriented parameter estimators). The result then would be a test statistics distributed approximately as chi-squared. To determine which georeferenced variable causes the residual forecasts to be heteroskedastic in the empirical-sampled spatiotemporal dataset, a malarialogist/experimenter can perform White's test manually in STATA; thus, regressing each X on the squared error or, by simply plotting the squared error versus each independent variable.

Note, that this solution applies only to large sampled district-level empirical datasets. For small malarial-related samples, bootstrapped standard errors may be preferable. For example, a malarialogist/experimenter can utilize STATA margins and *marginsplot* commands for calculation and presentation of results from seasonally regressed malaria-related hyperendemic transmission-oriented explanatory covariate coefficient datasets. In particular, through use of the *margins plot* command, the regressed explanatory covariates coefficients and their residual forecasts can be graphically visualized. STATA can then proceed to more-complicated district-level models where the effects of the independent variables may be nonlinear. After seasonally quantitating nonlinear effects in the empirical district-level malaria-related datasets, STATA can then render significance levels in the regressed district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive variables using both standard polynomial terms (i.e., squares and cubes of variables) as well as fractional polynomial models. Thereafter, the software can investigate the performance of these procedures with particular regard to over fitting. If there is excess Type I error rates in the residually forecasted uncertainty estimators this may be then quantitated. Modifications, χ^2or F approximations to likelihood ratio statistics may also be compared to fractional district-level malaria-related polynomial models. In all cases, the *margins plot* command in STATA can be employed to illustrate the effect that changing an independent variable (weekly rainfall measure) has on the dependent variable (district-level prevalence rate) in a predictive malarial regression-based risk model framework. Piecewise-linear models may be then presented as well; these are linear models in which the slope or intercept is allowed to change depending on the range of an independent variable. Additionally, STATA uses the contrast command when discussing categorical variables which can allow a malarialogist/experimenter to contrast predictions made for various levels of any sampled categorical variable.

Further, STATA can compute heteroskedasticity-consistent estimates of the OLS coefficient covariance matrix and then perform heteroskedasticity-robust hypothesis tests based on OLS district-level regressed seasonal-sampled malarial-related hyperendemic transmission-oriented explanatory coefficient estimates. For example, STATA can perform the Breusch-Pagan-Koenker (BPK) LM tests for mixed heteroskedastic errors in linear regression model. In statistics, the Breusch–Pagan test is used to test for heteroscedasticity in a linear regression model. The test evaluates whether the estimated variance of the residuals from a regression are dependent on the values of the independent

variables. For instance, suppose a malarialogist/experimenter estimates a seasonal predictive regression model employing $y=\beta+\beta_1 X+\mu$ and obtains from this fitted model a set of district-level malaria –related hyperendemic transmission-oriented explanatory covariate coefficients measurement measurement values. Ordinary least squares would constrain these values so that their mean is 0 and so, given the assumption that their variance does not depend on the independent variables, an estimate of this variance can be obtained from the average of the squared district-level explanatory covariate coefficients measurement values. If the assumption is not held to be true, a simple model might be that the variance is linearly related to independent variables. Such a district-level geopredictive malarial risk model can be examined by regressing the squared residuals on the independent variables, employing a regression equation of the form $\hat{u}^2 = \gamma_0 + \gamma_1 x + v.$ which is the basis of the Breach–Pagan test. If an F-test confirms that the independent variables and are jointly significant then the null hypothesis of homoscedasticity can be rejected. The Breach–Pagan test tests for conditional heteroscedasticity is a chi-squared test and the test statistic is $n\chi^2$ with *k* degrees of freedom (Rao 1973). If the Breusch–Pagan test reveals that that there is conditional heteroscedasticity in the residual forecasted hyperendemic transmission-oriented explanatory covariate coefficients they may be corrected by employing robust standard errors in STATA.

To identify an appropriate seasonal geopredictive ARIMA model, for an empirical seasonal-sampled dataset of district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive covariate coefficients, a malarialogist/experimenter may begin by identifying the order(s) of differencing in STATA needing to stationarize the series and to remove the gross features of seasonality, perhaps in conjunction with a variance-stabilizing transformation such as logging or deflating. STATA is a general-purpose statistical software package created for data management, which contain analytical tools for statistical analysis, graphics, simulations, and custom programming (www.stata.com). In seasonal predictive district-level malaria-related risk model construction, a variance-stabilizing transformation is a data transformation that is specifically chosen either to simplify considerations in graphical exploratory data analysis or, to allow the application of simple regression or other analysis of variance techniques (see Jacob et al. 2009d). In applied statistics, a variance-stabilizing transformation is a data transformation that is specifically chosen either to simplify considerations in graphical exploratory data analysis or, to allow the application of simple regression-based or analysis of variance techniques (Hosmer and Lemeshew 2000). In fact, the easiest way to think of STATA constructed seasonal ARIMA –related district-level malarial predictive model is as fine-tuned versions of random-walk and random-trend models: the fine-tuning consists of adding lags of the differenced series and/or lags of the forecast errors to the predictive equation to remove any last traces of serial correlation in the residual forecast errors. Lags of the differenced series appearing in a district-level malaria-related regression-based forecasting equation could then be categorized as "auto-regressive" terms in STATA while lags of the forecast errors would be the "moving average" terms.

A malarialogist /experimenter may employ STATA margins and margins plot commands for calculating residual forecast error from the empirical dataset of regressed district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive explanatory covariate coefficients. In particular, through use of the marginsplot command, the regressed explanatory covariates coefficients and their residually forecasted uncertainty estimators can be graphically visualized. STATA can then proceed to more-complicated district-level models where the effects of the seasonal-sampled independent variables (e.g., district-level monthly rainfall measurements) may be nonlinear. After seasonally quantitating nonlinear effects in the empirical district-level malaria-related empirical datasets, STATA can render significance levels in the regressed hyperendemic transmission-oriented explanatory covariate coefficients employing both standard polynomial terms (i.e., squares and cubes of variables) as well as fractional polynomial models. Thereafter, the software can investigate the performance of these procedures with particular regard to over fitting. If there is excess Type I error rates, for example, in the residual forecasts defining the district-level hyperendemic transmission-oriented explanatory covariate coefficients, these variables may be further quantitated. For example, modifications, χ^2or *F* approximations to likelihood ratio statistics may be compared to fractional polynomial model outputs. In all cases, the *marginsplot* command in STATA can illustrate the effect of changing every single seasonal-sampled malaria-related independent variable in a regression-based robust predictive hyperendemic transmission-oriented risk model on a district-level dependent variable (e.g., prevalence rates). Piecewise-linear models may then be constructed which are basically linear models in which the slope or intercept is allowed to change depending on the range of an independent variables. The contrast command in STATA would then allow a malarialogist or experimenter to easily contrast residuals forecasts made from various levels of the seasonal-sampled categorical variables.

Further, since seasonal ARIMA–related geopredictive models are, in theory, one most general class of models for forecasting a time series dataset, stationaries can be quantitated in STATA by transformations such as differencing and logging. By so doing, robust residual forecasts from vigorously regressed seasonal-sampled hyperendemic transmission-oriented district-level field/clinical/remote sampled malaria-related predictive autoregressive explanatory covariate coefficients can be generated. Logging is a series often has an effect very similar to deflating: it dampens exponential growth patterns and reduces heteroscedasticity(Cressie 1993). In statistics, a collection of random variables is heteroskedastic if there are sub-populations that have different variabilities (Hosmer and Lemeshew 2000).The possible existence of heteroscedasticity is a major concern in the application of regression analysis for geopredictive seasonal malarial uncertainty risk mode construction including the analysis of variance, as the presence of heteroscedasticity can invalidate tests of significance that assume that the modeling uncertainties are uncorrelated and normally distributed and that their variances do not vary with the effects being modeled (see Jacob et al. 2011c, Jacob et al. 2009d). Logging is not exactly the same as deflating--it does not eliminate an upward trend in the data--but it can straighten the trend out so that it can be better fitted by a linear model. Thus, if a malarialogist/experimenter logs the seasonal-sampled district-level data and then fits a model that implicitly or explicitly uses differencing for ARIMA-malarial related random walk, exponential smoothing and predictive model construction, then it would be redundant to deflate by any district-level sampled index, as long as the rate of change is slow (i.e. the percentage change measured in the sampled explanatory covariate coefficients is nearly the same as the percentage change in the residual uncertainty forecasts). For instance , since LOG function in STATA have the defining property that LOG (X*Y) = LOG(X) + LOG(Y), the logarithm of a product in a predictive seasonal-sampled district-level malarial-related risk model would equal the sum of the logarithms. Therefore, logging the seasonal-sampled explanatory hyperendemic transmission-oriented covariate coefficients would convert multiplicative relationships to additive relationships in the residual district-level uncertainty forecasts while simultaneously converting exponential malaria-related trends to linear trends.

Thereafter, by taking logarithms of the seasonal-sampled district-level predictive variables that are multiplicatively related and/or growing exponentially over time, a malarialogist/experimenter may be then able to explain and quantitate any erratic error behavior within an empirical dataset of linear district-level model residual forecasts. For instance, a graph of a seasonally regressed empirical dataset of log-transformed district-level parameter estimators may be used to convert the exponential malaria related prevalence district-level change pattern to a linear growth pattern, while simultaneously converting the multiplicative (i.e., proportional-variance) seasonal pattern to an additive (i.e., constant-variance) seasonal pattern in an uncertainty risk model framework. The logarithm transformation may then be applied only to the explanatory hyperendemic transmission-oriented district-level error prone covariate coefficients measurement values which would of course be strictly positive. The log of zero or a negative number cannot be performed (Rao 1973). Logging the seasonal-sampled district-level predictive malarial data before fitting a random walk model would then yield a geometric random walk—(i.e., a random walk with geometric) rather than just a linear growth projection of the sampled hyperendemic transmission-oriented explanatory covariate coefficients. A geometric random walk is the default forecasting model that is commonly used for data analyses (Cressie 1993).

Additionally, since there are two kinds of standardized logarithms in standard in STATA (e.g., "natural" logarithms and base-10 logarithms) it would be easy to test for any uncertainty in the residual hyperendemic transmission-oriented forecasts. Log transformation is often used to convert time series that are nonstationary in predictive district-level malaria-related modeling with respect to the innovation variance in a stationary time series (Jacob et al. 2009d). The usual approach is to take the log of the series in a DATA step and then apply it in a STATA–related ARIMA to transform the data. A DATA step in STATA would then transform the forecasts of the logs back to the original units of measurements where the confidence limits would also be transformed by employing the exponential function. As one alternative, a malarialogist/experimenter may simply exponenate the residually forecasted district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive estimators. This procedure would then render a forecast for the median of the series, but the antilog of the forecast log series may underpredict the mean of the original series. As such, to predict the expected value of the series, a malarialogist/experimenter would then just simply spatiotemporally quantitate the standard error of the forecast employing an AR(2) model to determine the log of a series *Y*. By so doing, the logarithm of a product in a predictive malarial-related seasonal risk model would then be *i* the sum of the logarithms of the numbers being multiplied. By

so doing, the logarithm of the ratio of any two district sampled explanatory hyperendemic transmission-oriented covariate coefficients would simply be the difference of the logarithms.

The logarithm of the *p*-th power of a sampled uncertainty explanatory covariate coefficient would thereafter be p times the logarithm of the covariate coefficient itself. In other words the logarithm of a *p*-th root would be the logarithm of the explanatory hyperendemic transmission oriented malaria-relayed district-level covariate coefficient divided by *p* in the model. In Statagraphics, the LOG function would then be the natural log, and its inverse which would be the EXP function. (EXP(Y) is the natural logarithm base, 2.718..., raised to the Yth power while the base-10 logarithm and its inverse would be LOG10 and EXP10 in Statgraphics (www.stata.com).

Interestingly, in Excel and many hand-held calculators, the natural logarithm function is written as LN instead, and LOG stands for base-10 logarithms. Regardless, when logarithmic transformations in STATA are applied to geopredictive malarial seasonal risk model exercised in conjunction with differencing and logging, the residual forecasts targeting the hyperendemic transmission-oriented district-level uncertainty explanatory covariate coefficients will convert absolute differences into relative (i.e., percentage) differences. Thus, the series DIFF (LOG(Y)) would represent the percentage change in Y from period to period in the district-level predictive risk model. Strictly speaking, the percentage change in Y at period t in the model would thereafter be defined as (Y(t)-Y(t-1))/Y(t-1), which would be approximately equal to LOG(Y(t)) - LOG(Y(t-1)), but the approximation would be almost exact, if the percentage change is small. In STATA graphics terms, this means that DIFF(Y)/LAG(Y,1) would be virtually identical to DIFF(LOG(Y)). For instance, for determining a robust residual product from a predictive malarial-related district-level risk model, a formula may be written as $\log_b(xy) = \log_b(x) + \log_b(y)$ or a quotient applying $\log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y)$ [e.g., $\log_2(64) = \log_2(2^6) = 6\log_2(2) = 6$] .The logarithm log(*x*) can then be computed from the logarithms of *x* and *b* with respect to an arbitrary base *k* using the following formula: $\log_b(x) = \frac{\log_k(x)}{\log_k(b)}$ in any district- level geopredictive spatiotemporal malarial risk –based uncertainty model

Geomathematically speaking, thereafter, DIFF (LOG(Y/CPI)) in the residual forecasts would be nearly identical DIFF (LOG(Y)): the only difference between the two is a very faint amount of noise due to fluctuations in the inflation rate**.** Further, to computate a natural logarithm to base-10 logarithm in a robust predictive seasonal district-level risk model, a malarialogist/experimenter would simply divide the residual forecasted hyperendemic transmission explanatory covariate coefficient measurement values by the conversion factor 2.303 in STATA (www.stata.com). For example, calculating Log (100) in a district-level seasonal model of malaria-related empirical repressors could yield Ln (100) = 4.60517, then Log (100) = Ln (100)/2.303 = 4.60517/2.303 = 1.9996. Thereafter, by calculating Log (1.6210^-4), the residual forecasts would yield Ln (1.6210^-4) = -8.728 where Log (1.6210^-4) = Ln (1.6210^-4)/2.303 = -8.728/2.303 = -3.790.

Conversely, to convert a natural antilog in a geopredictive district-level malarial uncertainty risk model to a base=10 antilog, the malarialogist/experimenter would simply multiply by the conversion factor 2.303 before tabulating the natural antilog. For instance, to calculate the base-10 antilog of -3 a malarialogist/experimenter would have to find InvLn (-3*2.303) = InvLn (-6.909) in the residually forecasted estimators for identifying robust filed/clinical/remote hyperendemic transmission-oriented uncertainty explanatory covariate coefficients. Then Antilog (-3) = InvLn (-6.909) = 9.9910^-4.) For instance , to calculate the base-10 antilog of -8.45 in a seasonal predictive malarial risk model residual forecasts a malarialogist/experimenter would simply calculate InvLn(-8.45*2.303) = InvLn(-19.460). By so doing, then AntiLog (-8.45) = InvLn (-19.460) = 3.53610^-9 in the residual forecasts.

Another interesting property of the logarithm in STATA is that errors in geopredicting the logged series can be interpreted as percentage errors as rendered by the original series; albeit the percentages are relative to the forecast values, not the actual values. Normally a malarialogist/experimenter would interpret the "percentage error" to be the error expressed as a percentage of the actual geosampled hyperendemic transmission-oriented explanatory covariate

coefficient measurement value, not the forecast value, although the statistical properties of percentage errors are usually very similar regardless of whether the percentages are calculated relative to actual values or forecasts. Thus, if a malarialogist/experimenter employs least-squares estimation to fit a linear forecasting model to logged district-level malaria-related data, they would be implicitly minimizing mean squared percentage error rather than mean squared error in the original units--which is probably a good thing if the log transformation was appropriate in the first place for the model. If the malarialogist/experimenter, thereafter, defines the error statistics in STATA in logged units, an interpretation can be rendered as percentages. For example, the standard deviation of the errors in geopredicting a logged series, in an empirical-sampled dataset of seasonal-sampled explanatory hyperendemic transmission-oriented uncertainty covariate coefficients in STATA would be essentially the standard deviation of the percentage errors for predicting the original series and the MAE.

Therefore, for geopredicting a logged district-level malarial series the MAPE would have to be quantitated for defining the original series from the sampled empirical datasets. In the forecasting procedure, in Statagraphics, the error statistics would then be shown on the Model Comparison report but in untransformed (i.e., original) units as to facilitate a comparison among model outputs, regardless of the transformations employed. This would be a very useful feature of the district-level malarial-related forecasting procedure in STATA as it is would be hard to quantitate a head-to-head comparison of model outputs without a log transformation. Therefore, whenever a geopredictive seasonal ARIMA-related malarial district-level model is fitted with a number of analytical tools in conjunction with a log transformation for quantitating the standard-error-of-the-estimate or white-noise-standard-deviation statistics on the Analysis Summary report in STATA the transformed (i.e., logged) errors in the residual forecasts would essentially be the root mean square percentage errors.

Additionally, a malarialogist/experimenter can utilize STATA margins and margins plot commands for calculation and presentation of results from seasonally regressed malaria-related hyperendemic transmission oriented uncertainty explanatory covariate coefficients. In particular, through use of the margins plot command, the regressed district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive explanatory covariates coefficients and their residual forecasts can be graphically visualized. STATA can then proceed to more-complicated district-level models where the effects of the independent variables may be nonlinear. After quantitating nonlinear effects in the district-level malaria-related empirical datasets, STATA can then render significance levels in the regressed endemic transmission-oriented predictive variables using standard polynomial terms (i.e., squares and cubes of variables) and fractional polynomial models. Thereafter, the statistical software package can investigate the performance of these procedures with particular regard to over fitting. If there is excess Type I error rates in the residual forecast targeting the hyperendemic transmission-oriented uncertainty explanatory covariate coefficients, this may be seasonally quantitated as well. Modifications, χ^2or *F* approximations to likelihood ratio statistics may then be compared to fractional polynomial models in all cases, where the margins plot command in STATA can be employed to illustrate the effect that changing an independent variable has on the dependent variable in a district-level predictive malaria-related regression-based risk model. Piecewise-linear models can then be constructed (i.e.,linear models in which the slope or intercept is allowed to change depending on the range of an independent variable). As the name suggests, this command would allow a malarialogist/experimenter to easily contrast forecasts made for various levels of regressed sampled categorical variables.

Further, residual forecasts rendered from a seasonal geopredictive malaria-related regression-based risk model constructed in STATA could characterize an empirical dataset of regressed time-series hyperendemic transmission-oriented explanatory covariate coefficients as having weak white noise, for example, since the sampled data would be a sequence of serially uncorrelated random variables with zero mean and finite variance. To run a non-linear regression in STATA, the manipulation of data is needed which requires the function *generate* is needed to generate a new variable:*generate [new var name]=[function statement] (e.g.* f, *generate weight2 = weight*weight* can be used to create the weight“ district-level prevalence rate ” squared. Upon creating the variables in the geopredictive malaria-related district-level model, the malarialogist/experimenter can then use them to run the non-linear regression. Conversely, strong white noise in a district-level seasonal malarial geopredictive model could also be determined to possess the quality of being i.d.d.. The term white noise is used for a discrete signal whose samples are regarded as a sequence of serially uncorrelated random variables with zero mean and finite variance (Griffith 2003). Depending on the context, it may be required that the time-series hyperendemic transmission-oriented

samples be independent and have the same probability distribution. In particular, if each sample has a normal distribution with zero mean, the signal is said to be Gaussian white noise.

Thus, a seasonal district-level statistical geopredictive autoregressive malarial risk model can be determined to be normally distributed with mean zero and the Gaussian white noise residual series which may be sufficiently quantized for estimating forecasted residual variance error estimates. Further, given a vector time-series empirical ecological dataset of district-level malaria-related explanatory district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive covariate coefficients, an AIC-like criterion can be employed in conjunction with any time-series error detection algorithm in STATA for spatiotemporally quantitating latent inconspicuous uncorrelated variables in a geopredictive ARIMA model regression-based framework. When it comes to dealing with residual random effects, STATA automatically calculates the F value of individual variables using MSE as the denominator which can produce Type I SS while simultaneously computing F values (www.stata.com).

However, to calculate Type III SS, STATA requires lengthy coding by implementing loops to carry out invasive residual analysis in a STATA environment. As such, after opening up the "regression" or "ANOVA" window the malarialogist/experimenter would have to run the analysis and store the residual forecast values in ythat (www.stats.com). To store the geopredicted value in yhat in STATA , use predict yhat; to store the residuals in res, use predict res, r. (www.stata.com) To run a random effect in STATA, the user can use the following codes with *xtmixed* statement:*xtmixed [mixed factors]* || *[random factors]:*or use the pull-down method by choosing the "GLM" options (refer to ANOVA section). One thing to note here is to always include the colon after stating the random factors or an error message will pop out. When it comes to dealing with random effect, STATA automatically calculates the F value of individual variables using MSE as the denominator and produce Type I SS, where it is usually not the right way to compute F values. However, to calculate Type III SS, STATA requires lengthy coding by implementing loops and therefore will not be discussed here.

To run an AVOVA geopredictive district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented gautoregressive malaria-related risk based analysis thereafter in STATA: *ANOVA [dependant var] [explanatory variable]* may be employed. For instance, to generate main effect and interaction plots with significance levels of a and two levels of b, the following coding would be required with an example of plot provided:*predict yhat,.gen b1 = xb if b==1,gen b2 = xb if b==2,* and *graph b1 b2 a* However, to transpose the data in a sense to interchange hyperendemic transmission oriented observational variables, the function *xpose* is needed. When the matrix is transposed, new names for the sampled explanatory hyperendemic transmission oriented covariate coefficients can be expected to be given to new variables. If the malariaolgist/experimenter does not want to give new variable names, the option "clear" has to be included or an error message will pop out.(www.stata.com).

Also, to carry out a chi-squared test in STATA employing an empirical dataset of hyperendemic transmission oriented covariate coefficients, *chitest* or *chitesti* function can be used where *chitest [group1] [group2] ,chitesti [obs 1] [obs2] … [obs1] [obs2]…*whereby each analysis, Pearson chi-Square and likelihood ratio chi-Square is calculated Then by plotting res, yhat can generate the residual plot thereafter. By so doing, a regression analysis output can be produced with an ANOVA table, model fit statistics (R^2, Adj R^2, Root MSE, etc.), and a table with the coefficients, standard errors, significance tests and confidence intervals of the respective sampled district-level covariate coefficients Unfortunately, any number of potential outlying estimators can occur in the residual forecasts during the transfer process.

One problem with least squares and other regression based analyses for geopredictive malaria-related district level modeling is that there are usually one or more large deviations,(i.e. cases whose values differ substantially from the other observations). Outliers occur because (a) extreme values of observed variables can distort estimates of regression coefficients, (b) they may reflect coding errors in the data. The degrees to which hyperendemic transmission oriented malaria-related outliers influence a geopredictive malaria-related district-level regression line will vary For instance , an extreme sampled value of y that is paired with an average value of X will have less effect than an extreme value of Y that is paired with a non-average value of X in a geopredictive malarial risk model. An observation with an extreme value on a predictor variable (or with extreme values on multiple Xs) is called a point with high *leverage*. Fox gives the useful formula Influence on Coefficients = Leverage x Discrepancy. By this he

means that outlying explanatory district-level malaria-related hyperendemic transmission oriented covariate coefficient measurement values on Y will have the greatest impact when (a) their corresponding X values are further away from the mean of X, and (b) the Y value is out of line with the rest of the Y values,(i.e. it does not fall on the same line that the other cases do). The leverage option (which can also be called hat) calculates the Hosmer and Lemeshew (2000) leverage or the diagonal element of the hat matrix (so named because its computation involves yhat). Univariate or multivariate X outliers are high-leverage observations (Cressie 1993). Since leverage is bounded by two limits: 1/n and 1 (Rao 1973) the closer the leverage rendered from the regressed empirical dataset of district-level malaria-related geopredictive hyperendemic transmission oriented is to unity, the more leverage the value has. When the leverage > 2k/n then there is high leverage in a predictive malaria-related model, while for small samples 3k/n can be used. (Jacob et al. 2009d).In some circumstances a district-level sampled hyperendemic transmission oriented point with leverage greater than (2k+2)/n may be examined.

One advantage of Stata over SPSS is that it includes so-called robust regression routines that are better able to handle outliers The STATA routines rreg and qreg resist the pull of outliers, giving them better than OLS efficiency in the face of nonnormal, heavy-tailed error distributions(www.stata.com). OLS tends to track outliers, fitting them at the expense of the rest of the sample. Over the long run, this can lead to greater sample-to-sample variation or inefficiency when time series district-level malarial data samples often contain outliers. Robust regression methods aim to achieve almost the efficiency of OLS with ideal data and substantially better than OLS efficiency in non-ideal (e.g., nonnormal errors) situations.

Another alternative is qreg, which stands for quantile regression (i.e.Least Absolute Value Models or minimum L1-norm models). The most common form of quantile regression in seasonal district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive modeling is median regression, where the goal is to estimate the median (rather than the mean) of the dependent variable, conditional on the values of the independent variables. Put another way, median regression constructed using an empirical datset of geopredictive malaria-related explanatory hyperendemic transmission oriented covariate coefficients will find a line through the data that minimizes the sum of the absolute residuals rather than the sum of the squares of the residuals as in ordinary regression. Hence, the term Least Absolute Value as opposed to Least Squares would be more valuable in seasonal malaria-related risk modeling. Medians are less affected by outliers than means are, (Hosmer and Lemeshew 2000), so qreg can do better than regress when there are extreme outliers.

Further, the rreg and qreg routines work best when it is the DV that has outliers rather than the IVs. Statistics With STATA, Version 8, p. 239). According to the STATA 12 Manual, "One of the most useful diagnostic graphs is provided by leverage-versus-residual-squared plot lvr2plot, a graph of leverage against the normalized residuals squared." According to the STATA 12 Manual, "Standardized and Studentized residuals are also attempts to adjust residuals for their standard errors .Studentized residuals can be interpreted as the t statistic for testing the significance of a dummy variable equal to 1 in the observation in question and 0 elsewhere (Belsley et al. 1980). Such a dummy variable in a geopredictive district-level model would effectively absorb the sampled hyperendemic transmission observation and so remove its influence in determining the other malarial-related covariate coefficients. Thus, the lines on the chart in STATA will reveal geopredictive malaria-related seasonal regressed average values of leverage and the normalized residuals squared. Points representing sampled district-level malaria-related data above the horizontal line will then have a higher-than-average leverage while points to the right of the vertical line will have larger-than-average residuals.

Presently, there are two other current malarial-related definitions for statistically seasonally describing georeferenced error distributions from regressed empirical sampled autoregressively forecasted malarial-related explanatory hyperendemic transmission-oriented covariate coefficients. Some authors, for example, employ statistically residually forecasted error distributions which do not have all their power moments finite in the model parameter estimators, while others choose distributions that do not have a variance. The optimal methodology may then include encompassing all the seasonal malarial district-level residual forecasted distributions as well as those derived by the alternative definitions such as log-normal distributions that possess all power moments in the regression model framework for spatiotemporally quantitating biased district-level explanatory hyperendemic transmission-oriented estimators. Regardless, the outputs rendered, from the seasonal geopredictive hyperendemic

transmission-oriented risk-based error distribution model would indicate that the forecasted residual distribution has high kurtosis.

Kurtosis is a descriptive statistics based on a relative concentration of scores in the center, the upper and lower ends (i.e., tails), and the shoulders of a distribution (Fotheringham et al., 2002). As such, higher kurtosis in a seasonal-sampled district-level geopredictive regression-based risk model constructed from an empirical ecological dataset of hyperendemic transmission-oriented malarial-related explanatory covariates coefficients, for example, would indicate more of the variance in the residual forecasts is due to infrequent extreme deviations, as opposed to frequent modestly-sized deviations in the regressed coefficients. Environmental-related seasonal malarial-related feature data attributes that have more kurtosis than the normal (i.e., fat-tailed) commonly has its extremes extended beyond that of the normal (Jacob et al. 2009d). Ideally, a malariologist/experimenter would prefer a seasonal distribution of the sampled covariate coefficients with low kurtosis (e.g., residual forecasts not far away from the mean). However, for a seasonal district-level malarial-related geopredicted error distribution to be normalized, the regressed field/clinical/remote sampled hyperendemic transmission oriented predictive explanatory uncertainty covariate coefficients would have to exhibit an excess kurtosis equal to 0. Alternatively, a seasonal geopredictive district-level malarial regression-based risk model targeting specific endemic transmission-oriented covariate coefficients with positive kurtosis in the residual forecasts would have to exhibit a peak in the middle and fat tails versus a normal distribution. Fat-tailed distributions have values of kurtosis that are greater than 3.0 (Hosmer and Lemeshew 2000). Thus, the extreme values would be positive in the seasonal district-level geopredictive regression–based malarial model residual forecasts. However, this is only possible when the scenes in the residual forecasts are positive.

In probability theory and statistics, scenes is a measure of the extent to which a probability distribution of a real-valued random variable "leans" to one side of the mean (Fotheringham 2002). The skewness value can be positive or negative, or even undefined. Unfortunately, the qualitative interpretation of the skew is complicated in seasonal geopredictive malarial-related endemic uncertainty risk mapping. For a unimodal distribution, for example, rendered from a regressed empirical dataset of district-level malarial-related endemic transmission-oriented uncertainty explanatory covariate coefficients, a negative skew would indicate that the tail on the left side of the probability density function is longer or, fatter than the right side as it does not distinguish these shapes. In statistics, a unimodal probability distribution is a probability distribution which has a single mode. As the term "mode" has multiple meanings in a malarial-related district level geopredictive model so does the term "unimodal". Strictly speaking, a mode of a discrete probability distribution in a geopredictive malarial-related district-level uncertainty risk model would be a value at which the probability mass function (pmf) takes its maximum value (e.g., seasonal-sampled hyperendemic transmission oriented explanatory covariate coefficient measurement indictor value) (see Jacob et al. 2011b). A mode of a continuous probability distribution is a value at which the probability density function (pdf) attains its maximum value (Hosmer and Lemeshew 2000). Note, that in both cases there can be more than one mode in the malarial-related model, since the maximum value of either the pmf or the pdf can be attained at more than one seasonal-sampled parameter estimator value.

Conversely, positive skew in a district-level forecasted malarial-related distribution would indicate that the tail on the right side is longer or fatter than the left side. In cases where one tail is long but the other tail is a fat, scene does not obey a simple rule (see Fothingham 2002). For example, a zero value in a seasonal geopredictive malarial-related district-level uncertainty model would indicate that the tails on both sides of the mean balance out, which is actually the case for a symmetrical distribution. On the other hand, for asymmetric distributions rendered from a dataset of regressed malarial-related endemic transmission-oriented uncertainty explanatory covariate coefficients, the asymmetries would eventually even out, thus one tail would be long but thin, and the other would be short but fat. Further, in multimodal distributions and discrete distributions rendered from the district-level regressed dataset of district-level malarial-related field/clinical/remote sampled hyperendemic transmission oriented predictive autoregressive uncertainty explanatory covariate coefficients, skewness would also be difficult to interpret. For example, as soon as the skewness is negative in the district-level malarial-related forecasts, the impact of a high excess kurtosis would adversely affect the extreme negative regressed endemic transmission-oriented uncertainty explanatory covariate coefficient measurement values in the residual error variance.

Frequently, adjusted version of Pearson's kurtosis have been employed to seasonally quantitate the excess kurtosis in empirical datasets of regressed ecological-related seasonal-sampled district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive explanatory covariate coefficients and to provide a comparison of the shape of a given model error distribution, to that of the normal distribution. Pearson (1905) introduced kurtosis as a measure of how flat the top of a symmetric distribution was when compared to a normal distribution of the same variance. He referred to more flat-topped distributions ($\gamma_2 < 0$) as "platykurtic," less flat-topped distributions ($\gamma_2 > 0$) as "leptokurtic," and equally flat-topped distributions as "mesokurtic" ($\gamma_2 \approx 0$). Kurtosis is actually more influenced by scores in the tails of the distribution than scores in the center of a distribution (Hosmer and Lemeshew, 2000). Accordingly, it is often appropriate to describe a leptokurtic distribution in a spatiotemporal predictive district-level regression-based uncertainty malarial-related risk model as "fat in the tails" and a platykurtic distribution as "thin in the tails". Distributions with negative or positive excess kurtosis are called leptokurtic distributions, respectively (Fotheringham 2002, 2000). Leptokurtic distributions are identified by peaks that are thin and tall (Hosmer and Lemeshew, 2000). Platykurtic curves, on the other hand, have shorter 'tails' than the normal curve of error and leptokurtic longer 'tails'. Skewed distributions are always leptokurtic (Hopkins and Weeks 1990). Pearson's measure of kurtosis, however, has been often criticized in literature regarding predictive seasonal malarial-related regression-based uncertainty risk models as the indices do not focus adequately on the central part of an error distribution. Although never proposed, an alternative measure of kurtosis for seasonal district-level malarial-related predictive regression-based uncertainty risk model, the measurement of kurtosis may adjust the regressed georeferenced hyperendemic transmission-oriented explanatory covariate coefficients by removing the effect of skewness in the residual forecasts using autocorrelation statistics.

Spatial autocorrelation measures offer you additional insight into the interdependence of spatial data. These measures quantitate the correlation of a time -series geopredictive malarial-related data [e.g., z(s)] with itself at different locations as such these statistics can be very useful whether information exists at exact locations (point-referenced district-level data) or, measurements that characterize an area type such as census tracts, zip codes, and so on (areal data). One measure of spatial autocorrelation provided by one measure of spatial autocorrelation provided by PROC VARIOGRAM is Moran's *I* statistic, which was introduced by Moran (1950) and is defined as $I = \frac{n}{(n-1)S^2W}\sum_i\sum_j w_{ij}v_iv_j$ where $S^2 = (n-1)^{-1}\sum_i v_i^2$, and $W = \sum_i \sum_{j\neq i} w_{ij}$. Another measure of spatial autocorrelation in PROC VARIOGRAM is Geary's c statistic (Geary 1954), defined as $c = \frac{1}{2S^2W}\sum_i\sum_j w_{ij}(z_i - z_j)^2$ These expressions indicate that Moran's *I* coefficient makes use of the centered variable, whereas the Geary's c expression uses the non-centered values in the summation. Moran's *I* is a measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation. Geary's C is also known as Geary's contiguity ratio, Geary's ratio, or the Geary index.

In this research Moran's *I* was 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 units (e.g., districts) indexed by i (e.g. field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive covariate coefficients) and *j* (seasonal-sampled measurement indicator values). X was the variable of interest (i.e., seasonal prevalence rates in Uganda); $\bar{X}$ was the mean of X; and w*ij* was an element of a matrix of spatial weights. We used the expected value of Moran's *I* under the null hypothesis of no spatial autocorrelation which was $E(I) = \frac{-1}{N-1}$ whose variance in the residual was equivalent to $\mathrm{Var}(I) = \frac{NS_4 - S_3S_5}{(N-1)(N-2)(N-3)(\sum_i\sum_j w_{ij})^2}$

where $S_1 = \frac{1}{2}\sum_i\sum_j (w_{ij} + w_{ji})^2$, $S_2 = \frac{\sum_i(\sum_j w_{ij} + \sum_j w_{ji})^2}{1}$, $S_3 = \frac{N^{-1}\sum_i (x_i - \bar{x})^4}{(N^{-1}\sum_i (x_i - \bar{x})^2)^2}$, $S_4 = \frac{(N^2 - 3N + 3)S_1 - NS_2 + 3(\sum_i\sum_j w_{ij})^2}{1}$, $S_5 = S_1 - 2NS_1 + \frac{6(\sum_i\sum_j w_{ij})^2}{1}$ in SAS/GIS. (See Griffith 2003) Negative (positive) values indicate negative (positive) spatial autocorrelation. Values range from −1

(indicating perfect dispersion) to +1 (perfect correlation). A zero value indicates a random spatial pattern. 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 (Cressie 1993). Moran's *I* is inversely related to Geary's C, but it is not identical which is also a measure of spatial autocorrelation (Griffith 2003). In this research, Geary's C was defined as $C = \frac{(N-1)\sum_i \sum_j w_{ij}(X_i - X_j)^2}{2W\sum_i (X_i - \bar{X})^2}$ where *N* was the number of districts in Uganda indexed by *i* and *j* ; X was the district-level prevalence rates ; $\bar{X}$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 (Cliff and Ord 1971).

Inference on autocorrelation statistic comes from approximate tests based on the asymptotic distribution of *I* and *c* , which both commonly tend to have a normal distribution as *n* increases (Griffth 2003). To this end, PROC VARIOGRAM calculates the means and variances of *I* and c. The outcome then depends on the assumption made regarding the distribution$Z(s)$. In particular, a malarialogist can choose to investigate any of the statistics under the normality (i.e., Gaussianity) or the randomization assumption in a geopredictive malarial seasonal uncertainty-based risk model. Cliff and Ord (1981) provided the equations for the means and variances of the $I$and $c$distributions, as described in the following. The normality assumption asserts that the random field $Z(s)$follows a normal distribution of constant mean ($\bar{Z}$) and variance, from which the $z_i$values are drawn. In this case, the I statistics yield $\mathrm{E}_g[I] = -\frac{1}{n-1}$ and $\mathrm{E}_g[I^2] = \frac{1}{(n+1)(n-1)W^2}(n^2 S_1 - nS_2 + 3W^2)$ where $S_1 = 0.5\sum_i \sum_{j\neq i}(w_{ij} + w_{ji})^2$ and $S_2 = \sum_i (\sum_j w_{ij} + \sum_j w_{ji})^2$. The corresponding moments for the C statistics are $\mathrm{E}_g[c] = 1$ and $\mathrm{Var}_g[c] = \frac{(2S_1 + S_2)(n-1) - 4W^2}{2(n+1)W^2}$. According to the randomization assumption, the I and C malarial-related observations are considered in relation to all the different values in the geopredictive risk model which $I$and C could take, respectively, if the $z_i$values were repeatedly randomly permuted around the domain *D*. The moments for the I statistics would then be $\mathrm{E}_r[I] = -\frac{1}{n-1}$ and $\mathrm{E}_r[I^2] = \frac{A_1 + A_2}{(n-1)(n-2)(n-3)W^2}$ where $A_1 = n[(n^2 - 3n + 3)S_1 - nS_2 + 3W^2], A_2 = -b_2[n(n-1)S_1 - 2nS_2 + 6W^2]$. The factor $b_2 = m_4/(m_2{}^2)$ would then be the coefficient of kurtosis that uses the sample moments $m_k = \frac{1}{n}\sum_i v_i^k$ for $k = 2, 4$.

Finally, the c statistics under the randomization assumption would be given by $\mathrm{E}_r[c] = 1$ and $\mathrm{Var}_r[c] = \frac{B_1 + B_2 + B_3}{n(n-2)(n-3)W^2}$ with $B_1 = (n-1)S_1[n^2 - 3n + 3 - (n-1)b_2]$, $B_2 = -\frac{1}{4}(n-1)S_2[n^2 + 3n - 6 - (n^2 - n + 2)b_2]$ and $B_3 = W^2[n^2 - 3 - b_2(n-1)^2]$. Thereafter, if a malarialogist specifies LAGDISTANCE= to be larger than the maximum data distance in a specific domain, the binary weighting scheme used by the VARIOGRAM procedure would lead to all weights $w_{ij} = 1$, $i \neq j$. In this extreme case the preceding definitions would show that the variances of the I and c statistics becomes zero under either the normality or the randomization assumption. A similar effect might even occur when co-located malarial-related observations exist in the model. The Moran's *I* and Geary's C statistics allow for the inclusion of such pairs in the computations. Hence, contrary to the semi variance analysis, PROC VARIOGRAM does not exclude pairs of collocated data from the autocorrelation statistics.

The ARIMA procedure provides the identification, parameter estimation, and forecasting of autoregressive integrated moving average (Box-Jenkins) models, seasonal ARIMA models, transfer function models, and intervention models. The ARIMA procedure offers complete ARIMA (Box-Jenkins) modeling with no limits on the order of autoregressive or moving average processes. Estimation can be done by exact maximum likelihood, conditional least squares, or unconditional least squares. In addition you can model intervention models, regression

models with ARMA errors, transfer function models with fully general rational transfer functions, and seasonal ARIMA models. PROC ARIMA's model identification diagnostics include plots of autocorrelation, partial autocorrelation, inverse autocorrelation, and cross-correlation functions.

In this research we used PROC ARIMA for constructing a tentative autoregressive moving average (ARMA) order identification predictive malarial risk model based on smallest canonical correlation and an extended sample autocorrelation function. ARIMA model-based interpolation of missing values were permitted. Forecasting is tied to parameter estimation methods (Cressie 1993). Finite memory forecasts were used for estimating by maximum likelihood and exact nonlinear least squares, while infinite memory forecasts were used for estimating conditional least squares. The ARIMA procedure offers a variety of model diagnostic statistics, including AIC ,BIC ,Ljung-Box chi-square test statistics for white noise residuals stationarity tests, including Augmented Dickey-Fuller (including seasonal unit root testing), Phillips-Perron, and random-walk with drift tests (www.sas.edu) The %DFTEST macro performs Dickey-Fuller tests for simple unit roots or seasonal unit roots in a time series. The %DFTEST macro is useful to test for stationarity and determine the order of differencing needed for the ARIMA modeling of a time series.

In this research, we also employed the eigenvector filtering approach promoted by Griffith (2003) and Gets and Griffith (2002) by focusing on specification of a mean response to force the spatially dependent parameter values of an auto-model to zero in SAS/GIS. The Griffith approach is closely tied to the covariance matrices so fundamental to regression analyses, which is especially important for identifying the presence of multicollinearity as the number of repressors increases. Thus, we assumed that the Griffith would allow for predictive autoregressive endemic district-level malarial-oriented regression model residuals to either avoid a spatial specification or duplicate it in a more explicit spatial form. In most predictive autoregressive arthropod-related risk uncertainty mapping this is the case as the eigenvectors for the geostatistical covariance and autoregressive inverse covariance matrices in a time series-related model would be approximately the same (see Jacob et al. 2011a, Jacob et al. 2010c,Jacob et al. 2009d). Our assumption was this approach would allow in-depth, invasive study of precisely how spatial autocorrelation impacts upon error correlation coefficients rendered from regressed georeferenced seasonal-sampled explanatory covariates. We also assumed that these eigenvectors would relate directly to the Moran's *I* statistic, as well as to spatial autoregressive model specifications.

The basis of our procedure was the decomposition of Moran's *I* statistic into orthogonal and uncorrelated uncertainty malarial map pattern components. This tool measures spatial autocorrelation (i.e., seasonal-sampled malarial-oriented data feature similarity) based not only on the georeferenced feature locations or attribute values alone but on both feature locations and feature values simultaneously. Given a set of sampled district-level malarial-related hyperendemic malarial-oriented predictive autoregressive features can then be associated sampled attributes; the statistic can evaluate whether the pattern expressed is clustered, dispersed, or random (see Cliff and Ord 1972). Critical *Z* score values use a 95% confidence level normally occur between -1.96 and +1.96 standard deviations for determination of null hypothesis acceptance/rejection (Blackwell 1985). Moran's *I* tool in SAS/GIS can calculate the Moran's *I* Index value and a Z score for evaluating the significance of a seasonal-sampled malarial-related index value. For example, Jacob et. al. (2008a) applied the Getis-Ord Gi* statistic and found a significant cluster in the Kangichiri rice-village agro ecosystem complex in the Mwea Rice Scheme, Kenya (Z score > 3.70, $p < 0.05$), with the clustering of habitats highest at a distance of 400 m from the village complex.

The authors then used a robust *Z* score for evaluating immature aquatic larval habitats of *Anopheles arabiensis*, a major malaria-related mosquito in a landscape–oriented epidemiological model to determine if, seasonal-sampled georeferenced predictor variables fell outside that range or, if the pattern exhibited by the explanatory covariate coefficients were classified too unusual to be rendered through random chance. The authors then assumed that if the latter was the case it was then possible to reject the null hypothesis and proceed with the determination of the actual causation of the clustering tendencies within the residual algorithmic predictive uncertainty autoregressive outputs (e.g., either a statistically significantly aquatic larval habitat cluster or a statistically significantly dispersed pattern). The authors did this by employing a robust ArcGIS 'hot spot' analysis and calculating the Getis-Ord Gi* statistic for each sampled georeferenced immature *An. arabiensis*-related spatial data feature attribute within the riceland epidemiological study site. Typically hot spots or hot spot areas are concentrations of incidents [e.g., aggregation of habitats based on spatiotemporal field-sampled count data] within a limited geographical area that appear over time

(McDonald 2008). This tool quantified each georeferenced seasonal-sampled attributed within the context of the sampled neighboring features. If a sampled georefernced habitat feature attribute value was high, and the values for all of its neighboring sampled feature attributes was also high, the location is considered a hot spot (Patz 2000). Yearly arthropod –borne infectious disease clusters in similar areas are indicative of endemicity (i.e., hotspots) (Hay 2000).

A weighted set for neighboring seasonal-sampled riceland features was then attained. The local sum for a georeferenced *An. arabiensis* aquatic larval habitat sampled attribute value and its neighbors was then compared proportionally to the sum of all the seasonal- sampled data at the riceland study site. When the local sum is much different than the expected local sum, and that difference is too large to be the result of random chance, a statistically significant *Z* score is the result (Hosmer and Lemeshew 2002). When the analysis was conducted in the neighboring riceland village agro-ecosystem , two clusters were noted (Z score > 3.70, $p < 0.05$)—but, only up to a maximum distance of 400 m for a northern cluster and 150 m for a southern cluster. Commonly, non-binary weights are allowed in Gi(d) and G*i statistics, and the correlations between nearby values of the statistics are thereafter derived and verified by simulation (see Getis and Ord, 1992). Environmental and geographic features were then added to the seasonal predictive malarial regression-based risk maps to determine possible environmental factors associated with outbreaks.

In this research, we assumed that our decomposition would make orthogonal the latent spatial correlation represented by the geographic configuration of the sample district-level locations in Uganda described by a given spatial weights matrix. Commonly in an auto-model generated from seasonal -sampled field and remote related malarial–data, the probability density mass/functions contain a linear combination of the dependent variables values at a nearby sampled location (see Jacob et al., 2005b). We also assumed we could account for redundant locational information by generating eigenvectors of a given geographic weights matrix. These corresponding eigenvectors were then employed as observational uncertainty explanatory covariate coefficients in a predictive risk seasonal district-level endemic transmission-oriented regression-based equation. The aim of the spatial filtering analyses was to control for latent autocorrelation error coefficients in the empirical ecological dataset of the spatiotemporal-sampled, georeferenced hyperendemic transmission-oriented explanatory covariate coefficients with a set of proxy variables, rather than to identify a global autocorrelation parameter for efficiently seasonally spatially sampling district-level dependent process. Global statistics summarize standard error and other time-series dependent parameter uncertainty estimates from multiple sampled georeferenced locations making it difficult for spatial assessment of autoregressive error at a single predictive sampled site (Jacob et al., 2009d; Jacob et al., 2008a).We then utilized the misspecification interpretation of spatial autocorrelation, which assumed that residual error correlation among the sampled district-level uncertainty estimators was induced by missing exogenous variables, which themselves may have been spatially correlated. Our assumption was that by quantitating inconspicuous latent autocorrelation error coefficients propagation in a seasonal-sampled predictive malarial uncertainty risk-based map employing robust spatial filter eigenvectors, we could derive unbiased estimators spatially targeting district-level hyperendemic transmission-oriented explanatory covariate coefficients. Further, we assumed we could thereafter generate accurate predictive residual auto variance estimates and mean squared error probabilities for temporally forecasting district-level endemic transmission zones in Uganda. These model outputs would then adjust for the spatially predictive autoregressive error estimates in a space-time district-level forecasting model based on crucially dependent assumptions of non-normality and non-homogeneity in the sampled data.

Initially, we concentrated on standard linear regression models $y = Xb + e$ where y was an ($n \times 1$) vector of the endogenous variable for the n georeferenced district-level sampled observations, where X was an ($n \times k$) matrix of k exogenous variables, including an ($n \times 1$) unity vector 1, b was the ($k \times 1$) vector of regression parameters, and e was an (n x 1) vector of random disturbances. We assumed that the autocorrelation coefficients among the regression disturbances were induced by exogenous spatially correlated factors, which were not incorporated into the district-level model. This lead to a model misspecifications by shifting parts of the relevant information from the mean response Xb or, first-order component into an($n \times n$) covariance structure of the disturbances or, second-order component cov(e). Alternatively, we allowed an underlying spatial process to induce spatial autocorrelation in the district-level predictive district-level malarial hyperendemic transmission-oriented model. By so doing, we assumed that an observed spatial pattern in the response variable(i.e., district-level Ugandan prevalence rates) could be decomposed into, preferably three, statistically independent components: (a) a systematic spatial trend component that was specified by a parsimonious set of exogenous variables with a substantive meaning for the problem under

investigation; (b) a stochastic signal that reflected either an underlying spatial process and/or a set of missing exogenous factors with an inherent spatial pattern; and, (c) the independent white-noise disturbances.

Thereafter, we validated all residual district-level model outputs by incorporating the spatiotemporal-sampled georeferenced specific weights from the geopredictive residual autoregressive error matrix by configuring the field - sampled district-level malarial data into a cumulative covariate correlation algorithm to evaluate the robustness of the distributions rendered by the linear-based transmission-oriented risk model. Our assumption was that by calculating analytic derivatives with non-linear parameter restrictions employing simultaneous linear-based system and nonlinear predictive regression equations with distributed lags and time series error processes, robust spatial forecasters of district-level malaria-related prevalence rates could be accurately quantitated in SAS/GIS.

In this research, the spatially filtered SAS/GIS formatted the seasonal-sampled district-level georeferenced malarial-oriented spatial data feature data attributes sampled in the Ugandan study site which was then integrated with a SAS application, using SAS/EIS, SAS/GIS. These application sessions, driven from SAS/EIS or SAS/AF, provided powerful SAS Component Language (SCL) mechanisms and data step processing capabilities for spatially manipulating the district-level malarial risk mapping georefernced uncertainty explanatory covariate coefficients. After refining the sampled explanatory district-level malarial-related observational geopredictor variables and making them suitable to be used in a Poissonian model, we input this data into a SAS/GIS employing standard regression specifications of seasonally-sampled district-level representing 2006, 2007, 2008, 2009, 2010 to check their linear integrity. Spatial filter analysis was then also performed in SAS/GIS. Spatial information/data from each of the Ugandan districts was then imported interactively and in a batch mode format into ArcGIS Geostatistical Analyst for an on-line geopredictive uncertainty risk mapping and analysis solution. For example, empirical Bayesian-oriented probabilistic estimation matrices are currently available in Geostatistical Wizard and as a geoprocessing tool in the Geostatistical Analyst toolbox in ArcGIS which can be exported into a SAS geodatabase (www.esri.com). The on-line solution was constructed to enable the forecasted data rendered from our model residuals to directly generate malarial district-level uncertainty maps configured to illustrate their choices of risk and contextual data set integrations. The Ugandan predictive risk mapping solution was then extended into the classical temporal-geographic modeling concepts to determine prolific high to low malaria risk foci within classified Thessian polygons at the district level including maps from 2006-2010 of the sampled observational variables to establish futuristic malaria risk trends.

We then attempted to construct a robust random unbiased error estimator using the residuals rendered from the spatial filter model to seasonally quantitate the uncertainty effects in the selected geosampled malarial-related georeferenced coefficients in order to capture district-level dependence in the empirical datasets. The articulated tessellations for the Ugandan study site based upon district geocodes were then digitally overlaid onto interpolated risk maps from the Malaria Atlas Project (http://www.map.ox.ac.uk) in ArcGIS. Mosquito vector arthropod species seasonal distributional data such as species range maps (i.e., extent-of-occurrence), district-level surveys and biodiversity atlases are a common source for risk–based cartographic displays of mosquito species-environment relationships (Hay 2000). In this research, we constructed multiple seasonal stochastic models using archived Kriged malarial mosquito species *An. arabiensis s.s* and *An. gambiae s.l.* distribution data. Contagious processes, such as conspecific attraction, may generate spatiotemporal error patterns in species abundance cannot be explained by simple regression-based hierarchical cluster-based models (see Griffith 2005).

In this research, the remotely-dependent explanatory predictor covariate intra-cluster error coefficients were derived using QuickBird 0.61m spatial resolution image data (www.digitalglobe.com) for constructing multiple malarial-related autoregressive malaria mosquito aquatic larval habitat uncertainty distribution models. Further, to improve seasonal district-level malarial predictive risk modeling we overlaid the spatial tessellation rendered from our model onto the historical datasets from Malaria Atlas Project (i.e., MAP). The MAP team has assembled a unique geospatial database on linked information based on medical intelligence and satellite-derived climate data to constrain the limits of malaria transmission employing the largest ever remote archive of community-based estimates of parasite prevalence (http://www.map.ox.ac.uk). These data have been assembled and analyzed by a team of geographers, statisticians, epidemiologists, biologists and public health specialists.

Therefore, the objectives of this research were to: (1) construct robust stepwise regression models using multiple georeferenced observational variables for spatiotemporally quantitating residual varying and constant uncertainty

covariate coefficients associated to sampled district-level parameter estimators (2) filter all latent autocorrelation uncertainty coefficients in the residual variance estimates using an eigenfunction error diagnostic decomposition algorithm, (3) and thereafter, construct robust predictive autoregressive malarial risk maps using existing data in MAP ATLAS for accurately forecasting field and remote sampled district-level malarial indices in Uganda.

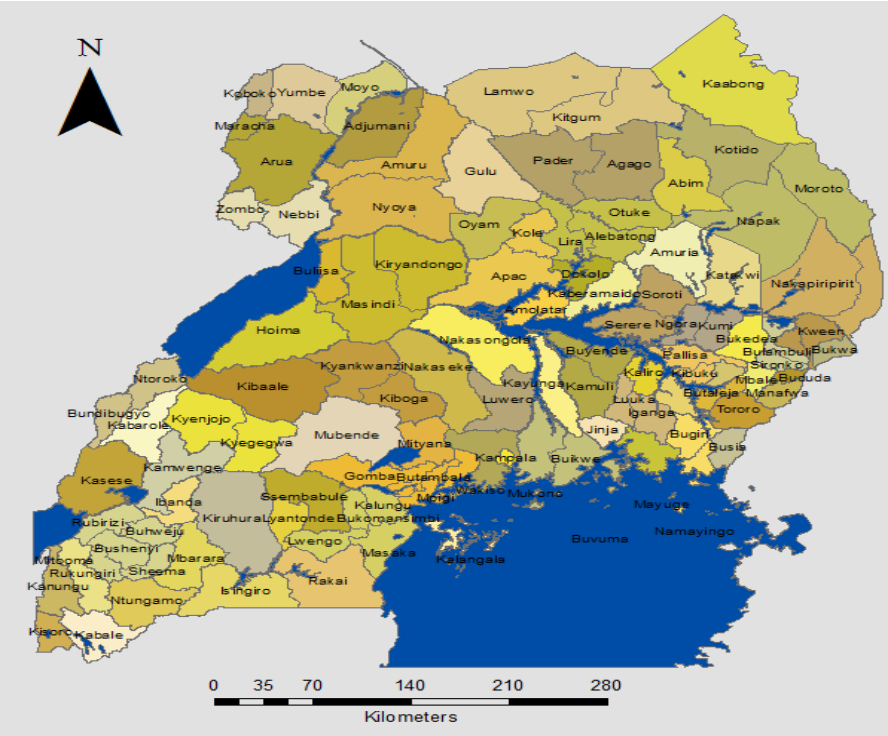
Materials and Methods

2.1. Study site: Uganda is a landlocked country in East Africa. The country is located on the East African plateau, lying mostly between latitudes 4°N and 2°S (a small area is north of 4°), and longitudes 29° and 35°E. It is about 1,100 meters (3,609 ft.) above sea level, and these slopes very steadily downwards to the Sudanese Plain to the north. However, much of the south is poorly drained, while the center is dominated by Lake Kyoga, which is also surrounded by extensive marshy areas. In many endemic areas, malaria prevalence in communities is maximum in areas bordering on marshes where rates can range from 1% to 15% according to age and season of the year (Trape et. al. 1992).

Although generally equatorial, the climate is not uniform as the altitude modifies the climate. Southern Uganda is wetter with rain generally spread throughout the year. Although *An. gambiae* is usually the predominant species in environments with high humidity and rainfall, *An. arabiensis* is more common in zones with less rainfall and both species occur sympatrically across a wide range of tropical Africa (Coetzee and le Sueur 2000). Larvae of *An. gambiae* are commonly found in clear, sunlit pools of water in small depressions such as foot or hoof prints, the edges of bore holes and burrow pits, roadside puddles formed by tire tracks, irrigation ditches and other man-made shallow water bodies (Gimnig et al. 2001, Gillies and De Meillon 1968, White 1972). *Anopheles gambiae* malaria vectors have also been found breeding in polluted water rich in organic matter (Sattler et al. 2005, Keating et al. 2004), in large bodies of water such as flood plains (Castro et al. 2010) and in pools of water along lake shores especially when there are fluctuations in water level as in Lake Victoria (Minakawa et al. 2008). At Entebbe on the northern shore of Lake Victoria, most rain falls from March to June and in the November/December period. Further to the north a dry season gradually emerges; at Gulu about 120 km from the South Sudanese border, November to February is much drier than the rest of the year.

Uganda is divided into districts, spread across four administrative regions: Northern, Eastern, Central (i.e., Kingdom of Buganda) and Western. The districts are subdivided into counties. A number of districts have been added in the past few years, and eight others were added on July 1, 2006 plus others were added throughout 2010 due to increased urbanization and local policy changes. Environmental alterations due to deforestation, swamp reclamation mainly for agriculture, excavation of sand and building stones, brick making and vegetation clearance may lead to an increase in aquatic larval habitats of malaria vectors, such as *An. gambiae s.l.* (Carlson et al. 2004, Fillinger et al. 2004). Presently there are over 100 districts in the Ugandan study site (.http://en.wikipedia.org/wiki/Uganda - cite_note-district-21) Most districts are named after their main commercial and administrative towns. (See Figure 1 for district-level administrative divisions in Uganda

Figure 1: Administrative Boundaries: Districts of Uganda per 2010 data



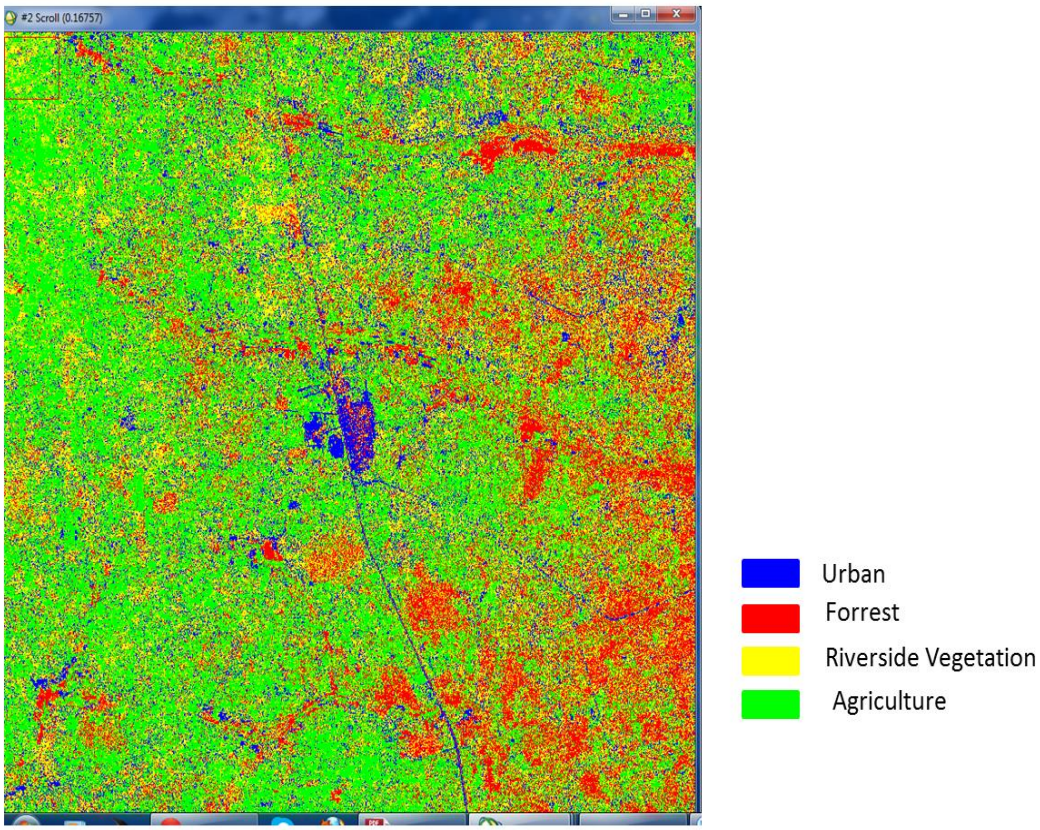
2.2 District mapping: Initially, base maps were generated using Google Earth ™ and global positioning systems (GPS) ground coordinates of the various districts at the Ugandan epidemiological study site. The GPS ground coordinates were acquired from a CSI max receiver which has a positional accuracy of +/- .178 (Jacob et al. 2009c)). Using a local GPS broadcaster can compensate for ionospheric and ephemeris effects which can improve horizontal accuracy significantly and can bring altitude error down in a spatiotemporal seasonal-sampled predictive vector arthropod larval habitat distribution model (Jensen 2005). Each georeferenced district in Uganda featured attribute was entered into the VCMS™ relational database software product (Clarke Mosquito Control Products, Roselle, IL). The VCMS™ database supports a mobile field data acquisition component module, called Mobile VCMS™ that synchronizes field-sampled data from industry standard Microsoft Windows Mobile™ devices and can support add-on GPS data collection. Mobile VCMS™ and its corresponding FieldBridge® middleware software component were used to support both wired and wireless synchronizing of the seasonal field-sampled data collected district-level data. The data was collected with the Mobile VCMS™ and then synchronized thereafter directly into a centralized VCMS™ relational repository database. Additional geocoding and spatial display of the spatiotemporal-sampled seasonal Ugandan data attributes was then mapped using the embedded VCMS™ GIS Interface Kit™ which was developed utilizing ESRI's MapObjects™ 2 technology. In this research, the VCMS™ database supported the export of all geoparameters using any combination of the time-series district-level estimators in order to further process and geospatially display specific data attributes in a stand-alone desktop GIS software package (i.e., ArcGIS 10.1®).

2.3 Grid-based algorithm: Thereafter, a digitized matrix was constructed by applying a mathematical algorithm in order to fit the continuous and bounded sampled district-level surfaces from a field and vegetated canopy -sampled attribute. GIS grid-based data files consist of columns and rows of uniform cells coded according to georeferenced data values (Jensen 2001). Each digitized grid cell within the matrix contained an attribute value as well as the district sampled geocoordinates. As such, the spatial location of each district cell was implicitly contained within the

ordering of the matrix . Multiple data layers were then created using different coded field/clinical/remote sampled hyperendemic transmission oriented covariate coefficients values for the various field attributes which were related to the same grid cell. Each polygon was assigned a unique identifier. Field attribute tables were then linked to the polygons. The polygons were used thereafter to define the district-level sampling frame. This allowed for multiple interactions enabling retrieval and transformation of the geosampled district-level data efficiently, regardless of spatial dimensionality of the data featured attribute.

Remote Sensing: QuickBird (www.digitalglobe.com) images were acquired in March 11th 2008, for the SJL study site. QuickBird multispectral products provided four discrete non-overlapping spectral bands covering a range from 0.45 micrometer (µm) to 0.72 µm, with an 11-bit collected information depth with a spatial resolution of 0.61m. QuickBird imagery was classified using the Iterative Self-Organizing Data Analysis Technique (ISODATA) unsupervised routine in ERDAS *Imagine* V.8.7™. The images were co-registered manually, using ground control points and georectified images from the QuickBird data. The satellite images were co-registered by applying a first order polynomial algorithm with a nearest neighbor resampling method. The Universal Transverse Mercator (UTM) Zone 37S datum WGS-84 projection was used for all of the spatial datasets.A land use land cover analyses was then generated as in Jacob et al. (Jacob et al. 2013b,

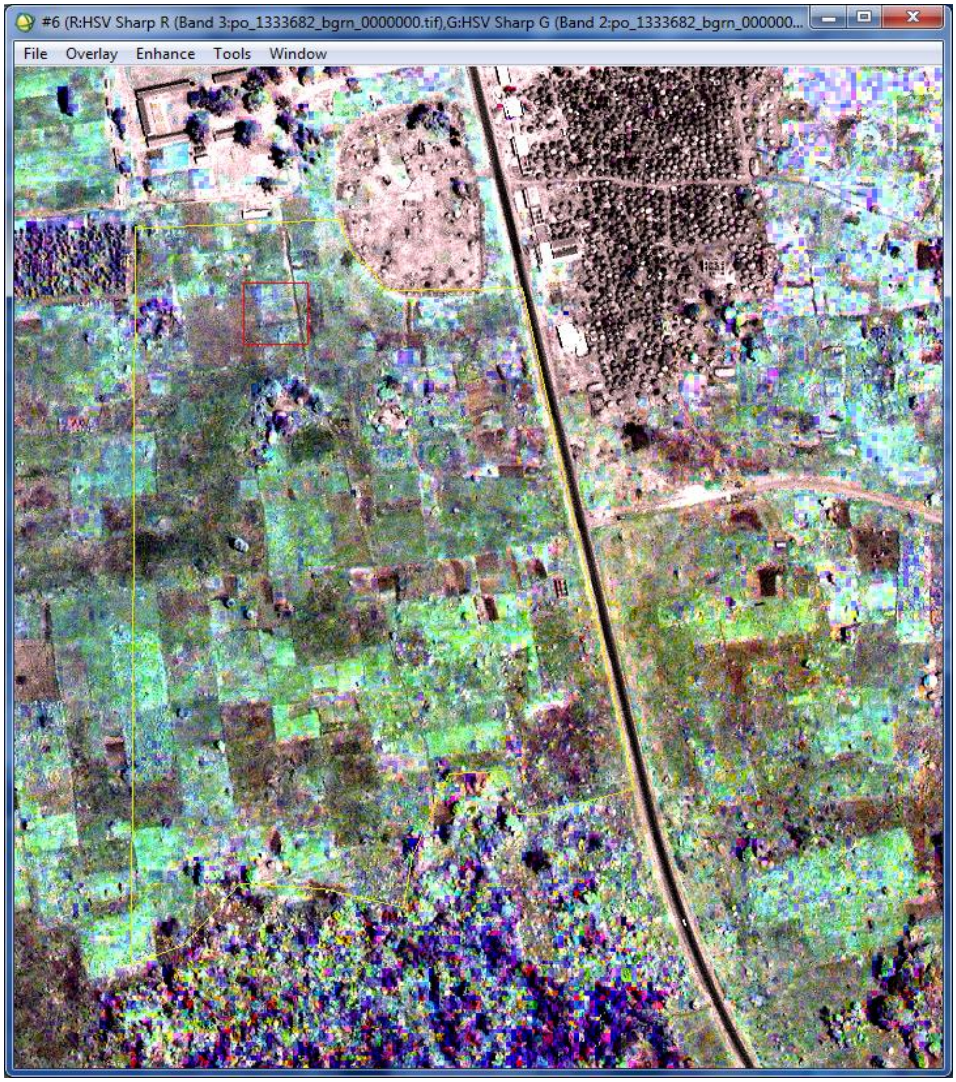
Figure 2: An ArcGIS land use land cover analyses for the Ugandian study site



2.4 Environmental parameters: Multiple district-level georeferenced explanatory predictor covariate coefficients were then examined extensively using: longitude, latitude, and altitude data. The criteria involved the centrographic measures of spatial mean . The attributes included district level monthly rainfall, humidity, population density and specific vegetation geoparameter estimates.

2.5 Vegetation Indices*:* The different modules in Spatial Analyst® extension of ArcGIS 10.1 and spatial modeller tools from ERDAS *Imagine* 9.1 were then used to perform VI calculations. NDVI was calculated using radiance, surface reflectance (*p*), and the apparent reflectance measured at the top of the atmosphere(TOA) employing the georeferenced district level vegetated canopy spectral covariate coefficients and the satellite spectral bands. The ratio of reflected vegetated canopy radiance from the red and NIR bands from each imager were then used to normalize illumination and topographic variation and to form the NDVI, which was then used as an indicator of the amount and vigor of canopy vegetation in the district-level epidemiological sites. We performed Raster modeling in ArcGIS 10.1 which included performing image differencing on the NDVI layers, classifying the layers into different classes and calculating a wetness index using the Raster Calculator. The difference of the QuickBird visible and NIR bands was then divided by their sum, which formed the functionally equivalent NDVI, over the vegetated canopy and terrestrial surfaces of the epidemiological study site. In this research,NDVI was computed directly without any bias or assumptions regarding plant physiognomy, canopy cover class, soil type, or climatic conditions, within a range from -1.0 to 1.0 using the QuickBird visible and NIR reflectances, (*p)*, as in Jacob et al., (2012)

Figure 3: An ArcGIS NDVI analyses for Gulu district of the Ugandian study site



2.6 Environmental Regression Parameters: Initially, the data analysis explored covariation between prevalence employing the formula adjusted cases)/population, in SAS/GIS which in this research was not the same as the reported number of probable and confirmed cases. In this research we employed variable Y, and the following variables: annual population density, density of clinics, and density of water bodies; monthly humidity, rainfall, and minimum and maximum temperature. Lines graphs were then generated in SAS/GIS using median rainfall and prevalence data. The graphs were generated for five different years (2006-2010) and across the different georeferenced sub-regions, following the traditional regional designations outlined in Figure 1.

We then generated histograms in SAS which was constructed using a 95% confidence level ascertained whether the proportions of the within cluster-based district-level field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented residual explanatory covariate coefficient estimates differed by sampled district locations. The SAS procedures (e.g, PROC REG) were the used to fit the district-level models. SAS/STAT procedures commonly perform at least one type of regression analysis (e.g., CATMOD, GENMOD, GLM, LOGISTIC, MIXED, NLIN, ORTHOREG, PROBIT, RSREG, and TRANSREG procedure). SAS/ETS and PROC REG procedures are specialized for applications in time-series or simultaneous systems (www.sas.edu). In this research, the regression models assumed independent Bernoulli outcomes denoted by Yi = 0 or 1, taken at the sampled Ugandan district sites $i = 1, 2, \ldots, n$,. The estimator measurement indicator values were then described by $\mathbf{X}$i, a 1-by-(K+1) vector of K values and a 1 for the intercept term which represented a sampled district-level site location *i*. The probability of a 1 being realized for the binary outcome data was provided by: $P(Yi = 1| \boldsymbol{X}i) = \exp(\boldsymbol{X}i\boldsymbol{\beta})/ [1 + \exp(\boldsymbol{X}i\boldsymbol{\beta})]$ (2.1) where $\boldsymbol{\beta}$ was the (*K*+1)-by-1 vector of non-redundant parameters and $P(Yi = 0| \boldsymbol{X}i) = 1 - P(Yi = 1|\boldsymbol{X}i)$. Jacob et al (2009d) used the simplest form of Equation (2.1) for qualitatively assessing and quantizing a constant probability across multiple randomized spatiotemporal sampled aquatic larval habitats of *An. arabiensis*, using [i.e., $P(Yi = 1| \boldsymbol{X}i) = P(Yi = 1| \alpha) = \exp(\alpha)/[1 + \exp(\alpha)]$ which rendered a constant α using a bivariate regression notation. This statistical procedure was performed by denoting β_0, where $P(Yi = 1| \alpha) \rightarrow 0$ as $\alpha \rightarrow -\infty$, $P(Yi = 1| \alpha) \rightarrow 0.5$ as $\alpha \rightarrow 0$, and $P(Yi = 1| \alpha) \rightarrow 1$ as $\alpha \rightarrow \infty$ in the multivariate regression model matrix framework.

A Poisson regression with statistical significance was also calculated by a 95% confidence level in PROC REG. The Poisson regression is a member of a class of generalized models which is an extension of traditional linear models which allows the mean of a population to depend on a linear predictor covariate coefficient estimate through a nonlinear link function while allowing the response probability distribution to be any member of an exponential family of distributions (see Hosmer and Lemeshew 2002). The following statements were used to estimate the Poisson regression model:

```
proc countreg data=one ;

  model y = x / dist=poisson ;

run;
```

The response variable (i.e., Y) represented district-level malarial prevalence which was numeric and had nonnegative field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients estimators integer values

In the Poisson 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, ..., where the log of the mean μ was assumed to be a linear function of the sampled independent variables. That is, log(μ) = intercept + b1*X1 +b2*X2 + ...+ b3*Xm, implied that μ was the exponential function of independent variables, where μ = exp(intercept + b1*X1

+b2*X2 + ...+ b3*Xm). The Poisson regression model was then rewritten in the following form: log(μ) = log(N) + intercept + b1*X1 +b2*X2 + ...+ b3*Xm, where *n* was the total number of explanatory covariates sampled in each district–level study site. The logarithm of variable *n* was used as an offset, that is, a quantitative regression variable with a constant coefficient of 1 which in this research represented each sampled independent observation. 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 parameter estimator error hierarchy rendered from the of regression model residuals in PROC GENMOD.

In this research the parameter $\lambda_i(\mathbf{X}_i)$ was both the mean and the variance of the Poisson distribution for a specific sampled district *i*. The sampled district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented exploratory observational exploratory predictors data was log-transformed before the regression-based 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 sampled district-level locations i=1, 2…n. The Poisson 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 spatiotemporal-sampled district-level covariate coefficients. This expression was written more compactly as $\log(\mathbb{E}(Y|x))$ 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*. Therefore, in our Poisson model, θ was an input vector *x* and the predicted mean of the associated Poisson distribution rendered from the sampled district-level explanatory covariate coefficient estimates which in this research was provided by $\mathbb{E}(Y|x) = e^{\theta x}$ but, only if X ε R^n was a vector of the independent variables. Thereafter, the Poisson model took the form $\log(\mathbb{E}(Y|x)) = a'x + b$ where a ε R^n $a \in \mathbb{R}^n$ and b ε R $b \in \mathbb{R}$. Positing salient error estimators using Poisson-derived regression estimates, the maximization of an auto-Gaussian log-likelihood function and a set of eigenvectors where lamda is the sub-space of R^n can identify and quantify malaria-related observational covariate coefficients (Jacob et al. 2011c). The Gaussian distribution is a continuous probability distribution that is often used as a first approximation to describe real-valued random variables that tend to cluster around a single mean value (Hosmer and Lemeshow 2002).

In our regression framework the district-level data were denoted by matrix $\mathbf{X}_i$, which was constructed employing a 1×*p* vector of district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented exploratory observational exploratory observational predictor values for a specific resampled location *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 parameters and the Poisson specified parameter estimators were then rendered by $\lambda_i(\mathbf{X}_i)=\mu_i(\mathbf{X}_i)/n_i(\mathbf{X}_i)$ (2.2). Thereafter, the Poisson regression models were generalized by introducing an unobserved heterogeneity term for the sampled observational variables (*i*). Thus, the district-level spatiotemporal-sampled data was assumed to differ randomly in a manner that was not fully accounted for by the explanatory covariate error coefficient estimates. These distributions were then formulated as $E(y_i|x_i,\tau_i) = \mu_i\tau_i = e^{x_i\beta+\varepsilon_i}$, where the unobserved heterogeneity term $\tau_i = e^{\varepsilon_i}$ was independent of the vector of regressors x_i; thus, the distribution of y_i conditional on x_i and τ_i was Poisson with a conditional variance of $\mu_i\tau_i$: $f(y_i|\mathbf{x}_i,\tau_i) = \dfrac{\exp(-\mu_i\tau_i)(\mu_i\tau_i)^{y_i}}{y_i!}$. We then let $g(\tau_i)$ be the probability density function (pdf) of τ_i. Then, the distribution $f(y_i|x_i)$, was no longer conditional on τ_i in the x. Thereafter, spatiotemporal-sampled district-level linear malarial model district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented exploratory observational exploratory predictors residuals were obtained by integrating $f(y_i|x_i,\tau_i)$ with respect to τ_i: $f(y_i|\mathbf{x}_i) = \int_0^\infty f(y_i|\mathbf{x}_i,\tau_i)g(\tau_i)d\tau_i$.

We noticed that the district-level geopredictive autoregressive malaria geopredictive model error residuals 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. 2009d). We assumed that τ_i followed a gamma (θ, θ) distribution with $E(\tau_i) = 1$ and $V(\tau_i) = 1/\theta$: $g(\tau_i) = \frac{\theta^\theta}{\Gamma(\theta)} \tau_i^{\theta-1} \exp(-\theta\tau_i)$ where $\Gamma(x) = \int_0^\infty z^{x-1} \exp(-z) dz$ was the gamma function and θ was a positive sampled malarial-related parameter. Thus, the density of y_i in the time series-dependent district-level regression models was $\mathbf{X}_i$ which in this research was further quantified using the equation: $f(y_i|\mathbf{x}_i) = \int_0^\infty f(y_i|\mathbf{x}_i, \tau_i) g(\tau_i) d\tau_i = \frac{\theta^\theta \mu_i^{y_i}}{y_i! \Gamma(\theta)} \int_0^\infty e^{-(\mu_i+\theta)\tau_i} \tau_i^{\theta+y_i-1} d\tau_i = \frac{\Gamma(y_i+\theta)}{y_i!\Gamma(\theta)} \left(\frac{\theta}{\theta+\mu_i}\right)^\theta \left(\frac{\mu_i}{\theta+\mu_i}\right)^{y_i} =$

$$\frac{\theta^\theta \mu_i^{y_i} \Gamma(y_i+\theta)}{y_i!\Gamma(\theta)(\theta+\mu_i)^{\theta+y_i}} \quad (2.3)$$

Unfortunately, extra-Poisson variation was detected in the residual variance estimates in the models. Evidence of overdispersion indicates inadequate fit of the Poisson model (Hosmer and Lemeshew 2002). A common way to deal with overdispersion for counts is to use a Generalized Linear Model (GLM) framework, where the most common approach is a ‘‘quasi-likelihood,’’ with Poisson-like assumptions (i.e., quasi-Poisson) or a negative binomial model (see Hosmer and Lemeshew 2002). 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 geopredicted 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 by making the by incorporating $\alpha = \frac{1}{\theta}$ $(\alpha > 0)$ in equation 2.1 as in Jacob et al (2010a). The negative binomial 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 Poisson-related malarial random 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 further estimate the spatiotemporal malarial district-level 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 $\frac{\partial \mathscr{L}}{\partial \alpha} = \sum_{i=1}^{N} \left\{ -\left(\sum_{j=0}^{y_i-1} \frac{\alpha^{-1}\mu_i}{(j\alpha+\mu_i)} \right) - \alpha^{-2}\mu_i \ln(1+\alpha) - \frac{(y_i+\alpha^{-1}\mu_i)}{1+\alpha} + \frac{y_i}{\alpha} \right\}$.

In this research, the 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 $\alpha = \frac{1}{\delta_i} = 0$. A Wald test of this hypothesis was also provided which then rendered the reported t statistic for the estimates in the negative binomial 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 predictive West Nile Virus (WNV) mosquito vector *Culex quinquefasciatus* habitat regression cluster-based model in Birmingham, Alabama. In their research, the NEGBIN2 model with $p = 2$, was the standard formulation of the negative binomial model. Although the formulation derived employed the same technique as in Jacob et al. (2010c) there other values of p were also input, $(\text{i.e., } -\infty < p < \infty)$, into the district-level models which interestingly had the same density $f(y_i|x_i)$ except that α^{-1} was replaced by $\alpha^{-1}\mu^{2-p}$.

Shapiro–Wilk diagnostic test: The Shapiro–Wilk test was then used to test the null hypothesis that the spatiotemporal-sampled cluster-based georeferenced explanatory covariate coefficient estimates $x_1, \ldots, x_n$. 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 predictive geo-autoregressive malaria mosquito larval habitat distribution model of *Anopheline gambiae s.l.* in multiple spatiotemporal datasets of georeferenced field and remote-sampled predictor covariate coefficients by constructing W statistic. In their research W represented the ratio of an optimal uncertainty error estimator of the residual variance based on the square of a linear combination of ordered statistic which in turn was based on the corrected sum of squares estimator of the variance. 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. Further, in Jacob et al. (2008b) the predictive autoregressive *An. gambiae s.l.* regression risk model were optimized employing maximize *f(x,y)* subject to g(x,y)=C. They then introduced a new explanatory observational variable (λ) into the model and studied the Lagrange function which was then defined as $\Lambda(x, y, \lambda) = f(x, y) + \lambda \cdot \big(g(x, y) - c\big)$. The model residuals revealed that when $f(x_0, y_0)$ was a maximum of $f(x, y)$ for any constrained problem in the model, then there existed λ_0 such that (x_0, y_0, λ_0) was a stationary point for the Lagrange function. Additionally, in the model the district-level stationary points were those points that where the partial derivatives of Λ were zero,(i.e. $\nabla\Lambda = 0$).

In this research 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 SAS/GIS. We noticed in our regression-based malaria model when $X_{(1)}$ was the *i*th order statistic, (i.e., the *i*th-smallest number in the district-level 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 district-level malarial-related 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 ordered sample values in

SAS/GIS based on the corresponding expected normal order statistics, which was linear in this research for the distributed district-level populations. After W was calculated, the hypothesis of normality was rejected in the geopredictive autoregressive district-level malaria risk model residual error matrix since W was less than a quantile from any sampled value in the model

These district-level data were then furthered analyzed via a Q-Qin SAS/GIS. A Q-Q plot is a plot of the quantiles of two distributions against each other, or a plot based on estimates of the quantiles; the pattern of points in the plot is then used to compare the two distributions as in Anselin (1995). The main step in constructing our Q-Q plot was estimating the quantiles spatially derived from the sampled district-level georeferenced explanatory malarial-related covariate coefficients. If one or both of the axes in a Q-Q plot is based on a theoretical distribution with a continuous cumulative distribution function (CDF), all quantiles are uniquely defined and can be obtained by inverting the cdf in SAS/GIS ((http://webhelp.esri.com/arcgisdesktop/). If one or both of the axes in a Q-Q plot is based on a theoretical distribution with a continuous CDF, then all quantiles are uniquely defined and can be obtained by inverting the CDF (Anselin 1995). If a theoretical malarial-related probability distribution with a discontinuous CDF is one of the two distributions being compared, some of the quantiles may not be defined, so an interpolated quantile may be plotted (see Fotheringham 2002). If the Q-Q plot is based on time-series data, there are multiple quantile estimators in use (Cressie 1993). Rules for forming Q-Q plots when quantiles must be estimated or interpolated are called plotting positions (Anselin 1995) Jacob et al. (2011c) constructed a probability distribution with a discontinuous cdf using interpolated quantile for urban malaria mosquito habitat mapping *An. gambiae s.l.* aquatic larval habitats in Kisumu and Malindi, Kenya.

To construct the district-level malarial regression Q-Q plot in this research it was necessary to use an interpolated quantile estimate so that quantiles corresponded to the respected underlying district-level probability distribution. In the model, given the cumulative probability distribution functions F and G, with associated quantile functions F^{-1} and G^{-1}, the inverse function of the cdf in the district-level models represented the quantile function. The Q-Q plot then drew the qth quantile of F against the qth quantile of G for a range of the sampled values of q. Thus, the Q-Q plot was a parametric curve, which was then indexed over [0,1] with the sampled malarial regression-based values in the real plane R^2. Then we employed the formula k / n for $k = 1, ..., n$, as these were the quantiles that the sampling 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. To fix this, we shifted the sampled district-level georeferenced field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficient estimates over, using $(k − 0.5) / n$, and spaced the sampled points evenly in the uniform distribution, using $k / (n + 1)$. By so doing, a probability plot was generated where the quantiles were the rankits, (i.e., the quantile of the expected value of the order statistic of a standard normal distribution). In GIS-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 were then compared the shapes of the distributions while providing a graphical view of how the properties such as georefernced habitat location, scale, and skewness were similar or different in the two distributions.

In terms of heuristics for the quantiles, the comparison district-level distributions we employed the formula $k/(n + 1)$ as in Jacob et al. (2011c) . Although several different formulas have been used or proposed as symmetrical plotting positions for malarial-related explanatory covariate coefficients such formulas 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 district-level model residuals could not generate an accurate depiction of the data as they were highly non-Gaussian. Although the district-level georefernced points plotted in the SAS/GIS-oriented Q-Q plot where non-decreasing when viewed from left to right as expected (see Anselin 1995), the non-normality inherent in the sampled district- level covariate coefficients could not be cartographically defined and hence displayed. If the two distributions being compared are identical, the Q-Q plot follows the 45° line $y = x$. (www.esri.com). Further, if sampled distributions agree after linearly transforming the values in one of the distributions, then the Q-Q plot follows some line, but not necessarily the line $y = x$. In our district-level malarial cluster-based regression model residuals 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 general trend of the Q-Q plot was not steeper than the line $y = x$, and as such, the distribution plotted on the vertical axis was less dispersed than the distribution plotted on the horizontal axis. Generally, Q-Q plots delineating time series-dependent georeferenced malarial–related regression-based covariate coefficients are often arced, or "S" shaped, indicating that one of the distributions is more skewed than the other, or that one of the distributions has heavier tails than the other(Jacob et al. 2011c, Jacob et al. 2009d).

Although the Q-Q plots rendered from the time series dependent predictor covariate coefficients was based on accurately tabulated quantiles, the Q-Q plot was not able to quantize which georeferenced district-level point in the Q-Q plot determined a given quantile. For example, it was not possible to determine the median of the district-level plotted distributions. Commonly Q-Q plots indicate deciles to make determinations such as this possible. The slope and position of the district-level malarial regressors between the quantiles 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 location, and the slope is a measure of scale (Cressie 1993). The distance between medians of relative district location was not reflected in the district-level Q-Q plots. The probability plot error correlation coefficients (i.e., the correlation coefficient between the paired sample quantiles) was thus not quantifiable. Commonly in a malarial-related cluster-based 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). Further, for malarial-related distributions with a single shape parameter, the probability plot correlation coefficient plot provides a method for estimating the shape parameter by computing the correlation coefficient for different sampled spatiotemporal-sampled values of the shape parameter, and thereafter uses the one with the best fit, just as if one were comparing distributions of different types (see Jacob et al. 2011c). In this research the use of Q-Q plots was not able to quantitatively assess nor compare the district-level distribution of the samples to the standard normal distribution [i.e., $N(0,1)$], as in a normal probability plot.

2.2 Autocorelation model: Initially, a misspecification perspective for performing a spatial autocorrelation estimation analysis using the district-level malarial indicators. The model was generated using the $y = X\beta + \varepsilon$ (i.e. regression equation) assuming the sampled data had autocorrelation disturbances. The model was also assumed that this data could be decomposed into a white-noise component, ε, and a set of unspecified and/or misspecified

$$y = XB + \underbrace{E\gamma + \varepsilon}_{=\varepsilon^*}$$

sub-models that had the structure . White noise in a malaria-based 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. 2008d). In this research, the misspecification term was ε_{ij}. Quantification of the topographic patterns generated from the distribution of the sampled district-level georeferenced district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented explanatory covariate coefficients was required to describe independent key dimensions of the underlying spatial processes in the sampled data and for defining a spatial pattern in the misspecification term.

A spatial autoregressive (SAR) model was then generated that used an district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented exploratory explanatory variable **Y**, as a function of nearby sampled district–level covariate coefficients. In this research, **Y** had an indicator value I (i.e., an autoregressive response) and/or the residuals of **Y** which were values of nearby sampled **Y** residuals (i.e., an SAR or spatial error specification). For spatiotemporal modeling malarial-related regression parameter estimators, the SAR model furnishes an alternative specification that frequently is written in terms of matrix **W** (see Jacob et al. 2008 b,c)**.** As such, its spatial covariance was 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 denoted 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., $\mathbf{W}^{T}\mathbf{W}$) – adjacent sampled 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 $0 \leq \rho \leq 1$.

In this research district-level distance measurements were defined in terms of an *n*-by-*n* geographic weights matrix, **C**, whose c_{ij} values were; 1 if the sampled district locations *i* and *j* were deemed nearby, and 0 otherwise. Adjusting this matrix by dividing each row entry by its row sum gave **C1**, where **1** was an *n*-by-**1** vector of ones which converted this matrix to matrix **W**. The resulting SAR model specification, with no sampled georeferenced explanatory district-level field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients present (i.e., the pure spatial autoregression specification), took on the following form: $\mathbf{Y} = \mu(1-\rho)\mathbf{1} + \rho\mathbf{W}\mathbf{Y} + \boldsymbol{\varepsilon}$, where μ was the scalar conditional mean of **Y**, and $\boldsymbol{\varepsilon}$ was an *n*-by-1 error vector whose parameters were statistically i.d.d. normally random variates. The spatial covariance matrix for analyzing the sampled district-level georeferenced covariate coefficients was then calculated using $E[(\mathbf{Y}-\mu\mathbf{1})'(\mathbf{Y}-\mu\mathbf{1})] = \boldsymbol{\Sigma} = [(\mathbf{I}-\rho\mathbf{W}')(\mathbf{I}-\rho\mathbf{W})]^{-1}\sigma^2$, where E ($\bullet$) denoted the calculus of expectations, **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 spatiotemporal-sampled field and remote district-level georeferenced parameter estimators of the malarial regression model, *c* was a constant and 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 (see Griffith 2003) which in this research involved the matrix (**I** - ρ **C**), where **I** was an *n*-by-*n* identity matrix. In an geo-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 in this research was to bring the spatially unlagged endogenous variable, y, exclusively on the left-hand side of the district-level malarial regression equation in order to decorrelate the sampled georeferenced explanatory predictor covariate error coefficients. In this research, 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 spatial process in the sampled ecological datasets was stationary (see Bivand, 1984). The simultaneous autoregressive error model was then rewritten as $y - \rho V y = X\beta - \rho V X\beta + \varepsilon$. Substituting this transformation rendered:

$$y = (I-\rho V)^{-1}[X\beta - \rho V(X\beta) + \varepsilon],$$

$$y = \sum_{k=0}^{\infty} \rho^k V^k (X\beta - \rho V X\beta + \varepsilon),$$

$$y = \sum_{k=0}^{\infty} \rho^k V^k X\beta - \sum_{k=0}^{\infty} \rho^{k+1} V^{k+1}(X\beta) + \sum_{k=0}^{\infty} \rho^k V^k \varepsilon,$$

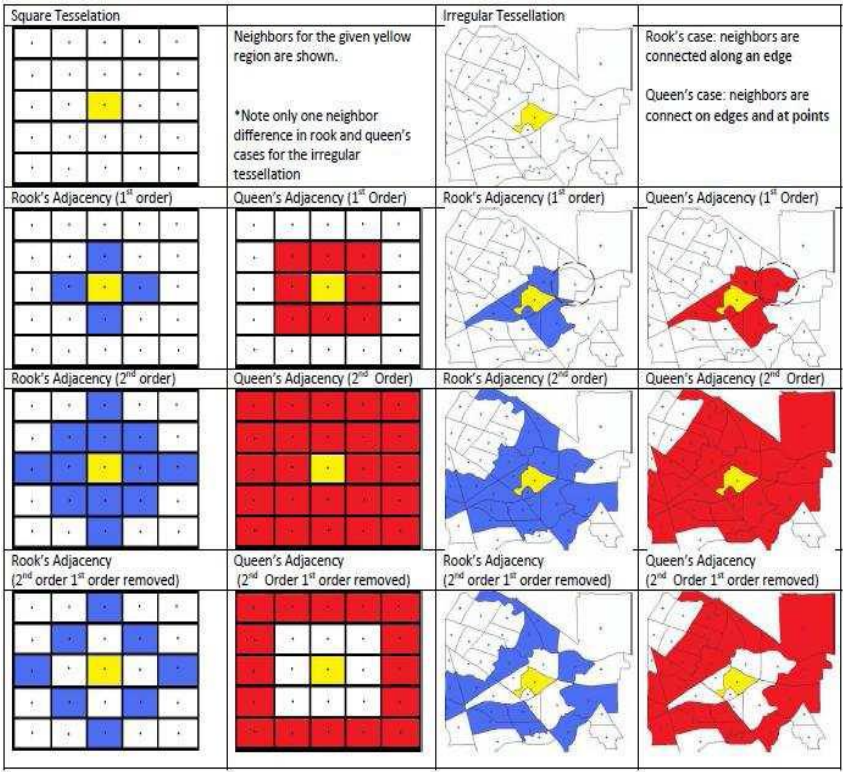
International Journal of Geographic Information System
Vol. 1, No. 1, November 2013, PP: 01 - 143
Available Online at http://acascipub.com/Journals.php

$$y = X\beta + \underbrace{\sum_{k=1}^{\infty}\rho^k V^k X\beta - \sum_{k=1}^{\infty}\rho^k V^k (X\beta)}_{=0} + \sum_{k=0}^{\infty}\rho^k V^k \varepsilon \quad ,$$

$$y = X\beta + \underbrace{\sum_{k=1}^{\infty}\rho^k V^k \varepsilon}_{misspecification\ term} + \varepsilon$$

As a part of deciphering the spatial surface of a sampled district-level malarial variable, it was important to consider how the sampled data feature attributes were connected. This was done by specifying the "case" and "order" of the connectivity. Generally speaking, regions can be connected with other neighboring regions either along an edge or at a shared point. Regions connected along edges and not at single points are referred to as having the rook's case adjacency; regions connected along edges as well as at single points are referred to as having the queen's case adjacency. We illustrated these relationships for both a regular square tessellation and an irregular tessellation per sampled district. For the square tessellation, the distinction between a point and an edge was easily visualized, while for an irregular tessellation the distinction was less important since it happened less often that polygons which met only at one point. As can be seen in the district's sampled there was only one region difference between the rook and queen specification for the irregular surface, while for the regular square tessellation there is always multiple neighbor difference

Figure 4: Rook's case (blue) and Queen's case (red) adjacency and order of adjacency used for the identifying clustering in the malarial –related estimators



International Journal of Geographic Information System
Vol. 1, No. 1, November 2013, PP: 01 - 143
Available Online at http://acascipub.com/Journals.php

We noticed that the misspecification term $\sum \rho^k V^k (k = 1, \ldots, \infty)$ in the district-level model remained uncorrelated with the exogenous variable, X, as the standard OLS assumption of the disturbances, ε, were uncorrelated with the predictor covariate error coefficients generated from the district-level parameter estimators (b). The spatial lag model on the other hand, 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 $y = X\beta + \underbrace{\sum_{k=1}^{\infty} \rho^k V^k (X\beta + \varepsilon)}_{misspecification\ term} + \varepsilon$. The misspecification term $\sum \rho^k V^k (X\beta + \varepsilon)(k = 1, \ldots, \infty)$ included the exogenous variables X. Consequently, the exogenous variables were correlated with the misspecification term. Under this condition, standard OLS results for the basic district-level malaria linear regression model $y = X\beta + \varepsilon$, generated from the sampled georeferenced predictor error covariate coefficients, provided biased estimates $\hat{\beta}$ of the underlying regression based district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented exploratory observational exploratory parameters estimators.

An autoregressive integrated moving average (ARIMA) model was then constructed in SAS/GIS based on generalization of an 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).In this research an ARIMA model was fitted to the time series dependent district-level data to predict future points in the series. Employing a time series of district-level field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficient data X_t where *t* was an integer index and the X_t a was the sampled predictor values, then an ARMA(*p*,*q*) model was 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 was the lag operator, the α_i were the parameter estimators of the autoregressive part of the model, the θ_i were the parameters of the moving average part and ε_t were error terms. The error terms ε_t are generally assumed to be i.d.d. variables in a robust malarial cluster-based regression model using a normal distribution with zero mean (Jacob et al. 2009d). We assume now the polynomial $\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right)$ had a unitary root of multiplicity *d*. In this research 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, and q*) process then expressed this polynomial factorization property which was given by the model expression $\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$ and thus was 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).

Initially, we constructed a time series district-level hyperendemic transmission-oriented regression model with AR (p) noise. Our time series {[Y.sub.t]} of interest followed a linear regression model of the form [Y.sub.t] = [x'.sub.t][beta] + [N.sub.t], t = 1,...,T, (1) where [x.sub.t] = ([x.sub.t1],...,[x.sub.tr])' was a r-dimensional vector of deterministic or stochastic district-level field and remote-sampled malarial regressors and [beta] were equal to ([[beta].sub.1],...,[[beta].sub.r])' which in this research represented the vector of unknown parameters to be estimated in the risk model. Initially, for quantitating the noise series in the endemic transmission oriented model we assumed a stationary process following an AR(p) model, [N.sub.t] = [[phi].sub.1][N.sub.t-1] +

[[phi].sub.2][N.sub.t-2] +...+[[phi].sub.p][N.sub.t-p] + [[epsilon].sub.t], (2) where {[[epsilon].sub.t]} was the white noise process with mean 0 and variance [[[sigma].sup.2].sub.[epsilon]]. The AR(p) district-level malarial hyperendemic transmission-oriented model was then written as [phi](B)[N.sub.t] = [[epsilon].sub.t], where [phi](B) = 1 - [[phi].sub.1]B-...-[[phi].sub.p][B.sup.p]. We noticed that in our model for {[N.sub.t]} to be stationary, all roots of [phi](B) = and t had be greater than 1 in absolute value. The autocovariance function of the hyperendemic district-level transmission-oriented model -[gamma](l) = cov([N.sub.t], [N.sub.t+l]) of {[N.sub.t]} which then satisfied the difference equation [gamma](l) = [[phi].sub.1][gamma](l - 1) + [[phi].sub.2][gamma](l - 2) +...+ [[phi].sub.p] [gamma](l - p), l [greater than or equal to] 1, with [gamma](0) = [[[sigma].sup.2].sub.[epsilon]]/[delta], where [delta] = 1 - [[phi].sub.1][[rho].sub.1] - [[phi].sub.2][[rho].sub.2]-...-[[phi].sub.p][[rho].sub.p] and [rho]l = [gamma](l)/[gamma](0). Given a sample of T observations, if Y =([Y.sub.1],...,[Y.sub.T])' and N = ([N.sub.1],...,[N.sub.T])' then T x 1 data and the noise vectors can be efficiently quantitated by letting [epsilon] = ([[[epsilon].sup.*].sub.1],..,[epsilon].sup.*].sub.p],in the hyperendemic district level transmission-oriented model then [[epsilon].sub.p+1],...,[[epsilon].sub.T])'.intentionally would bias estimates in the time series regression models employing REML.

In the REML approach a particular form of maximum likelihood estimation was employed which did not utilize base estimates for determining the maximum likelihood fit of all the information, but instead used a likelihood function calculated from a transformed set of spatiotemporal malarial related district-level field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients estimators, so that nuisance parameters had no effects. The case of variance component estimation, the original data set was replaced by a set of contrasts calculated from the sampled data, and the likelihood function was then calculated from the probability distribution of these contrasts, according to the hyperendemic district-level malarial transmission-oriented risk model for the complete data set. In particular, in this research the REML was used as a method for fitting linear mixed models. In contrast to the earlier MLE, REML can produce unbiased estimates of variance and covariance parameters (Cressie 1993). The idea underlying our REML estimation was put forward by (Bartlett 1937). The first description of the approach applied to estimating components of variance in unbalanced data was by Patterson (1971), although they did not use the term REML. A review of the early literature was given by Harville (1977) REML. Fortunately, REML estimation is available in a number of general-purpose statistical software packages, including Genstat (the REML directive), SAS (the MIXED procedure), SPSS (the MIXED command), STATA (the xtmixed command), and R (the lme4 and older nlme packages), as well as in more specialist packages such as MLwiN, HLM, ASReml, Statistical Parametric Mapping and CropStat

We then defined the T x r matrix X = [[x.sub.1],...,[x.sub.T]]', and assumed that X is of full rank r a which satisfies the Grenander conditions. These conditions on the regressors under which the OLS estimator are consistent but they are weaker than the assumption on the regressor X that $\lim_{n->infinity}$(X'X)/n is a fixed positive definite matrix, (see, e.g., Anderson 1971). Briefly, with [d.sub.ij](h) = [[[sigma].sup.T-h].sub.t=1] [x.sub.i,t+h][x.sub.jt],h = 0, 1, ..., the Grenander conditions assumed that [d.sub.ii](0) [right arrow] [infinity], [[x.sup.2].sub.i,T+1]/[d.sub.ii](0) [right arrow] 0, and lim [d.sub.ij](h)/[square root][d.sub.ii](0)[d.sub.jj](0) [equivalent] [r.sub.ij](h) exists as T [right arrow] [infinity], for i,j = 1,...,r, and h = 0, [+ or -]1, [+ or -]2,..., and thus we assumed the matrix of the limits of elements {[r.sub.ij](0)} was positive definite (nonsingular). Then the model was expressed in matrix form as Y = X[beta] + N, P'N = [epsilon], where the T x T transformation matrix P' was the lower triangular with its first p diagonal elements equal to [[delta].sup.1/2], [([delta]/[[delta].sub.1]).sup.1/2],...,[([delta]/[[delta].sub.p-1]). sup.1/2]. By so doing, its remaining diagonal elements was equivalent to 1, . Additionally elements in the (i,j)th position were equal to -[[[phi].sup.*].sub.i-j,i-1] for j = 1 ,..., i-1 and i = 2 ,..., . Further p, in the seasonal malarial district level endemic transmission-oriented risk model was equal to -[[phi].sub.i-j] for j = i-p ,..., i-1 when i was greater than p.

In the model the first p rows of P', the elements [[[phi].sup.*].sub.ik] = [[phi].sub.ik] [([delta]/[[delta].sub.k]).sup.1/2], i = 1 ,..., k, where [[phi].sub.1k],...,[[phi].sub.kk], for k = 1 ,..., p-1,which were solutions for coefficients [[phi].sub.1],...,[[phi].sub.k] w .Thereafter, the system of the first k ARMA equations iwith p set was equal to k, and [[delta].sub.k] = 1 - [[phi].sub.1k][[rho].sub.1] - [[phi].sub.2k][[rho].sub.2] - phi].sub.kk][[rho].sub.k].Thereafter , [[phi].sub.11] = [gamma](1)/[gamma](0) = [[rho].sub.1] and [[delta].sub.1] = 1 - [[rho].sup.2].sub.1]as cov([epsilon]) = [[[sigma].sup.2].sub.[epsilon]]I which followed the covariance matrix of N which in this research was cov(N) = cov([P'.sup.-1][epsilon]) = [[[sigma].sup.2].sub.[epsilon]][P'.sup.-1][P.sup.-1]=[[[sigma].sup.2].sub.[epsilon]]V,where [V.sup.-1] = PP'

Results

By employing the NDVI we were able to obtain a derivative of surface reflectance of the vegetated canopied explanatory covariates with respect to QuickBird (V)and (NIR) wavelengths. Initially, the Band Math function of ENVI 4.8 was used to calculate a NDVI. Therefater we color balanced the natural color QuickBird imagery collected over the Ugandan epidemiological study site. We then exported these layers as a GeoTIFF from ENVI, to ArcMap to complete the analysis of the riverine larval habitat vegetation parameter estimators extracted with a NDVI calculation. Our NDVI was calculated as : NDVI = (Band 4 – Band 3) / (Band 4 + Band 3).Once this value was calculated for every larval habitat pixel, we created a colorized image where all the healthy vegetation in the scene was red (typically a value of ~0.35 for NDVI) NDVI has a range of possible values from -1.0 to 1.0 (Tucker 1991).Next, we performed a filtering exercise and displayed the NDVI in ArcMap for determining the lowest NDVI value that was associated to all the green vegetation parameters at the study sites. This value could have easily been adjusted upward or downward as the fidelity of the analysis required. To decide this value, we added color-balanced natural color imagery as the top layer and then added the NDVI layer below. By clicking on multiple pixels with the Identify Tool, all district-level vegetation canopy-related explanatory covariate coefficients were not quantitated.

For the next step in the NDVI analysis, we created sequential classes of NDVI values and then color coded them accordingly. To complete this step, we navigated to the Symbology Tab under the Layer Properties of the NDVI TIFF file. We switched the symbology type to Classified and calculated histograms. Pressing the Classify button, we then set the Exclusion values to -1.0 to 0.09. This procedure allowed ArcMap to filter out all the vegetated canopied district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented explanatory covariate ceofficent values for classifying the data into clusters of similar values. We then classified by Natural Breaks (i.e.,jenks) with 3 classes to show a schema with less ability to separate riverine larval habitat unhealthy vegetation from the most vigorous. Finally, we created a copy of this NDVI layer following the same steps above but this time created a classification schema with 7 classes. For both of these schemas, we chose the same color ramp whereby unhealthy vegetation was in red; the healthiest vegetation in bright purple; and vegetation with intermediate health in yellow.

Initially, we constructed a Poisson regression model using the spatiotemporal-sampled district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented exploratory observational exploratory observational predictor measurement values. Our model was generalized by introducing an unobserved heterogeneity term for each sampled district-level observation i. The weights were then assumed to differ randomly in a manner that was not fully accounted for by the other spatiotemporal-sampled covariate coefficients. In this research this linear district-level process was formulated as $E(y_i|x_i,\tau_i) = \mu_i\tau_i = e^{x_i\beta+\varepsilon_i}$ 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 was Poisson with conditional mean and conditional variance$\mu_i\tau_i$: $f(y_i|x_i,\tau_i) = \frac{\exp(-\mu_i\tau_i)(\mu_i\tau_i)^{y_i}}{y_i!}$. We then let $g(\tau_i)$ be the pdf of τ_i. Then, the distribution $f(x_i|y_i)$ was no

longer conditional on τ_i. Instead it was obtained by integrating $f(x_i|y_i,\tau_i)$ with respect to τ_i: $f(y_i|x_i) = \int_0^\infty f(y_i|x_i,\tau_i)g(\tau_i)d\tau_i$. We found that an analytical solution to this integral existed in the district-level linear model when τ_i was assumed to follow a non-homogenous gamma distribution.

Our count values uncertainty model also assumed that y_i, was the vector of the explanatory uncertainty covariate coefficients x_i, which was also independently Poisson distributed with $P(Y_i = y_i|\mathbf{x}_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \quad y_i = 0,1,2,\ldots$ and the mean parameter — that is, the mean number of district-level events per 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 then and the coefficients for the k regressors were $\beta_1, \ldots, \beta_k$. Taking the exponential of $x_i'\beta$ ensured that the mean parameter μ_i was nonnegative. Thereafter, the conditional mean was provided $E(y_i|x_i) = \mu_i = \exp(x_i'\beta)$. The district-level predictive geo-autoregressive parameter estimators were then evaluated using $\ln[E(y_i|x_i)] = \ln(\mu_i) = x_i'\beta$. Note, that the conditional variance of the count random variable was equal to the conditional mean (i.e., equidispersion) in our model [e.g., $(y_i|x_i) = E(y_i|x_i) = \mu_i$]. In a log-linear model the logarithm of the conditional mean is linear (Haight 1967). The marginal effect of any district-level regressors in the model was then provided by $\frac{\partial E(y_i|x_i)}{\partial x_{ji}} = \exp(x_i'\beta)\beta_j = E(y_i|x_i)\beta_j$. Thus, a one-unit change in the jth regressor in the regression model led to a proportional change in the conditional mean $E(y_i|x_i)$ of β_j.

Given the spatiotemporal-sampled dataset of district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented exploratory observational exploratory parameter estimators (i.e;, θ) and an input vector *x*, the mean of the predicted Poisson distribution was then provided by $E(Y|x) = e^{\theta' x}$ and thus, the Poisson distribution's pmf of the sampled district-level explanatory uncertainty covariate coefficients was given by $p(y|x;\theta) = \frac{e^{y\theta' x} e^{-e^{\theta' x}}}{y!}$. The pmf in a spatiotemporal predictive hierarchical spatial malaria regression-based model is often the primary means of defining a discrete probability distribution, and such functions exist for either scalar or multivariate random variables, given that the distribution is discrete (Jacob et al. 2008b). Since the sampled data consisted of *m* vectors $x_i \in R^{n+1}, i = 1, \ldots, m$, along with a set of *m* values $y_1, \ldots, y_m \in R$ then, for the district-level parameter estimators θ, the probability of attaining this particular set of data was provided by $p(y_1, \ldots, y_m|x_1, \ldots, x_m; \theta) = \prod_{i=1}^{m} \frac{e^{y_i(\theta' x_i)} e^{-e^{\theta' x_i}}}{y_i!}$. By the method of maximum likelihood, we found the set of θ that made this probability as large as possible. To do this, the equation was first rewritten as a likelihood function in terms of θ: $L(\theta|X,Y) = \prod_{i=1}^{m} \frac{e^{y_i(\theta' x_i)} e^{-e^{\theta' x_i}}}{y_i!}$ Note, that the expression on the right hand side in our model had not actually changed.

Next, we used the log-likelihood: $\ell(\theta|X,Y) = \log L(\theta|X,Y) = \sum_{i=1}^{m} \left(y_i(\theta' x_i) - e^{\theta' x_i} - \log(y_i!)\right)$. Notice that the parameters θ only appeared in the first two terms of each term in the summation. Therefore, given that we were only

interested in finding the best value for θ we dropped the $y_i!$ and simply wrote $\ell(\theta|X,Y) = \sum_{i=1}^{m} \left(y_i(\theta' x_i) - e^{\theta' x_i}\right)$. To find a maximum, we solved an equation $\frac{\partial \ell(\theta|X,Y)}{\partial \theta} = 0$ which had no closed-form solution. However, the negative log-likelihood (LL)[i.e.,, $-\ell(\theta|X,Y)$] was a convex function, and so standard convex optimization and gradient descent techniques was applied to find the optimal value of θ .

We found that given a Poisson process, was given by the limit of a binomial distribution $P_p(n \mid N) = \frac{N!}{n!\,(N-n)!}\, p^n (1-p)^{N-n}$. in the district-level district-level malaria-related hyperendemic transmission oriented model thus; $P_{\nu/N}(n \mid N) = \frac{N!}{n!\,(N-n)!} \left(\frac{\nu}{N}\right)^n \left(1 - \frac{\nu}{N}\right)^{N-n}$, Letting the sample size N become large, the distribution then approached P when $\lim_{N\to\infty} P_p(n \mid N)$, $\lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{n!} \frac{\nu^n}{N^n} \left(1 - \frac{\nu}{N}\right)^N \left(1 - \frac{\nu}{N}\right)^{-n}$, $\lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{N^n} \frac{\nu^n}{n!} \left(1 - \frac{\nu}{N}\right)^N \left(1 - \frac{\nu}{N}\right)^{-n}$, $1 \cdot \frac{\nu^n}{n!} \cdot e^{-\nu} \cdot 1$ and $\frac{\nu^n e^{-\nu}}{n!}$, Note, that the sample size N had completely dropped out of the probability function, which in this research had the same functional form for all values of ν.

As expected, the Poisson distribution was normalized so that the sum of probabilities equaled 1, since $\sum_{n=0}^{\infty} P_\nu(n) = e^{-\nu} \sum_{n=0}^{\infty} \frac{\nu^n}{n!} = e^{-\nu}\, e^{\nu} = 1.$ The ratio of probabilities was then given 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 reached a maximum when $\frac{d P_\nu(n)}{d n} = \frac{e^{-\nu}\, n\,(\gamma - H_n + \ln \nu)}{n!} = 0,$ where γ was the Euler-Mascheroni constant and H_n was a harmonic number, leading to the transcendental equation $\gamma - H_n + \ln \nu = 0$, which in this research could not be solved exactly for n. The Euler–Mascheroni constant is a mathematical constant recurring in analysis and number theory (Hosmer and Lemeshow 2002).

We noticed that the moment-generating function of the Poisson distribution was given by $M=e-vevet=ev(et-1)$, $M = \nu e^t e^{\nu(e^t-1)}$ and $M=(vet)2ev(et-1)+vetev(et-1)$, when $R = \nu(e^t - 1)$, $R' = \nu e^t$ so $R = R'(0) = \nu$ The raw moments was then computed directly by summation, which yielded an unexpected connection with the exponential polynomial $\Phi_n(x)$ and Stirling numbers of the second kind, $\phi_n(x) = \sum_{k=0}^{\infty} \frac{e^{-x} x^k}{k!} k^n = \sum_{k=1}^{n} x^k S(n,k)$ which in this research was the Dobiński's formula. The expression was then given by Dobinski's formula as the *n*th moment of the Poisson distribution had an expected value 1. This then lead to $v1+v$,(1+3v+v2) and $\nu(1 + 7\nu + 6\nu^2 + \nu^3)$. The central moments was thereafter computed as $\nu(1 + 3\nu)$ so the mean, variance, skewness, and kurtosis were $\frac{\mu_1}{\sigma^1} = \frac{\nu}{\nu^{3/2}} = \nu^{-1/2}$, $\frac{\mu_4}{\sigma^4} - 3 = \frac{\nu(1+3\nu)}{\nu^2} - 3$ and $\frac{\nu+3\nu^2-3\nu^2}{\nu^2} = \nu^{-1}$ respectively.

The characteristic function for the Poisson distribution in the district-level Poisson geopredictive model was then evaluated using Φt=ev(eit−1) where the cumulant-generating function was $K(h) = v(e^h - 1) = v(h + \frac{1}{2}h^2 + \frac{1}{2!}h^3 + \cdots)$,so $k^r = v$. The mean deviation of the Poisson distribution in our mode was of the Ugandan district level prevalence rates was then given by.$MD=2e-vvv+1v!$ The Poisson distribution was then expressed in terms of $\lambda = \frac{v}{x}$, and the rate of changes was equal to $P_\gamma(n) = \frac{(\lambda x)^n e^{-\lambda x}}{n}$. The moment-generating function of the Poisson distribution generated from the sampled district-level explanatory variables was then provided by $M(t)\equiv ev1+v2(et-1)$. Given a random variable x and a probability distribution function $P(x)$, if there existed an $h > 0$ such that $M(t) = (e^{tx})$ for $|t| < h$, where (y)denotes the expectation value of y, then $M(t)$ is called the moment-generating function (Papoulis 1984). Commonly, for a continuous distribution in a robust malaria-related linear regression model$\int_{-\infty}^{\infty} e^{tx} P(x)\,dx$,$\int_{-\infty}^{\infty}\left(1 + tx + \frac{1}{2!}t^2x^2 + \ldots\right)P(x)\,dx$ and $1 + tm_1' + \frac{1}{2!}t^2 m_2' + \cdots$, where $m_r'$is the rth raw moment (Jacob et al. 2009d).

In this research for independent X and Y, the moment-generating function in our spatiotemporal malarial risk model satisfied $M_{x+y}(t) = (e^{t(x+y)}), (e^{tx}e^{ty}), (e^{tx})(e^{ty})$ and $M_x(t)M_y(t)$. If the independent variables $x_1, x_,\ldots, x_N$ have Poisson distributions with parameters $\mu_1, \mu_2, \ldots, \mu_N$, then $X = \sum_{j=1}^{N} \chi_j$ has a Poisson distribution with parameter $\mu = \sum_{j=1}^{N} \mu_j$. (Haight 1967). In our linearized data distribution this was evident since the cumulant-generating function was $K_i(h) = \mu_j(e^h - 1)$ and $K \equiv \sum_j K_j(h) = (e^h - 1)\sum_j \mu_j = \mu(e^h - 1).$

We then tested for overdispersion with a likelihood ratio test based on Poisson and negative binomial distributions. This test quantitated the equality of the mean and the variance imposed by the Poisson distribution against the alternative that the variance exceeded the mean in the district-level linear malarial model. For the negative binomial distribution, the variance in our spatiotemporal model was equal to the mean + k mean^2, when $k \geq 0$and the negative binomial distribution reduced to Poisson when k=0. It is important to remember that 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 then 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 df. To test the null hypothesis further at the significance level α, we then used the critical value of chi-sq distribution corresponding to significance level 2α, that is we rejected H_0 if LR statistic $> 2\ (1 - 2\alpha, 1\ df)$.

Next, we assumed that the district-level malarial model was based on the log of the mean, μ, which in this research was a linear function of independent variables, log(μ) = intercept + b1*X1 +b2*X2 + ...+ b3*Xm. This log-transformation implied that μ was the exponential function of independent variables where μ = exp(intercept + b1*X1 +b2*X2 + ...+ b3*Xm). Instead of assuming as before that the distribution of the district level parameter estimators [i.e.,Y], and the number of sampling occurrences was Poisson, we now assumed that Y had a negative

binomial distribution. That meant, in particular, relaxing 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 $k \geq 0$ was a dispersion parameter. The maximum likelihood method was then used to estimate k as well as the parameter estimators of the model for log(μ). 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. Instead of **dist = poisson**, **dist = nb** was used.

Results from both a Poisson and a negative binomial (i.e., a Poisson random variable with a gamma distrusted mean) in SAS/GIS revealed that the district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented exploratory observational predictors were highly significant, but furnished virtually no predictive power. In other words, the sizes of the population denominators were not sufficient to result in statistically significant relationships, while the detected relationships were inconsequential.

Inclusion of indicator variables denoting the time sequence and the district location spatial structure was then articulated with Thiessen polygons, (see Figure 1a) which also failed to reveal meaningful predictors. Further, Figure 1b implied presence of additional noise in the data for 2010, attributable to an expansion of the districts; thus, for this data analysis we retained the original 80 districts for space-time consistency.

Under the regression model and the normality assumption [epsilon] [sim] N(0,[[[sigma].sup.2].sub.[epsilon]]I), the log-likelihood function was l([beta],[phi],[[[sigma].sup.2].sub.[epsilon]]) = - T/2 log(2[pi]) - T/2 log([[[sigma].sup.2].sub.[epsilon]]) - 1/2 log [absolute val. of V] - 1/2[[[sigma].sup.2].sub.[epsilon]] S([beta],[phi]), where [phi] = ([[phi].sub.1],...,[[phi].sub.p])' and S([beta],[phi]) = (Y-X[beta])' [V.sup.-1] (Y-X[beta]) was the sum of squares function. Note that [absolute val. of V] = [absolute val. of [V.sub.p]], where [[[sigma].sup.2].sub.[epsilon]][V.sub.p] was the covariance matrix of p consecutive values from the AR(p). For example, in the AR(2) Ugandan malarial model, p = 2, and so [absolute val. of V] = [absolute val. of [V.sub.2]] = (1-[[[rho].sup.2].sub.1])/[[delta].sup.2].

To derive the likelihood equations, we employed S([beta],[phi]) which in this research was expressed as a quadratic function of the parameter [phi] as in Box, Jenkins, and Reinsel 1994, p. 298). We then used the following two equivalent expressions for S([beta],[phi]): S([beta],[phi]) = Y'[V.sup.-1]Y - 2[beta]'X'[V.sup.-1]Y + [beta]'X'[V.sup.-1]X[beta] and S([beta],[phi]) = [[[sigma].sup.T].sub.t=p+1] [[[epsilon].sup.2].sub.t] + [[[epsilon].sup.*2].sub.1] + ... + [[[epsilon].sup.*2].sub.p] = N'PP'N = [C.sub.00] - 2[phi]'[c.sub.p] + [phi]'[C.sub.p][phi], where [C.sub.00] = [[[sigma].sup.T].sub.t=1] [[N.sup.2].sub.t], [c.sub.p] = ([C.sub.10], [C.sub.20],...,[C.sub.p0])', [C.sub.p] when the p x p symmetric matrix with (i,j)th element [C.sub.ij], and the elements [C.sub.ij] were "symmetric" sums of squares and lagged cross-products of the [N.sub.t] ([N.sub.t] = [Y.sub.t] - [x'.sub.t][beta]),which in this research was provided parsimoniously by [C.sub.ij] = [[[sigma].sup.T-j].sub.t=i+1] [N.sub.t][N.sub.t-i+j] = [C.sub.ji], with T - i - j terms in the sum. The (i,j)th element of [C.sub.p] haf expected value E([C.sub.ij]) = (T - i - j)[gamma](i - j)i,j = 1,..., p.

From the foregoing expression, the first partial derivatives of S([beta],[phi]) were [partial]S/[partial][beta] = -2X'[V.sup.-1]Y + 2X'[V.sup.-1]X[beta] and [partial]S/[partial][phi] = 2([C.sub.p][phi]-[c.sub.p]), with E[[partial]S/[partial][beta]] = 0 and E[[partial]S/[partial][phi]] = 2{E([C.sub.p])[phi] - E([c.sub.p])}. We noticed that the *i*th element of E([C.sub.p])[phi] - E([c.sub.p]), [[[sigma].sup.p].sub.j=1] E([C.sub.ij])[[phi].sub.j] - E([C.sub.i0]), was equal to [[[sigma].sup.p].sub.j=1] (T - i - j)[gamma](i - j)[[phi].sub.j] - (T - i)[gamma](i) = -[[[sigma].sup.p].sub.j=1]j[gamma](i - j)[[phi].sub.j], using the autocorrelation equations (3). The likelihood equations were then [partial]l/[partial][phi] = - 1/2 [partial]/[partial][phi] log [absolute val. of V] - 1/[[[sigma].sup.2].sub.[epsilon]] ([C.sub.p][phi]-[c.sub.p]) = 0, (5) [partial]l/[partial][beta] = 1/[[[sigma].sup.2].sub.[epsilon]] (X'[V.sup.-1]Y - X'[V.sup.-1]X[beta]) = 0, (6)

The transformation from spherical coordinates (r, θ, φ) to Cartesian coordinates (x_1, x_2, x_3), in the autocorrelation model was then given by the function $F: \mathbf{R}^+ \times [0,\pi] \times [0,2\pi) \rightarrow \mathbf{R}^3$ with components: $x_1 = r \sin\theta \cos\phi, x_2 = r \sin\theta \sin\phi, x_3 = r \cos\theta.$ The Jacobian matrix for this coordinate change was $J_F(r,\theta,\phi) = \begin{bmatrix} \frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} & \frac{\partial x_1}{\partial \phi} \\ \frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta} & \frac{\partial x_2}{\partial \phi} \\ \frac{\partial x_3}{\partial r} & \frac{\partial x_3}{\partial \theta} & \frac{\partial x_3}{\partial \phi} \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi \\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi \\ \cos\theta & -r\sin\theta & 0 \end{bmatrix}.$ The determinant was $r^2 \sin\theta$.

An autoregressive model specification was then constructed in SAS/GIS to describe the autoregressive variance uncertainty estimates in the sampled data. A spatial filter (SF) model specification was also used to describe both Gaussian and Poisson random variables. The resulting SAR model specification took on the following form (3.1) where μ was the scalar conditional mean of Y, and ε was an n-by-1 error vector whose elements were statistically independent and identically distributed (i.i.d.) normally random variates. The spatial covariance matrix for equation (3.1)regressed the sampled district level covariate coefficients using E [(Y - μl)' (Y - μl)] = Σ = [(I - ρ W')(I - ρ W)]-1σ2, where E (●) denoted the calculus of expectations, I was the n-by-n identity matrix denoting the matrix transpose operation, and σ2 was the error variance.

In this research, two different spatial autoregressive parameters appeared in the spatial covariance matrix which for our Ugandan malarial-related SAR model specification became: (3.2)where the diagonal matrix of autoregressive parameters, <ρ >diag, contained two sampled parameters: ρ+ for those district-level sampled covariate coefficient pairs displaying positive spatial dependency, and ρ. for those pairs displaying negative spatial dependency. For example, by letting σ2 = 1 and employing a 2-by-2 regular square tessellation, for the vector , enabled positing a positive relationship between the sampled covariates, y1 and y2, whereby a negative relationship between covariates, y3 and y4, existed and, no relationship between covariates y1 and y3 and between y2 and y4 was noted. This covariance specification then yielded: (3.3)where I+ was a binary 0-1 indicator variable which denoted those district-level covariate coefficients displaying positive spatial dependency, and I- was a binary 0-1 indicator variable denoting those sampled habitats displaying negative spatial dependency, using I+ + I- = 1. Expressing the preceding 2-by-2 example in terms of equation yielded: We noticed that if either ρ+ = 0 (and hence I+ = 0 and I- = I) or ρ- = 0 (and hence I- = 0 and I+ = I), then equation (3.3) reduced to equation (3.1). This indicator variable classification was made in accordance with the quadrants of the corresponding Moran scatterplot generated using the sampled district level explanatory covariate coefficients.

To identify district level spatial clusters, Thiessen polygon surface partitioning were also generated to construct geographic neighbor matrices, which also were used in the spatial autocorrelation analysis. Entries in matrix were 1, if two sampled district-level georefernced endemic transmission-oriented explanatory covariate coefficients shared a common Thiessen polygon boundary and 0, otherwise. Next, the linkage structure for each surface was edited to remove unlikely geographic neighbors to identify pairs of sampled covariate coefficients sharing a common Thiessen polygon boundary. Attention was restricted to those district-level map patterns associated with at least a minimum level of spatial autocorrelation, which, for implementation purposes, was defined by |MCj/MCmax| > 0.25, where MCj denoted the jth value and MCmax, the maximum value of MC. This threshold value allowed two candidate sets of eigenvectors to be considered for substantial positive and substantial negative spatial autocorrelation respectively. Because larval/pupal counts were being analysed, a Poisson spatial filter model specification was employed in this research

The model specification was written as follows: where μi was the expected mean district location i, μ was an n-by-1 vector of expected explanatory covariate coefficient counts, LN denoted the natural logarithm (i.e., the generalized linearized model link function), α was an intercept term, and η was the negative binomial dispersion parameter. This

log-linear equation had no error term; rather, estimation was executed assuming a negative binomial random variable.

In terms of the eigenfunctions of the spatial weighted matrix ,rhe upper and lower bounds for a spatial matrix generated using Moran's indices (I) were given by λmax(n/1TW1) and λmin(n/1TW1) where λmax and λmin which in this research was the extreme eigenvalues of Ω = HWH [23]. Hence, in this research, the eigenvectors of Ω were vectors with unit norm maximizing Moran's I. The eigenvalues of this matrix were equal to Moran's I coefficients of spatial autocorrelation post-multiplied by a constant. Eigenvectors associated with high positive (or negative) eigenvalues have high positive (or negative) autocorrelation (Griffith 2003).

The diagonalization of the spatial weighting matrix generated from the field and remote-sampled district-level covariate coefficients consisted of finding the normalized vectors ui, stored as columns in the matrix U = [u1 ⋯ un], satisfying: where Λ = diag (λ1 ⋯ λ n), and for i ≠ j. Note that double cantering of Ω implied that the eigenvectors ui generated from the ecological sampled district-level malarial-related covariate coefficients were centered and at least one eigenvalue was equal to zero. Introducing these eigenvectors in the original formulation of Moran's index lead to considering the centered vector z = Hx.

In this research, r was the number of null eigenvalues of Ω (r ≥ 1). These eigenvalues and corresponding eigenvectors were removed from Λ and U respectively. Equation 2.8 was then strictly equivalent to:, Moreover, it was demonstrated that Moran's index for a given eigenvector ui was equal to I(ui) = (n/1T W1)λ i so the equation was rewritten: (3.9).The term cor2 (ui, z) represented the part of the variance of z that was explained by ui in the district-level model z = β i ui+ ei. This quantity was equal to. By definition, the eigenvectors ui were orthogonal, and therefore, regression coefficients of the linear models z = β i ui+ ei were those of the multiple regression model z = Uβ + ε = β iui + ⋯ + β n-r un-r + ε.

In terms of the distribution of the error residuals in the seasonal district level malarial-related autocovariance matrix, the maximum value of I was not obtained by all of the variation of z, as explained by the eigenvector u1, which in this research did not correspond to the highest eigenvalue λ1 in the autocorrelation error matrix. In this research, cor2 (ui, z) = 1 (and cor2 (ui, z) = 0 for i ≠ 1) and the maximum value of I, was not deduced for Equation (3.9), The minimum value of I in the error matrix was not obtained as all the variation of z was not explained by the eigenvector un-r corresponding to the lowest eigenvalue λn-r generated in the model. This minimum value [e.g. Imin = λn-r (n/1TW1)] was also not derived. By so doing, we were able to define a random effects estimate for each district.

A one-dimensional Wiener process was then generated in SAS. The unconditional probability density function at a fixed time *t*: $f_{W_t}(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.$ The expectation was zero: $E[W_t] = 0.$

The variance, using the computational formula, is $\operatorname{Var}(W_t) = E\left[W_t^2\right] - E^2[W_t] = E\left[W_t^2\right] - 0 = E\left[W_t^2\right] = t.$ The covariance andcorrelation: $\operatorname{cov}(W_s, W_t) = \min(s,t), \operatorname{corr}(W_s, W_t) = \frac{\operatorname{cov}(W_s, W_t)}{\sigma_{W_s}\sigma_{W_t}} = \frac{\min(s,t)}{\sqrt{st}} = \sqrt{\frac{\min(s,t)}{\max(s,t)}}.$ The results for the expectation and variance follow immediately from the definition that increments have a normal distribution, centered at zero. Thus $W_t = W_t - W_0 \sim N(0,t).$ The results for the covariance and correlation follow from the definition that non-overlapping increments are independent, of which only the property that they are uncorrelated is used. Suppose that $t_1 < t_2$.

$\operatorname{cov}(W_{t_1}, W_{t_2}) = E\left[(W_{t_1} - E[W_{t_1}]) \cdot (W_{t_2} - E[W_{t_2}])\right] = E\left[W_{t_1} \cdot W_{t_2}\right]$ ·Substituting $W_{t_2} = (W_{t_2} - W_{t_1}) + W_{t_1}$ we arrive at $E[W_{t_1} \cdot W_{t_2}] = E\left[W_{t_1} \cdot ((W_{t_2} - W_{t_1}) + W_{t_1})\right] = E\left[W_{t_1} \cdot (W_{t_2} - W_{t_1})\right] + E\left[W_{t_1}^2\right]$ · Since $W(t_1) = W(t_1) - W(t_0)$ and $W(t_2) - W(t_1)$, are independent, $E\left[W_{t_1} \cdot (W_{t_2} - W_{t_1})\right] = E[W_{t_1}] \cdot E[W_{t_2} - W_{t_1}] = 0.$ Thus $\operatorname{cov}(W_{t_1}, W_{t_2}) = E\left[W_{t_1}^2\right] = t_1$ ·The running maximum of the model then was qunatiated The joint distribution of the running maximum $M_t = \max_{0 \le s \le t} W_s$ and W_t was $f_{M_t, W_t}(m, w) = \frac{2(2m - w)}{t\sqrt{2\pi t}} e^{-\frac{(2m-w)^2}{2t}}, \quad m \ge 0, w \le m.$ To get the unconditional distribution of f_{M_t}, we integrated over $-\infty < w \le m$ as $f_{M_t}(m) = \int_{-\infty}^{m} f_{M_t, W_t}(m, w)\, dw = \int_{-\infty}^{m} \frac{2(2m - w)}{t\sqrt{2\pi t}} e^{-\frac{(2m-w)^2}{2t}}\, dw = \sqrt{\frac{2}{\pi t}} e^{-\frac{m^2}{2t}}$ The expectation then was $E[M_t] = \int_0^{\infty} m f_{M_t}(m)\, dm = \int_0^{\infty} m \sqrt{\frac{2}{\pi t}} e^{-\frac{m^2}{2t}}\, dm = \sqrt{\frac{2t}{\pi}}$.

A Brownian scaling was then generated . For every $c > 0$ the process $V_t = (1/\sqrt{c}) W_{ct}$ I in our geopredictive malaria-related district-level malaria-related model was another Wiener process. The time process $V_t = W_1 - W_{1-t}$ for $0 \le t \le 1$ is distributed like W_t for $0 \le t \le 1$. A class of Brownian martingales was then generated. If a polynomial $p(x, t)$ satisfies the PDE $\left(\frac{\partial}{\partial t} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\right) p(x, t) = 0$ then the stochastic process $M_t = p(W_t, t)$ is a martingale (Cressie 1993) In our model $W_t^2 - t$ is a martingale, which revealed that the quadratic variation of W on $[0, t]$ was equal to t. It followed that the expected time of first exit of W from $(-c, c)$ was equal to c^2. More generally, for every polynomial $p(x, t)$ in the geopredictive malaria-related model the following stochastic process was a martingale: $M_t = p(W_t, t) - \int_0^t a(W_s, s)\, \mathrm{d}s,$ where a was the polynomial $a(x, t) = \left(\frac{\partial}{\partial t} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\right) p(x, t).$ Quantitative properties of our model then quantitated using an iterated logarithm which revealed that $\limsup_{t \to +\infty} \frac{|w(t)|}{\sqrt{2t \log\log t}} = 1$ Local modulus of continuity: $\limsup_{\varepsilon \to 0+} \frac{|w(\varepsilon)|}{\sqrt{2\varepsilon \log\log(1/\varepsilon)}} = 1,$ The Global modulus of continuity was then $\limsup_{\varepsilon \to 0+} \sup_{0 \le s < t \le 1, t - s \le \varepsilon} \frac{|w(s) - w(t)|}{\sqrt{2\varepsilon \log(1/\varepsilon)}} = 1,$ The image of the Lebesgue measure on $[0, t]$ under the district-level malarial risk map w had a density $L_t(\cdot)$. Thereafter, $\int_0^t f(w(s))\, \mathrm{d}s = \int_{-\infty}^{+\infty} f(x) L_t(x)\, \mathrm{d}x$ was derived for a wide class of functions f (namely: all continuous functions; all locally integrable functions; all non-negative measurable functions) in the model residual forecasts. The density L_t was continuous. The number $L_t(x)$ I at x of w on $[0, t]$. It was strictly positive for all x of the interval (a, b) where a and b were the least and the greatest value of w on $[0, t]$, respectively. Treated as a function of two sampled explanatory hyperendemic transmission oriented variables x and t, was still continuous.

Treated as a function of *t* (while *x* is fixed), the local time was a singular function corresponding to a measure on the set of zeros of *w*.

Next, an ARIMA analysis of individual district time series revealed a conspicuous but not very prominent first-order temporal autoregressive structure in the district-level data. We initially derived the a posteriori estimate covariance matrix. Starting with our invariant on the error covariance we solved fro $\mathbf{P}_{k|k}$ in $\mathbf{P}_{k|k} = \operatorname{cov}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})$ We then substituted in the definition of $\hat{\mathbf{x}}_{k|k}\mathbf{P}_{k|k} = \operatorname{cov}(\mathbf{x}_k - (\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\tilde{\mathbf{y}}_k))$ and $\tilde{\mathbf{y}}_k$ in and $\mathbf{z}_k$ for

$\mathbf{P}_{k|k} = \operatorname{cov}(\mathbf{x}_k - (\hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})))$ We then collected the error vectors from $\mathbf{P}_{k|k} = \operatorname{cov}((I - \mathbf{K}_k\mathbf{H}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) - \mathbf{K}_k\mathbf{v}_k)$ Since the measurement error $\mathbf{v}_k$ in the geopredictive district level malaria-related hyperendemic trasnmission oriented coariate coefficients was uncorrelated with the other terms, this become $\mathbf{P}_{k|k} = \operatorname{cov}((I - \mathbf{K}_k\mathbf{H}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})) + \operatorname{cov}(\mathbf{K}_k\mathbf{v}_k)$. Thereafter by the properties of vector covariance this model became $\mathbf{P}_{k|k} = (I - \mathbf{K}_k\mathbf{H}_k)\operatorname{cov}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(I - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k\operatorname{cov}(\mathbf{v}_k)\mathbf{K}_k^{\mathrm{T}}$ Then using our invariant on $\mathbf{P}_{k|k-1}$ and the definition of $\mathbf{R}_k$ the model became $\mathbf{P}_{k|k} = (I - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}(I - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}}$ This formula was valid for any value of $\mathbf{K}_k$ in our model.

Thereafter we generated a Kalman gain derivation. The Kalman filter is a minimum mean-square error estimator. The error in the a posteriori state estimation is $\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}$ (Cressie 1993) Wethen saught to minimize the expected value of the square of the magnitude of this vector, $\mathrm{E}[\|\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}\|^2]$. We assumed this procedure would be equivalent to minimizing the trace of the a posteriori estimate covariance matrix $\mathbf{P}_{k|k}$. By expanding out the terms in the geopredictive regression-based equation above we got: $\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)\mathbf{K}_k^{\mathrm{T}}$
$= \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{K}_k^{\mathrm{T}} + \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^{\mathrm{T}}$ The trace was then minimized when its matrix derivative with respect to the gain matrix which in this research was zero. Using the gradient matrix rules and the symmetry of the matrices involved we found that $\frac{\partial \operatorname{tr}(\mathbf{P}_{k|k})}{\partial \mathbf{K}_k} = -2(\mathbf{H}_k\mathbf{P}_{k|k-1})^{\mathrm{T}} + 2\mathbf{K}_k\mathbf{S}_k = 0.$ Solving this for K_k yielded the Kalman gain: $\mathbf{K}_k\mathbf{S}_k = (\mathbf{H}_k\mathbf{P}_{k|k-1})^{\mathrm{T}} = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}$ and $\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\mathbf{S}_k^{-1}$ This gain, was the optimal Kalman gain for our geopredictive malaria-related district-level model for yielding the MMSE estimates

We then simplified *a* posteriori error covariance formula for quantitating the district-level seasonal-sampled hyperendemic transmission oriented explanatory covariate coefficients**.** The formula used to calculate the a posteriori error covariance can be simplified when the Kalman gain equals the optimal value derived from a forecasting model (Box and Jenkins 1976) Multiplying both sides of our Kalman gain formula on the right by $\mathbf{S}_k\mathbf{K}_k^T$, it followed that $\mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T$ Referring back to our expanded formula for the *a posteriori* error covariance, $\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{K}_k^T + \mathbf{K}_k\mathbf{S}_k\mathbf{K}_k^T$ was generated which required cancelling out two terms thus rendering $\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1} = (I - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}$. If arithmetic precision is unusually

low causing problems with numerical stability, or if a non-optimal Kalman gain is deliberately used simplifications cannot be applied; the a posteriori error covariance formula must be used (Griffith 2003).

In this research the modified bessel function of the second kind is the function $K_n(x)$ was computed which was based on the solutions to the modified Bessel differential equation. function $I_n(x)$which is one of the solutions to the modified Bessel differential equation and is closely related to the Bessel function of the first kind $J_n(x)$. The above plot shows $I_n(x)$for $n = 1$, 2, ..., 5. The modified Bessel function of the first kind is implemented in SAS/GIS as Bessell[*nu*, *z*]. The modified Bessel function of the first kind $I_n(z)$can be defined by the contour integral $I_n(z) = \frac{1}{2\pi i}\oint e^{(z/2)(t+1/t)}\, t^{-n-1}\, dt,$ where the contour encloses the origin and is traversed in a counterclockwise direction (Arfken 1985). In terms of $J_n(x)$, $I_n(x) \equiv i^{-n} J_n(i x) = e^{-n\pi i/2} J_n\left(x\, e^{i\pi/2}\right)$. For a geopredictive district-level malaria-related hyperendemic transmission oriented explanatory covariate coefficients measurement valus ν, $I_\nu(z) = \left(\frac{1}{2}z\right)^\nu \sum_{k=0}^{\infty} \frac{\left(\frac{1}{4}z^2\right)^k}{k!\,\Gamma(\nu+k+1)},$ the function was computed using where $\Gamma(z)$is the gamma function. An integral formula[i.e. $I_\nu(z) = \frac{1}{\pi}\int_0^\pi e^{z\cos\theta}\cos(\nu\theta)\,d\theta - \frac{\sin(\nu\pi)}{\pi}\int_0^\infty e^{-z\cosh t-\nu t}\,dt,$] was then employed [which simplifies for νan integer $n$to $I_n(z) = \frac{1}{\pi}\int_0^\pi e^{z\cos\theta}\cos(n\theta)\,d\theta$

The modified Bessel function of the second kind is implemented in *SAS* as BesselK[*nu*, *z*]. $K_n(x)$is closely related to the modified Bessel function of the first kind $I_n(x)$and Hankel function $H_n(x)$, $K_n(x) = \frac{1}{2}\pi i^{n+1} H_n^{(1)}(i x) = \frac{1}{2}\pi i^{n+1}\left[J_n(i x) + i N_n(i x)\right] \frac{\pi}{2}\frac{I_{-n}(x) - I_n(x)}{\sin(n\pi)}$ In our analyses a sum formula for $K_n(x)$ was tabulated as $K_n(z) = \frac{1}{2}\left(\frac{1}{2}z\right)^{-n}\sum_{k=0}^{n-1}\frac{(n-k-1)!}{k!}\left(-\frac{1}{4}z^2\right)^k + (-1)^{n+1}\ln\left(\frac{1}{2}z\right)I_n(z) + (-1)^n\frac{1}{2}\left(\frac{1}{2}z\right)^n\sum_{k=0}^{\infty}[\psi(k+1)+\psi(n+k+1)]\frac{\left(\frac{1}{4}z^2\right)^k}{k!\,(n+k)!},$ where

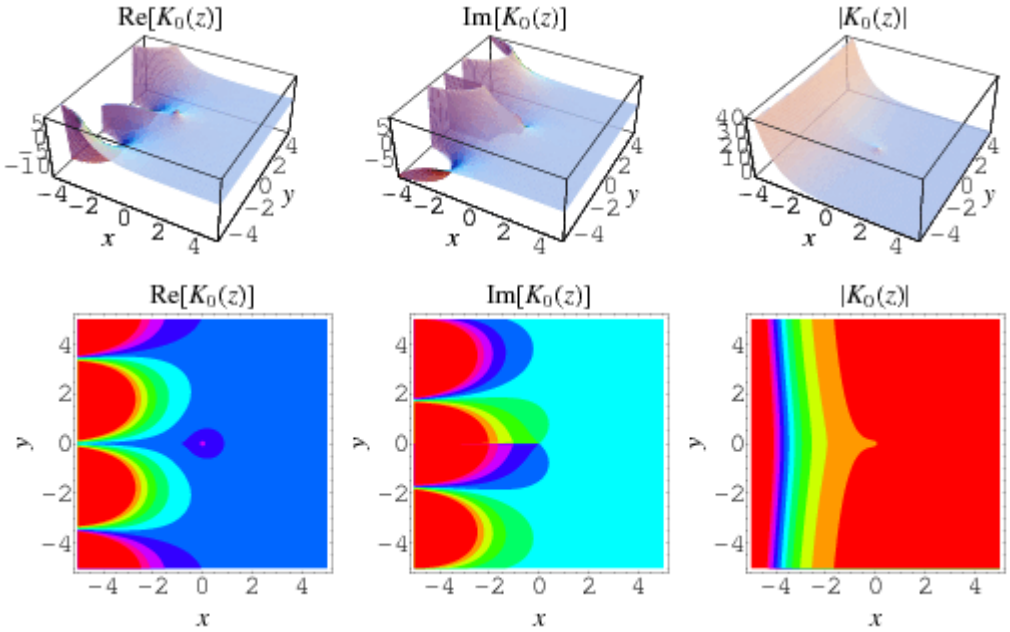
$K_n(z) = \frac{1}{2}\left(\frac{1}{2}z\right)^{-n}\sum_{k=0}^{n-1}\frac{(n-k-1)!}{k!}\left(-\frac{1}{4}z^2\right)^k + (-1)^{n+1}\ln\left(\frac{1}{2}z\right)I_n(z) + (-1)^n\frac{1}{2}\left(\frac{1}{2}z\right)^n\sum_{k=0}^{\infty}[\psi(k+1)+\psi(n+k+1)]\frac{\left(\frac{1}{4}z^2\right)^k}{k!\,(n+k)!},$

In these equations ψ was the digamma function in our model . We employed an = integral formula which was $K_\nu(z) = \frac{\Gamma\left(\nu+\frac{1}{2}\right)(2z)^\nu}{\sqrt{\pi}}\int_0^\infty \frac{\cos t\,dt}{\left(t^2+z^2\right)^{\nu+1/2}}$ which, for $\nu = 0$, simplified to $K_0(x) = \int_0^\infty \cos(x\sinh t)\,dt = \int_0^\infty \frac{\cos(x\,t)\,dt}{\sqrt{t^2+1}}.$ We identified other identied using $K_n(z) = \frac{\sqrt{\pi}}{\left(n-\frac{1}{2}\right)!}\left(\frac{1}{2}z\right)^n\int_1^\infty e^{-z\,x}\left(x^2-1\right)^{n-1/2}dx$ for $n > -1/2$and

$K_n(z) = \sqrt{\frac{\pi}{2z}}\frac{e^{-z}}{\left(n-\frac{1}{2}\right)!}\int_0^\infty e^{-t}\,t^{n-1/2}\left(1-\frac{t}{2z}\right)^{n-1/2}dt = \sqrt{\frac{\pi}{2z}}\frac{e^{-z}}{\left(n-\frac{1}{2}\right)!}\sum_{r=0}^{\infty}\frac{\left(n-\frac{1}{2}\right)!}{r!\,\left(n-r-\frac{1}{2}\right)!}(2z)^{-r}\int_0^\infty e^{-t}\,t^{n+r-1/2}\,dt.$

which then genertd he following graphical outputs in SAS/GIS.

International Journal of Geographic Information System
Vol. 1, No. 1, November 2013, PP: 01 - 143
Available Online at http://acascipub.com/Journals.php



The special case of $n = 0$ was then given $K_0(z)$ as the integrals $K_0(z) = \int_0^\infty \cos(z \sinh t)\, dt = \int_0^\infty \frac{\cos(z\,t)}{\sqrt{t^2+1}}\, dt$.

The Kalman filtering equations provided an estimate of the state $\hat{\mathbf{x}}_{k|k}$ and its error covariance $\mathbf{P}_{k|k}$ recursively in the geopredictive seasonal-sampled malaria-related risk model. 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 (that is, of a vector of random variables). Therefore each element of the predictive malaria-related seasonal district level hyperendemic transmission oriented model was based on the vector of a scalar random variable, either with a finite number of observed empirical values or with a finite or infinite number of potential values specified by a theoretical joint probability distribution of all the sampled field/clinical/remote random variables. Intuitively, the covariance matrix generalizes the notion of variance to multiple dimensions. As an example, the variation in a collection of random points in two-dimensional space cannot be characterized fully by a single number, nor would the variances in the *x* and *y* directions contain all of the necessary information; a 2×2 matrix would be necessary to fully characterize the two-dimensional variation

The estimate and its quality depend on the system parameters and the noise statistics fed as inputs to the estimator. This section analyzes the effect of uncertainties in the statistical inputs to the filter. In the absence of reliable statistics or the true values of noise covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, the expression $\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T$ no longer provides the actual error covariance. In other words, $\mathbf{P}_{k|k} \neq E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T]$. In most real time applications the covariance matrices that are used in designing the Kalman filter are different from the actual noise covariances matrices(Cressie 1993). This sensitivity analysis described the behavior of the estimation error covariance in our district level hyperendemic transmission oriented risk model when the noise covariance as well as the system matrices $\mathbf{F}_k$ and $\mathbf{H}_k$ were fed as inputs to the filter were incorrect. Thus, the sensitivity analysis described the robustness of sampled district-level predictive malaria-related estimator to misspecified statistical and parametric inputs to the estimator.

This research was limited to the error sensitivity analysis for the case of statistical uncertainties in the geopredictive malarial-related hyperendemic transmission-oriented risk model. Here the actual noise covariances were denoted by $\mathbf{Q}_k^a$and $\mathbf{R}_k^a$respectively, whereas the design sampled hyperendemic transmission oriented values used in the estimator are $\mathbf{Q}_k$and $\mathbf{R}_k$respectively. The actual error covariance in the model was denoted by $\mathbf{P}_{k|k}^a$and $\mathbf{P}_{k|k}$as computed by the Kalman filter. When $\mathbf{Q}_k \equiv \mathbf{Q}_k^aand\mathbf{R}_k \equiv \mathbf{Q}_k^a$, in our model outputs this meant that$\mathbf{P}_{k|k} = \mathbf{P}_{k|k}^a$. While computing the actual error covariance using$\mathbf{P}_{k|k}^a = E[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T]$, substituting for $\hat{\mathbf{x}}_{k|k}$and using the fact that $E[\mathbf{w}_k\mathbf{w}_k^T] = \mathbf{Q}_k^aandE[\mathbf{v}_k\mathbf{v}_k^T] = \mathbf{R}_k^a$, resulted in the following recursive equations for $\mathbf{P}_{k|k}^a$:$\mathbf{P}_{k|k-1}^a = \mathbf{F}_k\mathbf{P}_{k-1|k-1}^a\mathbf{F}_k^T + \mathbf{Q}_k^a$an $\mathbf{P}_{k|k}^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}^a(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k^a\mathbf{K}_k^T$ While computing $\mathbf{P}_{k|k}$, by design the filter implicitly assumed that $E[\mathbf{w}_k\mathbf{w}_k^T] = \mathbf{Q}_k$and $E[\mathbf{v}_k\mathbf{v}_k^T] = \mathbf{R}_k$. Note that the recursive expressions for $\mathbf{P}_{k|k}^a$and $\mathbf{P}_{k|k}$were identical except for the presence of $\mathbf{Q}_k^a$and $\mathbf{R}_k^a$in place of the design values $\mathbf{Q}_k$and $\mathbf{R}_k$respectively in our geopredictive malaria-related model.

ARIMA models used the observable non-stationary processes Xt that had some clearly identifiable trends which included a constant trend (i.e. zero average) which was modeled by$d = 0$, a linear trend (i.e. linear growth behavior) which was then modeled by $d = 1$ and a quadratic trend. Thereafter, this quadratic growth behavior was modeled by $d = 2$. In this research the ARIMA model was viewed as a "cascade" of two models. The first was non-stationary: $Y_t = (1 - L)^d X_t$ while the second was wide-sense stationary: $\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) Y_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$. Finally, forecasts techniques were formulated for the processY_t, and then having the sufficient number of initial conditions) X_t was forecasted via opportune integration steps.

Thereafter, a random effects term was specified with the sampled monthly time series data. This random effects specification revealed a non-constant mean across the sampled districts that were variable that represented a district-constant across time. This specification also represented a district-specific intercept term that was a random deviation from the overall intercept term as it was based on a draw from a normal frequency distribution. In this research, this random intercept represented the combined effect of all omitted district-specific predictor covariate coefficients that caused some districts to be more prone to the malaria prevalence than other districts. Inclusion of a random intercept assumed random heterogeneity in the districts' propensity or underlying risk of malaria prevalence that persisted throughout the entire duration of the time sequence under study. The Poisson mean response specification was mu = expel [a + re+ LN (population)], Y ~Poisson (mu). The mixed-model estimation results included a = -3.1876, re ~ n (0, s^2), mean re = -0.0010 s^2 = 0.2513, P(S-W) = 0.0005 and Pseudo-R^2 = 0.3103 for prevalence regressed on geopredicted prevalence. This random effects term displayed no spatial autocorrelation, and failed to closely conform to a bell-shaped curve. Its variance implied a substantial variability in the prevalence of malaria across districts. We noticed that he estimated model contained considerable over dispersion (i.e., excess Poisson variability): quasi-likelihood scale = 76.5648.

Figure 2 portrays scatterplots of observed versus predicted prevalence for selected months, and reflects the considerable amount of noise in the malarial prevalence data as well as the random effects term accounting for about a third of the variance in the space-time series of malarial prevalence. As with most statistical procedures, the random effects term in our model corresponded more closely with the data in the center of the time series. This goodness-of-fit feature implied that although the random effects term can be used for purposes for predicting

malarial at the district level in Uganda but it was less effective for quantitating data associated to a relatively lengthy time series.

Based on the spatiotemporal-sampled district level predictor covariate coefficients a random effects model was generated. The articulated tessellations for Uganda based upon district geocodes was then digitally overlaid onto interpolated data in ArcGIS® using species distribution (e.g., *Anopheles gambiae s.l.)* and Entomological Inoculation Rate (EIR), and remote sensing band data from the Malaria Atlas Project (http://www.map.ox.ac.uk). The MAP team has assembled a unique spatial database on linked information based on medical intelligence, satellite-derived climate data to constrain the limits of malaria transmission and the largest ever archive of community-based estimates of parasite prevalence (Hay and Snow 2006). The initial focus of MAP has been centered on predicting the endemicity of *Plasmodium falciparum*, the most deadly form of the malaria parasite, due to its global epidemiological significance and its better prospects for elimination and control but this database has been since improved (http://www.map.ox.ac.uk).We overlaid the spatial tessellations generated from the geopredictive district-level map onto MAP data.

Figure 5: A predicted spatial distribution of Plasmodium falciparum EIR map in 2012 for Uganda

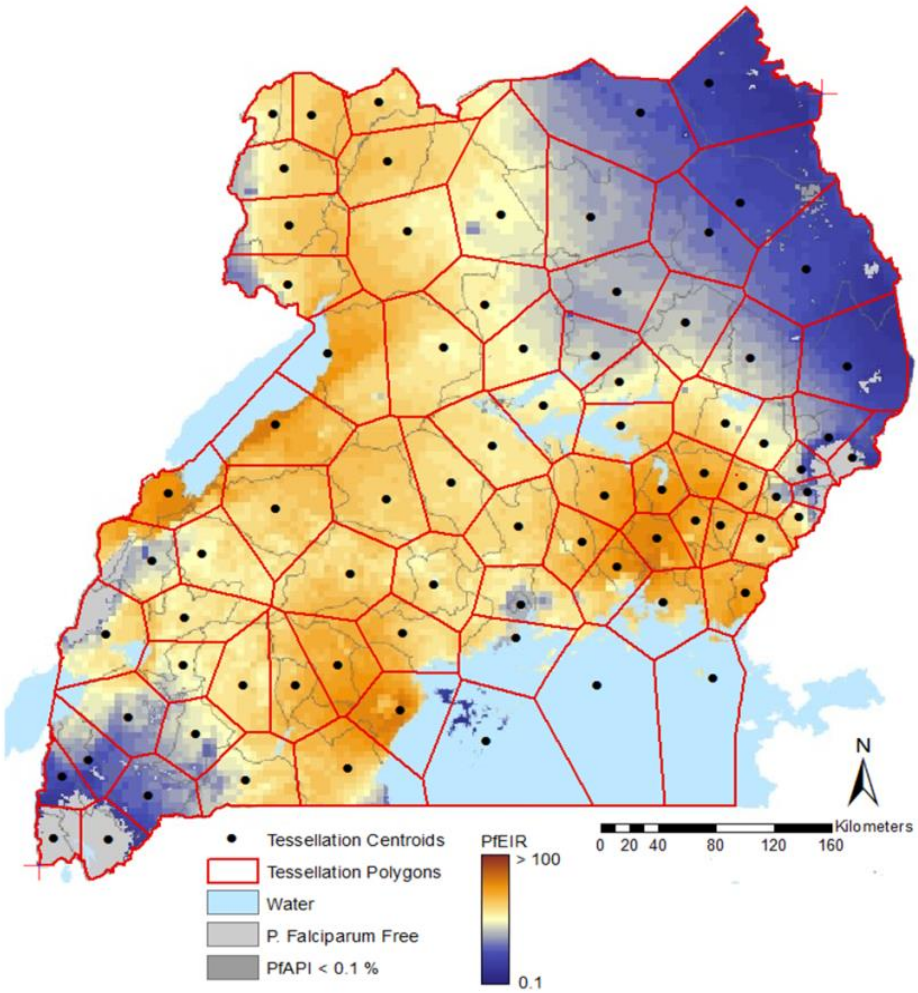


Figure 6: The spatially predicted distribution of *Anopheles gambiae s.l.* throughout the Ugandan study site

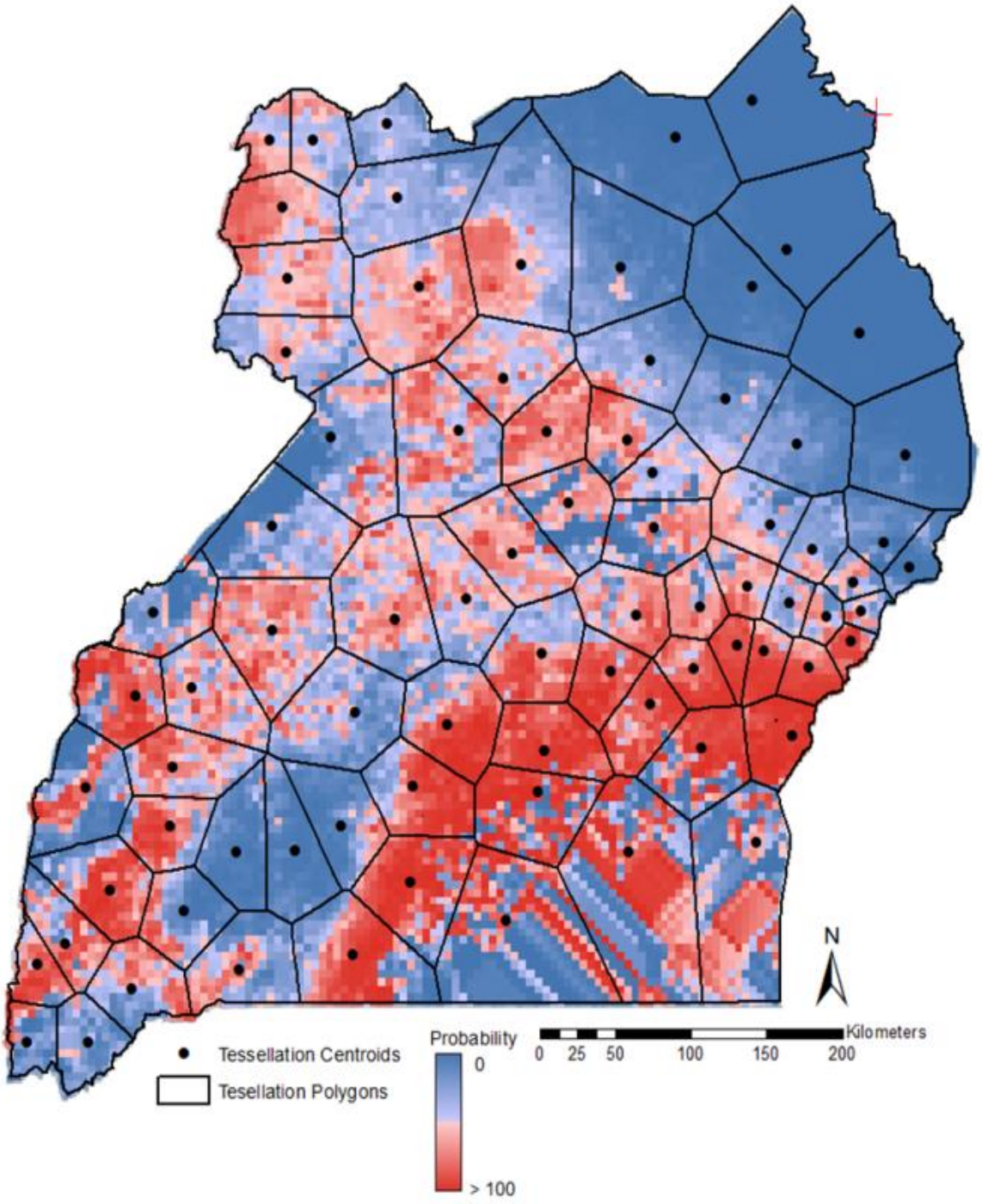


Figure 7: District-level risk map of *Anopheles arabiensis* for the Ugandan study site

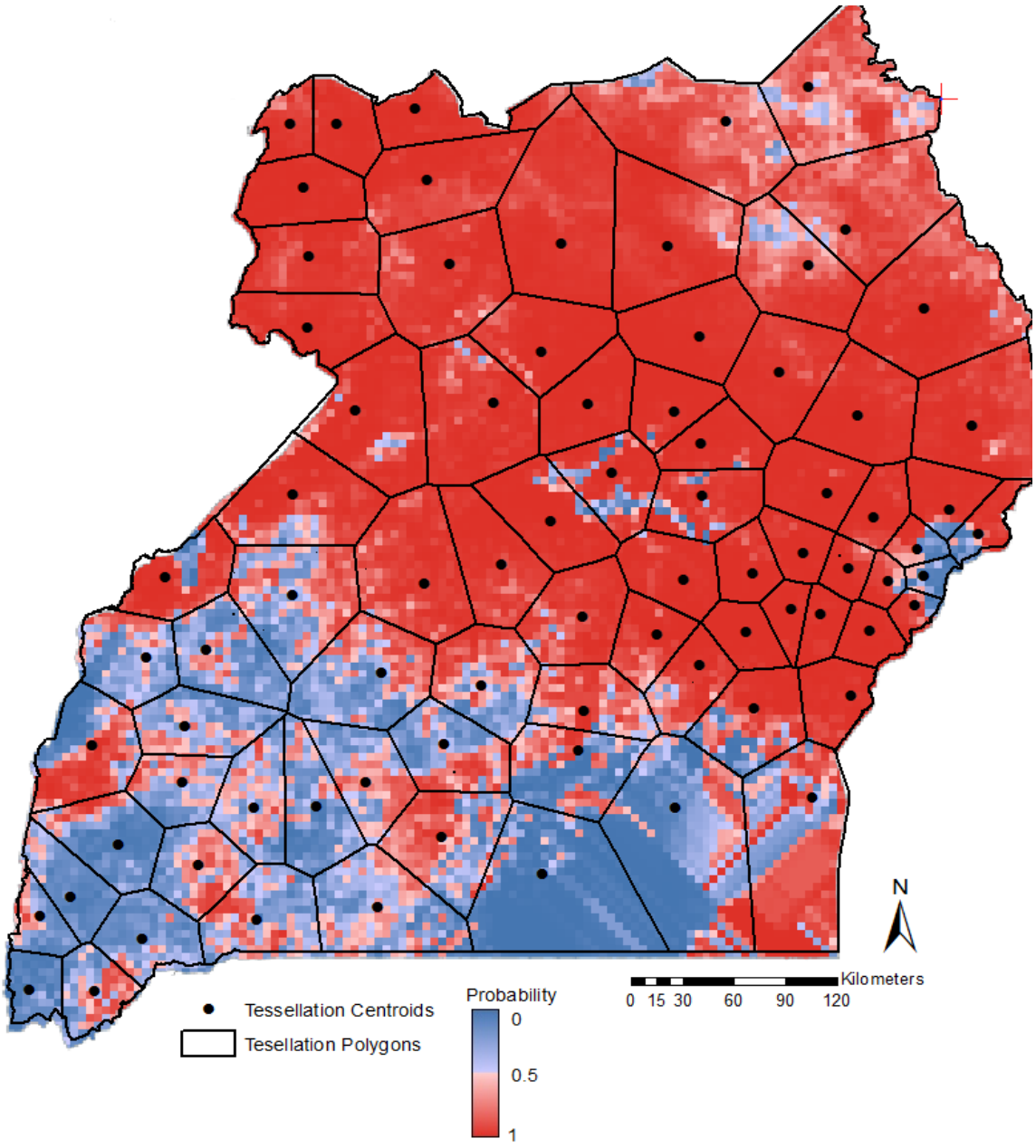


Figure 8: District-level band 1 malarial radiance risk map for the Ugandan study site

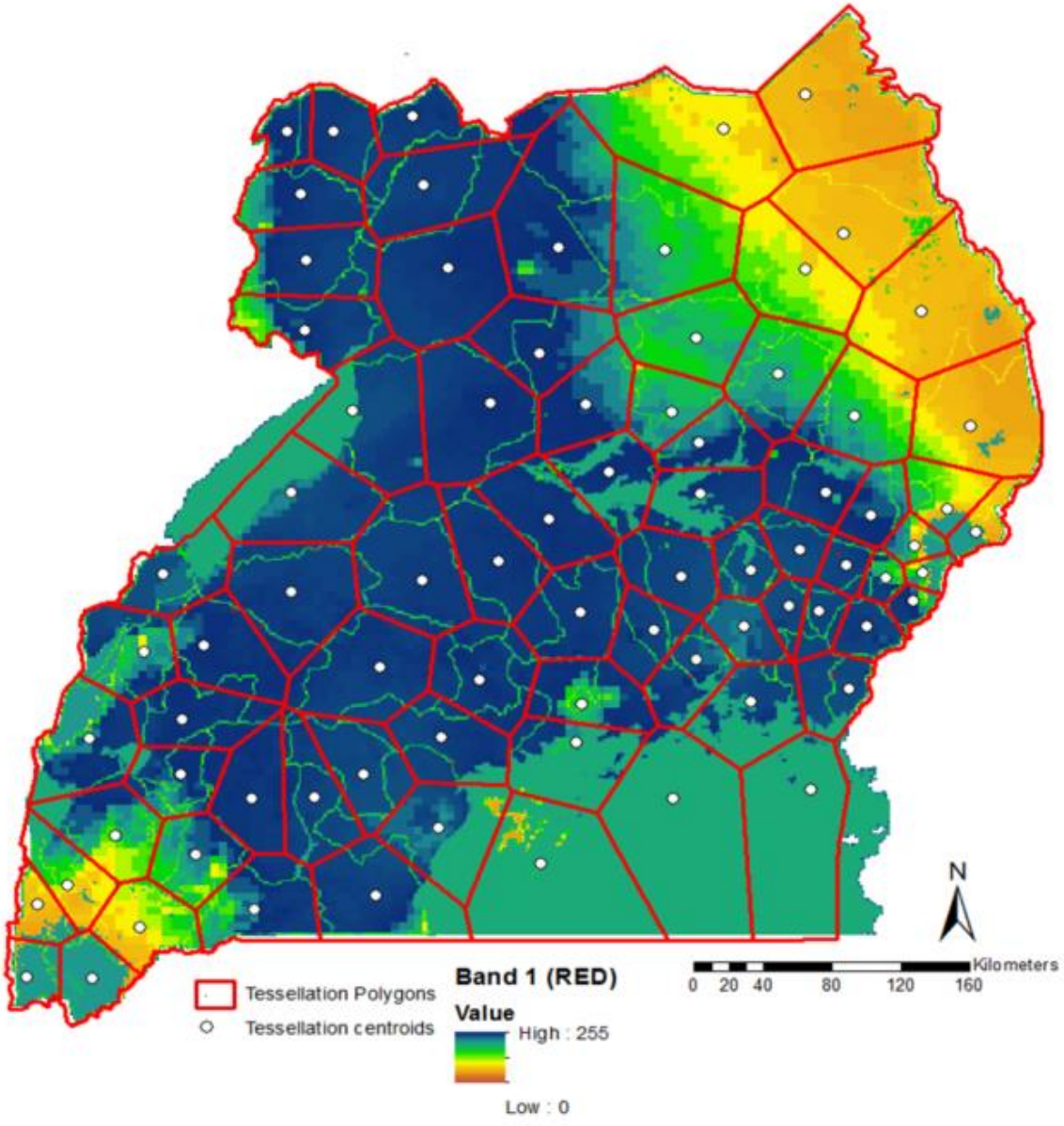


Figure 9: District-level band 2 malarial radiance risk map for the Ugandan study site

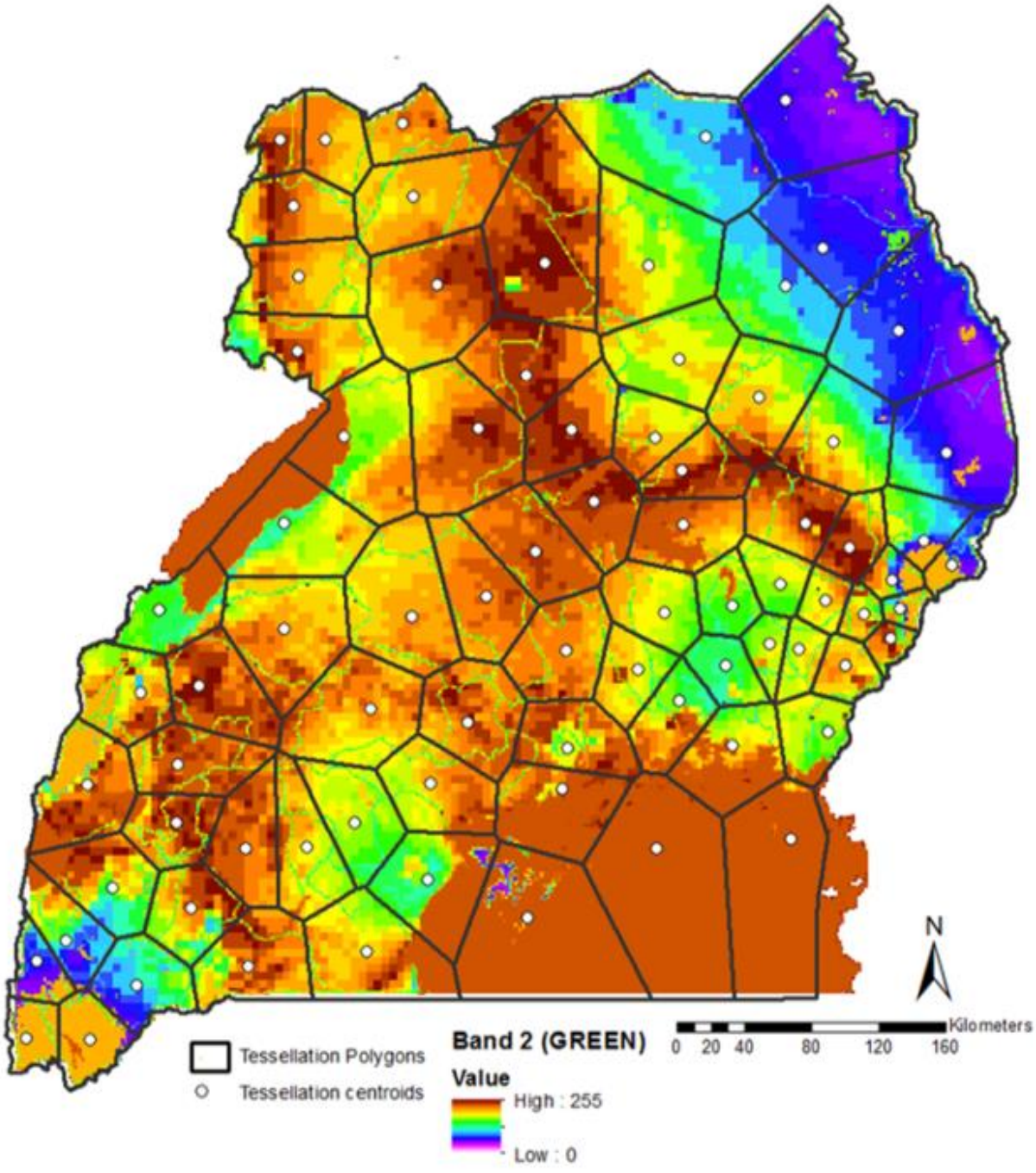
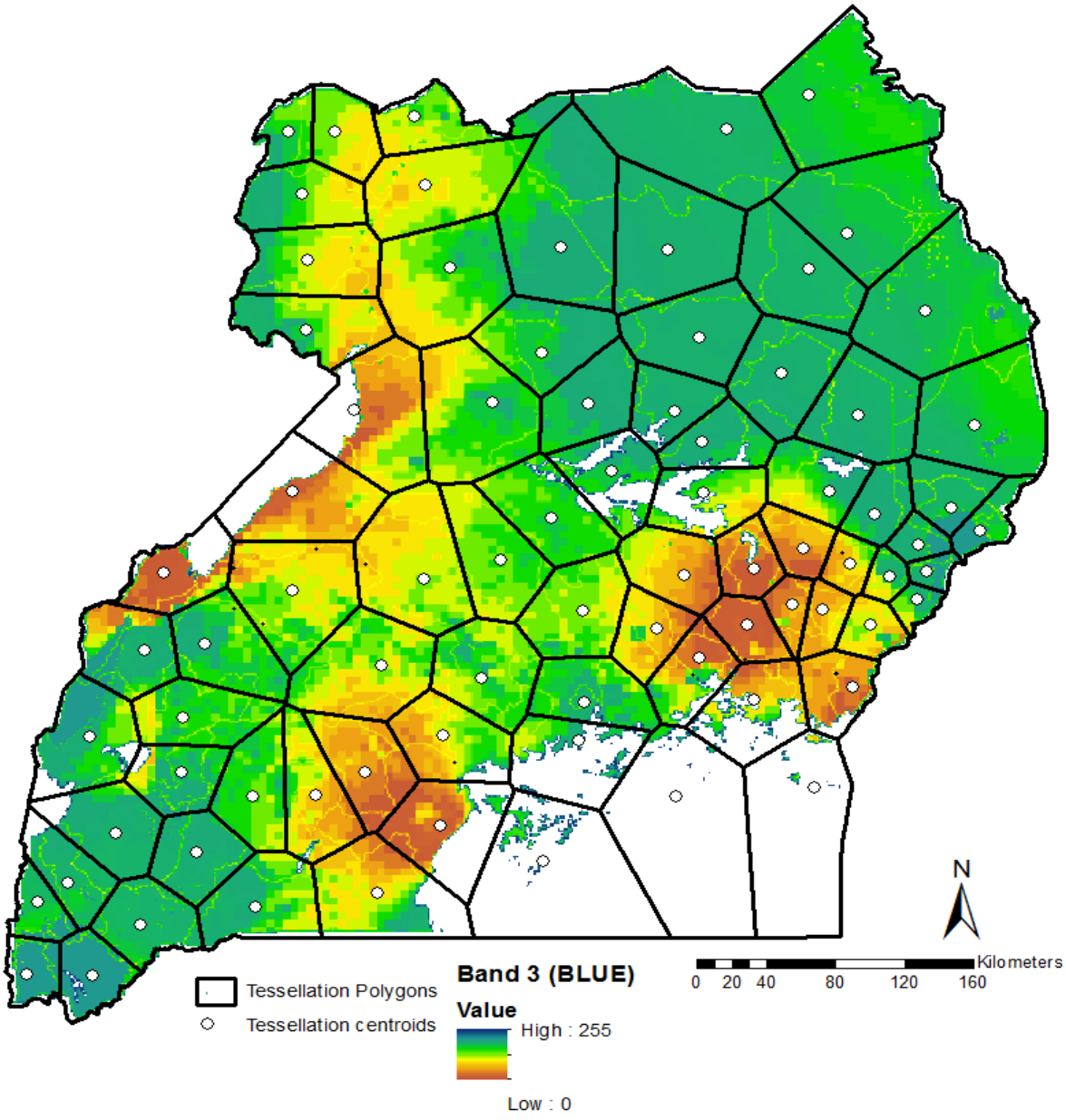
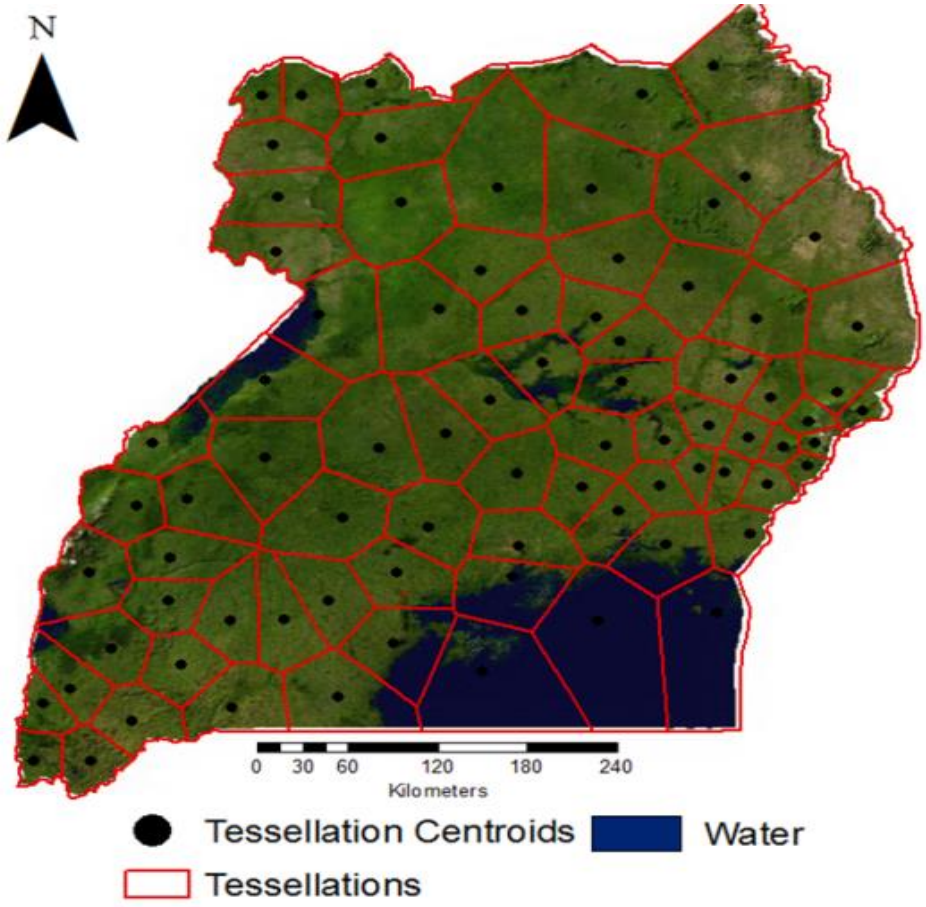


Figure 10: District-level band 3 malarial radiance risk map for the Ugandan study site



We then generated a Land Use Land Cover (LULC) model in ArcGIS®. The model was then draped with the district-level tessellations images. Traditional per-pixel spectral-based supervised LULC malaria-related classification require incorporation of textural images and multispectral images, spectral-spatial classifier, and segmentation-based data for developing control strategies in urban landscapes at the cluster -level (Jacob et al. 2003). In this research we employed the spatial information of the district level data as an ArcGIS segmentation-based classification method which significantly improved land cover classification performance especially for quantifying hydrological-related covariate coefficients at the district level.

Figure 11: A land use land cover model for the Ugandan study site

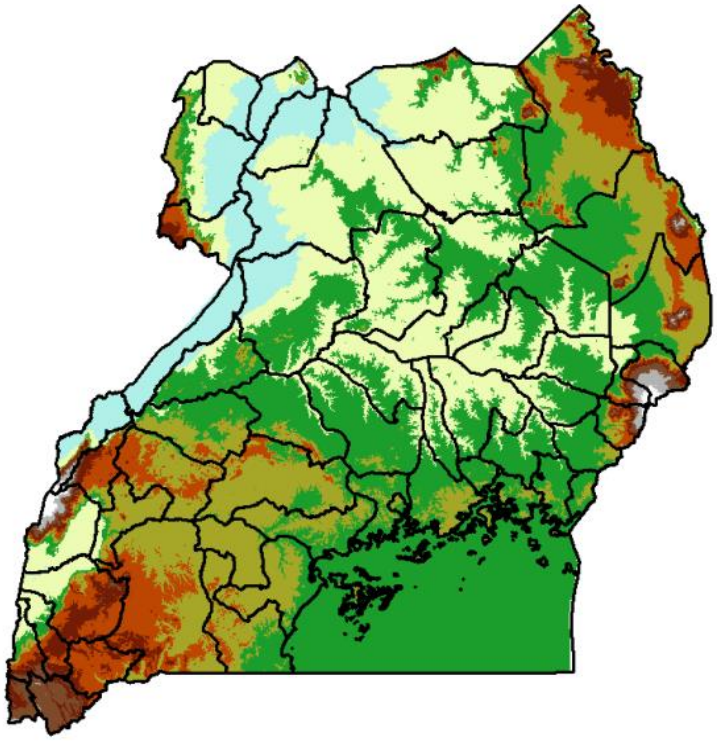


Next, to improve our random effects model we used Digital Elevation Model (DEM) stratification and then added the terrain-related statistic derived to improve our original random effects model. Flood and swamp water mosquito abundance can be predicted in real time using high resolution data through application of a dynamic hydrological model (Jacob et al. 2008a). These models can account for topographic variability and their control over soil moisture heterogeneity and runoff within a shed. Soil moisture levels can also be associated with local malaria mosquito biting rates on humans and entomologic inoculation rates (EIR) (Pats 1998). The probability distribution of the soil moisture deficit, i.e., statistics of topography was generated from the DEM data by using a multidirectional flow routing algorithm, which in this research was tied to an adaptive error correction (pit infill) scheme needed for low-relief areas in Uganda. In Jacob et al. (2008a) a robust DEM was employed to yield several catchment hydrological

variables including percent surface saturation, and total surface runoff for identification of urban malaria mosquito *An. gambiae s.l*. mosquitoes sampled in Gulu, Uganda.

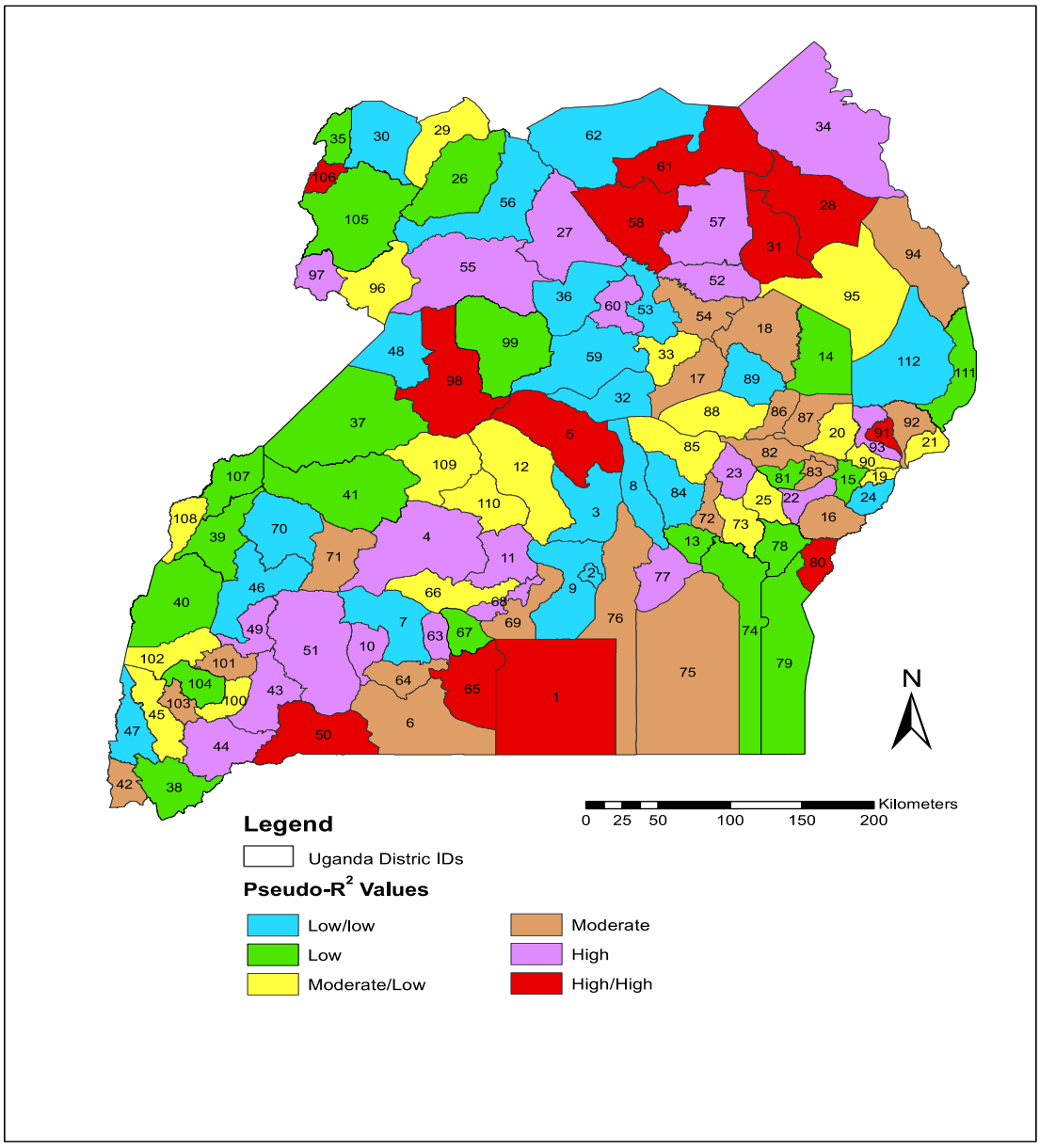
A Stream Raster Grid was then generated in ArcGIS®. Euclidian distance-to-nearest hydrological body was then calculated as the distance from a grid cell to a stream grid cell defined by a Stream Raster Grid. Flow distance-to-stream was employed to quantitate availability of the aquatic larval habitats and these were calculated as the distance from a grid cell moving downstream to a stream grid cell defined by the Stream Raster grid. The Terrain Analysis Using DEM (i.e., TauDEM) in ArcGIS® was then used to retrieve terrain-related geomorphological uncertainty parameter estimators. A three-dimensional model of the Ugandan study area was also constructed based on the DEM using Arc Scene extension of ArcGIS®. 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 *An. gambiae s.l.* aquatic larval habitats was 0.171%. The model revealed that for aquatic larval habitat count sampled at the Ugandan study site employing the sampled parameter estimator slope there was a negative correlation (-0.23) for a local district –level model (e.g., Abim) based on distance to stream. A DEM of the study area was downloaded from seamless United States Geological Survey (USGS, March 17th, 2013). In this research, the DEM was constructed based on a contour map of 1:50,000. We then overlaid the district-level tessellation in ArcGIS onto the DEM (see Figure 10).

Figure 12: Digital elevation model for the Ugandan study site



We then qualitatively assessed THE geomorphological terrain-related LULC statistics derived from the DEM and band radiance estimates to create more robust indices based on our primary model estimates. This model was then based on tabulated DEM, LULC parameter estimators ,satellite band radiance values at the district level , a model random effects term, and a regressed predicted district-level prevalence count. The Poisson mean response specification was then: $\mu = \exp[a + re + LN(population)]$, $Y \sim Poisson(\mu) + DEM$ (zonal statistic). The mixed-model estimation results included: a = -3.1876 re ~ n(0, s^2) mean re = -0.0010 s^2 = 0.2513 where P(S-W) = 0.0005 and the Pseudo-R^2 = 0.3103. (See Figure 11).

Figure 13: Prioritized districts based on random effects hierarchical linear-based malaria risk model at the Ugandan study site.



In our Ugandian malraial model the GLS estimator is unbiased, consistent, efficient, and asymptotically normal: GLS is equivalent to applying OLS to a linearly transformed version of time series-dependent data (Cressie 1993).

To attain factor Ω = BB′, for instance the linear endemic malarial transmission-oriented regression risk-based regression model which can be constructured ed employing the Cholesky decomposition. Cholesky decomposition or Cholesky triangle is a decomposition of a Hermitian, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose. Given a symmetric positive definite matrix A , the Cholesky decomposition is an upper triangular matrix with strictly positive diagonal entries such that Cholesky decomposition is commonly implemented in SAS/GIS as CholeskyDecomposition[m]. In a loose, metaphorical sense, decomposition this can be thought of as the matrix analogue of taking the square root of a number. Then if we multiply both sides of the equation Y = Xβ + ε by B−1, we get an equivalent linear model Y* = X*β + ε*, where Y* = B−1Y, X* = B−1X, and ε* = B−1ε. In this model Var[ε*] = B−1ΩB−1 = I. Thus, we can efficiently estimate β by applying OLS to the transformed linear endemic malarial transmission-oriented regression risk-based model regression model data, which would then simply require minimizing.By so doing, any effect of standardizing the scale of the errors would be "de-correlating". Since OLS is applied to data with homoscedastic linear endemic malarial transmission-oriented regression risk-based model regression model errors, the Gauss–Markov theorem applies, and therefore the GLS estimate is the best linear unbiased estimator for β (see Cressie 1993).In ArcGIS a table was generated which identifed and ranked the predicted cluster (i.e., district) (see Table 2).

Table 1: Ugandan districts listed by malaria risk priority as is shown on Figure 11 above

ID	Name	ID	Name	ID	Name	ID	Name	ID	Name	ID	Name
1	Busia	13	Zombo	33	Alebatong	53	Gomba	73	Ntoroko	93	Kanungu
2	Kitgum	14	Ntungamo	34	Budaka	54	Namutumba	74	Hoima	94	Amuru
3	Pader	15	Bukomansimbi	35	Ngora	55	Rubirizi	75	Namayingo	95	Buliisa
4	Kalangala	16	Bulambuli	36	Moroto	56	Bukedea	76	Jinja	96	Amolatar
5	Kotido	17	Gulu	37	Kyegegwa	57	Bukwa	77	Adjumani	97	Kamwenge
6	Abim	18	Buikwe	38	Buhweju	58	Rukungiri	78	Kibaale	98	Soroti
7	Isingiro	19	Otuke	39	Rakai	59	Buyende	79	Kasese	99	Kayunga
8	Masaka	20	Mityana	40	Luuka	60	Dokolo	80	Kibuku	100	Lira
9	Nakasongola	21	Mubende	41	Kisoro	61	Bundibugyo	81	Arua	101	Kyenjojo
10	Kapchorwa	22	Kole	42	Mitooma	62	Iganga	82	Kabale	102	Ssembabule
11	Maracha	23	Ibanda	43	Tororo	63	Napak	83	Mbale	103	Kamuli
12	Masindi	24	Nyoya	44	Amuria	64	Serere	84	Kiryandongo	104	Wakiso
		25	Agago	45	Kween	65	Sironko	85	Amudat	105	Manafwa
		26	Mbarara	46	Kumi	66	Sheema	86	Bugiri	106	Apac
		27	Butaleja	47	Lwengo	67	Nebbi	87	Kabarole	107	Lamwo
		28	Lyantonde	48	Mpigi	68	Kyankwanzi	88	Mayuge	108	Luwero
		29	Butambala	49	Mukono	69	Moyo	89	Bushenyi	109	Yumbe
		30	Kaliro	50	Buvuma	70	Bududa	90	Katakwi	110	Oyam
		31	Kaabong	51	Kaberamaido	71	Kiboga	91	Koboko	111	Kampala
		32	Kiruhura	52	Pallisa	72	Nakaseke	92	Kalungu	112	Nakapiripirit

Discussion

Initially, a geopredictive LULC analyses was conducted employing the time series district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients from the empirical sampled dataset. An unsupervised classification algorithm in ArcGIS for the Ugandan LULC district-level malarial risk map subdivided the sampled data attributes into two phases: (i) the "calibration" phase in which the algorithm identified a classification scheme based on signatures of different bands obtained from known "training" sites having known class labels (e.g., land cover , crop types), and (ii) the prediction phase, in which the classification scheme was

applied to other geolocations with unknown class membership. The algorithms revealed relationships, (e.g., "rules", "networks", and "likelihood" measures, between the input [e.g., district-level remote sampled malaria-related spectral reflectance objects at different visible and NIR bands in geopredictor space) and the output (i.e., the district-level LULC class label) so that either an appropriate discriminant function was maximized and a cost function accounting for misclassified observations was minimized. In other words, our seasonal LULC predictive malaria-related hyperendemic transmission oriented risk model followed the traditional modeling paradigm that attempted to find an "optimal" unbiased residual forecasted estimator employing the distance between the observed hyperendemic transmission oriented featured attributes and the classification response. Our model revealed land cover classes at each district sampled at the Ugandan epidemiological study site.

A different LULC approach may be proposed in the future for quantitating parcels of within the context of district-level land cover topographic classification for generating a robust geopredictive malaria-related risk model at a meso-scale using the modified nearest-neighbor (MNN) technique. The MNN algorithm is a hybrid algorithm in ArcGIS that combines algorithmic features of a dimensionality reduction algorithm. This algorithm can survey existing feature selection for district-level malaria-related hyperendemic transmission oriented classification and clustering of district-level sampled groups and then compare the group employing a categorizing framework based on search strategies, evaluation criteria, and data mining tasks. By so doing, the residual algorithmic outputs can reveal unattempted district-level hyperendemic transmission oriented combinations and provide guidelines in selection of other district-level malarial featured spatial/spectral objects.

Within a categorizing MNN framework, efforts toward building an integrated system for intelligent feature selection can also be developed in ArcGIS. A unifying platform may be proposed for instance. Given a set D of objects and a query object q, an MNN query returns from D, the set of objects that are among the k1 (P1) nearest neighbors (NNs) of q (see Jensen 2005). As such, many illustrative examples of individual district-level malaria-related predictive time series feature selection may be then presented to show how existing LULC topographic time series trend analyses data may be integrated into a meta algorithm. An added advantage of doing so is to help a malariaologist/experimenter employ a suitable algorithm without knowing specific details of each algorithm (e.g., methodology of latent uncertainty filed/clinical/remote hyperendemic transmission oriented forecast quantitation). Some seasonal district-level predictive malaria-related risk mapping applications may be then included to demonstrate the use of a feature selection in data mining. Further, the MNN derived predictive malarial-related district-level model could provide (i) an extremely flexible and parsimonious environment for a few georeferenced residual forecasted estimators (e.g., *k* the number of NNs) to be qualitatively/quantitatively assessed, (ii) an extremely attractive field map validatior attained without requiring preprocessing of the forecasted data nor, assumptions with respect to the distribution of the training data; and, (iii) robust regressed forecasts r as the single 1-NN rule would guarantee an asymptotic error rate at the most twice that of the Bayes probability of error.

For instance, suppose an unknown geoparameter district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented observational estimator θ is known to have a prior distribution π in seasonal geopredictive malarial risk model where $\delta = \delta(x)$ is an estimator of θ (based on some regressed district-level field/clinical/remote sampled explanatory covariate coefficients measurement values *x*), and l. In such a model $L(\theta, \delta)$ would be a loss function, such as squared error. The Bayes risk in the district-level model would then be of δ and as such defined as $E_{\pi}\{L(\theta, \delta)\}$, where the expectation would be taken over by the probability distribution of θ in the sampled data attributes. This SAS/GIS derived model would thereafter be able to define the risk function in the model as a function of δ. An estimator δ is said to be a Bayes estimator if it minimizes the Bayes risk among all estimators (Gilks 1996). Equivalently, the estimator which minimizes the posterior expected loss [i.e., $E\{L(\theta, \delta)|x\}$] in the predictive district-level malaria-related regression-based risk model would then minimize the Bayes risk for each *x* in the residually forecasted regression-based estimates.

Importantly, if the prior is improper in a time series district –level predictive malaria-related risk model, then the field/clinical/remote sampled malaria-related hyperendemic transmission oriented exploratory observational estimators which minimize the posterior would be the expected loss for each *x* in the generalized Bayes forecasted error estimates. The most common risk function used for Bayesian probabilistic estimation is the MSE, also called

squared error risk (Cressie 1993). The MSE is defined by $\mathrm{MSE} = E\left[(\widehat{\theta}(x) - \theta)^2\right]$, where the expectation is taken over the joint distribution of θ and x (Rao 1973). Thus, employing the MSE as risk, the Bayes estimate of the unknown geoparameter in a time series district-level SAS/GIS derived predictive malarial-related regression-based risk model would then be simply the mean of the posterior distribution, $\widehat{\theta}(x) = E[\theta|x] = \int \theta\pi(\theta|x)\,d\theta.$. This then would be expressed as a MMSE estimator.

In a SAS/GIS derived malaria-related district-level geopredictive time series risk model, a minimum mean square error (MMSE) estimator would be an estimation method which minimizes the MSE of the fitted empirical sampled dataset of field/clinical /remote sampled hyperendemic transmission measurement values of a dependent variable(e.g., district-level prevalence rate). The basic definition of a district-level predictive time series malarial-related MMSE can be then expressed in SAS/GIS, if a malarialogist/experimenter lets x be a $n \times 1$ unknown (hidden) random vector variable, and thereafter lets y be a $m \times 1$ known random vector variable (i.e., the measurement or district-sampled observation). These variables would not have to be of necessarily of the same dimension. As such, an estimator $\hat{x}(y)$ of x in a time series district-level geopredictive malaria-related regression-based risk model would be any function of the sampled field/clinical /remote hyperendemic trasnmission explanatory covariate coefficient measurement(y). The estimation error vector would then be subsequently given by $e = \hat{x} - x$ and its MSE would be given by the trace of error covariance matrix[i.e.,] where the expectation would be taken over both x and y. When x is a scalar variable then MSE expression simplifies to $E\left\{(\hat{x} - x)^2\right\}$ (Rao 1973).

Note that MSE in SAS/GIS can equivalently be defined in other ways, when constructing a district-level time series geopredictive malaria-related regression-based risk model since $\mathrm{tr}\left\{E\{ee^T\}\right\} = E\left\{\mathrm{tr}\{ee^T\}\right\} = E\{e^T e\} = \sum_{i=1}^{n} E\{e_i^2\}.$ For instance, the MMSE estimator could be defined by the field/clinical /remote hyperendemic transmission geopredictor achieving the most minimal MSE coefficient value. Under some weak regularity assumptions, the MMSE estimator in SAS/GIS could also be uniquely defined in a district-level geopredictive malarial risk model residually forecasted estimates by $\hat{x}_{\mathrm{MMSE}}(y) = E\left\{x|y\right\}$ (www.sas.com).In such circumstances, the MMSE estimator would be the conditional expectation of x in the malaria-related risk model given the known observed values of the sampled field/clinical /remote hyperendemic transmission explanatory covariates. The MMSE estimator would then be unbiased if $E\{\hat{x}_{\mathrm{MMSE}}(y)\} = E\left\{E\{x|y\}\right\} = E\{x\}$. Further, when x is a scalar in the district-level predictive malaria-related regression-based risk model, the estimator could be constrained to be of the form $\hat{x} = g(y)$ which could then be an optimal estimator, [i.e. $\hat{x}_{\mathrm{MMSE}} = g^*(y)$,]in the model but only if $E\{(\hat{x}_{\mathrm{MMSE}} - x)g(y)\} = 0$ for all $g(y)$ enclosed in linear subspace when $\mathcal{V} = \{g(y)|g : \mathbb{R}^m \to \mathbb{R}, E\{g(y)^2\} < +\infty\}$ (see Cressie 1993). Thus, since the MSE for estimation of a random vector is the sum of the MSEs of the coordinates (Rao 1973), finding the MMSE estimator of a random vector in a SAS/GIS derived geopredictive time series malarial-related district-level model would simply entail finding the MMSE estimators of the coordinates of X separately:[e.g., $E\{(g_i^*(y) - x_i)g_j(y)\} = 0$,]for all *i* and *j*. More succinctly put, in a robust time series SAS/GIS derived district-evel predictive malaria-related regression-based risk model $E\{(\hat{x}_{\mathrm{MMSE}} - x)\hat{x}^T\} = 0$. Also, if x and y are jointly Gaussian, then the district-level predictive malaria-related regression-based risk model MMSE estimator would be linear, (i.e., the field/clinical /remote sampled hyperendemic transmission regressors would fit the form $Wy + b$ for matrix W and constant b) As a consequence, to find the optimal hyperendemic transmission residually forecasted estimator, it would be sufficient

to construct a linear MMSE estimator from an empirical dataset district-level geopredictive malaria-related regression-based risk model in SAS/GIS using a dataset of empirical sampled field/clinical /remote explanatory covariate coefficient measurement values.

The term MMSE more specifically refers to estimation in a Bayesian setting with quadratic cost function (Gilks 1996). The basic idea behind the Bayesian probabilistic approach for modeling district-level geopredictive malarial –related hyperendemic transmission oriented covariate coefficient estimation in SAS/GIS stems from practical situations where some prior information about the field/clinical/remote sampled coefficients needs to be estimated. For instance, prior information about the range that a sampled district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented parameter estimator may be modified in SAS/GIS when a new field and remote sampled observation may be made available in an ecological regressed dataset. This is in contrast to the non-Bayesian malarial district-levels time series regression-based approach like minimum-variance unbiased estimator (MVUE) in other statistical packages (STATA, R) where absolutely nothing is assumed to be known about the sampled estimators in advance. In the Bayesian district-level geopredictive time series malarial risk model approach, prior information would be captured by the prior probability density function of the sampled geoparameter estimators which would then be directly based on the Bayes theorem.

In probability theory and statistics, Bayes' theorem (alternatively Bayes' law or Bayes' rule) is a result that is of high importance in the mathematical manipulation of conditional probabilities. It is a result that derives from the more basic axioms of probability. In probability theory, the probability P of some event E, denoted $P(E)$, is usually defined in such a way that P satisfies the Kolmogorov axioms (Cressie 1993). This axiom states that if a malarialogist/experimenter , for instance, lets $Q_i$denote anything subject to weighting in a geopredictive district-level malaria-related risk model by a normalized linear scheme of weights, the sum to unity in a set W then would be based on the Kolmogorov axioms for every $Q_i$in W, which would be simply a sampled field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate coefficient measurement value $Q(Q_i)$ where the Kolmogorov coefficient weight of Q_i is such that $Q(Q_i) + Q(\overline{Q}_i) = 1$. In this model scheme $\overline{Q}_i$ would denote the complement of Q_i W. For the mutually exclusive district-level subsets Q_1, Q_2, ... in W, would then be expressed a $Q(Q_1 \cup Q_2 \cup Q_3 \cup \ldots) = Q(Q_1) + Q(Q_2) + Q(Q_3) + \ldots$Qn.

These assumptions can be summarised further for accurate district-level geopredictive time series malaria-related risk modeling.For instance, if a malariologist/experimenter lets (Ω, F, P) be a measure space in a district-level geopredictive malaria-related risk model with $P(\Omega)=1$, then (Ω, F, P) would be a probability space, with sample space Ω, event space F and a probability measure P.Thus given an ecological empirical dataset of time series district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients, a field/clinical/remote sampling event E in a sample space S would either be finite with N coefficient elements or countably infinite with $N = \infty$elements. Thereafter, a malarialogist/experimenter could write $S \equiv \left(\bigcup_{i=1}^{N} E_i\right)$, and quantitate $P(E_i)$ if so desired, which then would be the probability of the field/clinical/remote sampling event E_i, being spatiotemporally defined such that $0 \le P(E_i) \le 1$ in the risk model forecasted estimes . By so doing, $P(E_1 \cup E_2) = P(E_1) + P(E_2)$ would be subsequently derived from the time series district-level malaria-related model derivatives where E_1 and E_2 are mutually exclusive. Further, the countable additivity in the model estimates could be defined as $P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$for $n = 1$, 2, ..., N I which then could also be employed to generate the hyperendemic trasnmission oriented covariate coefficient probablity distribution where E_1, E_2, .. would be mutually exclusive (i.e., $E_1 \cap E_2 = \emptyset$).

An alternative approach to formalising probability in a geopredictive malaria-related model may be given by Cox's axioms and functional equations. These models are based on the plausibility of a proposition for determining the plausibility of the proposition's negation; either by decreasing or increasing empirical sampled data(e.g., seasonal-sampled district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients values). Because "a double negative is an affirmative", this can be performed for a robust time series geopredictive district-level malaria-related risk model by a functional equation $f(f(x)) = x$, whereby the

function *f* that maps the probability of a proposition to the probability of the proposition's negation is an involution, (i.e., it is its own inverse). By so doing, the plausibility of the conjunction [*A* and *B*] of two propositions *A*, *B*,in the malarial-related risk model would depend only on the plausibility of *B* and that of *A* given that *B* is true. From this robust residually forecasted field/clinical/remote sampled malarial-related hyperendemic transmission oriented covariate coefficients estimates would be derived which would be able to infer whether the conjunctions of plausibilities tested in the risk model is associative. Because of the associative nature in propositional logic, this becomes a functional equation whereby the function of *g* in the district-level malarial risk model residual forecasts can be described as $P(A \text{ and } B) = g(P(A), P(B|A))$ which in theory is an associative binary operation. All strictly increasing associative binary operations performed on the district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients measurement values would thereafter be isomorphic to multiplication of other sampled coefficient values in the interval [0, 1].

For instance, suppose [*A* and *B*] is equivalent to [*C* and *D*] in a time series SAS/GIS derived district-level geopredictive malarial-related hyperendemic transmission oriented risk model. If a malarialogist/experimenter then acquires new sampled field /clinical/remote information *A* and then acquires further new information *B*, and then updates all probabilities each time in the empirical sampled parameter estimator dataset, the updated probabilities would be the same as if the malarialogist/experimenter first acquired new information *C* and then acquired further new information *D*. In view of the fact that multiplication of probabilities can be taken to be ordinary multiplication of the seasonal sampled district-level time series field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients measurement values, this data then subsequently could be regressed using a functional equation $y\, f\left(\frac{f(z)}{y}\right) = z\, f\left(\frac{f(y)}{z}\right)$ (see Cressie 1993). Further, Cox's theorem implies that any plausibility district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented risk model that meets the postulates would be equivalent to the subjective probability model, whereby the residually forecasted hyperendemic transmission oriented estimates can be converted to the probability estimates by rescaling.

There are other implications of Cox's postulate for geopredictive time series malaria-related district-level hyperendemic transmission oriented risk modeling. For instance, the laws of probability derivable from these postulates would be based on $w(A|B)$ whereby the "plausibility" of the proposition *A* (e.g. finding a high density count district-level sampled anopheline aquatic larval habitat,) given *B* (e.g.,district-level rainfall event), and *m* (low humidity) is some positive estimator measurement value . As such, A^C would then represent the absolute complement of *A* in the risk model residual forecasts statistically targeting the significant field/clinical/remote sampled hyperendemic transmission oriented predictors. Thereafter the forecasts certainty could be represented by $w(A|B) = 1$, $w^m(A|B) + w^m(A^C|B) = 1$ and/or $w(A, B|C) = w(A|C)\, w(B|A, C) = w(B|C)\, w(A|B, C)$.

It is important to note that the Cox postulates can only imply general properties in a robust SAS/GIS derived geopredictive time series district-level malarial-related risk model. These are equivalent to the usual laws of probability assuming some conventions, namely that the scale of the district-level time series field/clinical/remote sampled hyperendemic transmission oriented covariate coefficient measurement values derived from the empirical sampled dataset may be log transformed from zero to one, and the plausibility function can be conventionally denoted *P* or Pr, which then would be equal to w^m. By so doing, the malariologist/experimenter would have the ability to measure probabilities from one to infinity, with infinity representing certain falsehoods in the malarial risk model forecasts. With these conventions, a malarialogist/experimenter could also employ the laws of probability in a more familiar form when constructing a robust district-level geopredictive time series malarial-related risk model. As such, certain truths may be then accurately represented by $\Pr(A|B) = 1$ in the regressed empirical dataset of the district-level georferenced field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients as their certain falsehoods would be quantitated. For instance, the residual forecasted data could be quantitated by $\text{PR}(A|B) = 0$. $\Pr(A|B) + \Pr(A^C|B) = 1$ abds $\Pr(A, B|C) = \Pr(A|C)\, \Pr(B|A, C) = \Pr(B|C)\, \Pr(A|B, C)$. This residual uncertainty equation may also be considered for determining the precise statistical significance of specific seasonal-sampled hyperendemic transmission oriented covariate coefficients in the empirical sampled dataset.

Thereafter, the SAS-GIS constructed geopredictive malarial-related hyperendemic transmission oriented model residual forecasts derived from a spatiotemporal regression-based uncertainty matrix would yield, countable additivity probability estimates. The measure-theoretic formulation of Kolmogorov assumes that a probability measure is countably additive(Cressie 1993). For instance, suppose a malarialogist/experimenter lets $P_\theta$be a family of probability measures indexed by $\theta \in \Theta$ in a geopredictive time series district-level malaria-related risk model. For notational convenience, he or she may then assume $0 \in \Theta$, so that $P_0$is one of the probability measures in the empirical sampled dataset. Then $L(\theta) = E_0 \left[\frac{dP_\theta}{dP_0} \mid \mathcal{X}\right]$ would be the likelihood function, where the σ-algebra $\mathcal{X}$ would be able to describe the possible district-level field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented observations and E_0 would then denote expectation with respect to the measure P_0.

In such circumstances a malarialogist/experimenter could also consider the special case where the probability measure is described by a pdf [e.g. $p(x, y; \theta)$] in the SAS/GIS derived geopredictive time series field/clinical/remote sampled malaria-related data. Here, x would be a district-level hyperendemic transmission oriented covariate coefficient real-valued random variable observed, y would be a real-valued unobserved random variable, and θ would index the family of joint pdfs. The likelihood function when there is a "hidden variable" $y$is usually defined as $\theta \mapsto p(x; \theta)$where $p(x; \theta)$is the marginalised pdf obtained by integrating out the unknown variable$y$, that is, $p(x; \theta) = \int_{-\infty}^{\infty} p(x, y; \theta)\, dy$ (see Griffith 2003). As such, the malarialogist/experimenter would qunatiate the variables and their forecast uncertainity if the likelihood function equals $L(\theta)$when $\mathcal{X}$is the σ-algebra in each regressed district-level field/clinical/remote sampled explanatory hyperendemic transmission oriented random variable$x$. By so doing, the correspondence between the sampled district-level malarial measure and the pdf would be $P_\theta(A) = \int_A p(x, y; \theta)\, dx\, dy$for any measurable empirical dataset when $A \subset \mathbb{R}^2$ which would in actuality be the probability that (x, y)lies in A. In this case, the Radon-Nikodym derivative $\frac{dP_\theta}{dP_0}$ would simply be the ratio $\frac{p(x,y;\theta)}{p(x,y;0)}$.

In mathematics, the Radon–Nikodym theorem is a result in measure theory which states that, given a measurable space (*X*,Σ), a σ-finite measure νon (*X*,Σ) is absolutely continuous with respect to a σ-finite measure μon (*X*,Σ). If such are the circumstances in a SAS/GIS derived geopredictive district-level malaria-related model, then there would be a measurable function *f* on *X* .Thus, taking district-level field/clinical/remote sampled explanatory hyperendemic transmission oriented random variable values in [0,∞), such that $\nu(A) = \int_A f\, d\mu$ any measurable set *A* may be achieved. Further, the function *f* satisfying the above equality would be uniquely defined up to a μ-null set. That is, if *g* in the predictive time series district-level malaria-related risk model is another function which satisfies the same property, then *f* = *g* μ- would help resolve any latent uncertainty forecasted estimates. *f* is commonly written *dν*/*dμ* and is called the Radon–Nikodym derivative(Cressie 1993). The choice of notation and the name of the function in the geopredictive risk model would then reflect the fact that the function in the residual forecasts is analogous to a derivative in calculus in the sense that it would describe the rate of change of density of one measure with respect to another measure in the model. This would be analogous to the way the Jacobian determinant is used in multivariable integration (see Griffith 2003).

A similar theorem can be proven for signed and complex measures: namely, that if μ is a nonnegative σ-finite measure in a robust predictive time series malaria-related district-level risk model, and ν is a finite-valued signed or complex measure such that $|\nu| \ll \mu$, (i.e. ν is absolutely continuous with respect to μ), then there would be a μ-integrable real- or complex-valued field/clinical/remote sampled hyperendemic transmission oriented-related

function *g* on *X* such that for every measurable set *A*, $\nu(A) = \int_A g\, d\mu.$ This theorem would be very important in extending the ideas of probability theory from probability masses and probability densities defined over the time series empirical dataset of field/clinical/remote hyperendemic transmission oriented-related sampled explanatory covariate coefficient measurement values to probability measures defined over arbitrary sets. By so doing, a malarialogist/experimenter would be able to determine if and how it is possible to change from one sampled probability measure to another. Specifically, the malarialogist/experimenter would determine if the pdf of a field/clinical/remote sampled hyperendemic transmission oriented-related function random variable is the Radon–Nikodym derivative of the induced measure with respect to some base measure (e.g., the Lebesgue measure for continuous random variables). The derivative may then be employed to prove the existence of conditional expectation in time series malarial risk model function probability measures. The latter itself is a key concept in probability theory, as conditional probability is just a special case of it. Amongst other fields, financial mathematics uses this theorem extensively for converting actual probabilities into those of the risk neutral probabilities. Such changes of probability measure may be the cornerstone of rational interpolation of regressed district-level time series autoregressive predictive malarial-related empirical field/clinical/remote sampled data.

Further, suppose a malarialogist/experimenter lets ν, μ, and λ be σ-finite measures on the same measure space in a geopredictive SAS/GIS derived district-level time series malaria-related risk model. If $\nu \ll \lambda$ and $\mu \ll \lambda$ where ν and μ are absolutely continuous in respect to λ in the model's forecasts , then $\frac{d(\nu+\mu)}{d\lambda} = \frac{d\nu}{d\lambda} + \frac{d\mu}{d\lambda}$. If $\nu \ll \mu \ll \lambda$, then $\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$ .In particular, if $\mu \ll \nu$ and $\nu \ll \mu$, then $\frac{d\mu}{d\nu} = \left(\frac{d\nu}{d\mu}\right)^{-1}$ .If $\mu \ll \lambda$ and *g* is a μ-integrable function, then $\int_X g\, d\mu = \int_X g\frac{d\mu}{d\lambda}\, d\lambda.$ in the risk model resdiual forecasts.If ν is a finite signed or complex measure, then $\frac{d|\nu|}{d\mu} = \left|\frac{d\nu}{d\mu}\right|$. but only if μ and ν are measures over X, and $\mu \ll \nu$ (see Griffith 2003).The Kullback–Leibler divergence would then need to determined from μ to ν employing $D_{\mathrm{KL}}(\mu\|\nu) = \int_X \log\left(\frac{d\mu}{d\nu}\right)\, d\mu.$

In probability theory and information theory, the Kullback–Leibler divergence (also information divergence, information gain, relative entropy, or KLIC) is a non-symmetric measure of the difference between two probability distributions *P* and *Q*. Specifically, the Kullback–Leibler divergence of *Q* from *P*, denoted $D_{\mathrm{KL}}(P\|Q)$, is a measure of the information lost when *Q* is used to approximate *P*(see Griffith 2003). KL measures the expected number of extra bits required to code samples from *P* when using a code based on *Q* in a malarial geopredictive risk model rather than using a code based on *P* (see Jacob et al. 2009d). Typically *P* would represent the "true" distribution (e.g., district-level field/clinical/remote sampled explanatory hyperendemic transmission oriented observations), or a precisely calculated theoretical distribution of the paramter estimators. The measure *Q* typically would then represent a model, description, or approximation of *P*. For discrete probability distributions *P* and *Q*, the K–L divergence of *Q* from *P* in a district-level time series geopredictive malarial-risk model could then be defined by $D_{\mathrm{KL}}(P\|Q) = \sum_i \ln\left(\frac{P(i)}{Q(i)}\right) P(i).$ In other words, the risk model residual forecasts targeting the statistically important estimators (e.g. field/clinical/remote hyperendemic transmission oriented covariates) would be based on the expectation of the logarithmic difference between the probabilities *P* and *Q* in a robust predictive time series where the expectation would be quantitated employing the probabilities of P. The K–L divergence is only defined if, *P* and *Q* both sum to 1 and if, $Q(i) = 0$ implies $P(i) = 0$ for all i (i.e., absolute

continuity)(see Cressie 1993). Further, if the quantity $0 \ln 0$ appears in the a district-level malarial-related risk-based geopredictive equation it could then be interpreted as zero because $\lim_{x \to 0} x \ln(x) = 0$.

For distributions *P* and *Q* of a continuous district-level time series field/clinical/remote sampled explanatory hyperendemic transmission oriented real-valued random v variable, KL-divergence could also be defined to be the integral as $$D_{\mathrm{KL}}(P\|Q) = \int_{-\infty}^{\infty} \ln\left(\frac{p(x)}{q(x)}\right) p(x)\,\mathrm{d}x,$$ in SAS/GIS where *p* and *q* denote the densities of *P* and *Q*. More generally, if *P* and *Q* are probability measures over a set *X*, and *P* is absolutely continuous with respect to *Q*, then the Kullback–Leibler divergence from *P* to *Q* in a robust predictive seasonal district-level malaria-related risk model could be defined in SAS/GIS as $$D_{\mathrm{KL}}(P\|Q) = \int_X \ln\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\,\mathrm{d}P,$$ where $\frac{\mathrm{d}P}{\mathrm{d}Q}$ is the Radon–Nikodym derivative of *P* with respect to *Q*, provided the expression on the right-hand side exists. Equivalently, this can be written as $$D_{\mathrm{KL}}(P\|Q) = \int_X \ln\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\frac{\mathrm{d}P}{\mathrm{d}Q}\,\mathrm{d}Q,$$ which may be recognized as the entropy of *P* relative to *Q*. Continuing in this case, if μ is any field/clinical/remote sampled hyperendemic transmission oriented covariate coefficient measure on *X* in the risk model for which $p = \frac{\mathrm{d}P}{\mathrm{d}\mu}$ and $q = \frac{\mathrm{d}Q}{\mathrm{d}\mu}$ exist, then the Kullback–Leibler divergence from *P* to *Q* would be given as $$D_{\mathrm{KL}}(P\|Q) = \int_X p \ln \frac{p}{q}\,\mathrm{d}\mu.$$ The logarithms in these formulae would be then taken to base 2 if the district-level malaria-risk based information is measured in units of bits, or to base *e* if the information is measured in nats. Most formulas involving the KL divergence for robust time series geopredictive district-level malarial risk modeling holds irrespective of log base (see Jacob et al. 2009d).

Further, various conventions exist for referring to $D_{\mathrm{KL}}(P\|Q)$ in a robust geopredictive district-level malarial-related risk model in SAS/GIS. Often in such modeling this variable describes the divergence between *P* and *Q*; however in some cirucmstances it fails to convey the fundamental asymmetry in the relation. Sometimes the variables may be described as the divergence of *P* from, or with respect to *Q* often in the context of relative entropy, or information gain. However, for robust geopredictive time series malarial-related risk modeling, the divergence of *Q* from *P* would be the optimal language used, as this would best describe the idea that *P* is considered the underlying "true" or "best guess" distribution where expectations are calculated while *Q* is some divergent approximate distribution. Although it is often intuited as a metric or distance, the KL divergence is not a true metric — for instance, it is not symmetric: the KL from *P* to *Q* is generally not the same as the KL from *Q* to *P* (Griffith 2003). However, in a geopredictive malarial time series district-level model this would be in infinitesimal form, specifically its Hessian, would be a metric tensor (i.e. the Fisher information metric).

In mathematics, the Hessian matrix or Hessian is a square matrix of second-order partial derivatives of a function (Cressie 1993). The matrix describes the local curvature of a function of many variables. Thus given a dataset of field/clinical/remote hyperendemic transmission real-valued function $f(x_1, x_2, \ldots, x_n)$, if all second partial derivatives of *f* exist and are continuous over the domain of the function in a robust georpedictive district-level malaria-related risk model, then the Hessian matrix of *f* in SAS/GIS would be $H(f)_{ij}(\mathbf{x}) = D_i D_j f(\mathbf{x})$ where $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ and D_i is the differentiation operator with respect to the *i*th argument which would then be delineated as

International Journal of Geographic Information System
Vol. 1, No. 1, November 2013, PP: 01 - 143
Available Online at http://acascipub.com/Journals.php

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Because *f* is often clear from context, in Hessian matrices $H(f)(\mathbf{x})$ is frequently abbreviated to $H(\mathbf{x})$ (see Griffith 2003). Hessian matrices are used in large-scale optimization problems within Newton-type methods because they are the coefficient of the quadratic term of a local Taylor expansion of a function. That is,

$$y = f(\mathbf{x} + \Delta\mathbf{x}) \approx f(\mathbf{x}) + J(\mathbf{x})\Delta\mathbf{x} + \frac{1}{2}\Delta\mathbf{x}^{\mathrm{T}} H(\mathbf{x})\Delta\mathbf{x}$$

where *J* is the Jacobian matrix, which is a vector (i.e.,the gradient) for scalar-valued functions. The full Hessian matrix can be difficult to compute in practice; in such situations, quasi-Newton algorithms have been developed that use approximations to the Hessian. The best-known quasi-Newton algorithm is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm which is an iterative method for solving unconstrained nonlinear optimization problems (Cressie 1993). The Hessian matrix is related to the Jacobian matrix by $H(f)(\mathbf{x}) = J(\nabla f)(\mathbf{x})$ in a robust geopredictive malarial-related risk model (see, Jacob et al. 2011c Jacob et al. 2009d).

In SAS/GIS-related information geometry, the Fisher information metric is a particular Riemannian metric which can be defined on a smooth statistical manifold, (*i.e.*, a smooth manifold whose points are probability measures defined on a common probability space)(www.sas.com). Additionally, Fisher informatic metric can be used to tabulate the informational difference between district-level geopredictive malaria-related time series field/clinical/remote sampled hyperendemic transmission oriented measurements.(total anopheline aquatic larval habitat density counts)in SAS/GIS(see Jacob et al. 2008b.c) The metric is interesting in several respects. First, it can be understood to be the infinitesimal form of the relative entropy (*i.e.*, the Kullback–Leibler divergence); specifically, it is the Hessian of the divergence. Alternately, it can be understood as the metric induced by the flat space Euclidean metric, after appropriate changes of a variable(e.g., district-level field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate).

Interestingly, when extended to complex projective Hilbert space, the Fisher information metric becomes the Fubini–Study metric; when written in terms of mixed states, it is the quantum Bures metric (Griffith 2003). Considered purely as a matrix, the metric is known as the Fisher information matrix. Considered as a measurement technique, where it is used to estimate hidden geoparameters in terms of observed random variables, the metric is known as the observed information (Cressie 1993) Further, for $\alpha > 0, \alpha \neq 1$ the Rényi divergence of order α

$$D_\alpha(\mu\|\nu) = \frac{1}{\alpha - 1}\log\left(\int_X \left(\frac{d\mu}{d\nu}\right)^{\alpha-1} d\mu\right).$$

from μ to ν the metric could be defined by employing By so doing the assumption of σ-finiteness in the Radon–Nikodym theorem would makes he assumption that a field/clinical/remote sampled geopredictive malaria-related hyperendemic transmission oriented explanatory covariate coefficient measure μ with respect to the rate of change of ν is sigma-finite in the risk model .In mathematics , a positive (or signed) measure μ defined on a σ-algebra Σ of subsets of a set *X* is called finite if $\mu(X)$ is a finite real number (field/clinical/remote sampled malaria-related hyperendemic transmission oriented explanatory covariate coefficient)rather than ∞(Hosmer and Lemeshew 2000) The measure μ is called σ-finite if *X* is the

countable union of measurable sets with finite measure. A set in a measure space is said to have σ-finite measure if it is a countable union of sets with finite measure(Cressie 1993).

Unfortunately, there may be examples in SAS/GIS and other statistical packages where μ is not sigma-finite and the Radon–Nikodym theorem fails to hold in a model formulation. For instance, suppose a malarialogist/experimenter employs the Borel sigma-algebra on the real line for risk modeling a residual forecasted distribution of regressed field/clinical/remote empirical sampled hyperendemic trasnmission oriented covariate coefficients in SAS/GIS. In mathematics, a Borel set is any set in a topological space that can be formed from open sets (or, equivalently, from closed sets) through the operations of countable union, countable intersection, and relative complement. For a topological space X in a geopredictive malarial-related time series risk model then the collection of all Borel sets on X would form a σ-algebra term generated from the spatiotemporally quantitated sampled field/clinical/remote sampled estimators. Therefater , if a malarialogist/experimenter lets the district-level malaria-related risk counting measure, μ, of a Borel set A is defined as the number of elements of A ,when A is finite, and $+\infty$ otherwise ,the information metric will check whether μ is indeed a valid district-level measure in the model residual derivatives. This measure would not be is not sigma-finite, as not every Borel set is at most a countable union of finite sets (see Cressie 1993). Thereafter, if a malarialogist/experimenter lets ν be the usual Lebesgue measure on the quantitated Borel algebraic data then, ν would be absolutely continuous with respect to μ, since for a set A there would be $\mu(A) = 0$, but only if A is the empty set. By so doing, $\nu(A)$ would be zero in the forecasts targeting the statistically significant district-level field/clinical/remote sampled hyperendemic transmission oriented covariates. Assuming that the Radon–Nikodym theorem holds, that is, for some measurable district-level field/clinical/remote sampled hyperendemic transmission oriented covariates function, then f could be defined employing $\nu(A) = \int_A f \, \mathrm{d}\mu$ for all Borel sets. Taking A to be a singleton set, $A = \{a\}$, and using the above equality, a malarialogist/experimenter would then find $0 = f(a)$ for all the geosampled hyperendemic transmission oriented covariates coefficient measurement values a. This would imply that the function f in the district-level geopredictive malaria-related time series risk model residual forecasts based on the Lebesgue measure ν, is zero.

For finite district-level time series geopredictive malaria-related risk model forecasts measures μ and ν, the idea of residually quantitating functions f with $f \, d\mu \leq d\nu$ in SAS/GIS may be interesting. The supremum of all such functions, along with the monotone convergence theorem, then would furnish the Radon–Nikodym derivative in the risk model. The fact that the remaining part of μ is singular with respect to ν would then follow from the geosampled field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate coefficient finite measures. Once the result is established for these finite measures, extending to σ-finite, signed, and complex measures can be done naturally. For instance, suppose that μ and ν are both finite-valued nonnegative measures in an empirical sampled dataset of district-level geopredictive field/clinical/remote explanatory hyperendemic transmission oriented covariate coefficients. If a malarialogist/experimenter lets F then be the set of those measurable functions $f : X \to [0, +\infty)$ satisfying $\int_A f \, d\mu \leq \nu(A)$ for every $A \in \Sigma$,F then would not be empty in the risk model derivatives for it would contain at least the zero function. Thus, if a malarialogist/experimenter lets $f_1, f_2 \in F$, A be an arbitrary measurable set, $A_1 = \{x \in A \mid f_1(x) > f_2(x)\}$, and $A_2 = \{x \in A \mid f_2(x) \geq f_1(x)$ then $\int_A \max\{f_1, f_2\} \, d\mu = \int_{A_1} f_1 \, d\mu + \int_{A_2} f_2 \, d\mu \leq \nu(A_1) + \nu(A_2) = \nu(A),$ and therefore, $\max\{f_1, f_2\} \in F$.Now,if the malarialogist/experimenter lets $\{f_n\}$ be a sequence of functions in F such that $\lim_{n\to\infty} \int_X f_n \, d\mu = \sup_{f\in F} \int_X f \, d\mu.$,then replacing f_n with the maximum of the first n functions in the risk model, would quantitate any sequence $\{f_n\}$. Thereafter, by letting g be a function in the malarial risk model residual forecasts defined as $g(x) := \lim_{n\to\infty} f_n(x).$ the Lebesgue's monotone convergence theorem would hold when s $\int_A g \, d\mu = \lim_{n\to\infty} \int_A f_n \, d\mu \leq \nu(A)$ for each $A \in \Sigma$, and hence, $g \in F$.

International Journal of Geographic Information System
Vol. 1, No. 1, November 2013, PP: 01 - 143
Available Online at http://acascipub.com/Journals.php

In the mathematical field of real analysis, the monotone convergence theorem refers to a number of related theorems proving the convergence of monotonic sequences (sequences that are increasing or decreasing) that are also bounded (Schechter (997).. Informally, the theorems state that if a sequence is increasing and bounded above by a supremum, then the sequence will converge to the supremum; in the same way, if a sequence is decreasing and is bounded below by an infimum, it will converge to the infimum. Thus if $\{a_n\}$is a monotone sequence of field/clinical/remote district-level geopredictive malarial-related explanatory hyperendemic transmission oriented covariate coefficients (i.e., if $a_n \leq a_{n+1}$ or $a_n \geq a_{n+1}$ for every n ≥ 1), then this sequence will have a finite limit if and only if the sequence is bounded. This can be easily proven in a robust district-level geopredictive malarial-related model by proving that if an increasing sequence $\{a_n\}$is bounded above in the model then it is convergent and the limit is $\sup_n\{a_n\}$.Since $\{a_n\}$is non-empty and by assumption, it is bounded above, then, by the Least upper bound property of the time series of field/clinical/remote district-level predictive malarial-related explanatory hyperendemic transmission oriented covariate coefficient measurements, $c = \sup_n\{a_n\}$ exists and is finite. Now for every $\varepsilon > 0$, there exists $a_N$such that $a_N > c - \varepsilon$, since otherwise $c - \varepsilon$is an upper bound of $\{a_n\}$, which contradicts to $c$being $\sup_n\{a_n\}$. Then since $\{a_n\}$is increasing, $\forall n > N, |c - a_n| = c - a_n \leq c - a_N < \varepsilon$, hence by definition, the limit of $\{a_n\}$is $\sup_n\{a_n\}$. in the resdiual forecasts.This may be also revealed in the model forecasts in SAS/GIS , by the construction of g, $\int_X g\,d\mu = \sup_{f\in F}\int_X f\,d\mu.$ Now, since $g \in F$, $\nu_0(A) := \nu(A) - \int_A g\,d\mu$ defines a nonnegative measure on Σ(Cressie 1993) and supposing $\nu_0 \neq 0$; then μ would be finite if ε > 0 such that $\nu_0(X) > \varepsilon\,\mu(X)$ in the geopredictive district-level time series malarial model derivatives.

Further by letting (*P*, *N*) in SAS/GIS be a Hahn decomposition for the signed measure $\nu_0 - \varepsilon\,\mu$. every $A \in \Sigma$ will be expressed as $\nu_0(A \cap P) \geq \varepsilon\,\mu(A \cap P)$, and hence, $\nu(A) = \int_{\cdot} g\,d\mu + \nu_0(A) \geq \int_{\cdot} g\,d\mu + \nu_0(A \cap P)$and $\geq \int_A g\,d\mu + \varepsilon\mu(A \cap P) = \int_A (g + \varepsilon 1_P)\,d\mu.$ in the geopredictive time series district-level time series malarial model. In mathematics, the Hahn decomposition theorem, , states that given a measurable space (*X*,Σ) and a signed measure μ defined on the σ-algebra Σ, there exist two measurable sets *P* and *N* in Σ such that: 1) $P \cup N = X$ and $P \cap N = \emptyset$, and for each *E* in Σ such that $E \subseteq P$ one has $\mu(E) \geq 0$; that is, *P* is a positive set for μ and for or each *E* in Σ such that $E \subseteq N$ one has $\mu(E) \leq 0$; that is, *N* is a negative set for μ.

Moreover, this decomposition is essentially unique, in the sense that for any other pair (*P*', *N*') of measurable sets in SAS /GIS for fulfilling multiple conditions in residual forecasts targeting statistically important robust district-level geopredictive malarial-related risk model derivarives. By so doing, the symmetric differences $P\,\Delta\,P'$ and $N\,\Delta\,N'$ would be μ-null sets in the sense that every measurable subset of the explanatory regressed empirical geosampled field/clinical/remote field/clinical/remote hyperendemic transmission oriented covariate coefficients would have zero measure. The pair (*P*,*N*) would then be effectively quantitated as Hahn decomposition of the signed measure. Also, note that $\mu(P) > 0$ would be apparent in the district-level malarial risk model forecasts for if $\mu(P) = 0$ then ν would be absolutely continuous in relation to μ where $\nu_0(P) \leq \nu(P) = 0$, so $\nu_0(P) = 0$ and $\nu_0(X) - \varepsilon\mu(X) = (\nu_0 - \varepsilon\mu)(N) \leq 0,$ thus contradicting the fact that $\nu_0(X) > \varepsilon\,\mu(X)$.Then, $\int_X (g + \varepsilon 1_P)\,d\mu \leq \nu(X) < +\infty,$ $g + \varepsilon\,1_P \in F$ in the SAS/GIS derived model residual forecasts would satisfy $\int_X (g + \varepsilon 1_P)\,d\mu > \int_X g\,d\mu = \sup_{f\in F}\int_X f\,d\mu.$

However, robust residual quantitation of explanatory regressed georefernced field/clinical/remote sampled district-level hyperendemic transmission oriented covariate coefficients could be mispsecified in SAS if during the model construction stage the initial assumption $\nu_0 \neq 0$ is false. Then, since g would be μ-integrable, and the set $\{x \in X \mid g(x) = +\infty\}$ would be μ-null in the forecasted estimates. Therefore, if f is defined in a time series predictive district-level malaria-related SAS/GIS derived risk model as $$f(x) = \begin{cases} g(x) & \text{if } g(x) < \infty \\ 0 & \text{otherwise,} \end{cases}$$ and f has the desired properties, the residual uniqueness of the model can be determined if a malarialogist/experimenter lets $f, g : X \to [0, +\infty)$ be measurable functions satisfying $$\nu(A) = \int_A f \, d\mu = \int_A g \, d\mu$$ for every measurable set A.By so doing, $g - f$ and $$\int_A (g - f) \, d\mu = 0.$$ would be μ-integrable in the optimal residual forecasts In particular, for $A = \{x \in X \mid f(x) > g(x)\}$, or $\{x \in X \mid f(x) < g(x)\}$. It follows then that $$\int_X (g - f)^+ \, d\mu = 0 = \int_X (g - f)^- \, d\mu,$$ and so, that $(g-f)^+ = 0$ μ-would be reflected in the forecasts; the same is true for $(g - f)^-$, and thus, $f = g$ μ for targeting the statistically important field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate coefficients.

For σ-finite positive measure edit in a geopredictive district-level SAS/GIS derived malaria-related time series model, if μ and ν are σ-finite, then X can be written as the union of a sequence $\{B_n\}_n$ of disjoint sets in Σ, each of which has finite measure under both μ and ν. For each n, there is a Σ-measurable function $f_n : B_n \to [0, +\infty)$ such that $$\nu(A) = \int_A f_n \, d\mu$$ for each Σ-measurable subset A of B_n. (Griffith 2003).The union f of these functions would then be then the required function for proper regeression of the sampled field/remote/clinical explanatory hyperendemic transmission oriented covariate coefficients. As for the uniqueness in the residually targeted forecasts, since each of the f_n would be μ-almost everywhere in the derivatives, then so would be f. If ν is a σ-finite signed predictive district-level malarial-related measure, then it can be Hahn–Jordan decomposed as $\nu = \nu^+ - \nu^-$ where one of the measures is finite. Applying the previous result to these two measures, an malarialogist/experimenter would obtain two functions, $g, h : X \to [0, +\infty)$, satisfying the Radon–Nikodym theorem for ν^+ and ν^- respectively in SAS/GIS when constructing the district-level geopredictive model. Further, at least one of the measures would be μ-integrable (i.e., its integral with respect to μ is finite). It is clear then that $f = g - h$ satisfies the required properties in a robust geopredictive district-level malaria-related risk model, including uniqueness, then both g and h in the residual forecasts would also be unique. If ν is a complex measure in the targeted field/clinical/remote sampled hyperendemic transmission oriented covariate coefficients, they can be decomposed as $\nu = \nu_1 + i\nu_2$, where both ν_1 and ν_2 are finite-valued signed measures. Applying the above argument, in SAS/GIS, a malarialogist/experimenter would obtain two functions, $g, h : X \to [0, +\infty)$, satisfying the required properties for ν_1 and ν_2, respectively in the residual forecasts. Clearly, $f = g + i\,h$ is a required function in a robust SAS/GIS derived geopredictive time series district-level malaria-related risk model.

The conditional expectation with respect to X in the Radon-Nikodym derivative in SAS/GIS would then be under the distribution $p(x, y; 0)$ in the predictive malarial-related time series district-level model which would then verify that $L(\theta)$ is indeed the likelihood function in the model residual forecasts. This verification does not make $L(\theta) = E_0 \left[\frac{dP_\theta}{dP_0} \mid \mathcal{X} \right]$ any less mysterious. Instead, it can be understood directly as follows. From the definition of conditional expectation, it is straightforward to verify that $L(\theta) = \left. \frac{dP_\theta}{dP_0} \right|_{\mathcal{X}}$ in a SAS/GIS constructed district-level time series geopredictive malarial risk model employing any $\mathcal{X}$-measurable set of field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate coefficients A, $P_\theta(A) = \int_A \left. \frac{dP_\theta}{dP_0} \right|_{\mathcal{X}} dP_0$

would possess a likelihood function whereby the "probability" of the observed dataset A contain the actual observations and thus $P_\theta(A)$ would vary with θ. This would work if $P_\theta(A) > 0$but otherwise it would be necessary for the malarialogist/experimenter to determine at how $P_\theta(A)$varies when $A$is an arbitrarily small but non-negligible set centered on a true geosampled district-level field/clinical/remote hyperendemic transmission oriented observation. It may be impossible to make a perfect observation correct to infinitely many significant figures; in the risk model instead, an observation of x could signify for instance , that $1.0 \leq x \leq 1.1$, hence A can be chosen to be the event that $1.0 \leq x \leq 1.1$instead of a negligible event (e.g. $x = 1.05$). It follows from the integral representation $P_\theta(A) = \int_A \left.\frac{dP_\theta}{dP_0}\right|_{\mathcal{X}} dP_0$ that $\left.\frac{dP_\theta}{dP_0}\right|_{\mathcal{X}}$ would then be able to describe the behavior of $P_\theta(A)$as *A* shrinks down from a range of district-level time series geopredictive malarial risk model outcomes to a single outcome. Importantly, the subscript $\mathcal{X}$means $L(\theta) = \left.\frac{dP_\theta}{dP_0}\right|_{\mathcal{X}}$is $\mathcal{X}$-measurable in a SAS/GIS constructed model therefore, $L(\theta)$depends only on what is observed (e.g., district-level geopredictive field/clinical/remote-sampled covariate coefficients) and not on any other hidden variables.

Importantly, unlike non-Bayesian approach where the time series malarial-related geoparameter hyperendemic transmission oriented estimators of interest would be assumed to be deterministic, the Bayesian estimator seeks to estimate a geoparameter that is itself a random variable. Further, Bayesian probabilistic estimation can also deal with situations where the sequence of geosampled district-level observations are not necessarily independent. Thus, Bayesian estimation of district-level field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate coefficients from and empirical ecological sampled dataset would provide yet another alternative to the MVUE for optimal district-level malaria-related risk modeling. The Bayes risk, in this case, would be the quantitated posterior variance where there is no inherent reason to prefer one prior probability distribution over another for risk modeling the sampled field/clinical/remote sampled hyperendemic transmission oriented estimators. On occasion a conjugate prior is sometimes may be chosen for simplicity for Bayesian error modeling (Cressie 1993).

A conjugate prior is defined as a prior distribution belonging to some parametric family, for which the resulting posterior distribution also belongs to the same family. This is an important property, for risk modeling georferenced district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients since the Bayes estimator, as well as its statistical properties (variance, confidence interval, etc.), can all be derived from the posterior distribution. Conjugate priors are especially useful for sequential estimation, for instance where the posterior of the current district-level malaria-related district-level field/clinical/remote sampled hyperendemic transmission oriented covariate coefficients measurement value is used as the prior in the next sampled measurement value. In sequential estimation, unless a conjugate prior is used, the posterior distribution typically becomes more complex with each added measurement, and the Bayes estimator cannot usually be calculated without resorting to numerical methods (Gilks 1996).

Regardless, there are some examples of conjugate priors that may be employed in SAS/GIS for time series district-level malarial risk modeling exercises For instance, if $x|\theta$ is normal, $x|\theta \sim N(\theta,\sigma^2)$ in the Bayesian malarial model , and the prior is normal, $\theta \sim N(\mu,\tau^2)$, then the posterior may be also normal and the Bayes estimator under MSE would be given by $\widehat{\theta}(x) = \frac{\sigma^2}{\sigma^2+\tau^2}\mu + \frac{\tau^2}{\sigma^2+\tau^2}x.$ If $x_1,...,x_n$ are i.i.d. Poisson random variables $x_i|\theta \sim P(\theta)$ in the risk model and , if the prior is Gamma distributed $\theta \sim G(a,b)$, then the posterior would be Gamma distributed, and the Bayes estimator under MSE would be given by $\widehat{\theta}(X) = \frac{n\overline{X}+a}{n+\frac{1}{b}}.$ If $x_1,...,x_n$ are i.i.d. uniformly distributed $x_i|\theta \sim U(0,\theta)$, and if the prior is Pareto distributed $\theta \sim Pa(\theta_0,a)$, then the posterior in the risk

model would also be Pareto distributed, and the Bayes estimator under MSE would then be given by $\hat{\theta}(X) = \frac{(a+n)\max(\theta_0, x_1, ..., x_n)}{a+n-1}$.

The Pareto distribution with pdf and distribution function $P(x)=\frac{a\,b^a}{x^{a+1}}$ and $D(x)=1-\left(\frac{b}{x}\right)^a$ defined over the interval $x \geq b$. This may be implemented in SAS/GIS as ParetoDistribution[*k*, *alpha*]. The n th raw moment then would be $\mu_n' = \frac{a\,b^n}{a-n}$ for $a > n$, giving the first few as $\mu_1'=\frac{a\,b}{a-1}$, $\mu_2'=\frac{a\,b^2}{a-2}$, $\mu_3'=\frac{a\,b^3}{a-3}$ and $\mu_4'=\frac{a\,b^4}{a-4}$. The *n* th central moment in the geopredictive malaria-related district level field/clinical/remote sampled hyperendemic transmission oriented covariate coefficient measurement values (n), for instance would be then be quantitated employing $\mu_n=\frac{a\,b^n\,\Gamma(a-n)\,{}_2\tilde{F}_1\left(a-n,-n;1+a-n;\frac{a}{a-1}\right)}{=(1-a)^{a-n}\,(-a)^{n-a}\,a\,b^n\,B\left(\frac{a}{a-1};a-n,n+1\right)}$, for $a > n$ where $\Gamma(z)$ is a gamma function, ${}_2\tilde{F}_1(a, b; c; z)$ is a regularized hypergeometric function, and $B(z; a, b)$ is a beta function, which then would render $\mu_2=\frac{a\,b^2}{(a-1)^2\,(a-2)}$, $\mu_3=\frac{2\,a\,(a+1)\,b^3}{(a-1)^3\,(a-2)\,(a-3)}$ and $\mu_4=\frac{3\,a\,(3\,a^3+a+2)\,b^4}{(a-1)^4\,(a-2)\,(a-3)\,(a-4)}$. Given a hypergeometric or generalized hypergeometric function ${}_pF_q\left(a_1, \ldots, a_p; b_1, \ldots, b_q; z\right)$, the corresponding regularized hypergeometric function in a SAS/GIS derived geopredictive malaria-related risk model can be defined by ${}_p\tilde{F}_q\left(a_1, \ldots, a_p; b_1, \ldots, b_q; z\right) \equiv \frac{{}_pF_q\left(a_1, \ldots, a_p; b_1, \ldots, b_q; z\right)}{\Gamma(b_1)\cdots\Gamma(b_q)}$, where $\Gamma(z)$ is a gamma function. Regularized hypergeometric functions can be implemented in SAS/GIS as the functions Hypergeometric0F1Regularized[*b*, *z*], Hypergeometric1F1Regularized[*a*, *b*, *z*], Hypergeometric2F1Regularized[*a*, *b*, *c*, *z*], and in general, HypergeometricPFQRegularized[{*a1*, ...*ap*}, {*b1*, ..., *bq*}, *z*]. The mean, variance, skewness, and kurtosis thereafter in the malaria-related geopredictive time series risk model could be quantitated as $\mu=\frac{a\,b}{a-1}$, $\sigma^2=\frac{a\,b^2}{(a-1)^2\,(a-2)}$, $\gamma_1=\sqrt{\frac{a-2}{a}}\,\frac{2\,(a+1)}{a-3}$, $\gamma_2=\frac{6\left(a^3+a^2-6\,a-2\right)}{a\,(a-3)\,(a-4)}$.

NDVI geoparameter estimators were then generated in ArcGIS using the QuickBird data. For the malaria-related NDVI value, the total areas were determined for specific surface vegetation-cover classes associated to the geosampled district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate (e.g., floating and emergent vegetation LULC and for a hybrid vegetation-cover class consisting of open water and submersed vegetation LULC). The vegetation- canopy cover class comprising the largest total area was then assigned to NDVI value. The equation ArcGIS Image Server employed to generate the output (i.e., NDVI = arctangent ((IR – R)/(IR+R)) produced a single-band dataset from the satellite data product. The differential reflection in the red and infrared (IR) bands from the imager enabled quantifying density and intensity of green vegetation growth at various LULC district-level sample sites using the spectral reflectivity of solar radiation

We then generated a correlation error matrix in ArcGIS to determine the accuracy of the district-level geopredictive malaria-related vegetation-related LULC predictors. The row in the matrix represented the QuickBird NDVI geoparameter estimators constructed from the satellite data products, while the columns represented the reference data (ground truth, *in-situ* sampled data). We generated measures of thematic accuracy including overall classification accuracy and percentage of omission. In this research, the vegetated canopy-related district-level explanatory hyperendemic transmission oriented parameter estimator percentage of omission was based on the percentage of QuickBird pixels that were in a given NDVI class but that which were not identified. Additionally, we generated the commission error which indicated pixels that were not identified, but were within a particular NDVI

class. A residual normalized uncertainty matrix output was then generated using the user's and producer's accuracy-combined measures and the field-verified estimates of the NDVI thematic vegetation canopied geopredictive district-level variables. As primary accuracy measures, these relative entropy change measures were normalized by the arithmetic mean of the vegetated canopied entropies generated by the proxy map variables. The overall classification accuracy of the QuickBird NDVI thematic maps revealed the highest level of accuracy (i.e., 93% with a kappa value of .89).The spatial variation of the NDVI geoparameters were then measured for determining surfaces of the georeferenced canopy vegetated district level estimators. It was found that emissivity was highly correlated with QuickBird NDVI after logarithmic transformation, with an average correlation coefficient of R = 0.94 throughout the study site districts.

The results of this analysis revealed that that there is more variability in district-level vegetated canopied malarial-related explanatory hyperendemic transmission oriented covariate coefficients across a single field . As more classes were added, the ability to determine areas that were healthier from those that were less healthy was enhanced. However, there may be certainly a limit to the fidelity that is needed to be explored in robust geopredictive district-level malarial risk analyses by comparing field measurements of plant productivity to QuickBird NDVI values. By so doing, a local abatement district manager could determine what range of NDVI values and at what time in the phonological development cycle correspond to areas of less productive immature growth for georeferenced seasonal anopheline mosquito aquatic larval aquatic habitats, for instance. Once identified, these geolocations could be easily seasonally mapped based on NDVI signatures. With QuickBirds ability to collect imagery over large areas on a daily basis, its value in high precision vegetated canopied district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented explanatory covariate coefficients measurements mapping cannot be understated.

In this research a traditional spatiotemporal linear hyperendemic district-level geopredictive malarial transmission-oriented risk-based regression model was constructed in PROC REG employing the seasonal-sampled district-level field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients measurements. The model was described as $\{y_i, x_{ij}\}_{i=1..n,j=1..p}$ for forecasting the district-level statistical units. The response values was placed in a vector $Y = (y_1, ..., y_n)'$in the model and the spatiotemporal-field/clinical/remote sampled explanatory covariate coefficient measurement values were placed in the design matrix $X = [[x_{ij}]]$ in SAS/GIS, where x_{ij} was the value of the jth variable for the ith sampled district-level statistical sample unit. The model assumed that the conditional mean of Y provided X was a linear function of X, whereas, the conditional variance of Y was provided by X a known matrix Ω. In this research this relationship was written as$Y = X\beta + \varepsilon, \quad \mathrm{E}[\varepsilon|X] = 0, \ \mathrm{Var}[\varepsilon|X] = \Omega.$Here β was a vector of unknown district-level "regression based coefficients" in the linear hyperendemic malarial transmission-oriented risk-based model that had to be estimated from the seasonal–sampled district-level data. We noted that when b was a candidate estimate for β in the model, then the residual vector for b was quantitated by $Y - Xb$.

The straightforward derivation of the linear district-level malarial model, however, from the negative binomial probability distribution function did not equate with the Poisson–gamma mixture-based version of the negative binomial. Rather, canonical link and inverse canonical link were converted to log form. A GLM-based negative binomial was then produced that yielded identical geoparameter estimators based on the regressed field/clinical/remote explanatory hyperendemic transmission-oriented covariate coefficient to those calculated by the mixture-based model. As a non-canonical linked model however, the standard errors did differ slightly from the mixture model. A ML estimator in SAS/GIS employed an observed information matrix to produce standard errors. The GLM algorithm produced standard errors, based on the expected information matrix using the difference in standard errors in the negative binomial analyses. The GLM negative binomial algorithm was, thereafter, amended to allow production of standard errors based on the geosampled district-level field/clinical/remote sampled malarial related hyperendemic transmission oriented explanatory covariate coefficients measurements. The amended GLM-based negative binomial produced identical estimates and standard errors to that of the mixture-based negative binomial analyses. The log-negative binomial data was then imported into an ArcGIS database, using the spatial analytical tools in SAS/GIS.

An autoregressive model specification was then constructed in SAS/GIS to describe the variance uncertainty estimates in the regressed district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented covariate coefficients. The Jacobian generalized the gradient of a scalar-valued function of multiple georefenced district-level geopredictor variables which were generalized by the derivative of a scalar-valued function. A more complex specification was then posited by generalizing binary indicator variables. We used F: $R^n \rightarrow R^m$ as a function from Euclidean n-space to Euclidean m-space which was generated using the linearized seasonal- sampled district-level hyperendemic transmission oriented explanatory covariate coefficients. The function was provided by m covariate (i.e., component functions), $y_1(x_1, xn)$, $y_m(x_1, xn)$. The partial derivatives of all these functions were organized in an m-by-n matrix, whereby the Jacobian matrix J of F, followed $J = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$.This matrix was denoted by $J_F(x_1,..., x_n)$ and $\frac{\partial(y_1,\ldots,y_m)}{\partial(x_1,\ldots,x_n)}$. The i th row (i = 1,..., m) of this matrix was the gradient of the i^{th} component function y_i:(∇ y_i). In this analyses p was a geosampled district-level malarial-related estimator in R^n where F was differentiable at p. As such, its derivative was given by $J_F(p)$. The model described by $J_F(p)$) was the best linear approximation of F near the sampled district-level point p, in the sense that: $F(x) = F(p) + J_F(p)(x - p) + o(\|x - p\|)$. The spatially adjusted models identified the clustering patterns of the geosampled district-level field/clinical/remote sampled hyperendemic transmission oriented explanatory covariate coefficients measurements in the seasonal empirical ecological dataset. The residually forecasted estimates accounted for all conditional heteroskedastic error terms in the model.

We then attempted a method of generalized differencing for quantitating the residual autocorrelation error coefficients in our district-level seasonal regression-based risk map for remotely targeting areas of district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented explanatory covariate coefficients measurements. By so doing, the generalized differencing allowed a transformed equation to be developed from which optimal linear unbiased uncertainty estimates obtained from an OLS. A full specification of the first order autocorrelation error [$Y_T = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \rho \varepsilon_{t-1} + VT$ (eqn 4.4)] was then generated in ArcGIS Geostatistical Analyst. In this research we lagged the equation and multiplied it by ρ which rendered $\rho Y_{t-1} = \rho \beta_0 + \rho \beta_1 X_{1t-1} + \rho \beta_2 X_{2t-1} + \rho^2 \varepsilon_{t-2} + \rho v_{t-1}$ (eqn 4.5). We subtracted Eqn 4. 5 from Eqn 4.4. which then yielded $Y_t - \rho Y_{t-1} = \beta_0 - \rho \beta_0 + \beta_1 X_{1t} - \rho \beta_1 X_{1t-1} + \beta_2 X_{2t} - \rho \beta_2 X_{2t-1} + \rho \varepsilon_{t-1} - \rho^2 \varepsilon_{t-2} + v_t - \rho v_{t-}$ By so doing, however, $Y_t - \rho Y_{t-1} = \beta_0(1 - \rho) + \beta_1(X_{1t} - \rho X_{1t-1}) + \beta_2 (X_{2t} - \rho X_{2t-1}) + v_t$ could not be quantitated. Theoretically, the model should have rendered $\rho \varepsilon_{t-1} - \rho^2 \varepsilon_{t-2} + \mathbf{v_t} - \rho v_{t-1}$ which then would have been equivalent to $\boldsymbol{\varepsilon_t} - \rho \varepsilon_{t-1}$ which, in turn, would have represented the first order error in the geopredicted district-level malarial-related autoregressive risk model. In previous malarial-related research (Jacob et al. 2011c and Jacob et al. 2009d) quantified multiple seasonal-sampled geopredictive malaria-related covariate coefficients employing :

$$Y_t^* = \beta_0^* + \beta_1 X_1^* + \beta_2 X_2^* + v_t$$

where

$$Y_t^* = (Y_t - \rho Y_{t-1})$$
$$\beta_0^* = \beta_0(1-\rho)$$
$$X_1^* = (X_{1,t} - \rho X_{1,t-1})$$
$$X_2^* = (X_{2,t} - \rho X_{2,t-1})$$

In this research we could not determine the eigenvectors $[e_1, \ldots, e_n]_{Lag}$ to filter spatial autocorrelation in the generic autoregressive model from each sampled district-level estimator. Ordinarily, the next step would have been to attempt to identify suitable and parsimonious subsets of eigenvectors E_{SAR} or E_{Lag} from either sampled model specification (2.1) or (2.2). This would have identified a particular subset of the geopredictive autoregressive

district-level malarial risk model error matrix eigenvectors which may have been delineated as suitable if the residuals $\hat{\varepsilon}$ of the resulting spatially filtered model residual forecasts estimates became stochastically independent with respect to the underlying sampled spatial structure V.

As such, the dependency in our model was then analyzed using random effect specifications. Random effects model specifications address samples for which observations are selected in a highly structured rather than random way, and involve repeated measures in frequentist analyses (Haight 1967). Commonly, time-series malarial related data furnishes a repeated measures context (Jacob et al. 2005b). An average for each time series exists for a space-time dataset (Griffith 2003). In this research, this average ignored both spatial and serial error correlation coefficients in the space-time district-level malaria-related series. A random effects model essentially works with these averages, adjusting them in accordance with the correlational structure latent in their parent space-time series, as well as their simultaneous estimation (Griffith 2003). Instead, in this research, the random effects model specification was achieved by fitting a distribution with as few geoparameter district-level field/clinical/remote sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients estimators as possible (e.g., a mean and a variance for a bell-shaped curve), rather than *n* means (i.e., fixed effects) for the *n* sampled district-level geolocational attributes. Consequently, we were able to distinguish a relationship which existed between the time series means and the random effects. This random effects specification included *n* indicator variables, each for a separate specific district local intercept where one local intercept was arbitrarily set to 0 to eliminate perfect multicollinearity within the global mean. Here, the local mean for district 112 was set to 0. The estimated global mean was -3.6723, the mean of the random effects term was -0.0010, and the mean of the local means was 0.4837; the sum of these three values was 3.1876, which in this research was exactly the same as the random effects global mean. The scatterplot of the random effects versus the local intercepts corresponded to a straightly line with no dispersion about it.

By using a random effect specification we were able to determine malarial prevalence at the district-level throughout the Ugandan study site. The following equation was thereafter employed to forecast the expected value of the prevalence of malaria for district: prevalence = $\exp[-3.1876 + (\text{random effect})_i]$. Although, the number of degrees of freedom in our models where so large that the CIs had a width approaching 0, we were still able to successfully construct a predictive district-level spatiotemporal malaria risk model. We then added geomorphological land cover statistics derived from the DEM and band radiance estimates to create more robust indices based on our primary model estimates.

This model was then based on tabulated accessory geomorphological hyperendemic transmission oriented estimators and band radiance values at the district-level, a model random effects term, and a regressed district-level prevalence count. The Poisson mean response specification was then: $\mu = \exp[a + re + LN(population)]$, Y ~Poisson(mu) +DEM (zonal statistic). The mixed-model estimation results included: a = -3.1876 re ~ n(0, s^2) mean re = -0.0010 s^2 = 0.2513 where P(S-W) = 0.0005 and the Pseudo-R^2 = 0.3103. By using a random effect specification and a DEM district-level covariate along with the quantitated band radiance data at the district-level, we were able to forecast prevalence at the district-level throughout the study site. The goodness-of-fit feature implied that although the random effects term combined with the DEM statistics could be used for geopredictive purposes cases, as counts, was still the response variable, supporting the use of a Poisson probability model specification. In order to describe prevalence at the district-level at the study site the following equation was then generated in SAS/GIS to forecast the expected value of the prevalence of malaria for each district at the study site: district: prevalence =exp[-3.1876 + (random effect)i] +mean DEM. For instance, the forecasted value for Abim was exp(-3.1876 + 0.89982) = 0.1015 , 95% CI = 90.10114, 0.10185) + a log-transformed DEM mean value of 1189.9. This random intercept represented the combined effect of spatiotemporal collected data (e.g., median rainfall) that caused districts to be more prone to the malaria prevalence than other districts.

In this research we revealed that quantiating geospatial autoregressive correlation in ArcGIS employing district-level SAS-derived malaria-related regression-based field/clinical/remote hyperendemic transmission oriented attributes can be expressed in a Pearson product–moment correlation coefficient formula. Pearson's correlation coefficient between two geosampled district-level hyperendemic transmission oriented variables can then be defined as the covariance of the two variables divided by the product of their standard deviations. Further, for identifying outliers in a spatiotemporal dataset of autoregressively predicted malarial-related residual forecasts

univariate analyses (e.g., show min, max, skewness and kurtosis), in SAS/GIS may help generate scatterplots, residual plots. Normal probability plots, regression for outlier diagnostics including standardized residuals, Hat diagonals, Cook's D stats, etc.) can then be examined (e.g., by using the "INFLUENCE" and "R" options in PROC REG's MODEL statement).

Importantly fdasave, fdause, can describe and convert time series district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive uncertainty estimators datasets of to and from statistical format for new device applications [e.g., Personal Digital Assistants (I-phone)] using SAS XPORT Transportformat. The primary intent of these commands is to assist malarialogists/experimenters for making data submissions tbut the commands are general enough for use in transferring data between SAS/GIS and Stata. To save the data in memory in the format, a malarialogist/experimenter simply need type. fdasave *filename* fdasave *filename*. Any district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive uncertainty estimators can then be labeled using the rename option. In any case, Stata will create *filename*.xpt as an XPORT file containing the data and, if needed, will also create formats.xpf—an additional XPORTfile—containing the value-label definitions. These files can be easily read into SAS/GIS.To read a SAS XPORT Transport residually forecasted district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive uncertainty estimator file into Stata the malarialogist/experimenter would simply have to ,type. fdause *filename* Stata will read into memory the XPORT file *filename*.xpt containing the sampled district-level data and, if available, will also read the value-label definitions stored in formats.xpf or FORMATS.xpf. The residually forecasted district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive uncertainty estimators will then be part of the contents of a SAS XPORT Transport file.

Further, fdasave can overwrite existing *filename*.xpt, formats.xpf, and *filename*.sasfiles. Thereafter vallabfile(xpf | sascode | both | none) can be employed to specify whether and how time series district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive autoregressive uncertainty estimators value labels are to be stored. SAS XPORT Transport files do not really have value labels (www.sas.com). In preparing geopredictive time series district level malaria-related datasets for submission, value-label definitions should be provided in one of two ways: 1. In an additional SAS XPORT Transport file whose data contain the value-label definitions 2. In a SAS command file that will create the value labelsfdasave can create either or both of these files. The vallabfile(xpf) can then specify that value labels be written into a separate SAS XPORT .Thus, fdasave will creates two files during the geopredictive time series modeling stage 1) *filename*.xpt, containingthe time series district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive autoregressive uncertainty estimators data, and 2)formats.xpf, containing the value labels. Fortunately no formats.xpf file is created if there are no value labels. SAS-based malarialogists/experimenters can easily use the resulting .xpt and .xpf XPORT files (See http://www.sas.com/govedu/fda/macro.html for SAS-provided macros for reading the XPORT files.

The SAS macro from exp() will read the XPORT files into SAS if so desired thereafter. By so doing, the SAS macro to exp() t will create XPORT files. When obtaining the macros, the malarialogist/experimenter must remember to save the macros at SAS's web page as a plain-text file. If the SAS macro file is saved as C:\project\macros.mac and the files mydat.xpt formats. xpf created by fdasave would be in C:\project\. The following SAS commands would then create the corresponding SAS/GIS dataset and format library and list the time series district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive data using SAS commands including "C:\project\macros.mac" ;%from exp(C:\project, C:\project) ;libname library 'C:\project' ;
data _null_ ; set library.mydat ; put _all_ ; run ; proc print data = library.mydat ;and,quit. Theereater vallabfile will specify which value labels can be written into a SAS command file(i.r., ,*filename*.sas), containing SAS proc format and related commands. For instance, fdasave may create two files when constructing a robust predictive malarial-related risk model: *filename*.xpt, containing the time series district-level field/clinical/remote sampled hyperendemic transmission oriented geopredictive autoregressive estimators and *filename*.sas, containing the value labels. Additionally, SAS-based malarialogists/experimenters may wish to edit the resulting *filename*.sas file to change the "libname datapath" and "libnamexptfile xport" lines to correspond to a specific district-geolocation. Fortunately fdasave will set the district-level geolocation to the current working directory at the time fdasave was issued. No .sas file will be created if there are no value labels(www.sas.edu).

Alternatively,Vallabfil e(both) may specify that both the actions described above be taken while other filesare created: *filename*.xpt, containing the time series district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented predictive data; formats.xpf, containing the value labels in XPORT format; and *filename*.sas, containing the value labels in SAS command-file format.vallabfile(none) specifies that value-label definitions not be saved.

SAS XPORT Transport file may contain one or more separate time series district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented geopredictive data datasets, known as members.It is rare for a SAS XPORT Transport file to contain more than one member. Seehttp://support.sas.com/techsup/technote/ts140.html for the SAS technical document describing the layout of the SAS XPORT Transport file.A SAS XPORT time series district-level hyperendemic transmission oriented geopredictive data dataset (member) however is subject to certain restrictions:1. The dataset may contain only 9,999 variables. 2. The names of the field/clinical/remote sampled hyperendemic transmission oriented variables and value labels may not be longer than eight characters and are case insensitive;(e.g., myvar, Myvar, MyVar, and MYVAR are all the same name.)3. Variable labels may not be longer than 40 characters. The contents of a sampled time series district-level field/clinical/remote hyperendemic transmission oriented geopredictive variable may be numeric or string:a. Numeric variables may be integer or floating but may not be smaller than 5.398e– 79 or greater than 9.046e+74, absolutely. Numeric variables may contain missing which may be ., . , .a, .b, . . . , .z. b. String variables may not exceed 200 characters. String variables are recorded in a"padded" format, meaning that, when variables are read, it cannot be determined whether the variable had trailing blanks.5. Value labels are not written in the XPORT dataset. Therefor suppose a malarialogist/experimenter uses the variable 0 and 1, for representing district-sampled anopheline aquatic habitat larval data where the values are labeled as(0=0 larval density count , and 1> everything 0 count). When the dataset is written in SAS XPORT Transport format, the malarialogist/experimenter would record that the variable label as associated with the larval count variable, but this may not be recorded as the association with the value labels 0 and 1. Value-label definitions are typically stored in a second XPORT dataset or in a text file containing SAS commands (www.sas.com).Instead the malarialogist/experimenter may use the vallabfile() option of fdasave to produce these datasets or files. By so doing, Value labels and formats can then be recorded in the same position in an XPORT file, meaning that names corresponding to formats can be used in SAS/GIS.

SAS/GIS software provides an interactive GIS within the SAS System enables viewing the sampled malarial data in its spatial context (www.esri.com). The ODS GRAPHICS of PROC REG in SAS/GIS, for instance, can automatically produce residually forecasted field/clinical/remote hyperendemic transmission-oriented district-level outlier diagnostic plots, Thereafter, a malarialogist/experimenter can use the free outlier attribute data in SAS/GIS as themes for layers within the module for generating robust risk-based district-level estimators . Under the null hypothesis of no autocorrelation, this data would be asymptotically distributed as χ^2 with *k* degrees of freedom. Responses to nonzero autocorrelation can then be devised including GLS and the Newey–West estimator (Heteroskedasticity and Autocorrelation Consistent (HAC) in various SAS/GIS-related modules (e,g., PROC NL MIXED) for deriving Pearson's correlation coefficients between any two tiem series district-level hyperendemic malarial transmission-oriented risk-based model variables which can then be defined as the covariance of the two variables divided by the product of their standard deviations.

By graphically portraying the relationship between two quantitative variables measured for the same district-level hyperendemic malarial transmission-oriented risk-based forecasted model observation, a scatterplot in SAS/GIS may then relate to the numerical values rendered by a correlation coefficient formula. This would be essential for ascertaining a viable residual diagnostic predictive error estimator within the framework of a time series hyperendemic transmission-oriented ARIMA malarial-related model as the residuals would be tabulated employing an average of the sampled numerical specifications between the georeferenced specific estimators for defining all possible pairs of sampled district-level geolocations. But since these time-series are unobservable, the assumption invoked would be exchangeability whereby, the set of time series can be permuted without affecting results in the geopredicted SAS/GIS constructed district-level malarial model residual forecasts.

The order in which a time series mechanism generates the field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate coefficients measurement indicator values across an interpolated spatiotemporal

district-level autocovariate regression-based risk map would thus be irrelevant. Instead district-level areas with statistically higher transmission rates, (e.g., positive correlated clusters) based on seasonal-sampled parasitological indicators (e.g., prevalence rates) in SAS/GIS can then be mathematically calculated and precisely targeted using a robust eigendecomposition spatial filter algorithm. By so doing, autocorrelation coefficients representing district-level time series malarial-related residually forecasted epidemiological data may be quantitated for spatially adjusting district –level seasonal geopredictive SAS/GIS hyperendemic transmission-related risk map data feature attributes in geospace.

Further, if the district-level seasonal predictive malarial model regression errors u_i are uncorrelated in SAS/GIS, the residually forecasted estimates, targeting the district-level hyperendemic transmission-oriented explanatory covariate coefficients would be robust. Even if the residual forecasts have distinct variances σ_i^2, then $\Sigma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ can be estimated with $\hat{\sigma}_i^2 = \hat{u}_i^2$ in SAS/GIS. This would provide a heteroskedasticity-consistent estimator [e.g., $v_{HCE}[\hat{\beta}_{OLS}] = (X'X)^{-1}X' \operatorname{diag}(\hat{u}_1^2, \ldots, \hat{u}_n^2)X(X'X)^{-1}$]. For instance, the heteroscedasticity consistent covariance matrix estimator (HCCME), also well-known as the sandwich, or empirical covariance matrix estimator, has been popular in recent years in SAS/GIS since it gives the consistent uncertainty estimation of the covariance matrix of the sampled geoparameter estimates even when the heteroscedasticity structure might be unknown or misspecified.

White (1980) first proposed the concept of HCCME, known as HC_0. However, the small-sample performance of HC_0 was not a powerful uncertainty estimator in seasonal geopredictive time series models. Davidson and MacKinnon (1993) then introduced more improvements of HC_0, namely HC_1, HC_2 and HC_3, with the degrees of freedom or leverage adjustment. Cribari-Neto (2004) then proposed HC_4 to deal with cases that have points of high leverage.

Presently, HCCME in SAS/GIS can be expressed in the following general "sandwich" form: $\Sigma = B^{-1}MB^{-1}$ where B, which stands for "bread", which is the Hessian matrix and M, which stands for "meat", which is the outer product of gradient (OPG) with or without any adjustment. The Hessian matrix is the square matrix of second-order partial derivatives of a function; that is, it describes the local curvature of a function of many variables (Cressie 1993). For HC_0, M is the OPG without adjustment; that is, $M_{\mathrm{HC0}} = \sum_{t=1}^{T} g_t g_t'$ where T for instance is the district level predictive hyperendemic transmission oriented malarial-related data sample size and g_t is the gradient vector of tth observation. For HC_1, M then would be the OPG with the degrees of freedom correction; that is, $M_{\mathrm{HC1}} = \frac{T}{T-k}\sum_{t=1}^{T} g_t g_t'$ where k is the number of sampled district-level geoparameters. For HC_2, HC_3, and HC_4, the adjustment then would be related to leverage, namely, $M_{\mathrm{HC2}} = \sum_{t=1}^{T} \frac{g_t g_t'}{1-h_{tt}}$ $M_{\mathrm{HC3}} = \sum_{t=1}^{T} \frac{g_t g_t'}{(1-h_{tt})^2}$ $M_{\mathrm{HC4}} = \sum_{t=1}^{T} \frac{g_t g_t'}{(1-h_{tt})^{\min(4, Th_{tt}/k)}}$ .The leverage h_{tt} can then be defined as $h_{tt} \equiv j_t'(\sum_{t=1}^{T} j_t j_t')^{-1} j_t$, where j_t is defined as follows: For an OLS geopredictive district-level malaria-related time series risk model, j_t would then be the tth observed field/clinical/remote-related regressors in column vector form. For an ARIMA error model, j_t would then be the derivative vector of the tth residual with respect to time series district-level explanatory hyperendemic transmission oriented geoparameters. Thereafter, a generalized method of moments (GMM) framework can be constructed in SAS/GIS for further quantitation of the geosampled explanatory hyperendemic transmission oriented covariate coefficients.

Generalized method of moments is a general estimation principle. A generic method for estimating geoparameters in SAS models is GMM (Cressie 1993). The method in SAS/GIS requires that a certain number of moment conditions be specified for constructing a robust district-level geopredictive time series malarial-related risk model. The moment conditions would then be functions of the model sampled district-level parameter estimators and the malarial district level data, such that their expectation would be zero. Suppose the available district-level data consists of T i.i.d. field/clinical/remote geosampled explanatory hyperendemic transmission oriented observations $\{Y_t\}_{t=1,\ldots,T}$, where each observation Y_t is an n-dimensional multivariate random variable. The model data may then be defined by an unknown geoparameter $\theta \in \Theta$. The goal of the estimation problem in the risk model then would be to find the "true" value of this geoparameter, θ_0, or at least a reasonably close estimate. In order to apply GMM there

would also exist a vector-valued function $g(Y,\theta)$ such that $m(\theta_0) \equiv \mathrm{E}[\, g(Y_t, \theta_0)\,] = 0$, where E denotes expectation, and Y_t is a generic geosampled malaria-related district-level explanatory hyperendemic transmission oriented observation, which would then be assumed to be i.i.d. Moreover, function $m(\theta)$ in the sampled dataset would not be equal to zero for $\theta \neq \theta_0$, or otherwise the geoparameter θ will not be point-identified. The basic idea for constructing a GMM geopredictive district-level seasonal malarial-related risk model in SAS/GIS would then be to replace the theoretical expected value E[·] with its empirical analog — sample average: $\hat{m}(\theta) = \hat{\mathrm{E}}\big[\, g(Y_t, \theta)\,\big] \equiv \frac{1}{T}\sum_{t=1}^{T} g(Y_t, \theta)$ and then to minimize the norm of this expression with respect to θ.

By the law of large numbers, $\hat{m}(\theta) \approx \mathrm{E}[g(Y_t,\theta)] = m(\theta)$ for any large predictive district-level seasonal sampled explanatory hyperendemic trasnmission oriented district-level field/clinical/remote sampled covariate coefficients measurements values of T thereafter a malarialogist/experimenter should expect that $\hat{m}(\theta_0) \approx m(\theta_0) = 0$. The generalized method of moments would look for a number $\hat{\theta}$ which would make $\hat{m}(\hat{\theta})$ as close to zero as possible. Mathematically, this is equivalent to minimizing a certain norm of $\hat{m}(\theta)$ (norm of m) in a SAS/GIS database whereby $\|m\|$, would measures the distance between m and zero. The properties of the resulting district-level estimator will depend on the particular choice of the norm function, and as such the GMM would consider an entire family of norms, defined as $\|\hat{m}(\theta)\|_W^2 = \hat{m}(\theta)' W \hat{m}(\theta)$, where W is a positive-definite weighting matrix, and m' denotes transposition.

In practice, the weighting matrix W usually (denoted as $\hat{W}$) . in SAS/GIS may be computed based on any available empirical sampled dataset of field/clinical/remote geosampled explanatory hyperendemic transmission oriented observations Thus, the GMM estimator can be written as $\hat{\theta} = \arg\min_{\theta \in \Theta} \left(\frac{1}{T}\sum_{t=1}^{T} g(Y_t, \theta) \right)' \hat{W} \left(\frac{1}{T}\sum_{t=1}^{T} g(Y_t, \theta) \right)$. Under suitable conditions this estimator is consistent, asymptotically normal, and with right choice of weighting matrix $\hat{W}$ asymptotically efficient. The GMM method would then minimize a certain norm of the sample averages of the moment conditions in the district-level empirical datset. The GMM estimators may then be found to be asymptotically normal, and efficient in the class of all time series estimators that do not use any extra information aside from that contained in the moment conditions for accurate quantitation of district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented explanatory covariate coefficients measurements.

Dickey-Fuller unit root tests based on regression models may be also constructed in SAS/GIS which can employ $y_t = \beta_0 + \beta_1 t + \alpha y_{t-1} + \varepsilon_t$ where ε_t is assumed to be white noise (Cressie 1993) for quantitating statistically significant district-level malaria-related hyperendemic transmission oriented estimators. The testing procedure for the Dickey–Fuller test may also be applied to the model $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$, where α is a constant, β the coefficient on a time trend and p the lag order of the autoregressive process. Imposing the constraints $\alpha = 0$ and $\beta = 0$ corresponds to modeling a random walk and using the constraint $\beta = 0$ corresponds to modeling a random walk with a drift (e.g., Cressie 1993). Further, by including lags of the order p in any time series malarial-related geopredictive formulation will allow for higher-order autoregressive processes in the sampled district-level data. This means that the lag length p has to be determined when applying the test. One possible approach is to test down from high orders and examine the field/clinical/remote sampled explanatory hyperendemic transmission covariate coefficient t-values. An alternative approach may be to examine information criteria such as BIC .

In statistics, the Bayesian information criterion (BIC) or Schwarz criterion (also SBC, SBIC) is a criterion for model selection among a finite set of models (Cressie 1993). The BIC is an asymptotic result derived under the

assumptions that the data distribution is in the exponential family. A robust geopredictive district-level time series risk model may be constructed in SAS/GIS if malarialogist/experimenter then lets:x= the observed sampled field/clinical/remote district level data; n= the number of data points in x, the number of hyperendemic transmission oriented observations, or equivalently, the sample size; and, k= the number of free geoparameters to be estimated. If the model under consideration is a linear regression, k would then be the number of regressors, including the intercept; $p(x|M)$= the marginal likelihood of the observed data given by the model M; (that is, the integral of the likelihood function $p(x|\theta, M)$times the prior probability distribution $p(\theta|M)$over the geoparameters $\theta$of the model $M$for fixed observed sampled district-level data x);and, $\hat{L}$= the maximized value of the likelihood function of the model M, (i.e. $\hat{L} = p(x|\hat{\theta}, M)$, where $\hat{\theta}$are the geoparameter estimator values that maximize the likelihood function). The formula for the BIC is:$-2 \cdot \ln p(x|M) \approx \mathrm{BIC} = -2 \cdot \ln \hat{L} + k \ln(n)$.(see Akaike 1977).Under the assumption that the model errors or disturbances are i.d.d. according to a normal distribution and that the boundary condition that the derivative of the log likelihood with respect to the true variance is zero, this becomes expressed as $\mathrm{BIC} = n \cdot \ln(\widehat{\sigma_e^2}) + k \cdot \ln(n)$where $\widehat{\sigma_e^2}$is the error variance. The error variance in this case is defined as $\widehat{\sigma_e^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2.$

By so doing, a malarialogist/experimenter may point out from probability theory that $\widehat{\sigma_e^2}$is a biased field/clinical/remote sampled estimator for the true variance, σ^2.This may be facilitated by a malarialogist/experimenter by letting $\widehat{\widehat{\sigma_e^2}}$denote the unbiased form of approximating the error variance in the district-level geosampled malarial-related district-level explanatory hyperendemic transmission oriented covariate coefficients which may be defined by $\widehat{\widehat{\sigma_e^2}} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2.$ Additionally, under the assumption of normality the following version may be more tractable$\mathrm{BIC} = \chi^2 + k \cdot \ln(n)$. Note that there is a constant added that follows from transition from log-likelihood to χ^2; however, in using the BIC to determine the "best" district-level malarial risk model the constant becomes trivial. Given any two estimated models, the model with the lower value of BIC is the one to be preferred (Gilks 1996). The BIC is an increasing function of $\sigma_e^2$and an increasing function of *k* (Griffith 2003).That is, unexplained variation in the dependent variable (e.g., district-level prevalence rate) in a robust geopredictive malaria-related district-level time series model and the number of field/clinical/remote sampled explanatory variables increase the value of BIC. Hence, lower BIC implies either fewer explanatory variables, better fit, or both in the risk model output. It is important to keep in mind that the BIC can be used to compare estimated models only when the numerical values of the dependent variable are identical for all estimates being compared (Cressie 1993). Additionally, the geopredictive risk model outputs being compared need not be nested, unlike the case when the residual derivatives are being compared using an F or likelihood ratio test.The unit root test may then be carried out under the null hypothesis $\gamma = 0$against the alternative hypothesis of $\gamma < 0$.Once a value for the test statistic $DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$is computed for a predictive malaria-related district-level time series risk model it can be compared to the relevant critical value for the Dickey–Fuller Test. If the test statistic is less than the larger negative)critical value in the risk model residual forecasts then the null hypothesis of $\gamma = 0$is rejected and no unit root is present.

Fortunately, there are two popular ways to account for general serial correlation between these type of errors in SAS/GIS. One is the augmented Dickey-Fuller (ADF) test, which uses the lagged difference in the regression model. This was originally proposed by Dickey and Fuller (1979) and later studied by Said and Dickey (1984) and Phillips and Perron (1988). Another method was proposed by Phillips and Perron (1988); it is called Phillips-Perron (PP) test. The tests adopt the original Dickey-Fuller regression with intercept, but modifies the test statistics to take account of the serial correlation and heteroscedasticity. In a geopredictive district-level SAS/GIS derived malaria-related time series risk model this would be a nonparametric specification as no specific form of the serial correlation of the residually forecasted district-level field/clinical/remote sampled malaria-related hyperendemic transmission oriented explanatory covariate coefficients measurement errors would be assumed.

A problem of the Dickey-Fuller and Phillips-Perron unit root tests is that they are subject to size distortion and low power. It is reported in Schwert (1989) that the size distortion is significant when the series contains a large MA geoparameter estimation. Among some more recent unit root tests that improve upon the size distortion and the low power are the tests described by Elliott, Rothenberg, and Stock (1996) and Ng and Perron (2001). These tests would involve a step of detrending the test statistics in SAS/GIS prior to constructing the geopredictive seasonal district-level malarial risk model constructed from seasonal-sampled field/clinical/remote explanatory hyperendemic transmission oriented covariate coefficients measurements. Most autoregressive testing procedures specify the unit root processes in SAS products as the null hypothesis(www.sas.com).

Further by constructing augmented Dickey-Fuller (ADF) and Phillips-Perron unit root tests, a group of time series geopredictive malarial-related field/clinical/remote sampled hyperendemic transmission oriented regression-based risk model may be a linked together by some long-run equilibrium relationship. Statistically, this phenomenon can be modeled by quantitating cointegration when constructing the risk model from an empirical sampled dataset of predictive time series district-level malaria-related field/clinical/remote hyperendemic transmission oriented explanatory covariate coefficients measurements. It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models. The ADF statistic, used in the test, is a negative number. Thus the more negative it is in the district level model output, the stronger the rejection of the hypothesis that there is a unit root at some level of confidence. The testing procedure for the ADF test is the same as for the Dickey–Fuller test but it is applied to the model $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$, where α is a constant, β the coefficient on a time trend and p the lag order of the autoregressive process (Cressie 1993). Imposing the constraints $\alpha = 0$ and $\beta = 0$ corresponds to modeling a random walk and using the constraint $\beta = 0$ corresponds to modeling a random walk with a drift.

When several nonstationary processes $\mathbf{z}_t = (z_{1t}, \cdots, z_{kt})'$ are cointegrated, there exists a $(k \times 1)$ cointegrating vector $\mathbf{c}$ such that $\mathbf{c}'\mathbf{z}_t$ is stationary and $\mathbf{c}$ is a nonzero vector (Griffith 2003). One way to test the relationship of cointegration is the residual based cointegration test in a district- level time series geopredictive malarial risk model is by employing the model $y_t = \beta_1 + \mathbf{x}_t'\beta + u_t$ where $y_t = z_{1t}$, $\mathbf{x}_t = (z_{2t}, \cdots, z_{kt})'$, and $\beta = (\beta_2, \ldots, \beta_k)'$. The OLS residuals from the risk model could then be employed to test for the null hypothesis of no cointegration in the empirical field /clinical/remote sampled hyperendemic transmission oriented geoparameter estimator dataset.

An Engle-Granger Cointegration test may also be employed for constructing a robust predictive malarial-related district-level time series risk model in SAS/GIS from empirical sampled field/clinical/remote sampled hyperendemic transmission oriented explanatory covariate coefficients measurements. Common unit root tests have the null hypothesis that there is an autoregressive unit root $H_0 : \alpha = 1$, and the alternative is $H_a : |\alpha| < 1$, where α is the autoregressive coefficient of the time series $y_t = \alpha y_{t-1} + \varepsilon_t$. This is referred to as the zero mean model.

For the zero-mean seasonal-sampled predictive district-level malaria-related risk model constructed from an empirical dataset of district-level field/clinical/remote sampled time series malaria-related hyperendemic transmission oriented explanatory covariate coefficients measurements, the asymptotic distributions of the Dickey-Fuller test statistics would be quantitated as $T(\hat{\alpha} - 1) \Rightarrow \left(\int_0^1 W(r)dW(r)\right)\left(\int_0^1 W(r)^2 dr\right)^{-1}$ $DF_\tau \Rightarrow \left(\int_0^1 W(r)dW(r)\right)\left(\int_0^1 W(r)^2 dr\right)^{-1/2}$ For the constant mean model, the

asymptotic distributions in the malaria-related risk model would then be $T(\hat{\alpha}-1) \Rightarrow \left([W(1)^2-1]/2 - W(1)\int_0^1 W(r)dr\right)\left(\int_0^1 W(r)^2dr - \left(\int_0^1 W(r)dr\right)^2\right)^{-1} DF_\tau^{-1/2}$
$\Rightarrow \left([W(1)^2-1]/2 - W(1)\int_0^1 W(r)dr\right)\left(\int_0^1 W(r)^2dr - \left(\int_0^1 W(r)dr\right)^2\right)$ For the predictive malaria-related trend model, the asymptotic distributions could thereafter be derived using:

$$T(\alpha-1) \Rightarrow \left[W(r)dW + 12\left(\int_0^1 rW(r)dr - \frac{1}{2}\int_0^1 W(r)dr\right)\left(\int_0^1 W(r)dr - \frac{1}{2}W(1)\right) - W(1)\int_0^1 W(r)dr\right] D^{-1}$$

$$DF_\tau \Rightarrow \left[W(r)dW + 12\left(\int_0^1 rW(r)dr - \frac{1}{2}\int_0^1 W(r)dr\right)\left(\int_0^1 W(r)dr - \frac{1}{2}W(1)\right) - W(1)\int_0^1 W(r)dr\right] D^{1/2}$$

where $D = \int_0^1 W(r)^2dr - 12\left(\int_0^1 r(W(r)dr\right)^2 + 12\int_0^1 W(r)dr\int_0^1 rW(r)dr - 4\left(\int_0^1 W(r)dr\right)^2$ if so desired.

The null hypothesis of the Dickey-Fuller test for a robust geopredictive malaria-related time series district-level risk model then would be a random walk, possibly with drift. By so doing, the differenced process in the residual forecasts targeting the statistically significant hyperendemic transmission oriented covariate coefficients would then be not serially correlated under the null of 1 (Cressie1993). Interestingly, nonstationary multivariate time series district-level malaria-related explanatory hyperendemic transmission oriented covariate coefficients can thereafter be tested for cointegration, which means that a linear combination of these time series district-level data would be assumed stationary. Formally, denoting the series by $\mathbf{z}_t = (z_{1t}, \cdots, z_{kt})'$ would then render robust residual field/clinical/remote related hyperendemic transmission oriented residual forecasts. The null hypothesis of cointegration would be that that there exists a vector c such that $\mathbf{c}'\mathbf{z}_t$ is stationary. Residual-based cointegration tests may be then studied in depth for robust district-level malarial related geopredictive time series uncertainty modeling. The first step in this uncertainty regression model, would be $y_t = \mathbf{x}_t'\beta + u_t$ where $y_t = z_{1t}$, $\mathbf{x}_t = (z_{2t}, \cdots, z_{kt})'$, and $\beta = (\beta_2, \ldots, \beta_k)'$. This regression may also include an intercept or an intercept with a linear trend. The residuals would then be tested for the existence of an autoregressive unit root.

Engle and Granger (1987) proposed ADR type regression without an intercept on the residuals to test the unit root. This series may then be expressed as the sum of the deterministic trend, random walk r_t, and stationary error u_t in SAS/GIS; that is, $y_t = \mu + \delta t + r_t + u_t r_t r_t = r_{t-1} + e_t$ where $e_t \sim$ i.i.d. $(0, \sigma_e^2)$, and an intercept μ. The null hypothesis of trend stationary would then be specified by $H_0 : \sigma_e^2 = 0$ in the risk model residual forecasts targeting the district level statistically significant hyperendemic transmission oriented explanatory covariates while the null of level stationary would be quantitated with the model restriction $\delta = 0$. Under the alternative that $\sigma_e^2 \neq 0$ in the district-level risk model there would also be a random walk component in the observed series y_t.

Interestingly, when the first step OLS does not include an intercept, in the geopredictive district-level malaria-related risk model the asymptotic distribution of the ADF test statistic DF_τ may be given by $DF_\tau \Rightarrow \int_0^1 \frac{Q(r)}{(\int_0^1 Q^2)^{1/2}} dS$ $Q(r) = W_1(r) - \int_0^1 W_1 W_2'\left(\int_0^1 W_2 W_2'\right)^{-1} W_2(r)$, $S(r) = \frac{Q(r)}{(\kappa'\kappa)^{1/2}}$ and $\kappa' = \left(1, -\int_0^1 W_1 W_2'\left(\int_0^1 W_2 W_2'\right)^{-1}\right)$ where $W(r)$ is a k vector standard Brownian motion and $W(r) = \left(W_1(r), W_2(r)\right)$ is a partition such that $W_1(r)$ is a scalar and $W_2(r)$ is $k-1$ dimensional. Further, by including lags of the order *p*, the ADF formulation would allow for higher-order residual autoregressive processes to be quantitated in the hyperendemic transmission oriented residual forecasts. The unit root test would then be carried out under the null hypothesis $\gamma = 0$ against the alternative hypothesis of $\gamma < 0$. Once a sampled district-level value for the test

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$ statistic is computed it can be compared to the relevant critical value for the Dickey–Fuller Test. If the test statistic is less than the larger negative critical value in the risk model output, then the null hypothesis of $\gamma = 0$ would be rejected and no unit root would be present in the residually forecasted estimates.

The asymptotic distributions of the test statistics in a robust geopredictive time-series malarial-related district-level hyperendemic transmission oriented risk model could then be further elaborated in SAS/GIS. In the stepwise regression, quantitation of the empirical sampled data would include an intercept, where $W(r)$ would be replaced by the demeaned Brownian motion[i.e., $\overline{W}(r) = W(r) - \int_0^1 W(r)dr$]. If the first step regression in the model construction process in SAS/GIS includes a time trend, then $W(r)$ could be replaced by the detrended Brownian motion. The critical values of the asymptotic distributions are tabulated in Phillips and Ouliaris (1990) and MacKinnon (1991).

Besides the ADF test, there is another popular unit root test that may be valid under general serial correlation and heteroscedasticity for constructing a robust time series district-level geopredictive malarial-related risk models Phillips (1997) and Phillips and Perron (1988), for instance, constructed tests using the AR(1) type regressions, , with corrected estimation of the long run variance of Δy_t. A unit root test tests whether a time series variable is non-stationary using an autoregressive model (Cressie 1993). These tests use the existence of a unit root as the null hypothesis. As such a malarialogist/experimenter may then consider the driftless random walk process $y_t = y_{t-1} + u_t$where the disturbances in the sampled hyperendemic transmission oriented geoparamter estimator dataset might be serially correlated with possible heteroscedasticity. Phillips and Perron (1988) proposed the unit root test of the OLS regression model, $y_t = \rho y_{t-1} + u_t$.

Also by denoting the OLS residual by $\hat{u}_t$ in a time series geopredictive malaria-related hyperendemic transmission oriented district-level risk model the asymptotic variance of $\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2$can be estimated by using the truncation lag l[i.e. $\hat{\lambda} = \sum_{j=0}^{l} \kappa_j[1 - j/(l+1)]\hat{\gamma}_j$] where $\kappa_0 = 1$, $\kappa_j = 2$for $j > 0$, and $\hat{\gamma}_j = \frac{1}{T}\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}$. This is a consistent estimator as suggested by Newey and West (1987). The variance of u_t in the time series geopredictive district-level malarial model can then be estimated by $s^2 = \frac{1}{T-k}\sum_{t=1}^{T}\hat{u}_t^2$by letting $\hat{\sigma}^2$ be the variance estimate of the OLS estimator $\hat{\rho}$. Then the time series geopredictive district-level malaria-related $\hat{Z}_\rho$test (i.e.,zero mean case) may be written as $\hat{Z}_\rho = T(\hat{\rho} - 1) - \frac{1}{2}T^2\hat{\sigma}^2(\hat{\lambda} - \gamma_0)/s^2$ The $\hat{Z}_\rho$statistic is just the ordinary Dickey-Fuller $\hat{Z}_\alpha$statistic with a correction term that accounts for the serial correlation(Cressie 1993).The correction term would then shift to zero asymptotically if there is no serial correlation in the residual forecasts targeting the statistically important explanatory georeferenced field/clinical/remote-sampled district-level hyperendemic transmission oriented covariate coefficients.

Thereafter, by letting $\tau_\rho$be the τstatistic for $\hat{\rho}$ in the district level malarial risk model the $\hat{Z}_t$ test may be written using $\hat{Z}_\tau = (\gamma_0/\hat{\lambda})^{1/2}t_{\hat{\rho}} - \frac{1}{2}T\hat{\sigma}(\hat{\lambda} - \gamma_0)/(s\hat{\lambda}^{1/2})$To incorporate a constant intercept, then, the regression - based risk model $y_t = \mu + \rho y_{t-1} + u_t$ may be employed where the null hypothesis in the series is a driftless random walk with nonzero unconditional mean. Additionally, to incorporate a time trend in the regression–based malarial-related risk model, $y_t = \mu + \delta t + \rho y_{t-1} + u_t$ may also be employed under the null the series as a random walk with drift for accurately quantitating the empirical sampled predictive district-level malaria-related risk model hyperendemic transmission oriented explanatory covariate coefficients measurements efficiently. The limiting distributions of the test statistics for the zero mean case would then be $\hat{Z}_\rho \Rightarrow \frac{\frac{1}{2}\{B(1)^2 - 1\}}{\int_0^1 [B(s)]^2 ds}$ and

$\hat{Z}_\tau \Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\}}{\{\int_0^1 [B(x)]^2 dx\}^{1/2}}$ where B(.) is a standard Brownian motion. The limiting distributions of the test statistics for the intercept case in the district-level geopredictive risk model residual forecasts then would be $\hat{Z}_\rho \Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\} - B(1)\int_0^1 B(x)dx}{\int_0^1 [B(x)]^2 dx - \left[\int_0^1 B(x)dx\right]^2}$ and $\hat{Z}_\tau \Rightarrow \frac{\frac{1}{2}\{[B(1)]^2 - 1\} - B(1)\int_0^1 B(x)dx}{\{\int_0^1 [B(x)]^2 dx - \left[\int_0^1 B(x)dx\right]^2\}^{1/2}}$

Finally, the limiting distributions of the test statistics for the trend case in a robust geopredictive time series district-level malaria-related risk model can then be derived as $\begin{bmatrix} 0 & c & 0 \end{bmatrix} V^{-1} \begin{bmatrix} B(1) \\ \left(B(1)^2 - 1\right)/2 \\ B(1) - \int_0^1 B(x)dx \end{bmatrix}$ where $\hat{Z}_\rho$ and $c = \frac{1}{\sqrt{Q}}$ for $\hat{Z}_\tau$, or $V = \begin{bmatrix} 1 & \int_0^1 B(x)dx & 1/2 \\ \int_0^1 B(x)dx & \int_0^1 B(x)^2 dx & \int_0^1 xB(x)dx \\ 1/2 & \int_0^1 xB(x)dx & 1/3 \end{bmatrix} Q = \begin{bmatrix} 0 & c & 0 \end{bmatrix} V^{-1} \begin{bmatrix} 0 & c & 0 \end{bmatrix}^T$ When several variables $\mathbf{z}_t = (z_{1t}, \cdots, z_{kt})'$ are cointegrated, there exists a $(k \times 1)$ cointegrating vector $\mathbf{c}$ such that $\mathbf{c}'\mathbf{z}_t$ is stationary and $\mathbf{c}$ is a nonzero vector (Cressie 1993). The residual based cointegration test would assume the following regression based geopredictive district-level time series malarial-related model: $y_t = \beta_1 + \mathbf{x}_t'\beta + u_t$ where $y_t = z_{1t}$, $\mathbf{x}_t = (z_{2t}, \cdots, z_{kt})'$, and $\beta = (\beta_2, \ldots \beta_k)'$. As such, the malaraiolgist/experimenter could estimate the consistent cointegrating vector by using OLS, if all the predictive time series district-level malaria-related risk model residual forecasts rendered from regressed field/clinical/remote empirical sampled malaria-related explanatory hyperendemic transmission oriented covariate coefficients measurements are difference stationary — that is, 1. The estimated cointegrating vector in the risk model then would be $\hat{\mathbf{c}} = (1, -\hat{\beta}_2, \cdots, -\hat{\beta}_k)'$.

The Phillips-Ouliaris test is computed using the OLS residuals from the preceding regression model, and it uses the PP unit root tests $\hat{Z}_\rho$ and $\hat{Z}_\tau$ developed in Phillips (1997), although in Phillips and Ouliaris (1990) the asymptotic distributions of some other leading unit root tests were also derived. The null hypothesis is no cointegration (Griffith 2003). Thus, a malarialogist/experimenter would need only to refer to the tables by Phillips and Ouliaris (1990) to obtain the p-value of the cointegration test for robust geopredictive time series district-level malaria-related risk modeling. Before applying the cointegration test the unit root test for malarial-related risk modeling, the sampled data attributes may however need to be tested. Unfortunately, cointegration tests can give conflicting results for different choices of the regression-based variables used to determine the optimal residually forecasted district-level hyperendemic transmission oriented covariates. There are other cointegration tests that are invariant to the order of the variables, including Johansen (1991) and Stock and Watson (1988) that may also be employed for district-level malarial-related geopredictive risk modeling empirical sampled explanatory hyperendemic transmission oriented covariate coefficients .

As mentioned earlier, ADF for an empirical dataset of geopredictive time series risk modeling explanatory hyperendemic transmission oriented regressors may suffer severe size distortion and low power. There is a class of newer tests that improves both size and power, sometimes called efficient unit root tests, among which Elliott, Rothenberg, and Stock (1996) and Ng and Perron (2001) are prominent. Elliott, Rothenberg, and Stock (1996) considered the data generating process $y_t = \beta' z_t + u_t$ and $u_t = \alpha u_{t-1} + v_t, t = 1, \ldots, T$ where $\{z_t\}$ was either $\{t\}$ or $\{(1, t)\}$ and $\{v_t\}$ were unobserved stationary zero-mean processes with positive spectral density at zero frequency. By so doing, the null hypothesis was $H_0 : \alpha = 1$, and the alternative was $H_a : |\alpha| < 1$. These models could be employed for accurately targeting statistically significant district-level hyperendemic transmission oriented explanatory covariate coefficients measurements in predictive malaria-related risk model residually forecasted estimates.

The key idea of Elliott, Rothenberg, and Stock (1996) was to study the asymptotic power and asymptotic power envelope of some new tests. Asymptotic power may be defined with a sequence of local alternatives (Griffith 2003).

For a fixed alternative hypothesis, the power of a test usually goes to one when sample size goes to infinity; however, this does not say anything about the finite sample performance(Cressie 1993). On the other hand, when regressing seasonal sampled district level malaria-related data the alternative moves closer to the null as the sample size increases and the power does not necessarily have to converge to one. The local to unity alternatives in the model may be then quantitated employing $\alpha = 1 + \frac{c}{T}$ and the power against the local alternatives which would have a limit as Tgoes to infinity, (i.e. asymptotic power). This value in a robust predictive malaria-related risk model would be strictly between 0 and 1. Asymptotic power indicates the adequacy of a test to distinguish small deviations from the null hypothesis (Rao 1973). The asymptotic power for the risk model may then be defined as $y_\alpha = (y_1, (1-\alpha L)y_2, \ldots, (1-\alpha L)y_T)$ and $= (z_1, (1-\alpha L)z_2, \ldots, (1-\alpha L)z_T)$

Thereafter, by letting $S(\alpha)$be the sum of squared residuals from a least squares regression of $y\alpha$on $z\alpha$ in a robust geopredictive district-level time series malaria-related risk model constructed in SAS/GIS, the field/clinical/remote sampled malaria-related hyperendemic transmission oriented explanatory covariate sample point may be optimally tested against the local alternative $\bar{\alpha} = 1 + \bar{c}/T$ which would have the form $P_T^{GLS} = \frac{S(\bar{\alpha}) - \bar{\alpha}S(1)}{\hat{\omega}^2}$ where $\hat{\omega}^2$is an estimator for $\omega^2 = \sum_{k=-\infty}^{\infty} E v_t v_{t-k}$. Note that the test rejects the null when $P_T$is small. The asymptotic power function for the point optimal test constructed with $\bar{c}$under local alternatives with c would then be denoted by $\pi(c, \bar{c})$ in the residually forecasted explanatory hyperendemic transmission oriented covariate coefficients. Then the power envelope would be $\pi(c, c)$ in the predictive malaria-related district-level risk model constructed from an empirical dataset of field/clinical/remote sampled explanatory covariate coefficients as the test formed with $\bar{c}$ would be the most powerful against the alternative $c = \bar{c}$. In other words, the asymptotic function $\pi(c, \bar{c})$, is always below the power envelope $\pi(c)$except that at one point $c = \bar{c}$they are tangent. Elliott, Rothenberg, and Stock (1996) show that choosing some specific values for $\bar{c}$ can cause the asymptotic power function $\pi(c, \bar{c})$of the point optimal test to be very close to the power envelope. Coincidentally, this is also true for the DF-GLS test. Elliott, Rothenberg, and Stock (1996) who proposed the DF-GLS test, as given by the $t$statistic for testing $\psi_0 = 0$in the regression $\Delta y_t^d = \psi_0 y_{t-1}^d + \sum_{j=1}^{p} \psi_j \Delta y_{t-j}^d + \varepsilon_{tp}$ where $y_t^d$is obtained in a first step detrending $y_t^d = y_t - \hat{\beta}_\alpha' z_t$ and $\hat{\beta}\alpha$is least squares regression coefficient of $y\alpha$ on $z\alpha$. DF-GLS is indeed a superior unit root test, according to Schwert (1989), and Elliott, Rothenberg, and Stock (1996). In terms of the size of the test, DF-GLS may be then almost as good as the ADF $t$test $\mathrm{DF}_\tau$and better than other tests such as the PP $\hat{Z}_\rho$and $\hat{Z}_\tau$test Stock (1994) for quantizing field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate coefficients. In addition, the power of the DF-GLS is larger than the ADF $t$test and ρ-test.

Additionally, regarding the lag length selection, Elliott, Rothenberg, and Stock (1996) favored the Schwartz BIC. The optimal selection of the lag length $p$and the estimation of $\omega^2$is further discussed in Ng and Perron (2001). The lag length is selected from the interval $[0, p_{max}]$for some fixed $p_{max}$by employing the modified AIC, $\mathrm{MAIC}(p) = \log(\hat{\sigma}_p^2) + \frac{2(\tau_T(p) + p)}{T - p_{max}}$ where $\tau_T(p) = (\hat{\sigma}_p^2)^{-1} \hat{\psi}_0^2 \sum_{t=p_{max}+1}^{T} (y_{t-1}^d)^2$ and $\hat{\sigma}_p^2 = (T - p_{max})^{-1} \sum_{t=p_{max}+1}^{T} \hat{\varepsilon}_{tp}^2$. For fixed lag length p, an estimate of $\omega^2$is given by $\hat{\omega}^2 = \frac{(T-p)^{-1} \sum_{t=p+1}^{T} \hat{\varepsilon}_{tp}^2}{\left(1 - \sum_{j=1}^{p} \hat{\psi}_j\right)^2}$

Ng and Perron (2001) also applied GLS detrending to obtain the following M-tests: $MZ_\alpha = (T^{-1}(y_T^d)^2 - \hat{\lambda}^2) - \left(2T^{-2} \sum_{t=1}^{T} (y_{t-1}^d)^2\right)^{-1}$ $MSB = \left(\frac{\sum_{t=1}^{T} (y_{t-1}^d)^2}{T^2 \hat{\omega}^2}\right)^{1/2}$ $MZ_t = MZ_\alpha \times MSB$

which may be applicable for quantitating district-level predictive time series field/clinical/remote hyperendemic transmission oriented covariate coefficients. The modified point optimal tests using the GLS detrended predictive malaria district-level hyperendemic transmission oriented explanatory covariate coefficient data would then be $MP_T^{GLS} = \frac{\bar{c}T^{-2}\sum_{t=1}^{T}(y_{t-1}^{d})^2 - \bar{c}T^{-1}(y_T^d)^2}{\hat{\omega}^2}$ for $z_t = 1$ and $MP_T^{GLS} = \frac{\bar{c}T^{-2}\sum_{t=1}^{T}(y_{t-1}^{d})^2 - (1-\bar{c})T^{-1}(y_T^d)^2}{\hat{\omega}^2}$ for $z_t = (1,t)$ The DF-GLS test and the MZ_t test have the same limiting distribution

$$\text{DF-GLS} \approx MZ_t \Rightarrow 0.5\frac{(J_c(1)^2-1)}{\left(\int_0^1 J_c(r)^2dr\right)^{1/2}} \quad \text{for } z_t = 1 \qquad \text{DF-GLS} \approx MZ_t \Rightarrow 0.5\frac{(V_{c,\bar{c}}(1)^2-1)}{\left(\int_0^1 V_{c,\bar{c}}(r)^2dr\right)^{1/2}} \quad \text{for } z_t = (1,t)$$

(Cressie 1993). The point optimal test and the modified point optimal test have the same limiting distribution $P_T^{GLS} \approx MP_T^{GLS} \Rightarrow \bar{c}^2\int_0^1 J_c(r)^2dr - \bar{c}J_c(1)^2$ for $z_t = 1$ and $P_T^{GLS} \approx MP_T^{GLS} \Rightarrow \bar{c}^2\int_0^1 V_{c,\bar{c}}(r)^2dr + (1-\bar{c})V_{c,\bar{c}}(1)^2$ for $z_t = (1,t)$ where $W(r)$ is a standard Brownian motion and $J_c(r)$ is an Ornstein-Uhlenbeck process defined by $dJ_c(r) = cJ_c(r)dr + dW(r)$ with $J_c(0) = 0$, $V_{c,\bar{c}}(r) = J_c(r) - r\left[\lambda J_c(1) + 3(1-\lambda)\int_0^1 sJ_c(s)ds\right]$, and $\lambda = (1-\bar{c})/(1-\bar{c}+\bar{c}^2/3)$.

The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was introduced in Kwiatkowski et al. (1992) to test the null hypothesis that an observable series is stationary around a deterministic trend. SAS/GIS can test for the null hypothesis using KPSS that x is level or trend stationary (www.sas.edu) Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests any be used for testing a null hypothesis that an observable predictive malarial-related time series is stationary around a deterministic trend. Such models were proposed in 1982 by Alok Bhargava in his Ph.D. thesis where several John von Neumann or Durbin–Watson type finite sample tests for unit roots were developed (see Bhargava, 1986). Later, Denis Kwiatkowski, Peter C.B. Phillips, Peter Schmidt and Yongcheol Shin (1992) proposed a test of the null hypothesis that an observable series is trend stationary (i.e., stationary around a deterministic trend). The series was expressed as the sum of deterministic trend, random walk, and stationary error. The test was the Lagrange multiplier test of the hypothesis that the random walk has zero variance. KPSS type tests are intended to complement unit root tests, such as the Dickey–Fuller tests (Cressie 1993) .

Lagrange multiplier test is a statistical test of a simple null hypothesis that a parameter of interest θ is equal to some particular value θ_0.(Cressie 1993) It is the most powerful test when the true value of θ is close to θ_0. The main advantage of the Lagrange multiplier test for quantitating district-level time series geopredictive time series field/clinical/remote hyperendemic transmission oriented covariate coefficients is that it does not require an estimate of the information under the alternative hypothesis or unconstrained maximum likelihood. This makes testing field/clinical/remote sampled malaria-related hyperendemic transmission oriented explanatory covariates in SAS/GIS, feasible when the unconstrained maximum likelihood estimate is a boundary point in geoparameter space. By testing both the unit root hypothesis and the stationary hypothesis, a malarialogist/experimenter may then distinguish sampled time series district-level malarial-related data that appear to be stationary, series that appear to have a unit root, and series for which the data (or the tests) are not sufficiently informative to be certain whether the residuals are stationary or integrated.Using kpss.test(x, null = c("Level", "Trend"), lshort = TRUE)where x is a numeric vector and where null indicates the null hypothesis and must be one of "Level" (default) or "Trend" and lshort for indicating where the long or short version of the truncation lag geoparameter estimator is employed, a robust geopredictive district-level malarial-related models.

Under stronger assumptions of normality and i.i.d. of u_t and e_t, a one-sided LM test of the null, a random walk ($e_t = 0, \forall t$) can be constructed in SAS/GIS as follows: $\widehat{LM} = \frac{1}{T^2}\sum_{t=1}^{T}\frac{S_t^2}{s^2(l)} = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 + \frac{2}{T}\sum_{s=1}^{l}w(s,l)\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s}$ $S_t = \sum_{\tau=1}^{t}\hat{u}_\tau = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 + \frac{2}{T}\sum_{s=1}^{l}w(s,l)\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s}$ $s^2(l) = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2 + \frac{2}{T}\sum_{s=1}^{l}w(s,l)\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s}$ $S_t = \sum_{\tau=1}^{t}\hat{u}_\tau$ Notice that under the null hypothesis, $\hat{u}_t$ can be estimated by OLS regression of y_t on an intercept and the time trend. Following the original work of Kwiatkowski, Phillips,

Schmidt, and Shin (1991) under the null ($\sigma_e^2 = 0$), $\widehat{LM}$statistic converges asymptotically to three different distributions depending on whether the model is trend-stationary, level-stationary ($\delta = 0$), or zero-mean stationary ($\delta = 0, \mu = 0$).

The trend-stationary model is denoted by subscript $\tau$and the level-stationary model is denoted by subscript μ (Cressie 1993). The case when there is no trend and zero intercept is denoted as 0. The last case, although rarely used in practice, is considered in Hobijn, Franses, and Ooms (2004).By so doing $y_t = u_t : \widehat{LM}_0 \xrightarrow{D} \int_0^1 B^2(r)dr$ $y_t = \mu + u_t : \widehat{LM}_\mu \xrightarrow{D} \int_0^1 V^2(r)dr$ $y_t = \mu + \delta t + u_t : \widehat{LM}_\tau \xrightarrow{D} \int_0^1 V_2^2(r)dr$ with $V(r) = B(r) - rB(1)$ and $V_2(r) = B(r) + (2r - 3r^2)B(1) + (-6r + 6r^2)\int_0^1 B(s)ds$ where $B(r)$is the Brownian motion (i.e. Wiener process), and $\xrightarrow{D}$is convergence in a district-level geopredicted malaria-related model distribution. Note that $V(r)$is a standard Brownian bridge, $V_2(r)$is a Brownian bridge of a second-level.

Importantly, when using the notation of the $\widehat{LM}$ statistic to compute the long-run variance $s(l)$ in a geopredictive district-level time series malaria-related risk model, the window width $l$and the kernel type $w(\cdot,\cdot)$ would be employed to qunatiate the field/clinical/remote sampled explanatory covariate coefficients using the KERNEL option in SAS/GIS. Further, employing the Newey-West/Bartlett (KERNEL=NW BART), default $w(s,l) = 1 - \frac{s}{l+1}$ Quadratic spectral (KERNEL=QS) [e.g., $w(s,l) = \tilde{w}\left(\frac{s}{l}\right) = \tilde{w}(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos\left(\frac{6}{5}\pi x\right)\right)$] can then specify the number of lags, l, in three different ways in a robust geopredictive malaria-related district-level model: [e.g., Schwert (SCHW = c) $l = \text{floor}\left\{c\left(\frac{T}{100}\right)^{1/4}\right\}$ (default for NW, c=4) Manual (LAG = l) Automatic selection (AUTO) (default for QS) Hobijn, Franses, and Ooms (2004)].

Table 2: The kernel function and a formula for optimal window width where $T$is the number of geopredictive district –level malaria-related observations in (AUTO) in SAS/GIS

NW Kernel	QS Kernel
$l = \min(T, \text{floor}(\hat{\gamma}T^{1/3}))$	$l = \min(T, \text{floor}(\hat{\gamma}T^{1/5}))$
$\hat{\gamma} = 1.1447\left\{\left(\frac{\hat{s}^{(1)}}{\hat{s}^{(0)}}\right)^2\right\}^{1/3}$	$\hat{\gamma} = 1.3221\left\{\left(\frac{\hat{s}^{(2)}}{\hat{s}^{(0)}}\right)^2\right\}^{1/5}$
$\hat{s}^{(j)} = \delta_{0,j}\hat{\gamma}_0 + 2\sum_{i=1}^{n} i^j \hat{\gamma}_i$	
$n = \text{floor}(T^{2/9})$	$n = \text{floor}(T^{2/25})$

where $\delta_{0,j} = 1$if $j = 0$and 0, otherwise; $\hat{\gamma}_i = \frac{1}{T}\sum_{t=1}^{T-i} u_t u_{t+i}$.

Broock, Dechert, and Scheinkman (1987) propose a test (BDS test) of independence based on the correlation dimension. Broock et al. (1996) show that the first-order asymptotic distribution of the test statistic is independent of the estimation error provided that the parameters of the model under test can be estimated $\sqrt{n}$-consistently. Hence, the BDS test can be used as a geopredictive district-level seasonal malarial risk model error estimator selection tool and as a specification test. Given the sample size T, the embedding dimension m, and the value of the radius r, the

BDS statistic for the risk model would be $S_{\mathrm{BDS}}(T,m,r)=\sqrt{T-m+1}\dfrac{c_{m,m,T}(r)-c_{1,m,T}^{m}(r)}{\sigma_{m,T}(r)}$ where $c_{m,n,N}(r)=\dfrac{2}{(N-n+1)(N-n)}\sum_{s=n}^{N}\sum_{t=s+1}^{N}\prod_{j=0}^{m-1}I_r(z_{s-j},z_{t-j})$ $I_r(z_s,z_t)=\begin{cases}1 & \text{if } |z_s-z_t|<r\\ 0 & \text{otherwise}\end{cases}$ $\sigma_{m,T}^2(r)$ $=4\left(k^m+2\sum_{j=1}^{m-1}k^{m-j}c^{2j}+(m-1)^2c^{2m}-m^2kc^{2m-2}\right)_{c=c_{1,1,T}(r)}$ The statistic has a standard normal distribution if the sample size is large enough. For small sample size, the distribution can be approximately obtained through simulation. Kanzler (1999) has a comprehensive discussion on the implementation and empirical performance of BDS test.

The runs test and turning point test are two widely used tests for independence (Cromwell, Labys, and Terraza; 1994) which may be applicable for geopredictive district-level malarial time series risk modeling. The runs test needs several steps. First,the malarialogist/experimenter would convert the original time series field/clinical/remote sampled explanatory hyperendemic transmission oriented covariate coefficients into the sequence of signs, $\{++--\ldots+---\}$, that is, map $\{z_t\}$into $\{sign(z_t-z_M)\}$where $z_M$is the sample mean of $z_t$and $sign(x)$is "+", if $x$is nonnegative and "_" if $x$is negative. Second, the malarialogist/experimenter would count the number of runs, R, in the sequence. A run of a sequence is a maximal non-empty segment of the sequence that consists of adjacent equal elements (Cressie 1993). For instance, the following sequence contains $R=8$runs: $\underbrace{+++}_{1}\underbrace{---}_{1}\underbrace{++}_{1}\underbrace{--}_{1}\underbrace{+}_{1}\underbrace{-}_{1}\underbrace{+++++}_{1}\underbrace{--}_{1}$ in SAS (www.sas.edu) Third, the number of pluses and minuses in the sequence would have to be noted and then denoted as N_+andN_-, respectively in the geopredictive malaria-related risk model. Finally, the malarialoglist/experimenter would have to compute the statistic of runs test employing $S_{\mathrm{runs}}=\dfrac{R-\mu}{\sigma}$ where $\mu=\dfrac{2N_+N_-}{T}+1\sigma^2=\dfrac{(\mu-1)(\mu-2)}{T-1}$ The statistic of the turning point test in the geopredictive seasonal district-level risk model would then be defined as follows: $S_{\mathrm{TP}}=\dfrac{\sum_{t=2}^{T-1}TP_t-2(T-2)/3}{\sqrt{(16T-29)/90}}$, for instance, where the indicator function of the turning point $TP_t$is 1 if $z_t>z_{t\pm1}$or $z_t<z_{t\pm1}$(that is, both the previous and next values are greater or less than the current sampled value); otherwise, 0. The statistics of both the runs test and the turning point test would then possess a standard normal distribution under the null hypothesis of independence.

Increasing the scope of data collection may, however, may require adaptations in statistical analysis in order to accommodate larger and more complex endmember-oriented ecological data in ArcGIS. For example, evaluating spectrally extracted canopied endmember data employing multivariate ordination techniques applying different spatial-scaled observational covariate coefficients may help describe malaria- related ecohydrological features. River meandering (i.e., lateral migration of the channel) and avulsion (i.e. channel cut-off) , for instance, may then create a mosaic of landscapes and vegetative diversity that may be key to immature anopheline productivity in specific district geolocations. The ability of the river to meander, avulse, and generate new floodplain surfaces may be crucial to supporting diverse urban and non-urban habitats and healthy populations of immature *Anopheles* and therefore may be the basis of many metrics used as indicators of riverine malaria-related ecosystem health.

Additionally, employing nonparametric multivariate smoothing of presence–absence riverine-based endmember data may be highly effective for detecting patterns in heterogeneous spectrally decomposed riverine larval habitat assemblage data. Recently, Jacob et al. (2013) a0.6m^2 QuickBird spectral signature characteristic of *Similium damnosum s.l.* a black fly vector of onchocerciasis (‘river blindness’) was generated using multiple unmixing algorithms, vegetation indices and object based classifiers. The model was developed to forecast *S. damnosum s.l.* larval habitats based on the 0.6m^2 signature (i.e.,34% red, 11% blue and 55% green). Since the model encompassed the habitat within canopied objects (e.g. Precambrian rock, turbid water spectral components),it was designated as

a black rock-rapid) model. The model successfully identified positive aquatic sites in hyperendemic areas employing a stochastically endmember interpolated map in Togo (i.e., a sensitivity and specificity approaching 100%) and in Uganda (i.e.,sensitivity of 80% and a specificity of 92% with a statistically significance of $p<0.0001$; Fisher's Exact test).

Although, traditionally spectrally continuous analysis has presented challenges in localizing and extracting endmember patterns from noisy ecological –related endmember data employing, a robust, parsimonious object-based canopy-oriented riverine larval habitat analyses over a range of spatial scales in ENVI may provide the opportunity to pre-evaluate unknown immature *Anopheles* distributional patterns in Uganda. This would then perhaps render a more robust signature using the object-based classification. While district level LULC have certain advantages over other terrestrial environments in terms of ecological analysis (e.g., relatively defined spatial boundaries), patterns in flowing waters within the districts may be more difficult to visualize and measure and may be constantly in motion relative to the aquatic and terrestrial landscapes in which they are embedded. Moreover, in scaling up to obtain a finer remote resolution perspective (e.g., World view 3 satellite data systems) , the mosaic two-dimensional structure of district-level riverine environments and other hydrological networks at the scale of meters may quickly become condensed into a one-dimensional line or network of lines at the scale of kilometers. DigitalGlobe's next satellite WorldView-3 is in a phased development process for an advanced fourth-generation satellite scheduled to launch in mid-2014 and will offer 0.31 meter resolution panchromatic and eight-band multi-spectral imagery (www.digitalglobe.com) Thus, examining spectral variability of district-level canopy-oriented immature *Anopheles* assemblages and observing the relative influences of temperature and riverine channel a morphology on these assemblage structure may determine the signature dependency on decomposed sub-pixel emissivities and the spatial scale of the end member analysis.

Conclusion

In conclusion in this research we constructed time series-dependent linear and non-linear residual diagnostic error estimation models in SAS/GIS® using multiple district-level georeferenced malaria-related observational predictors sampled from 2006 to 2010 in Uganda. Initially, a Poisson and a negative binomial (i.e., a Poisson random variable with a gamma distrusted mean) was constructed in PROC REG which revealed that the residuals derived from the models were significant, but furnished virtually no predictive power. The seasonal-sampled georeferenced explanatory covariate coefficients variables and the district locational spatial structure was then with Thiessen polygons in ArcGIS. However this process failed to reveal unbiased estimators. A spatial eigenvector filtering algorithm in SAS/GIS was then generated. Thereafter, an Autoregressive Integrated Moving Average (ARIMA) model was constructed in ArcGIS® which rendered a conspicuous but not very prominent first-order residual spatiotemporal autoregressive structure in the sampled individual district-level time-series-dependent data. Additionally, the estimated model residuals, contained considerable overdispersion (i.e., excess Poisson variability): quasi-likelihood scale = 76.5648. Further, a malarial district-level data in ArcGIS® by overlaying the sub-meter resolution tessellations rendered from the predictive random effects risk model onto Map Atlas Data which efficiently mapped endemicity, entomological inoculation rates and the interpolated distribution of two known malaria mosquitoes species at the study site (*Anopheline arabiensis* and *Anopheline gambiae s.l*). We then constructed a DEM in ArcGIS® to create more robust indices based on the primary random effect model estimates. By doing so, the Poisson mean response specification was tabulated as: $\mu = \exp[a + re + LN(population)]$, $Y \sim Poisson(\mu)$ +DEM (zonal statistic). The mixed-model estimation results included: $a = -3.1876$ $re \sim n(0, s^2)$ mean re = -0.0010 $s^2 = 0.2513$ where P(S-W) = 0.0005 and the Pseudo-R^2 = 0.3103. A random effect intercept can robustly quantitate spatiotemporal residual predictor error covariate coefficients in a ArcGIS® and SAS/GIS® based malarial-related regression-based model for predicting and prioritizing district-level prevalence rates.

Acknowledgement

We would like to thank Ms. Samia Mckeever, at the College of Public Health, Department of Global Health University of South Florida for her contribution in generating this manuscript.

References

[1] A. Anderson, Ordination methods in ecology, The Journal of Ecology, 713-726(1971).

[2] A. Baddeley, I. Bárány, R. Schneider, Spatial point processes and their applications, Stochastic Geometry: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 13–18, 2004, 1-75 (2007).

[3] A.D. Cliff, K. Ord, Evaluating the Percentage Points of a Spatial Autocorrelation Coefficient*, Geographical Analysis, 3; 51-62(1971).

[4] A.D. Cliff, J.K. Ord; 'Families of Frequency Distributions', Pion; London, England, (1972).

[5] A.D. Cliff, J.K. Ord; 'Spatial Processes: models and Applications', Pion; London, England,(1981).

[6] A. D. R. MacQuarrie, C. Tsai; 'Regression and Time Series Model Selection'. World ScientificPublishing Company, Incorporated; New Jersey (1998).

[7] A. Getis, J.K. Ord, The analysis of spatial association by use of distance statistics, Geographical analysis, 24; 189-206(1992).

[8] A. Mallet, F. Mentre., J.L. Steimer, F. Lokiec; "Nonparametric Maximum Likelihood EstimationPopulation Pharmacokinetics, which Application to Cyclosporine", Journal ofPharmacokinetics and Biofarmaceutics; (1988)

[9] A. Michaelides, S. Ng; 'Estimating the Rational Expectations Model of Speculative Storage: AMonte Carlo Comparison of Three Simulation Estimators', Journal of Econometrics; (2000).

[10] A. Papoulis, S.U. Pillai, Probability, random variables, and stochastic processes, Tata McGraw-Hill Education, (2002).

[11] A.R. Gallant, G.E. Tauchen; 'Which Moments To Match?' Sargent Reading Group Presentation; (1996).

[12] A.R. Gallant, G.E. Tauchen; 'Efficient Method of Moments"; (2001)

[13] A.S. Fotheringham, C. Brunsdon, M. Charlton, Quantitative geography: perspectives on spatial data analysis, Sage, (2000).

[14] B.K. Slinker, S.A. Glantz; 'Multiple regression for physio-logical data analysis: The problem of multi-colinearity', Amer. J. Physiiol, Pg 249; (1985).

[15] B.S. Everitt, A. Skrondal; 'The Cambridge Dictionary of Statistics', 4th Edition. Cambridge University Press; Cambridge (2002).

[16] B.G. Jacob, D.A. Griffith., E.J. Muturi, E.X. Caamano, J.I. Githure, R.J. Novak; 'Aheteroskedastic error covariance matrix estimator using a first-order conditional autoregressive Markov simulation for deriving asymptotical efficient estimates fromecological sampled Anopheles arabinoses aquatic habitat covariates', Malaria Journal;(2009).

[17] B.G. Jacob, K.L. Arheart, D.A. Griffith, C.M. Mbogo, A.K. Githeko, J. Regens, J.I. Githure, R.J. Novak, J.C. Beier; 'Evaluation of Environmental Data for Identification of Anopheles(Diptera: Culicidae) Aquatic Larval Habitats in Kisumu and Malindi, Kenya', Journal of Medical Entomology, Vol 42, Pg 751-755; (2005).

[18] B.G. Jacob, J. Mwangangi, E.J. Muturi, J. Shililu, E. Kabiru, C. Mbogo, J. Githure, R.J.Novak; 'Environmental covariates of Anopheles Arabinoses in a rice agro-ecosystem', Journal of American Mosquito Control Association, 23(4) 13-22; (2007).

[19] B.G. Jacob, J.M. Mwangangi, C.B. Mbogo, R.J. Novak; 'A Taxonomy of Unmixing Algorithms Using Li-Strahler Geometric- Optical Model and other Spectral Endmember ExtractionTechniques for Decomposing a QuickBird Visible and Near Infra-red Pixel of an Anopheles arabiensis Habitat', Open Remote Sensing; (2011);

[20] B.G. Jacob, D.A. Grifith, J.M. Mwangnagi, C. Mbogo, R.J. Novak; 'Uniform Convergence of Ergodic Markov Chains Using Gaussian Quadratures in SAS PROC NLMIXEDforCalculating Marginal Likelihoods in Space Time-Varying Coefficients Of Urban Anopheles gambiae s.l. aquatic Habitats', Acta Paristology of China, Vol 14: 3, Pg 41-53;(2010).

[21] B.G. Jacob, R.J. Novak, L. Toe, M.S. Sanfo, A.N. Afriyie, M.A. Ibrahim, D.A. Griffith, T.R. Unnasch, 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; 117-133(2012).

[22] B.J. Jacob, E.J. Muturi, J. Ephantus, S. Muriu, J. Shililu, J. Mwangangi, C. B. Mbogo, J. Githure,R.J. Novak; 'Effect of rice cultivation on malaria transmission in central Kenya',American Journal of Tropical Medicine and Hygiene, Vol. 78, Pg 270-275; (2008a).

[23] B.J. Jacob, J.L. Regens, C.M. Mbogo, A.K. Githeko, J. Keating, C.M. Swalm, J.T. Gunter, J.Githure, J.C. Beier; 'Occurrence and distribution of Anopheles (Diptera: Culicidae) larval habitats on land cover change site in urban Kisumu and urban Malinda, Kenya', Journal of Medical Entomology, Vol 40, Pg 777-784; (2003).

[24] C. Brunsdon, A. Fotheringham, M. Charlton, Geographically weighted summary statistics—a framework for localised exploratory data analysis, Computers, Environment and Urban Systems, 26 (2002) 501-524(2002) .

[25] C.R. Rao;. 'Linear statistical inference and its applications', in, John Wiley & Sons (1973).

[25] C.R. Henderson; 'Biometrics: Best Linear Unbiased Estimation and Prediction under a SelectionModel', Vol. 31, No. 2, International Biometric Society; Washington, D.C (1975).

[26] C. Meyer, Matrix analysis and applied linear algebra book and solutions manual, Siam(2000).

[27] D.A. Griffith; 'A Spatial Filtering Specification for the Auto-Poissonmodel', Statistics and Probability, Pg. 58; (2003).

[28] D.A. Harville. 'Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems', Journal of the American Statistical Association. Vol. 72; (1977).

[29] D. Duffie, K. J. Singleton; 'Simulated Moments Estimation of Markov Models of Asset Prices',Vol 61. No. 4., Econometrica; New York (1993).

[30] D. Poskitt, Identification of echelon canonical forms for vector linear processes using least squares, The Annals of Statistics, 195-215(1992).

[31] E. Sentana, Quadratic ARCH models, The Review of Economic Studies, 62 ; 639-661 (1995).

[32] E.J. Hannan, L. Kavalieris, A method for autoregressive-moving average estimation, Biometrika, 71) 273-280(1984).

[33] F.A. Haight; 'Handbook of the Poission Distribution'; Wiley, (1967).

[34] F.J. Breidt, N.-J Hsu; 'Best Mean Square Prediction for Moving Averages', Statistica Sinica, pg1; (2005).

[35] F. Cribari-Neto, Asymptotic inference under heteroskedasticity of unknown form, Computational Statistics & Data Analysis, 45 ;215-233(2004).

[36] F.E. Grubbs, Procedures for detecting outlying observations in samples, Technometrics, 11 1-21(1969).

[37] G.E.P. Box, G.M. Jenkins, G.C. Reinsel; 'Time Series Analysis: Forecasting and Control (Wiley Series in Probability and Statistics)'; Wiley, (1994).

[38] G.E.P Box, P, GM Jenkins ;" Time series analysis: forecasting and control, Francisco Holden-Day (1976).

[39] G.E.P. Box, D.A. Pierce; 'Distribution of residual autocorrelations in autoregressive-integratedmoving average time series models', J. American Statist. Assoc; (1970).

[40] Box, G. E. P. and Pierce, D. A., 'Distribution of residual correlations in autoregressive-integrated moving average time series models'. Journal of the American Statistical Association, 65, 1509–1526 (1970).

[40] G.M. Ljung, G.E. Box : ' On a measure of lack of fit in time series models ', Biometrika, 65 297-303 (1978) .

[42] G. Gennote, T.A. Marsh; 'Variations in Economic Uncertainty and Risk Premiums on Capital Assets', European Economic Revue, Vol. 37; (1993)

[43] G. Lawrence, R. Jagannathan, D. Runkle, On the Relation Between Expected Vale and Volatility of the Nominal Excess Returns on Stocks, Journal of Finance, 48 1779-1802. (1993) .

[44] G. Steinbrecher, W.T. Shaw, 'Quantile mechanics', European journal of applied mathematics, 19) 87-112 (2008).

[45] G. White, S.A. Magayuka, P. Boreham, Comparative studies on sibling species of the Anopheles gambiae Giles complex (Dipt., Culicidae): bionomics and vectorial activity of species A and species B at Segera, Tanzania, Bulletin of entomological research, 62 ;295-317(1972).

[46] H. Akaike; 'Maximum likelihood identification of Gaussian autoregressive moving averagemodel', Biometrica, Biometrika Trust; London (1973).

[47] H.D. Patterson, R. Thompson; 'Recovery of Inter-Block Information when Block Sizes areUnequal', Biometrika, Pg. 58. (1971).

[48] H. Lütkepohl, Forecasting contemporaneously aggregated vector ARMA processes, Journal of Business & Economic Statistics, 2 (1984) 201-214(1984).

[49] H. Lütkepohl, New introduction to multiple time series analysis, (2005).

[50] H. White, A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica: Journal of the Econometric Society, 817-838(1980).

[51] I. Domowitz, C.S. Hakkio, Conditional variance and the risk premium in the foreign exchange market, Journal of International Economics, 1947-6 (1985).

[52] I. Titov, R. McDonald, A joint model of text and aspect ratings for sentiment summarization, Urbana, 51; 61801(2008).

[53] J.A. Castro-Hermida, I. García-Presedo, M. González-Warleta, M. Mezo, < i> Cryptosporidium</i> and< i> Giardia</i> detection in water bodies of Galicia, Spain, Water research, 44; 5887-5896(2010).

[54] J.A. Patz, D. Campbell-Lendrum, T. Holloway, J.A. Foley; 'The impact of regional climate change on human health', Nature, Vol 438; (2005).

[55] J.A. Patz, D. Engelberg, J. Last, The effects of changing weather on public health, Annual review of public health, 21; 271-307(2000).

[56] J.A. Patz, K. Strzepek, S. Lele; 'Predicting key malaria transmission factors, biting and entomological inoculation rates, using modeled soil moisture in Kenya', Tropical Medicine and International Health; (1998).

[57] J.A. Suykens, J. Vandewalle, 'Least squares support vector machine classifiers', Neural processing letters, 9 293-300(1999).

[58] J.C. Carlson, B.D. Byrd, F.X. Omlin, Field assessments in western Kenya link malaria vectors to environmentally disturbed habitats during the dry season, BMC Public Health, 4 33(2004).

[59] J.E. Gimnig, M. Ombok, L. Kamau, W.A. Hawley, Characteristics of larval anopheline (Diptera: Culicidae) habitats in Western Kenya, Journal of Medical Entomology, 38; 282-288(2001).

[60] J. Fernández-Villaverde, J.F. Rubio-Ramírez, Estimating macroeconomic models: A likelihood approach, The Review of Economic Studies, 74 1059-1087(2007).

[61] J.-F. Trape, E. Lefebvre-Zante, F. Legros, G. Ndiaye, H. Bouganali, P. Druilhe, G. Salem, Vector density gradients and the epidemiology of urban malaria in Dakar, Senegal, American journal of tropical medicine and hygiene, 47 181-189(1992).

[62] J. Harris, H. Stocker; 'Handbook of Mathematics and Computational Science', Springer; New York (1998).

[63] J. Hay; Causes and Consequences of Word Structure. Ph.D. Dissertation. Northwestern University. Chicago, Illinois. (2000).

[64] J. Huang, Y. Pawitan; 'Quasi-likelihood Estimation of Noninvertible Moving Average Processes', Scandinavian Journal Of Statistics, pg 27; (2000).

[65] J.F. Breidt, N.-J. Hsu; 'Best Mean Square Prediction for Moving Averages", Statistical Sincia; (2005).

[66] J.M. Zhu; 'Study on the feasibility for ARIMA model application to predict malaria incidence in an unstable malaria area'. (2007).

[67] J.-M. Zakoian, Threshold heteroskedastic models, Journal of Economic Dynamics and control, 18;931-955 (1994).

[68] J. Keating, C.M. Mbogo, J. Mwangangi, J.G. Nzovu, W. Gu, J.L. Regens, G. Yan, J.I. Githure, J.C. Beier, Anopheles gambiae sl and Anopheles funestus mosquito distributions at 30 villages along the Kenyan coast, Journal of medical entomology, 42 ;241(2005).

[69] J.Y. Campbell, A. Lo, A.C. MacKinlay; 'The Econometrics of Finacial Markets", Princeton University Press; Princeton, N.J., (1997).

[70] K.D. Hopkins, D.L. Weeks, Tests for normality and measures of skewness and kurtosis: Their place in research reporting, Educational and Psychological Measurement, 50; 717-729(1990) .

[71] K. Pearson, " Das Fehlergesetz und Seine Verallgemeinerungen Durch Fechner und Pearson." A Rejoinder, Biometrika, 4; 169-212(1905).

[72] K.S. Lii, M. Rosenblatt; 'Asymptotic normality of cumulant spectral estimates', J. Theoretical Probability; (1990).

[73] K.S. Lii, M Rosenblatt; 'An Approximate Maximum Likelihood Estimation for Non-Minimum Phase Moving Average Processes', J. Multivariate Analysis; (1992).

[74 L. Anselin, Spatial econometrics: methods and models, Springer, (1988).

[75] L. Anselin, Local indicators of spatial association—LISA, Geographical analysis, (1995).

[76] L. Anselin; 'Spatial Econometrics: Methods and Models', Kluwer Academic Publishers; London (1998).

[77] L. Anselin, A. Getis; 'Spatial Statistical Analysis and Geographic Information Systems', Annals of Regional Science; (1992).

[78] L. Anselin; 'Exploring Spatial Data with GeoDATM: A Workbook', Spatial Analysis Laboratory; (2005).

[79] L.F. Olsen, W.M. Schaffer; 'Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics', Vol 249, No. 4968, Science 3; (1990).

[80] L.J. Christiano, M. Eichenbaum, R. Vigfusson, Assessing structural vars, in: NBER Macroeconomics Annual, Volume 21, MIT Press, 2007, pp. 1-106(2006).

[81] M.A. Akivis, M. Akivis, V.V. Goldberg, An introduction to linear algebra and tensors, DoverPublications. com, (1972).

[82] M.A. Sattler, D. Mtasiwa, M. Kiama, Z. Premji, M. Tanner, G.F. Killeen, C. Lengeler, Habitat characterization and spatial distribution of Anopheles sp. mosquito larvae in Dar es Salaam (Tanzania) during an extended dry period, Malaria journal, 4 (2005) .

[83] M. Coezee, D. le Sueur; 'Distribution of African Malaria Mosquitoes Belonging to the Anopheles Gambiae Complex', Parasitol Today; (2000).

[84] M.S. Bartlett, Properties of sufficiency and statistical tests, Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences, 160 ;268-282(1937).

[85] M.T. Gillies, B. De Meillon, The Anophelinae of Africa south of the Sahara (Ethiopian zoogeographical region), The Anophelinae of Africa south of the Sahara (Ethiopian Zoogeographical Region). (1968).

[86] M. Hazewinkel; 'Encyclopedia of Mathematics', Springer; (2001).

[87] M.L. Higgins, A.K. Bera, A class of nonlinear ARCH models, International Economic Review, 137-158 (1992).

[88] M.G Ljung,. and Box, G. E. P., 'On a measure of lack of fit in time series models'. Biometrika 65, 297–303(1978).

[89] M. Lanne, J. Luoto, P. Saikkonen; 'Optimal Forecasting of Noncausal Autoregressive Time Series', MPRA Paper, University Library of Munich; Germany (2010).

[90] M.S. Bartlett; 'Propertires of Sufficiency and Statistical Tests', Proceeding of the Royal Society A: Mathematical, Physical and Engineering Sciences, Pg 160; (1937).

[91] M. T. Grillies, B. De Meillon; 'The Anophelinae of Africa south of the Sahara (Ethiopian Zoogeographical Region)', Publications of the South African Institute for Medical Research, No. 54.; (1968)

[92] M. Rosenblatt; 'Gaussian and Non-Gaussian Linear Time Series and Random Fields', SpringerVerlag; New York. (2000)

[93] N. A. C. Cressie; 'Statistics for Spatial Data', Revised Edition, J. Wiley; New York (1993).

[94] N. Cressie, J. Kornak, Spatial statistics in the presence of location error with an application to remote sensing of the environment, Statistical science, 436-456(2003).

[95] N.D.-J. Schwartz, Linear operators, Part I, (1958).

[96] N.E. Coulson, R.P. Robins, Aggregate economic activity and the variance of inflation: another look, Economics Letters, 17 71-75(1985).

[97] N. Minakawa, G. Sonye, G.O. Dida, K. Futami, S. Kaneko, Recent reduction in the water level of Lake Victoria has created more habitats for Anopheles funestus, Malar J, 7; 119(2008.

[98] P.A. Moran, Notes on continuous stochastic phenomena, Biometrika, 37; 17-23(1950) .)

[99] P.J. Brockwell, R.A. Davis; 'Time Series: Theory and Methods'. 2nd Ed, Springer-Verlag; New York. (2006).

[100] P. Brockwell, et Davis, RA (1991), Time Series: Theory and Methods, in, Springer-Verlag, New York.

[101] P. Shaman, R. Stine; 'The Bias of Autoregressive Coefficient Estimators', Journal of the American Statistical Association. 83, 842-848. (1998).

[102] R. Bivand, Regression modeling with spatial dependence: an application of some class selection and estimation methods, Geographical Analysis, 16; 25-37(1984).

[103] R. Courant, D. Hilbert, Methods of mathematical physics, republished by John Wiley and Sons, New York, (1989).

[104 R. Davidson, J.G. MacKinnon, Estimation and inference in econometrics, OUP Catalogue, (1993).

[105] R.C. Geary, The contiguity ratio and statistical mapping, The incorporated statistician, 5; 115-146(1954).

[106] R. Farebrother, Remark AS R53: A Remark on Algorithms AS 106, AS 153 and AS 155: The Distribution of a Linear Combination of χ 2 Random Variables, Journal of the Royal Statistical Society. Series C (Applied Statistics), 33 366-369(1984).

[107] R.A. Horn, C.R. Johnson; 'Topics in Matrix Analysis', Cambridge University Press; Cambridge (1995).

[108] R. Ehlers, S. Brooks,' Model uncertainty in integrated ARMA processes', (2002).

[109] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, Econometrica: Journal of the Econometric Society, 987-1007. (1982)

[110] R.F. Engle, Estimates of the Variance of US Inflation Based upon the ARCH Model, Journal of Money, Credit and Banking, 15 286-301 (1983).

[111] R.F. Engle, D. Kraft, 'Multiperiod forecast error variances of inflation estimated from ARCH models', Applied Time Series Analysis of Economic Data, 293-302(1983) .

[112] R.F. Engle, D.M. Lilien, R.P. Robins,' Estimating time varying risk premia in the term structure: the ARCH-M model, Econometrica' , Journal of the Econometric Society, 391-407 (1987).

[113] R.J. Spiegel, B.K. Bose, Fuzzy logic integrated electrical control to improve variable speed wind turbine efficiency and performance, in, Google Patents, (1997).

[114] R.L. Plackett; 'Some Theorems In Least Squares', Biometrika; (1950).

[54] R.P. Haining; 'Spatial Data Analysis: Theory and Practice', Cambridge University Press; Cambridge, (2003).

[115] R.M. Neal, 'Probabilistic inference using Markov chain Monte Carlo methods', (1993).

[116] R.S. Bivand; 'Spatial Econometrics Functions in R; Classes and Methods', Journal of Geographical Systems, Pg; (2002).

[117] R.W. Farebrother, The Durbin-Watson test for serial correlation when there is no intercept in the regression, Econometrica: Journal of the Econometric Society, 1553-1563(1980).

[118] S. Axler; 'Linear Algebra Done Right', 2nd Ed. Berlin; New York (1997).

[119] S. Glantz, B. Slinker, One-way analysis of variance, Primer of Applied Regression and Analysis of Variance. 2nd ed. New York, NY: McGraw-Hill, Inc, 274-303(2001).

[120] S.L. Hay, R.W. Snow; 'The Malaria Atlas Project: Developing Global Maps of Malaria Risk',PLoS Med; (2006).

[121] S.P. Brooks, P. Giudici, G.O. Roberts, 'Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions', Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65 3-39(2003).

[122] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, Journal of econometrics, 31 307-327 (1986).

[123] T.F. Cooley, M. Dwyer, Business cycle analysis without much theory A look at structural VARs, Journal of Econometrics, 83 57-88(1998).

[124] U. Fillinger, G. Sonye, G.F. Killeen, B.G. Knols, N. Becker, The practical importance of permanent and semipermanent habitats for controlling aquatic stages of Anopheles gambiae sensu lato mosquitoes: operational observations from a rural town in western Kenya, Tropical Medicine & International Health, 9; 1274-1289(2004).

[125] V.V. Chari, P.J. Kehoe, E.R. McGrattan, New Keynesian models: Not yet useful for policy analysis, in, National Bureau of Economic Research, (2008).

[126] W.M. Bolstad, Topic Index, Understanding Computational Bayesian Statistics, 313-315 (2010).

[127] W. Enders; 'Applied Econometric Time Series', New York; Wiley (1995)

[128] W. K. Cheang, G.C. Reinsel; Bias reduction of autoregressive estimates in time series regressionmodel through restricted maximum likelihood, Journal of the American StatisticalAssociation. 95, 1173-1184. (2000).

[129] W. W. R. Gilks, S. Richardson, D. J. Spiegelhalter; 'Markov Chain Monte Carlo in Pratica',CRC Press; United States, (1996)

[130]Harvey, A. C. Time Series Models. 2nd Edition, Harvester Wheatsheaf, NY, pp. 44, 45(1993).