International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-31
Available Online at http://acascipub.com/Journals.php

Research article

Accounting for first-order differential flux extinction through dimensionless, radiation-based discontinuous vegetated canopies employing near infra-red and red wavelength proxy spectral irradiance for approximating unobserved isoline convergence and soil-perturbed responses in capture point endmember signatures positively autocorrelated to endemic *Naegleria fowleri* sample sites

Kori A. Conklin[a], Toni Panaou[a], Benjamin G. Jacob[a]*

[a]College of Public Health, University of South Florida, Tampa, FL, USA

*Corresponding Author. Email: bjacob1@health.usf.edu

Abstract

Naegleria fowleri is a free-living thermophilic amoeba that can survive in soil and fresh water bodies in its trophozoite and flagellated forms. It is responsible for a confirmed 134 deaths in the United States in the past 60 years with a fatality rate of over 97%. Though some research has been conducted concerning this amoeba, currently no contributions exist in the literature which has revealed how environmental co-factors determine amoebic prevalence. This project employed digital elevation data to create geoclassified landscape maps as well as biogeophysical proxy remote sensing indices (e.g., Soil Adjusted Vegetation maps) of Orange County, Florida to optimally forecast potential *N. fowleri* sample sites. These maps were employed to construct a Poisson probability regression model to predict the appearance of the amoeba *N. fowleri* in natural and man-made ecohydrological bodies in Florida. Our assumption was that by examining cofactors to amoebic growth with statistical analysis tools within an ArcGIS cyberenvironment, (e.g., eigenfunction decomposition clustering algorithms) a researcher may be able to visually asses the risk of infection in a given eco-georeferenceable ecohydrological body. The results revealed that linear and nonlinear probability models could not accurately prognosticate covariates associated with infection risk levels. Although the regression model did not forecast any significant vulnerability cofactors, a Gaussian process model utilizing a spectral vegetation index proxy target endmember signature yielded an unknown, un-geosampled, eco-georeferenceable hotspot, *N. fowleri* infection geographic location at the eco-epidemiological study site.

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-31
Available Online at http://acascipub.com/Journals.php

Keywords: *Naegleria fowleri,* amoeba, Poissonian, GIS, Florida

1. Introduction:

In the warm, late summer months between June and October, a killer swims in fresh water lakes and rivers in the United States. The amoeba *Naegleria fowleri* is commonly found in the sediment of natural warm bodies of water throughout the world, but it has also been observed in ditches, poorly maintained swimming pools, geothermal hot springs, warm water discharge from industrial plants, ground soil, and in rare cases, water heaters (Naegleria fowleri 2016). The *N. fowleri* amoeba is a thermophile, meaning its optimum growth temperature is from 80°F (26.7°C) to 115°F (46°C) or higher (Yoder 2009). *N. fowleri,* or as it is sometimes referred "the brain-eating amoeba", can only be contracted by the intake of infected fresh water through the nose, and once infected the amoeba makes its way to the brain, namely the frontal cortex, in search of a food source (Marciano-Cabral 2007).

In response to amoebic presence, the body's natural response causes inflammation leading to death, which is diagnosed as Primary Amoebic Meningoencephalitis (PAM) (Capewell 2014). The onset of symptoms can take only a few hours to a day to begin and the damage to the brain can kill its host in as little as 2 to 15 days, with early symptoms of this infection usually misdiagnosed as bacterial or viral meningitis which can also cause; severe headaches, fever, nausea, vomiting, and stiffness of the neck (Yoder 2009). Unfortunately when the progression of the infection worsens it is too late to treat. Hallucinations, seizures, and confusion mark the point in which the host has received permanent brain damage and the only way to diagnose this infection is to test samples of the spinal fluid for the amoeba (Capewell 2014). If diagnosed early, there is a protocol involving an experimental drug, miltefosine, which is currently being produced in Orlando, Florida that has been shown to kill the amoeba in lab tests and in humans in combination with induced chemical coma and therapeutic hypothermia which reduce swelling in the brain (Naegleria fowleri 2016). The Center for Disease Control (CDC) began watch on this amoeba in 1989 and in that time there have been 34 documented cases of PAM infection in Florida (Yoder 2009). The CDC has collected general data about confirmed cases, and have shown that PAM infections are most common among males aged 5-14 (Naegleria fowleri 2016). This project is designed to look at cases of PAM by *N. fowleri* in Florida employing biogeophysical proxy vegetation indices and temperature as cofactors for amoebic prevalence.

A normalized difference vegetation index (NDVI) was generated using near infer-red (NIR) and red reflectance data to eco-geographically classify areas of vegetation at known locations of *N. fowleri* in Florida. VI utilizes red and NIR canopy reflectance or radiances in the form of ratios (Tucker 1979) or in linear combination (Richardson and Wiegand 1977). Brown et al. (2008) employed canonical correlation analyses to determine if a significant relationship existed between NDVI, disease/water stress index and distance to water at four local West Nile Virus competent vectors (*Culex. pipiens, Cilex. restuans, Culex. salinarius, and Aedes. vexans*). Their model determined a significant relationship existed between the geographically sampled (henceforth geosampled), explanatory, paramterizable, land use land cover (LULC) covariates and the mosquito habitat regressors (0.93, P=0.03). The final model, residual, diagnostic output explained 86% of the variance in the environmental and mosquito measures. Diuk-Wasser et al. (2006) developed multivariate regression models to predict high and low adult mosquito abundance sites for determining arboviralactivity in Fairfield County, Connecticut USA. The eco-geographic, prognosticative models included non-forested LULC for *Cx. pipiens*, surface water and distance to estuaries for *Cx. salinarius,* surface water and grasslands/agriculture LULCs for *Ae. vexans*, distance to palustrine habitats for *Culiseta melanura*, and seasonal difference in the NDVI parameters.

NDVI indices have been found to be well correlated with various vegetation LULC parameters including green leaf area, biomass, percent green cover, productivity, and photosynthetic activity (Colwell, 1974; Hatfield et al., 1984; Asrar et al., 1984; Sellers, 1985). The NDVI is generated by converting raw data into an entirely new image using algorithms to calculate the color value of each pixel [http://chesapeake.towson.edu/data/all_image.asp]. This type of product is especially useful in multi-spectral remote sensing since transformations can be created that highlight relationships and differences in spectral intensity across multiple bands of the electromagnetic spectrum.

NDVI is calculated as a ratio between measured reflectivity in the red and near infra-red (NIR) portions. These two spectral bands are chosen because they are most affected by the absorption of chlorophyll in leafy green vegetation and by the density of green vegetation on the LULC surface (http://earthobservatory.nasa.gov) As a simple transformation of two spectral bands, NDVI are computed directly without any bias or assumptions regarding

plant physiognomy, LULC, soil type, or climatic conditions (Jackson and Huete 1991). NDVI calculated as: NDVI = $(\rho_{NIR} - \rho_R)/ (\rho_{NIR} + \rho_R)$.

To account for changing soil brightness, SAVI may be calculated utilizing an adjustment factor L that effectively shifts the origin of vegetation isolines in NIR/ (visible) VIS reflectance space. Because the NDVI does not account for variations in soil brightness (Huete et al. 1994) a darkening of the soil following a rainfall or periodic flooding will cause a change in NDVI that will be interpreted as a change in vegetation (Asner et al. 2003). SAVI is calculated using radiance, surface reflectance (r), using reflectance values in the red (R), and NIR spectral bands. The L factor is determined by the relative percentage of vegetation and is dependent on whether the soil is light or dark; it is used as an exponent assigned to the red band value in the denominator and as a multiplier (L+1) of the first term (Huete et al. 1992).

Limitations, however, exist as a result of atmospheric influences and soft substrate differences. Atmospheric turbidity generally inhibits reliable measures of vegetation and may delay the detection of an onset of stress in discontinuous LULC canopies. An atmospherically resistant vegetation index (ARVI) was developed in Jacob et al. (2006) for remotely sensing vegetation LULC from the sub-meter resolution [0.61 meter (m)] QuickBird VIS and NIR LULC data. The index took advantage of the presence of the blue channel (0.45 to 0.52 μm) in the QuickBird sensor, in addition to the red and the NIR channels that composed the NDVI. The resistance of the ARVI to atmospheric effects (in comparison to the NDVI) is accomplished by a self-correction process for the atmospheric effect on the red channel, using the difference in the radiance between the blue and the red channels to correct the radiance in the red channel (Kaufman and Tanre 1992). Aerosols, absorbing gases such as water vapor, and undetected clouds affect upwelling radiances measured by satellite instruments (Hay et al. 1998). In Jacob et al. (2006) the ARVI was calculated employing irradiance, surface wavelength reflectance (r), and reflectance values in the QuickBird blue channel (0.05 to 0.06 μm), red (R), and NIR spectral bands. The ARVI was defined as ARVI = $(\rho_{NIR} - \rho_{RB})/(\rho_{NIR} + \rho_{RB})$in ArcGIS where the subscript RB denoted the R and blue bands (B) and γ was the gamma value which was defined as:$\rho_{RB} = \rho_R - \gamma(\rho_B - \rho_R)$A single value of γ = 1.0 was used to substantially reduce the sensitivity of atmospheric effects. Gamma correction, or often simply gamma, is the name of a nonlinear operation used to encode and decode luminance or tristimulus values in video or still image systems (Jensen 2005).

Thereafter the NDVI, SAVI, and ARVI calculated from the QuickBird satellite information were successfully overlaid onto eco-georeferenced, field-based data of the three grid-stratified, riceland, agro-ecosystem, eco-epidemiological agro-irrigated, village complex, study sites. The vegetation indices (VI's) were used to select all paddy and canal habitats with low, intermediate and heavy vegetated values. A database was generated for each study site with the mean, minimum, maximum, and standard deviations for NDVI, SAVI, and ARVI aggregated to the riceland agro-village complex level. The VI datasets for the three study sites were then merged with entomological datasets representing, eco-georeferenced, malaria, mosquito *Anopheles arabiensis*, aquatic larval habitat, capture point, breeding foci, geo-spectrotemporally geosampled at the eco-epidemiological, study sites. The NDVI was sensitive to the presence of vegetation LULCs and were not affected differently by ecological changes at the three study sites. The change in the soil background caused by the transition in LULC throughout the crop season did not alter the red and NIR rice plant reflectance and calculated SAVI. Visually the data suggest that there was no higher soil influences in the SAVI as compared with the NDVI for all rice stages.

In these analyses NDVI, SAVI, ARVI equations could not identify LULC change for making inferences of *An. arabiensis* aquatic,larval abundance and distribution. Many studies have found NDVI's to be unstable varying with soil, sun-view geometry, atmospheric conditions and the presence of dead material as well as changes within soil moisture (Sellers 1985; Jackson and Pinter 1986; Myneni et al. 1992). Factors that reduce reflectivity of soils in the visible region include soil moisture or self-shadow (Karnieli et al. 2001). NDVI equation has a simple open loop structure (no feedback) which renders it susceptible to large sources of error (Liu and Huete 1995).

The SAVI may also exhibit asymptotic (saturated) signals over riceland areas decreasing atmospheric visibility with changing LULC during the crop season. Baret and Guyot (1991) expressed inconsistencies in SAVI especially in soils in which the slope is exactly unity and the intercept is zero. Bausch (1993) tested a step-wise variable L function in the SAVI but found no significant reduction in noise reduction.

It was of interest to determine how the blue band inclusion into the VI would identify the riceland LULC's for making inferences of *An. arabeinsis* immature abundance. The resistance of the ARVI to atmospheric effects, in comparison to NDVI was accomplished by a self-correction process for the atmospheric effect on the QuickBird red channel in ArcGIS using the difference between the imager's blue and red channels to correct the radiance in the red

channel. The results suggest that the ARVI was not able to normalize atmospheric conditions in the eco—georeferenced, ento-ecoepidemiological, riceland study sites. The percent atmospheric and noise in the ARVI was at the rice height of 0–1 for all LULC's. Overall the NDVI was not associated to rice height much higher than the SAVI for identifying LULC's in all three study sites. NDVI and SAVI exhibited decreasing percent error due to increasing rice height. At rice height beyond 50 cm all the NDVI and SAVI are the same.

Regardless, prediction of VI associated with endemic areas of *Naegleria fowleri* foci may be remarkably accurate bio-geophysical proxy VI, endmember signature may be employable to infer vegetation LULC, properties by isolating the attributes of vegetation from other materials (e.g., soil or water) and for mapping unknown geosampled, eco-georeferenceable foci in a stochastic or deterministic interpolator. The appeal of a VI is its simplicity and its relationship either empirically or theoretically to biophysical LULC variables (Bannari et al. 1995). VI's have been proven to be well correlated with various vegetation parameters such as green biomass (Tucker et al 1986), chlorophyll concentration (Buschmann and Nagel 1993), leaf area index (LAI) (Asrar et al. 1984),foliar loss and damage (Olgemann 1990), photosynthetic activity (Sellers 1985) and carbon fluxes (Tucker et al. 1986). Also, they have been found to be useful for different image analyses like crop classification (Ehrlich and Lambin 1996) and crop phenology (Co et al. 1985).

Soil background conditions may exert considerable influence on partial discontinuous canopy, wavelength, LULC irradiance, spectra and the calculated VI signature may influence *N. fowleri* sites. Soft brightness influences have been noted in numerous studies where, for a given amount of vegetation, darker soil LULC geographically classified (henceforth geoclassified) substrates resulted in higher VI values when the ratio vegetation index (RVI = MR/red) or the NDVI --(NIR - red)/(NIR + red) = (RVI -1)/(RVI+I)] were used as vegetation measures (Colwell, 1974). Huete et al. (1985) also found an opposite soil brightness influence on the Perpendicular Vegetation Index (PVI) such that brighter soils resulted in higher index values for a given quantity of incomplete vegetation-related LULC. The PVI of Richardson and Wiegand (1977) may be emplo the red and NIR bands to calculate the perpendicular distance between the vegetation spot on the NIR-Red scatterplot and the soil line. Hence, in LULC areas where there are considerable soil brightness variations in an eco-georeferenceable, *N. fowleri* sample site, moisture differences, roughness variations, shadow, or organic matter differences, there may be soil-induced influences on proxy VI signature, endmember LULC grid-stratifiable values. Soil influences on incomplete canopy spectra are partly due to a dependency of the soil background signal on the optical properties of the overlying canopy (Jackson et al., 1980; Huete, 1987).

Differences in red and NIR flux transfers through a canopy for an eco-georeferenced, eco-epidemiological, *N. fowleri,* capture point, sample site may result in a complex soil-vegetation LULC interaction, which would make it difficult to subtract or correct for soil background influences. A non-continuous. vegetated canopy, grid-stratified, LULC geolocation intermittently surrounding a *N. fowleri* site may scatter and transmit a significant amount of NIR flux towards the soil surface, irradiating the soil underneath as well as in between individual plants. The amoeba may subsequently reflect part of this scattered and transmitted flux back toward the sensor in a manner dependent upon the optical properties of the soft surface. By contrast, red light may be strongly absorbed by the uppermost leaf layers of the canopy, and irradiance at the soil surface which may be limited at a capture point, *N. fowleri* eco-epidemiological, sample site so that radiance may be received directly from the sun and sky through eco-georefereneceable canopy gaps.

Gap dynamics refers to the pattern of plant growth that occurs following the creation of a forest gap, (e.g., a local area of natural disturbance that results in an opening in the canopy of an eco-georefereneced, eco-epidemiological *N. fowleri* sample site) Gap dynamics are a typical characteristic of both temperate and tropical forests and have a wide variety of causes and effects on urban and forest life (Kricher 2011). Gaps are the result of natural disturbances in forests, ranging from a large branch breaking off and dropping from a tree, to a tree dying then falling over, bringing its roots to the surface of the ground, to landslides bringing down large groups of trees. Because of the range of causes, gaps, therefore, have a wide range of sizes, including small and large gaps. Regardless of size, gaps may allow an increase in light as well as LULC changes in moisture and wind levels, leading to differences in microclimate conditions at an *N. fowleri* sample site compared to those from below the closed canopy; which may be generally cooler and more shaded. For gap dynamics to occur in naturally disturbed areas, either primary or secondarysuccession must occur (Hubbell and Foster 1986).

Ecological secondary succession is much more common and pertains to the process of vegetation replacement after a natural disturbance. Secondary succession results in second-growth or secondary forest, which

currently covers more of the tropics than old-growth forest (Bazzaz and Pickett 1980). Since gaps let in more light they could create diverse microclimates for an *N. fowleri* capture point as they provide the ideal geolocation and conditions for rapid plant reproduction and growth. In fact, most plant species in the tropics are dependent, at least in part, on gaps to complete their life cycles (Bazzaz and Pickett 1980).

The major ecophysiological processes of vegetation LULC including photosynthesis and evapotranspiration are determined by the vegetation biophysical parameters that describe the canopy structure (Tucker 1991).The NDVI is one of the most extensively applied vegetation indices related to Leaf area index (LAI) and primary production is one of the principal eco-biophysical parameters in climate, weather, and ecological studies, and has been routinely estimated from remote sensing measurements. LAI is defined as one half the total radiation intercepting leaf area per unit ground horizontal surface area (Gonsamo and Pellikka 2008). Several numerical models require a continuous field of high spatial and temporal resolution LAI measurements due to heterogeneity and size of vegetation or natural agricultural patches, and the large seasonal dynamics of vegetation. To fulfill these needs, the LAI retrieval methods in ArcGIS included the processing of the remotely sensed data expected to be efficient and convenient for the end users.

Generally speaking, the success of LAI estimation from remotely sensed data remains cumbersome and there is always a need to calibrate remotely retrieved parameters with ground based observation (Jensen 2005). Jacob et al. (2015) demonstrated the feasibility of the large scale LAI inversion algorithms using red and NIR reflectance obtained from sub-meter spatial resolution satellite imagery [e.g., panchromatic sub-meter resolution(i.e., (QuickBird) data for optimally summarizing riverine, agro-village, trailing vegetation, Precambrian rock, tributary habitats of *Similium damnosum* s.l., a black fly vector of onchocerciasis (‘river blindness”) The algorithms were developed based on the principle commonly employed for ground-based optical determination of LAI by applying Beer-Lambert’s law(see Appendix 1) and by assuming extinction coefficient for the gap fraction retrieved from spectral vegetation indices (SVIs). The Beer–Lambert law, also known as Beer's law, the Lambert–Beer law, or the Beer–Lambert–Bouguer law relates the attenuation of light to the properties of the material through which the light is traveling (Jensen 2005).

The results from Jacob et al. (2015) suggest that a regressed NDVI–LAI relationship can vary both seasonally and inter-annually in tune with the variations in phenological development of the trailing vegetation of a ento-ecoepidemiological, hyerproductive, capture point, eco-georeferenceable, *S. damnsoum* s.l. riverine tributary, breeding foci in response to temporal variations of environmental conditions. Strong linear relationships were obtained during the canopy trailing vegetation production and canopy leaf senescence periods sample frame but the relationship is poor during periods of maximum LAI, apparently due to the saturation of NDVI at high values of LAI. The NDVI–LAI relationship was found to be poor, R^2 varied from 0.31 to 0.44 for different sources of NDVI, when all the data were pooled across the sample time frames apparently due to different leaf area development patterns seasonally. The strong exponential relationships between NDVI and biomass and NDVI and LAI suggest that NDVI saturates for higher values of biomass (> 100 g/m^2) and LAI (> 2 m^2/m^2) (Huete et al. 2002). The ento-ecoepidemiological, capture point *S. damnsoum* s.l. immature habitat was also affected by background NDVI, but this was minimized by applying relative NDVI.

Heilman and Kress (1987) investigated the differential rates of radiant flux penetration in incomplete cotton canopies and found the spectral response reflected from the soil surface to mimic that of green vegetation. Thus, regardless of the vegetation index tested, the soil-reflected signal may be indistinguishable from that of the vegetation signature for an eco-georeferenced, seasonal, *N. fowleri* sample site, foci. The soft component may render VIs from the amoeba "soft-dependent" because its magnitude may vary with the reflectance properties of the underlying soil. Huete et al. (1985) found that the sensitivity of vegetation indices to soil background was greatest in canopies with intermediate levels of vegetation cover (50% green cover). With low vegetation, grid-stratifiable, LULC amounts, there may not be enough discontinuous, canopy ‘greenness' at a *N. fowleri* , eco-epidemiological, eco-georeferenceable, sample site to impart a robust scattered, soil-reflected endmember (sub-pixel) signature while, at very high vegetation densities, there may not be enough soil signal emerging from the canopy to be of significance. At intermediate levels of LULC, geoclassified vegetation, however, significant scattering and transmission of NIR flux through the canopy may produce a soil-reflected spectral endmember LULC signature of a capture point *N. fowleri* sample site that may be strongly iterated. This reference signature may reveal ungeosampled, unknown sample sites of the amoeba when applied as a dependent variable in a stochastic or deterministic, explanatorial interpolator.

Even though ground based validation of satellite-derived indices of biomass is well 65 explored in other ecosystems (Soenen et al., 2010, Gamon et al., 1995) Pontailer et al. 2003 66 only a few studies have related plot level spectral reflectance indices to aboveground biomass) there has been no contribution in the literature on mapping *N. fowleri* .The purpose of this study was to model the soil brightness dependency of partially vegetated LULC discontinuous canopy spectra emitted from an eco-georeferenced, capture point, *N. fowleri* sample site in order to develop a simple "global" VI that could account for dynamic soil-vegetation spectral behavior. We assumed that significant biophysical, eco-georferenceable, plant endmember ,LULC signature, geoclassified, parameters may be optimally forecasted in an ArcGIS cyberenvironment so as to identify unknown, ungeosampled *N. fowleri* sample site sites.

The transformation technique presented in this study is an important step toward the establishment of simple global VI signature models that can adequately describe dynamic soft-vegetation systems for eco-graphically cartographically delineating a eco-georferenecable, eco-epidemiological, *N. fowleri* capture point, on a geoclassified grid-stratfied, remotely sensed LULC site. A SAVI eco-georeferenced *N. fowleri* LULC site map in particular may be useful as the model may correct for over brightness of soil in the data bands.

To assess the importance of the vegetation and temperature co-factors, linear and nonlinear regression models were constructed to determine the statistical significance of any perimeters within a Poissonian regression probability framework associated with optimally parameterizable covariate estimators of *N. fowleri*. In statistics, Poisson regression is a generalized linear model (GLM) form of regression analysis employed to model count data and contingency tables(Cameron and Trivedi 1998*)*. In statistics, the GLM is a flexible generalization of ordinary linear regression (OLS) that allows for response variables that have error distribution models other than a normal distribution. In statistics, a contingency table (also known as a cross tabulation or crosstab) is a type of table in a matrix format that displays the (multivariate) frequency distribution of the variables (Christensen 1997) In statistics, ordinary least squares (OLS) or linear least squares is a method for estimating the unknown parameters in a linear regression model, with the goal of minimizing the sum of the squares of the differences between the observed responses (values of the variable being predicted) in the given dataset and those predicted by a linear function of a set of explanatory variables(Freedman 2005). We assumed that cartographically, this relationship could be robustly illustratable as the sum of the squared vertical distances between each eco-georeferenced *N. fowleri* sample site, grid-stratifiable,LULC data point in a remotely sensed dataset and the corresponding endmember capture point on the regression line.

The Poissonian regression assumed the *N. fowleri* response variable *Y* had a Poisson distribution, and assumed the logarithm of its expected value was able to be modeled by a linear combination of known environmental geosampled LULC parameters (e.g., discontinuous canopy spectral, wavelength frequencies). A Poisson regression model is sometimes known as a log-linear model, especially when employed to model contingency tables (Haight 1967). A log-linear model is a mathematical model that takes the form of a function whose logarithm equals a linear combination of the parameters of the model, which makes it possible to apply (possibly multivariate) linear regression(Christensen 1997).

Lastly an endmember, geoclassified LULC spectral signature of a confirmed infection geolocation was taken and run in ArcGIS using a hotspot analysis to create a kriged georeferenced *N. fowleri* map. In statistics, kriging or Gaussian process regression is a method for which iterable, quantitative, interpolative values (soil adjusted vegetation *N. fowleri* signal) are modeled by a Gaussian process governed by prior covariances, as opposed to a piecewise-polynomial splinechosen to optimize smoothness of the fitted values (Cressie 1993).In mathematical field of numerical analysis, spline interpolation is a form of interpolation where the interpolant is a special type of piecewise polynomial called a spline (Cressie 1993) In mathematics, a polynomial is an expression consisting of variables (or indeterminates) and coefficients that involves only the operations of addition, subtraction, multiplication, and non-negative integer exponents of variables. (e.g., polynomial of a single indeterminate x is $x^2 - 4x + 7$) A spline interpolation is often preferred over polynomial interpolationas the interpolation error can be made small even when employing low degree endmember signature, grid-stratified, geoclassified, LULC polynomials for the spline (Jacob et al. 2015), Spline interpolation avoids the problem of Runge's phenomenon, in which oscillation can occur between points when interpolating using high degree polynomials (Ronald S. (2004). Our assumption was that under suitable assumptions on the priors, kriging could render an optimal linear unbiased capture point for an *N. fowleri* sample site; eco-georeferenceable geolocation in ArcGIS employing a discontinuous proxy vegetation-related signature dependent variable (see Jacob et al. 2016,

Grifftih 2005). Interpolating methods based on other criteria such as smoothness of prevalence statistics or case distribution data may yield the most likely intermediate values for optimally identifying unknown, ungeosampled *N. fowleri*, sample site geolocations. 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).

The basic idea of kriging a *N. fowleri,* vegetation, proxy, endmember signature is to prognosticate the value of a function at a given capture point by computing a weighted average of the known values of the function in the neighborhood of the point. The method is mathematically closely related to regression analysis. Both theories derive a best linear unbiased estimator, based on assumptions on covariances, make use of Gauss–Markov theorem to prove independence of the estimate and error, and make use of very similar formulae.

In statistics, the Gauss–Markov theorem, states that in a linear regression model in which the errors have expectation zero and are uncorrelated and have equal variances, the best linear unbiasedestimator of the coefficients may be rendered by the ordinary least squares (OLS) estimator, provided it exists (Aitken 1935). In statistics, OLS or linear least squares is a method for estimating the unknown parameters in a linear regression model, with the goal of minimizing the sum of the squares of the differences between the observed responses (values of the variable being predicted) in the given dataset and those predicted by a linear function of a set of explanatory variables(e.g., empirical, *N. fowleri* vegetation proxy signature,regressors). Visually this may be seen as the sum of the squared vertical distances between each data point (georeferenced *N. fowleri* capture point) in the dataset and the corresponding point on the regression line – the smaller the differences, the better the model fits the data. According to Jacob et al. 2009 the OLS estimator is consistent in an infectious epidemiological, linear dependent , forecast, vulnerability model when the regressors are exogenous, 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 provides minimum-variance mean-unbiased estimation when the errors have finite variances. Under the additional assumption that the errors are normally distributed, OLS is the maximum likelihood estimator (Christensen, R., 1997).

In this research "best" was defined as giving the lowest variance of the estimate, as compared to other unbiased, linear estimators in the *N. fowleri* vegetation proxy signature, iterative, interpolative, explanatory model. The errors do not need to be normal, nor do they need to be independent and identically distributed (only uncorrelated with mean zero and homoscedastic with finite variance) in a geospectrotempoeal, endmember, signature interpolator (Jacob et al. 2011). The requirement that the estimator be unbiased cannot be dropped in a VI biogeophysical, LULC model, since biased estimators exist with lower variance (Hay 1997). Kriging may be made for estimation of a single realization of a *N. fowleri,* LULC endemic signature random field, while regression models may be based on multiple observations of a multivariate county-level geosampled dataset.

Henceforth, our objective the objectives of this study were to: (a) construct LULC base maps from Google Earth ^{TM} data (Figure 1b) establish the sensitivities and dynamic ranges of NDVI, and SAVI signature maps (b) examine spectral reflectance of vegetated surrounding using Landsat data, and (d) determine the environmental variables associated with distribution and abundance of *N. fowleri* at the county-level in Florida.

International Journal of Geographic Information System
Vol. 4, No. 3, October 2017, pp. 1-31
Available Online at http://acascipub.com/Journals.php

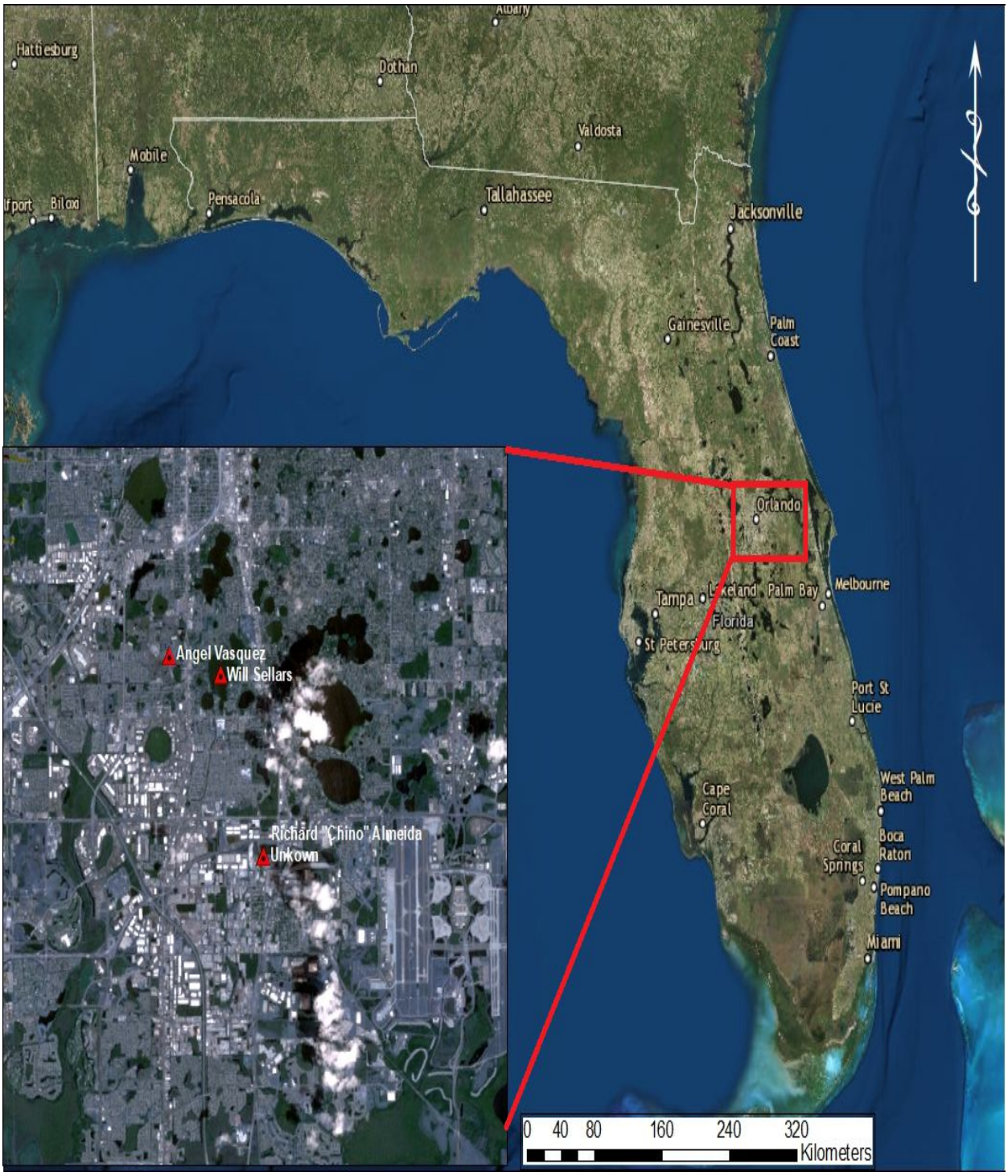


Figure 1. A map of Orange County, Florida and the four confirmed cases being studied

2.Materials and Methods:

2.1Study site;

The geosampled, county-level, data for this project was acquired by phone interview from the Hillsborough Public Health office in Tampa, Florida, as counties with registered cases. This presented two problems the first being, due to the rarity of this type of infection, thorough documentation of cases was not implemented until the early 2000's; so the data used in this project only included 4 confirmed cases for Orange County, Florida from 2000-2015. The second being that when a patient died from a PAM infection they were registered to their county of residence, not to the county in which they initially contracted the amoeba. Thus upon receiving this 'general data', research into news outlets of cases matching the counties of the patients was conducted to find the location of infection. Many different news reports were analyzed, spanning the entirety of the patients' infection, to obtain the most correct data and details of the case.

After finding the most likely water source of infection from the 'general data' and news data, longitudes and latitudes, in decimal degrees, were ascertained via Google Earth based on the location of infection. These points were converted to an excel data table to be used in Statistical Analysis Software (SAS), and into a shape file to be used in GIS.

2.2 Remote Sensing data:

Landsat 7 images of the study site, encompassing the visible and near inferred (NIR) bands were acquired from https://landsat.usgs.gov/. TheLandsat Enhanced Thematic Mapper sensoris carried on Landsat 7, and images consist of eight spectral bands with a spatial resolution of 30 meters for Bands 1 to 7. The resolution for Band 8 (panchromatic) is 15 meters (see Table 1). All bands can collect one of two gain settings (high or low) for increased radiometric sensitivity and dynamic range, while Band 6 collects both high and low gain for all scenes. The approximate scene size is 170 km north-south by 183 km east-west (106 mi by 114 mi).

Table 1 Landsat 7 * ETM+ wave band data

Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 - Blue	0.45-0.52	30
Band 2 - Green	0.52-0.60	30
Band 3 - Red	0.63-0.69	30
Band 4 - Near Infrared (NIR)	0.77-0.90	30
Band 5 - Shortwave Infrared (SWIR) 1	1.55-1.75	30
Band 6 - Thermal	10.40-12.50	60 * (30)
Band 7 - Shortwave Infrared (SWIR) 2	2.09-2.35	30
Band 8 - Panchromatic	.52-.90	15

* ETM+ Band 6 is acquired at 60-meter resolution, but products are resampled to 30-meter pixels.

The main features of Landsat 7 Plus (ETM+)data include 1) panchromatic band with "15 m (49 ft)" spatial resolution (band 8), 2) Visible (reflected light) bands in the spectrum of blue, green, red, near-infrared (NIR), and mid-infrared (MIR) with 30 m (98 ft) spatial resolution (bands 1-5, 7), 3) A thermal infrared channel with 60 m spatial resolution (band 6); and, 4) Full aperture, 5% absolute radiometric calibration.

Digital elevation data for the study site was acquired from http://earthexplorer.usgs.gov. Characteristics of drainage networks and drainage basin physiographic parameters have been used in hydrologic calculation and modeling flood and swamp vector arthropod abundance, in real time, using high resolution data (Jacob et al. 2008, Shaman et al. 2002). Automated generation of drainage networks has become increasingly popular with the use of GIS and availability of digital elevation models (DEMs) (www.esri.com). These models account for topographic variability and their control over soil moisture heterogeneity and runoff within a shed. For example, Patz et al. (1998) used a water balance ArcGIS model to hindcast weekly soil moisture levels in the Lake Victoria basin. These soil moisture levels were then associated with local human biting rates and entomologic inoculation rates. Jacob et al. (2010), evaluated environmental factors such as elevation range to determine human Eastern Equine Encephilitis virus (EEEV) risk in the greater Chicago area. The model yielded several catchment hydrological variables including percent surface saturation, and total surface runoff for identification of potential, productive, *Culex erectus*aquatic, larval habitats. Hence we assumed that an LULC -epidemiologic approach may be used to examine whether digital elevation and vegetation patterns account for some of the variance in densities of *N. fowleri* prevalence in Orange County, Florida.

2.3 Regression Analyses

The relationship between the geosampled, county-level, *N. fowleri* indicators was investigated by single variable regression analysis in PROC REG using SAS. Prevalence data (binomial fractions) was employed in a regression model; as is standard practice for the analysis of the county-level data. Poisson probability regression analyses were also employed to infer the relationship between the geosampled, time-series, explanatorily, data variables and the archived county-level anthropogenic characteristics (i.e., independent variables).

Regression analyses assumes independent counts (i.e., *Ni*), taken at multiple, geosampled, georeferenced, county-level sub-geolocations, $i = 1, 2, \boxtimes, n.$ The geo-spatiotemporal related county-level counts were then described by a set of variables denoted by matrix, $\mathbf{x}_i$, where a $1 \times p$ was a vector of covariate coefficient indicator values for a interpretively geosampled, endemic transmission-oriented, explanative *N. fowleri* foci *i*. The expected value of the data was given by $\mu_i(\mathbf{X}_i) = n_i(\mathbf{X}_i)\exp(\mathbf{X}_i\beta),$ where β was the vector of the parameterizable non-redundant covariates in the endemic, transmission-oriented, interpretively operationalizable, eco-epidemiological, prognosticative, county-level, forecast, vulnerability model where the Poisson rates were given by $\lambda_i(\mathbf{X}_i) = \mu_i(\mathbf{X}_i)/n_i(\mathbf{X}_i).$ The rates parameter $\lambda_i(\mathbf{X}_i)$ was both the mean and the variance of the Poisson distribution for each geosampled, endemic, *N. fowleri*-related, eco-georeferencable, county-level, zip code geolocation *i*. The dependent variable was water temperature at the time of infection. The Poisson probability regression model assumed that the geosampled, expository predictors were equally dispersed, that is, that the conditional variance equaled the condition mean. Partial correlations were then defined after introducing the concept of conditional distributions.

Jacob et al. (2009) found, if over-dispersion is a feature in an asymptotical, seasonal, explanatory, predictive, risk–related, endemic, transmission-oriented, distribution model, then an alternative model with additional free parameters may provide a better fit; which is why a Poisson mixture model with a negative binomial distribution was employed here. The mean of the Poisson distribution was, itself, a random variable drawn from the gamma distribution; thereby, introducing an additional free parameter in the empirical-geosampled, georeferencable, county-level, endemic, *Naegleria* distribution model. The family of negative binomial distributions is a two-parameter family which uses several parameterizations for treating over dispersion data (Hosmer 1980). The Poisson distribution has one free parameter and does not allow for the variance to be adjusted independently of the mean (Hansen, 2007).

A parameterization technique was then employed in PROC REG such that any two eco-georferenced, county-level, explanatorial regressable, geo-spatiotemporal geosampled, operationalizable, sociodemiographic county level prognosticative variables p and r were within the ranges of $0 < p < 1$ and $r > 0$. Under this parameterization, the probability mass function (pmf) of the asymptotical, geosampled, asymptotically, normalized, time series, clinical, field and remote-specified, explanatorial, predictor variables with a NegBin distribution took the following form: for $k = f(k; r, p) = \binom{k + r - 1}{k} \cdot p^r \cdot (1 - p)^k \; 0, 1, 2,$ where $\binom{k + r - 1}{k} = \frac{\Gamma(k + r)}{k!\,\Gamma(r)} = (-1)^k - \binom{-r}{k}$ and $\Gamma(r) = (r - 1)!$.

The negative binomial regression was conducted using an alternative parameterization for optimally regressively geo-spatiotemporally quantitating the, explanatorial, geosampled, county-level, eco-georferenced, *N. fowleri* LULC data employing the mean ($\lambda : \lambda = r \cdot \left(p^{-1} - 1\right), = \frac{r}{r + \lambda}$) which also implemented the mass function which then became:

$$g(k) = \frac{\lambda^k}{k!} \cdot \frac{\Gamma(r + k)}{\Gamma(r)(r + \lambda)^k} \cdot \frac{1}{\left(1 + \frac{\lambda}{r}\right)^r},$$

where λ and r were the parameters. Under this parameterization, we were able to generate: $r \underrightarrow{\lim} \infty g(k) = \frac{\lambda^k}{k!} \cdot 1 \cdot \frac{1}{\exp(\lambda)}$ which was noted, resembled the mass function of a Poisson-distributed random variable with Poisson rate. In other words, the alternatively, parameterized, negative, binomial non-homogeneous gamma distributed mean distribution generated from the regressed county-level, endemic *N. fowleri* explanatorial, asymptotical, operationalized, time-series, eco-georeferencable, parameterizable county-wide, sociodemiographic covariates converged to the Poisson distribution, and r controlled the deviation from the Poisson. This made the negative binomial habitat, gamma distributed model suitable as a robust alternative to the Poisson regression-based framework for risk modeling the endemic, geosampled, transmission-oriented, risk-related, explanatorily, operationalizable, interpolative, time-series, dependent predictors.

The negative binomial distribution of the asymptotical, geosampled, georeferencable explanatory, operationalizable, time-series, parameterizable county-level *N. fowleri* covariates arose as a continuous mixture of Poisson distributions where the mixing distribution of the Poisson rate was a gamma distribution. The mass function of the negative binomial distribution of the geosampled endemic, transmission-oriented, county level, sociodemiographic explanators then was written as:

$$f(k) = \int_0^\infty Poisson\,(k|\lambda) \cdot Gamma\,(\lambda|r, (1 - p)/p)d\lambda$$

$$= \int_0^\infty \frac{\lambda^k}{k!} \exp(-\lambda) \cdot \frac{\lambda^{r-1} \exp(-\lambda p/(1 - p))}{\Gamma(r)((1 - p)/p)^r} d\lambda$$

$$= \frac{1}{k!\,\Gamma(r)} p^r \frac{1}{(1 - p)^r} \int_0^\infty \lambda^{(r+k)-1} \exp(-\lambda/(1 - p))d\lambda$$

$$= \frac{1}{k!\,\Gamma(r)} p^r \frac{1}{(1 - p)^r} (1 - p)^{r+k} \Gamma(r + k)$$

$$= \frac{\Gamma(r + k)}{k!\,\Gamma(r)} p^r (1 - p)^k.$$

2.4 LULC model:

Spatial analysis was performed by synthesizing a gridded, LULC map employing high resolution, cloud-free Landsat 7 imagery available, from Earth Explorer, of the study area. The resulting LULC supervised classification showed the study site to contain four unique classification areas; (1) Urban, (2) Vegetation, (3) Ocean, and (4) Lakes. To assess the accuracy of the LULC classes, a method of empirically analysis was conducted by selecting the LULC classes and comparing them with reference data, in the form of google satellite imagery obtained online from Earth Explorer. To perform this analysis, a random selection of 25 pixels in each LULC class were assessed and compared to the satellite classification as in (Jacob 2008).

2.5 Vegetation Indices:

Disregarding weather conditions, land cover classification, plant physiognomy, and soil type, NDVI and SAVI were calculated in ArcMap 10.5.1® using Landsat 7 satellite captured images. Landsat 7 data is captured in 8 different bands of light, for the purpose of the conducted calculations the 3rd band, which corresponds to the red band of visible light, and the 4th band, which corresponds to the near infrared band, were used in both the NDVI and SAVI calculations utilizing the equations:

$$SAVI = \frac{NIR - RED}{(NIR + RED + L)} \times (1 + L)$$

NI d 1.0, with values closer to -1 d values closer to 1.0 indicati 991). Soil adjusted vegetation indices in areas where vegetative cover is e is exposed, the reflectance of light in the red and near infrared spectra ca... influence vegetation index values (Huete 1994).

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} \times \rho_{RED}}$$

In previous research, Jacob et al. (2016), constructed multiple NDVI and NDVI-variant geographic maps using QuickBird visible and NIR data and ecogeoreferenced West Nile mosquito vector *Cx. pipiens* explanatory larval habitat predictor covariates geosampled in a mosquito abatement district in northern Illinois. This paper presented a model for the gap probability of a discontinuous vegetation canopy, such as forest, or shrubland. The negative exponential attenuation of light within individual plant canopies where a capture point *Cx. pipiens*was geosampled was assumed, and as such the problem of modeling the gap probability then become the problem of estimating the distribution of habitat distances within canopies using theNDVI and NDVI-variant. This, was not difficult since the distribution of individual canopy sizes and shapes was known and individual canopies were randomly distributed but did not overlap. A comparison of modeled gap probabilities with observed gap probabilities for a pine stand showed good agreement for zenith angles of illumination up to about 45". Above 45", the fit worsened, presumably because the horizontal branch structure of the pine canopy at the *Cx. pripiens* capture point was less attenuating as the illumination angle approached the horizon. The paper derives two simple indexes that are functions of leaf area index, leaf angle distribution, and count density (number per square unit of area) and size (base radius and height) of plant canopies associated to *Cx, pipien* capture points (assumed to be spherical or ellipsoidal). These indexes were used 1) to assess quantitatively the difference between continuous and discontinuous models of the same translucent vegetation*Cx pipiens* habitat canopy, thus revealing when the use of the more complex, discontinuous model is seasonally warranted; and 2) identify cases in which a simpler, discontinuous but opaque model will yield good results. Their models revealed that NDVI and SAVI parameters could quantify prolific *Cx. pipien*habitats based on geo-spatiotemporal field geosampled, eco-georeferenced,immature count data. Here, SAVI were computed, employing the equation above for this investigation. An adjustment factor (L) of 0.5 was used, as it was shown to reduce soil-induced noise throughout a range of vegetation densities.

3. Results

The LULC map shown in Figure 2 was created in ArcGIS using DEM data and the supervised classification tool. All four of the geosampled, geospatially county-level, *N. fowleri* cases occurred in an area of water with surrounding vegetation. Recently,Jacob et al. (2015)] modeled the reflectance of a black fly vector of onchocerciasis ,*Similium damnsoums* employing trailing vegetation as a function of fixed signatures of foliage, grass, litter, and bare soil under sunlit and shaded conditions as weighted by their areal proportions within a sub-

meter resolution scene. The areal proportions were determined by a three-layer model (tree, shrub, and ground) A three-dimensional geometrical model account fully for the shadowing of each layer upon the next and for the disappearance of shadows as illumination and viewing positions converge (hotspot effect).Here,a resultant map revealed that the, *N. fowleri* cases occurred in predominantly urban areas, this is most likely due to the nature of the study site rather than its significance with this amoeba. Orange County, Florida was chosen as the experimental location because it has the highest number of confirmed cases per zip code in Florida.

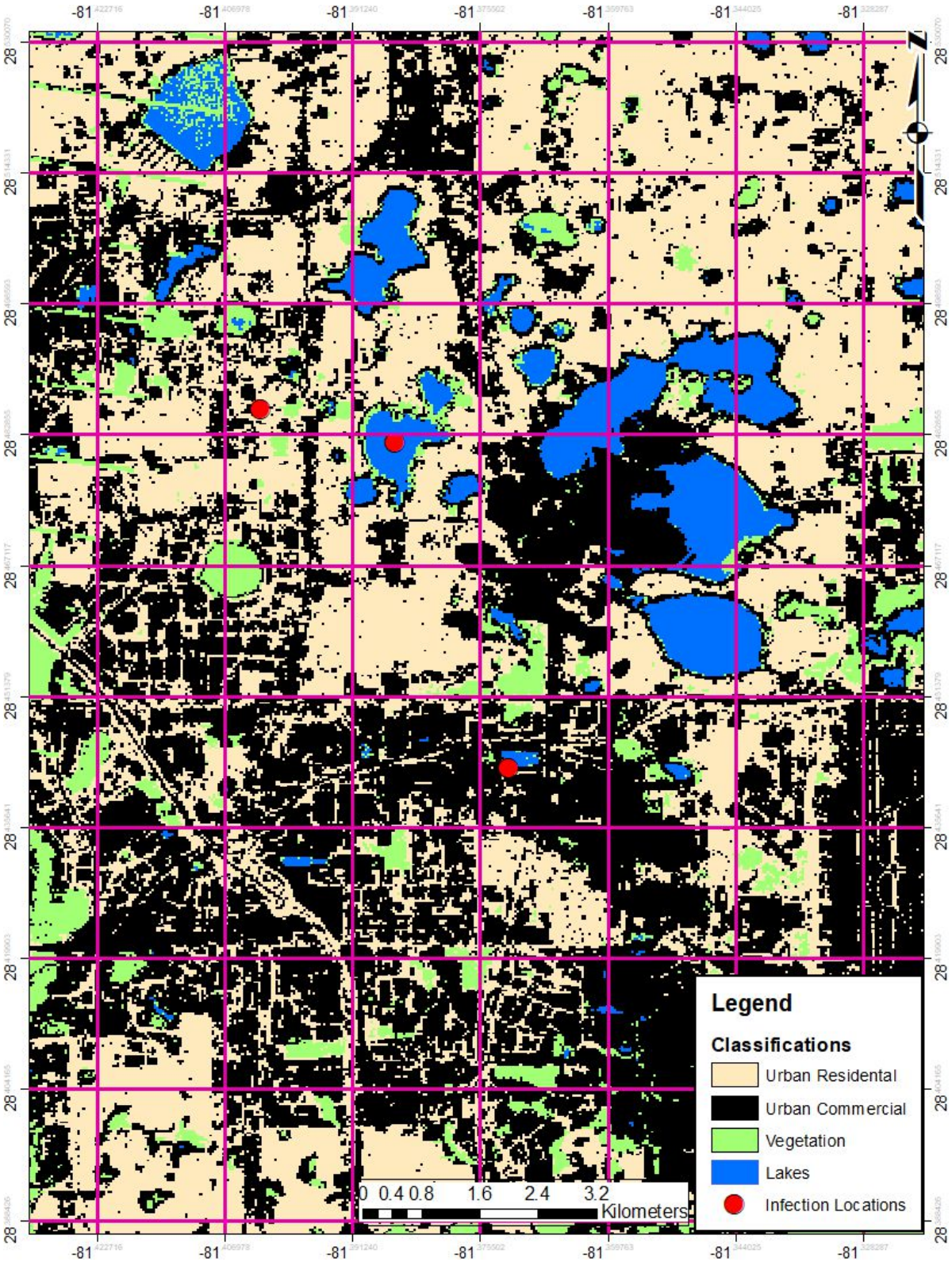


Figure 2: Supervised Classification of Naegleria fowleri cases in Orange County, Florida 2000-2015.

With conformation of the possible importance of vegetation as a cofactor for prevalence an NDVI map was modeled as shown in Figure 3. The NDVI map gave each geosampled, county-level positive infection location a NDVI score which was employed as a explicative regressor in the linear model for this experiment. NDVI scores are values given to a pixel using NIR and red reflectance bands to calculate the amount of light in each wavelength was absorbed by the chlorophyll in vegetation; by the nature of the equation shown above the NDVI score will fall between -1 and 1 with a positive score indicating more abortion of light in the NIR band and thus more green vegetation (Jacob 2016).

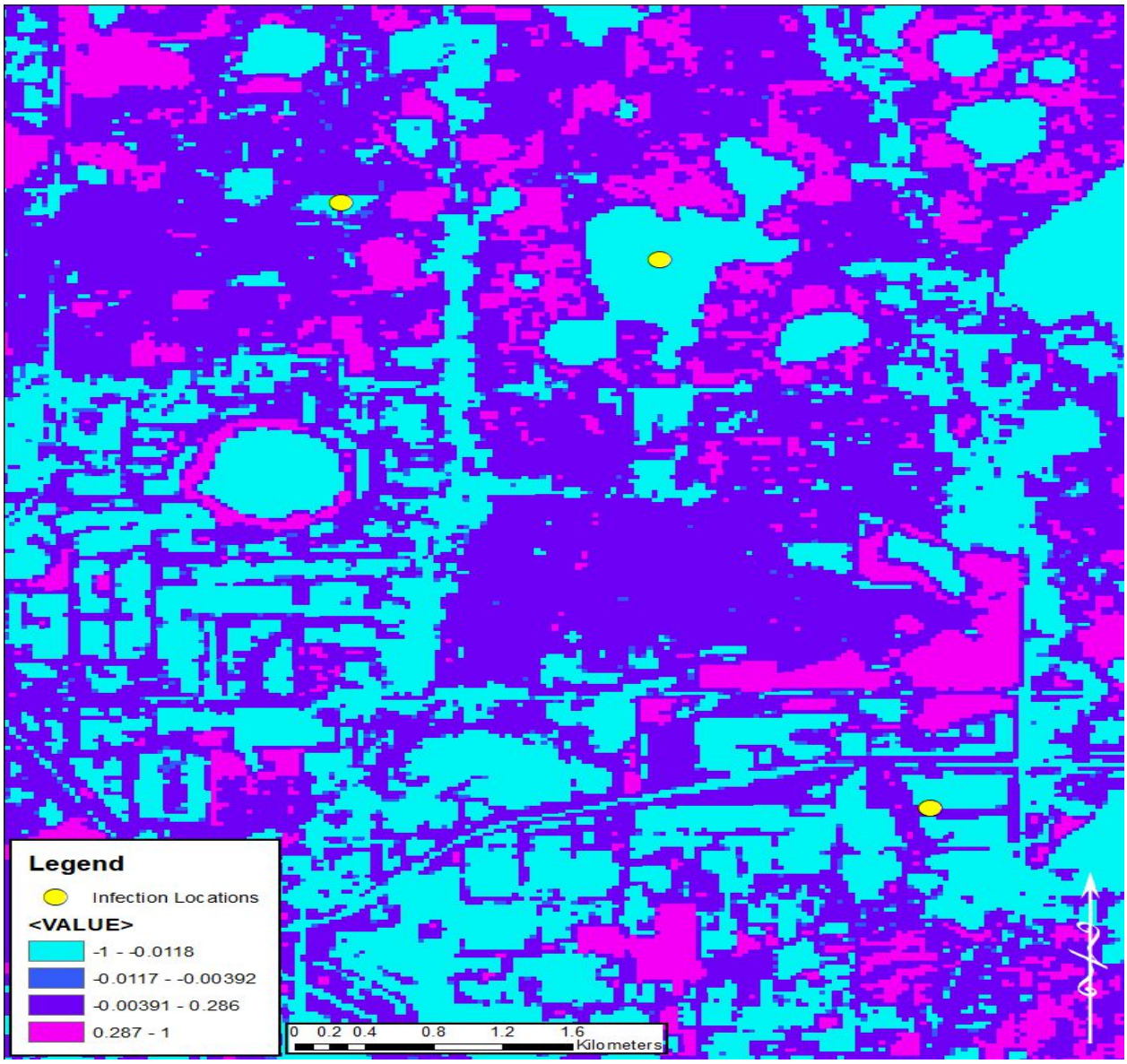


Figure 3. NDVI map of *N fowleri* cases in Orange County, Florida 2000-2015.

SAVI vegetation maps such as the one shown in Figure 4 were adjusted for lower lying vegetation which is commonly observed in urban commercial and urban residential areas in Florida. From prior knowledge of the study site it was understood that an SAVI model would more accurately represent the type of relatively low-lying, sparse, over dispersed vegetation found in a large urban county. SAVI, like NDVI, is an index that measures the levels of absorption of light in the NIR and red bands (Tucker 1991, Huete 1994). However, the SAVI models were corrected

with an L value to mitigate the reflectance of bear soil and urban neighborhoods that may otherwise disrupt the calculation of the georeferenced *N. fowleri* sites. Both NDVI and SAVI scores were important data figures in the effort to understand if vegetation is indeed a cofactor to amoeba prevalence.

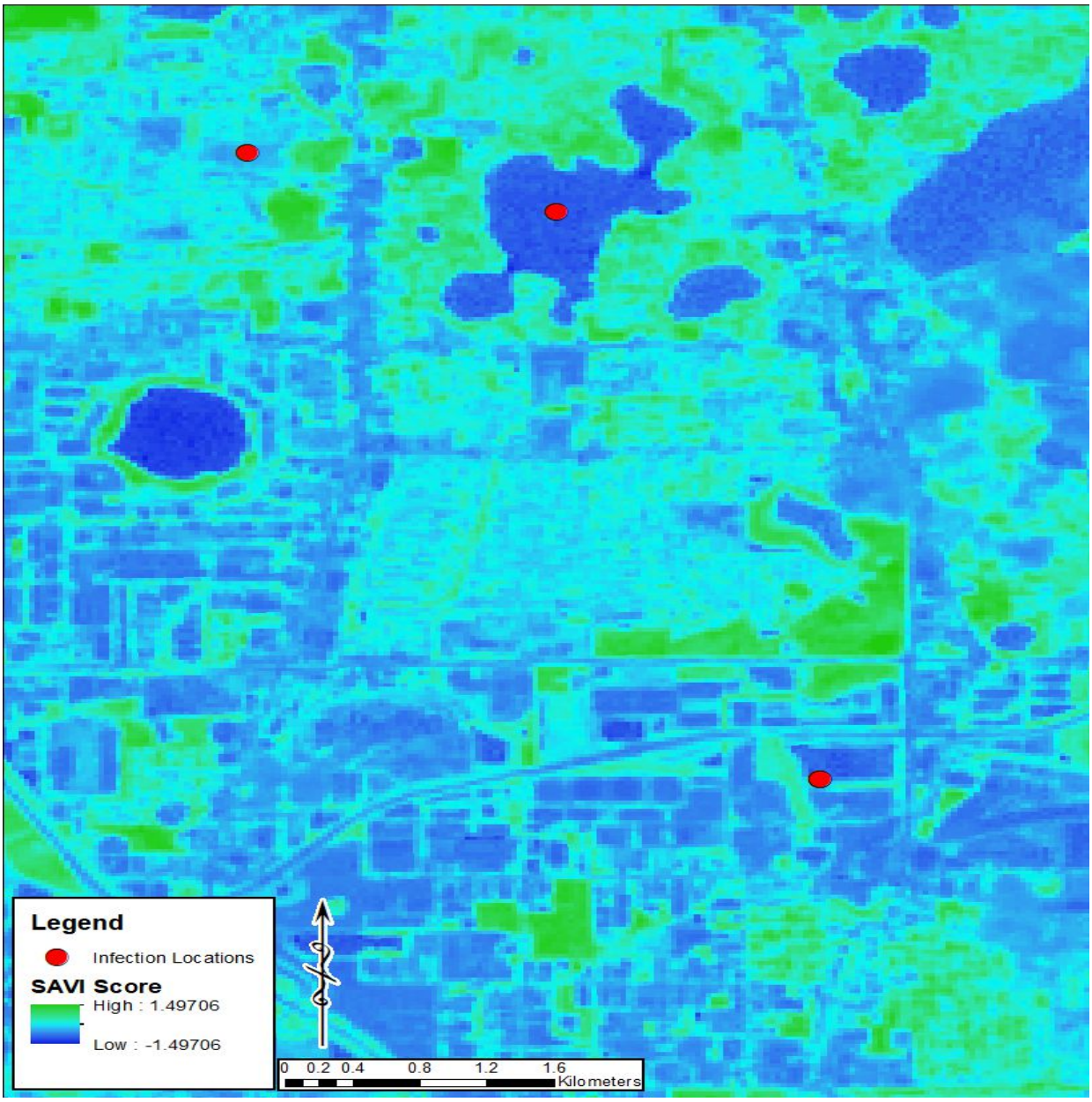


Figure 4. SAVI map of *N. fowleri* cases in Orange County, Florida 2000-2015

The relationship between county-level prevalence and each explanatory, individual, potential, *N. fowleri*geosampled regressor was investigated by employing single variable regression analysis in PROC NL MIXED. The first line of the code began the PROC REG command. The second line specified the fixed portion of each, endemic, epidemiological risk model, [i.e., the model without the random intercept, value (i.e., *xb*)]. The third line of code created a value (i.e., rand) that was equal to the fixed part of the model (*xb*) plus a random intercept term *u*. The model statement specified that the parameterizable unbiased, explanative, covariate, geo-spatiotemporal

estimators were distributed (~) normally with a mean of *x*b and variance *s*2. The random statement defined the random effect *u* whilst optimally quantitating the normally, distributed, operationalizable, time series data with mean zero and a variance term. In so doing, *s*2*u* was optimally solved.

The level 2 units were identified by subject = *id* in PROC REG. Importantly, the last two lines of the command in PROC REG were predictive statements. While traditionally only a single set of predicted values are created when constructing time-series, clinical, field and/or remote-specified, vector, iterative interpolative, endemic *N. fowleri*, elucidative, geosampled, georeferencable, forecasting epidemiological, zip-code-level, sociodemographic, explanatory, probabilisc risk models, we generated two. The predict statement in SAS rendered the explanatorial, county-level, time-series, predicted values for the fixed portion of the model. The model identified *x*b, and output a dataset called output-fixed. The second predict statement generated the time-series, regressed, endemic, county-level *N. fowleri*, transmission-oriented, zip-code, eco-georferenceable, polygon predicted values that included the estimate of the random intercept in addition to the quantitated, fixed portion, randomized estimates.

We employed the regression line $(\mathbf{y}_i - \bar{\mathbf{y}}) = (\hat{\mathbf{y}}_i - \bar{\mathbf{y}}) + (\mathbf{y}_i - \hat{\mathbf{y}}_i)$ to generate a pseudo R^2 value where the first term was the total variation in the response *y* (county-level *N. fowleri* prevalence) and the second term was the variation in mean response based on the geosampled, asymptotical, normalized, parameterizable, sociodemographic, zip-code-level, covariate estimators. The third term was the rendered residually forecasted, elucidative regressed geosampled, endemic, transmission-oriented, derivative values in the operationalizable, time-series, interpolative, endemic transmission-oriented, VI,endmember,LULC signature risk model derivatives. Squaring each of these terms and adding over all of the geosampled, georeferencable, county-level*N. fowleri,* sociodemographic observations generated the equation $\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$. This equation was written as SST = SSM + SSE, where SS was notation for sum of squares and *T*, *M*, and *E* were the notation for total quantized model error estimates. The square of the sample correlation was then equal to the ratio of the estimates while the sum of squares was related to the total sum of squares: r^2 = SSM/SST. This formalized the interpretation of R^2 for explaining the fraction of variability in the geosampled, county-level epidemiological, zip code,polygon, eco-georfeenced, LULC, data explained by the regression model. The sample variance s_y^2 was equal to $\sum \frac{(y_i - \bar{y})^2}{n-1}$, which in turn was equal to the SST/df, the total sum of squares divided by the total df.

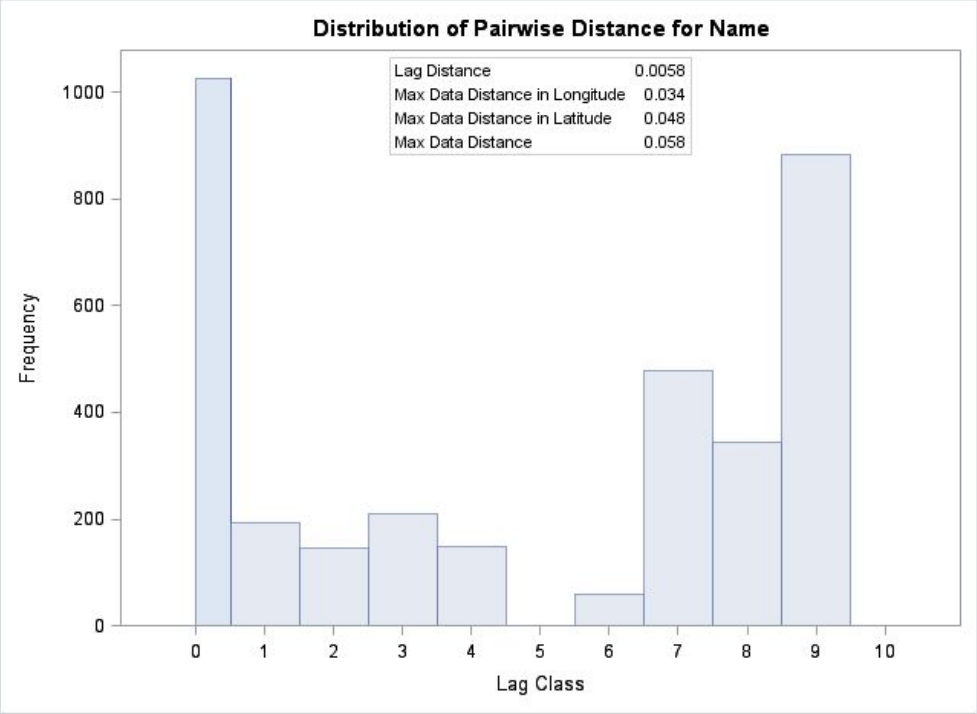
Next a regression equation was constructed by employing the mean square model $(\text{i.e., MSM}) = \sum \frac{(\hat{y}_i - \bar{y})^2}{l}$, which was equal to the SSM/df. The corresponding MSE was $\sum \frac{(y_i - \hat{y}_i)^2}{n-2}$ which was determined to be equal to SSE/df and the quantitated, time-series, operationalizable, county-level, *N. fowleri*, endemic, transmission-oriented, explicatory, georeferencable estimate of the variance about the regression line $(\text{i.e., } \sigma^2)$. The MSE is an estimate of σ^2 for determining whether or not the null hypothesis is true (Draper and Smith 1981).

For robustly, parsimoniously, quantizing, the geo-spatiotemporal, geosampled, operationalizable, endemic, transmission oriented, explanatory, interpolative, endemic, county-level *N. fowleri,* georeferencable prognosticators (*p*) a DFM, was generated which we noted was equal to *p* and the error degrees of freedom (DFE). This product was also equal to $(n - p - 1)$, and the total degrees of freedom (DFT) which was subsequently equal to $(n - 1)$. The sum of DFM and DFE was determined. The eco-epidemiological, *N. fowleri* geosampled dataset had binary and dichotomous predictors; the model reviled a non-Gaussian curve as shown in Figure 5.

Non-normality is common in logistic regression due to the violations of assumption (non-homoskedastic propogagational parameters)here are four principal assumptions which justify the use of linear regression models for purposes of inference or prediction:(i) linearity and additivity of the relationship between dependent and independent variables: (a) The expected value of dependent variable is a straight-line function of each independent

variable, holding the others fixed.(b) The slope of that line does not depend on the values of the other variables. (c) The effects of different independent variables on the expected value of the dependent variable are additive.(ii) statistical independence of the errors (in particular, no correlation between consecutive errors in the case of time series data)(iii) homoscedasticity (constant variance) of the errors (a) versus time (in the case of time series data)(b) versus the prediction (c) versus any independent variable(iv) normality of the error distribution.(Christensen, R., 1997). If for example multiorrelinearity exists in a logistic regression, *N. fowleri* frequency landscape model, the vulnerable forecasts will be misspecified. Even if an epidemiologist or other researcher is able tease out noisy variables in a forecast vulnerability, eco-epidemiologcal, *N. fowleri* model, the residuals would not be able to be eco-cartographically robustly illustratable, (e.g., percentage of discontinuous chlorophyll-A in a sparsely canopied georeferenced *N. fowleri* sampled, capture point).

The results of the Poissonian model are shown in Figure 5. In probability theory and statistics, the Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known constant rate and independently of the time since the last event(Christensen, R., 1997).. The lag classes' of the regression model show a two-tailed distribution of data. This may be because of non-normality caused by violations of assumption or it may be because of the nature of the data itself.



Autocorrelation Statistics

Assumption	Coefficient	Observed	Expected	Std Dev	Z	Pr > |Z|
Normality	Moran's I	-0.0431	-0.012	0.0320	-0.972	0.3311
Normality	Geary's c	1.0865	1.000	0.0335	2.583	0.0098

Figure 5. Summary of the eigenfucntion decomposition autocorrelation

An ordinary kriging interpolation stochastic algorithm was employed to reveal unknown, un-gepsampled, *N. fowleri*, eco-georeferenceable, capture point. We made the assumption that there was a ``regionalized variable" *Z* definable at each *N. fowleri*, sample point on a county map which had *N* data locations $x_i,\ i = 1 \ldots N$ and a value z_i at each geolocation x_i. We derived the best linear unbiased estimator of *z*, which (i.e., z^*), at a geolocation capture point *x*. In the process of doing so, we discovered the necessity of making certain assumptions (ie.., intrinsic hypothesis') for quantitating the eco-georeferenced, capture point. Note that the weights λ_i and z_i were actually functions of both *x* and x_i. Our strategy was similar to Jacob et al. (2015). A linear estimate of the z_i was employed to determine the λ_i such that $\mathrm{Var}(z(x) - z^*(x))$ was a minimum, and such that the proxy *N. fowleri* signature LULC estimator was unbiased: $E[z(x) - z^*(x)] = 0.$ For the latter condition, we introduced the assumption that $E[z(y) - z(x)] = 0$ for all sampled *N. fowleri*, capture point, geolocations *x* and *y*. In so doing, a sufficient condition for unbiasedness became $\sum_{i=1}^{N} \lambda_i = 1.$

$$z^*(x) = \sum_{i=1}^{N} \lambda_i(x; \{x_j\}) z_i$$

Since we considered only one eco-georeferenced, capture point, *N. fowleri*, geolocation *x*, we suppressed the dependence of λ_i on *x* and x_i. Because the weights were not fixed, but had to be recalculated at each geolocation we employed $z(x_i)$ as a response variable. Taking advantage of this constraint, the model revealed

$$E[z - \sum_{i=1}^{N} \lambda_i z_i]^2 = E\left[\sum_{i=1}^{N} \lambda_i (z - z_i)\right]^2 .$$

[3.1]. We minimized the function employing

$$E\left[\sum_{i=1}^{N} \lambda_i (z - z_i)\right]^2 + 2\mu(1 - \sum_{i=1}^{N} \lambda_i),$$

where $z = z(x)$ was the true value at the county-level geolocation *x*. Expanding the first term we found that the geosampled, forecast, vulnerability *N. fowleri*, eco-epidemiological, autorrelated model revealed

$$E\left[\sum_{i=1}^{N} \lambda_i (z - z_i)\right]^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j E[(z - z_i)(z - z_j)]$$

. Note now the model looked like:

$$E[(z - z_i)(z - z_j)] = \frac{1}{2}\left[E[z - z_j]^2 + E[z - z_i]^2 - E[z_i - z_j]^2\right],$$

We replaced $E[(z - z_i)(z - z_j)]$ by theeco-georferenced, capture point, *N. fowleri*, model endmember, grid-stratified, LULC, expansion employing the sum(3.1),and simplified the model to

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j E[(z - z_i)(z - z_j)] = 2\sum_{i=1}^{N} \lambda_i \frac{1}{2} E[z - z_j]^2 + \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \frac{1}{2} E[z_i - z_j]^2$$

Differentiating with respect to λ_k (for each $k = 1 \ldots N$) and μ led to the following linear system of equations:

$$\begin{cases} \sum_{i=1}^{N} \frac{1}{2} E[z_k - z_i]^2 \lambda_i + \mu = \frac{1}{2} E[z_k - z]^2, & k = 1 \ldots N \\ \sum_{i=1}^{N} \lambda_i = 1. \end{cases}$$

Without further assumptions we iteratively quantitatively interpolated the *N. fowleri* endmember LULC signature employing the quantities $\frac{1}{2} E[z_k - z]^2$ and $\frac{1}{2} E[z_k - z_i]^2$ which were optimally derived from a theoretical variogram. In spatial statistics the theoretical variogram is a function describing the degree of spatial dependence of a spatial random field or stochastic process (Cressie 1993). We made the assumption that $\gamma(x - y) - \frac{1}{2} E[z(x) - z(y)]^2$ existed in the model and was independent of an capture point, *N. fowleri* county-level, eco-georeferenceable, geolocation (γ). This expression was then written as $\gamma(\underline{h}) = \frac{1}{2} E[z(\underline{x} \mid \underline{h}) \quad z(\underline{x})]^2$, where e indicated a vector quantity (e.g. $\underline{h}$) in ArcGIS.

The two assumptions we made (the one above, plus the earlier assumption that the mean is constant) together comprised the so-called intrinsic hypothesis, which must be satisfied in order to derive the ordinary kriging equations (Cressie 2003). Putting it all together, this system derived a eco-georeferenceable, capture point, *N. fowleri* matrix which revealed

$$\begin{bmatrix} \gamma(x_1 - x_1) & \gamma(x_1 - x_2) & \cdots & \gamma(x_1 - x_N) & 1 \\ \gamma(x_2 - x_1) & \gamma(x_2 - x_2) & \cdots & \gamma(x_2 - x_N) & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ \gamma(x_N - x_1) & \gamma(x_N - x_2) & \cdots & \gamma(x_N - x_N) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(x_1 - x) \\ \gamma(x_2 - x) \\ \vdots \\ \gamma(x_N - x) \\ 1 \end{bmatrix},$$

which we rewrote concisely as $\begin{bmatrix} \Gamma & \underline{1} \\ \underline{1}^T & 0 \end{bmatrix} \begin{bmatrix} \underline{\lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \underline{\gamma^x} \\ 1 \end{bmatrix}$ where Γ was the matrix of variograms ($\Gamma_{ij} = \gamma(x_i - x_j)$), $\underline{1}$ was a column vector of 1's, and $\underline{\gamma^x}$ was a vector of variogram values relating the position at which one could optimally estimate (x) a capture point, *N. fowleri* data for determining unknown sample sitegeolocations using (x_i): $\gamma_i^x = \gamma(x_i - x)$ .Our strategy (expressed in mathematical terms) was as follows: $z^*(x) = \sum_{i=1}^{N} \lambda_i(x; \{x_j\}) z_i$ .The autocorrelation model was able to distinguish the irradiance frequencies rendered from the georeferenced sample capture point as shown in Figure 6.

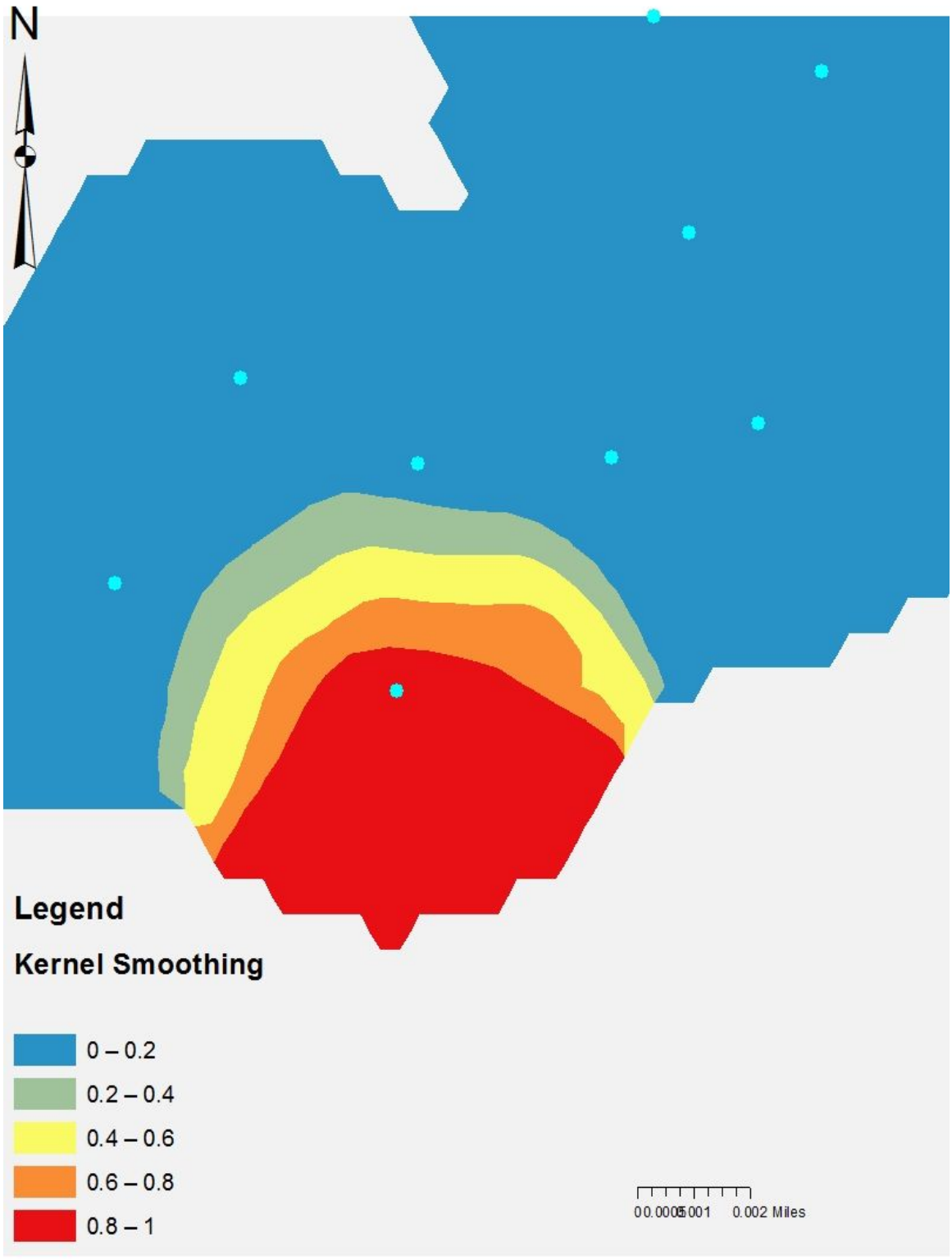


Figure 6. Probability prediction model for unknown, unsampled hot spot of *N fowleri* in Orange County, Florida.

4. Discussion

Initially, Poisson regression was employed to model the geo-spatiotemporal zip code data at the county study site. Unfortunately, overdispersion in the zip code regression-related LULC coefficients suggested that the Poisson model was inappropriate for differentiating the county endemic, *N. fowleri,* prognostic ative, explanatory, eco-georefernced, covariate, coefficient signature estimates. The Poisson probability regression residuals indicated an inappropriate model fit due to over dispersion caused by outliers.

The extra-Poisson noise was detected in the variance estimates in the *N. fowleri,* model. A modification of the iterated re-weighted least square scheme and/or a negative binomial, non-homogenous, regression-based framework conveniently accommodates the extra-Poisson variation employing the sampled environments covariates (see Haight 1967). Operationally these models consisted of making iterated weighted least square fit to approximate normally distributed, explanatorily dependent, geo-spatiotemporal, geosampled, *N. fowleri* -related, eco-georeferenecable, explanatory, predictor, covariate coefficients based on observed rates and their logarithm. Unfortunately, the variance of explanatory, diagnostic, capture point observations in log-linear equations are commonly assumed to be constant [Griffth 2003]. Subsequently, we assumed that introducing an extra-binomial variation scheme in a forecast, vulnerability, epidemiological, -related linear-logistic, *N. fowleri,* forecast model may be fitted for a Poissonian procedure. In so doing the outliers in the he Poisson-gamma mixture (negative binomial) distribution, *N. fowleri,* model was teased out and standard deviation was equivalent to the mean in the summary statement.

NDVI was calculated using radiance, surface reflectance (*p*), or apparent reflectance (measured at the top of the atmosphere) values in the Landsat^{TM} red and NIR spectral bands. The ratio of reflected radiance from the red and NIR bands was used to normalize illumination and topographic variation and to form the NDVI *N. fowleri* eco-epidemiological data using the eco-geoerferenced, grid-stratified, LULC data. The difference of the Landsat bands was divided by their sum, which formed the functionally equivalent NDVI. This NDVI, over terrestrial surfaces of the study site, was constrained between
0 and 1. The difference in reflectance was divided by the sum of the two reflectances. Raster modeling in ArcMap® included performing image differencing on NDVI layers, classifying the layers into different classes and calculating a wetness index using the Raster Calculator. NDVI was computed directly for the *N. fowleri* capture points without any bias or assumptions regarding plant physiognomy, land cover class, soil type or climatic conditions. A *N. fowleri* proxy signature was generated to account for changing soil brightness, SAVI was also calculated, utilizing an adjustment factor L that effectively shifted the origin of vegetation isolines in NIR/VIS reflectance space.

The SAVI utilized a constant *L* to remove the soil background noise in the *N. fowleri*sampled site. For high vegetation cover, the value of *L* is 0.0, and *L* is 1.0 for low vegetation cover. For intermediate vegetation cover, L = 0.5 is the value which is most widely used in generating SAVI (Huete, 1988). The appearance of *L* in the multiplier causes of SAVI have a range identical to the NDVI (-1.0 to1.0) The net result was an NDVI *N. fowleri* signature with an origin not at the point of zero. Red and NIR SAVI was calculated where *L* was a possible amoeba height adjustment factor that accounted for differential red and NIR. Since one of the objectives of this article was to find a self-adjustable L so as to increase the SAVI vegetation sensitivity for generating a robust *N. fowleri* signature, we experimentally increased the dynamic range and reduced the soil background effect In this research, *L* = 0.5 was used for the SAVI equation.

A modified SAVI (MSAVI) that replaces the constant L in the SAVI equation with a variable L function may be constructed for obtaining an iteratively interpolative eco-georferenceable, *N. fowleri*, proxy endmember, LULC signature. The L function may be derivable by induction or by employing the product of the NDVI and weighted difference vegetation index (WDVI). Results based on ground and aircraft-measured, discontinuous canopies may be presented at a *N. fowleri*, eco-epidemiological,capture point. The MSAVI may be shown to increase the dynamic range of the vegetation signal while further minimizing the soil background influences, resulting in greater vegetation sensitivity as defined by a "vegetation signal" to "soil noise" ratio.

Because it cannot be concluded if vegetation is an important cofactor to *N. fowleri* prevalence, future research may include time-series NDVI and SAVI models that may reveal even more robust landscape, explanative parameterizable,LULC covariants, that can be introduced into a stochastic or deterministic iteratable interpolator; in

so doing, more precision targeting of unknown, ungeosampled, eco-georeferencable, *N. fowleri* may be identified within a county-level, grid-stratified, high-spatial resolution WorldView 3 0.31m VIS and NIR data image.

A krige procedure computed the empirical (also known as sample or experimental) semivariogram from a set of *N. fowleri* capture point measurements. Semivariograms are used in the first steps of spatial prediction as tools that provide insight into the spatial continuity and structure of a random process (Cressie 1993). Naturally occurring randomness is accounted for by describing a process in terms of the spatial random field (SRF) concept (Christakos 1992). An SRF is a collection of random variables throughout your spatial domain of prediction (Griffith 2003).

Based on our sample, geospatial predictions providedby the *N. fowleri* iteratively interpolated endmember LULC Landsat ^{TM} spectral values un-geosampled, eco-georeferenceable, *N. fowleri* capture points may be optimally identified through techniques such as ordinary kriging. Our model required the use of a covariance model.

The covariance regression *N. fowleri* capture point, eco-epidemiological, endmember signature model parameterized the covariance matrix of a multivariate response vector as a parsimonious quadratic function of explanatory LULC variables. The approach is analogous to the mean regression model, and is similar to a factor analysis model in which the factor loadings depend on the explanatory variables. Employing a random-effects representation, parameter estimation for the model may be straightforward employing either an EM-algorithm or a Markov Chain Monte Carlo (MCMC) approximation via Gibbs sampling. The proposed methodology may provide a simple but flexible representation of heteroscedasticity across the levels of an explanatory, geo-spectrotemporal, geosampled, *N. fowleri*, eco-epidemiological, eco-georeferenceable, capture point, prognosticative, LULC, endmember signature, dependent variable. In so doing the interpolator may improve estimation of the mean function and render better calibrated prediction LULC regions when compared to a homoscedastic *N. fowleri*, county-level, and vulnerability mode. It may be possible that the empirical semivariance can provide an eco-georeferencable, *N. fowleri* estimate employing the theoretical semivariance, (Appendix 2) which may then be employed to seasonally characterize the spatial structure of the amebae.

Let $\{Z(\boldsymbol{s}), \boldsymbol{s} \in D \subset \mathscr{R}^2\}$ be a SRF with n measured orthogonally eigendecomposed,*N. fowleri* LULC endmember signature values $z_i = Z(\boldsymbol{s}_i)$ at respective eco-georeferenced, geolocations $\boldsymbol{s}_i$, $i = 1, \ldots, n$. An epidemiologist or researcher could employ the VARIOGRAM procedure for attaining insight into the spatial continuity and structure of $Z(\boldsymbol{s})$. A good measure of the spatial continuity of $Z(\boldsymbol{s})$ in *N. fowleri* the may be parsimoniously robustly definable by means of the variance of the difference $Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j)$, where $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ are eco-georferenceable ArcGIS derived county-level, geolocations in D. Specifically, $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ may be spatial LULC grid-stratifiable increments such that $\boldsymbol{h} = \boldsymbol{s}_j - \boldsymbol{s}_i$. In such circumstanceds then the variance function based on the increments $\boldsymbol{h}$ would be independent of the actual geolocations $\boldsymbol{s}_i$, $\boldsymbol{s}_j$. (N. *fowleri* capture point). Most commonly, the continuity measure used in common eco-epidemiological, endmember kriging practice is one half of this variance, better known as the semivariance function, or, equivalently, $\gamma_z(\boldsymbol{h}) = \frac{1}{2}\left(\mathrm{E}\{[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})]^2\} - \{\mathrm{E}[Z(\boldsymbol{s}+\boldsymbol{h})] - \mathrm{E}[Z(\boldsymbol{s})]\}^2\right)$The plot of semivariance as a function of $\boldsymbol{h}$ is the semivariogram (Cressie 1993).

Further, assume that the SRF $Z(\boldsymbol{s})$ is free of nonrandom (or systematic) surface trends, then, the expected *N. fowleri* capture point endmember signature, grid-stratifiable LULC value $\mathrm{E}[Z(\boldsymbol{s})]$ of $Z(\boldsymbol{s})$ would be a constant for all $\boldsymbol{s} \in \mathscr{R}^2$.As such the semivariance expression in the iteratively interpolated amoeba signature may be simplified to the $\gamma_z(\boldsymbol{h}) = \frac{1}{2}\mathrm{E}\{[Z(\boldsymbol{s}+\boldsymbol{h}) - Z(\boldsymbol{s})]^2\}$Given the preceding assumption, an epidemiologist or othe researcher could compute an estimate $\hat{\gamma}_z(\boldsymbol{h})$ of the semivariance $\gamma_z(\boldsymbol{h})$ from a finite set of eco-georeferenced, *N.*

fowleri capture points in a practical way by using the formula $\hat{\gamma}(h) = \frac{1}{2\,|N(h)|}\sum_{N(h)}[Z(s_i) - Z(s_j)]^2$ in ArcGIS Spatial Analyst ™where the sets $N(h)$ may contain all the neighboring capture oints pairs at distance h, $N(h) = \{i, j : s_i - s_j = h\}$ and $|N(h)|$ is the number of such pairs (i, j).The expression for $\hat{\gamma}(h)$ is called the empiricalsemivariance (Matheron,1963). This is the quantity that PROC VARIOGRAM would compute and its corresponding plot would be the empirical, *N. fowleri*, capture point semivariogram.

The empirical semivariance $\hat{\gamma}(h)$ is also referred to as classical. This name is used so that it can be distinguished from the robust semivariance estimate $\hat{\gamma}(h)$ and the corresponding robust semivariogram. The robust semivariance was introduced by Cressie and Hawkins (1980) to weaken the effect that outliers in the observations might have on the semivariance. It is described by Cressie (1993, p. 75) as $\hat{\gamma}(h) = \frac{\Psi^4(h)}{2[0.457 + 0.494/N(h)]}$ In the preceding expression the *N. fowleri* capture point endmember signature,LULC parameter $\Psi(h)$ may be defined as $\Psi(h) = \frac{1}{N(h)}\sum_{P_iP_j \in N(h)}[Z(s_i) - Z(s_j)]^{\frac{1}{2}}$

According to Cressie (1985), the estimate $\hat{\gamma}(h)$ has approximate variance $\mathrm{Var}[\hat{\gamma}(h)] \simeq \frac{2[\gamma(h)]^2}{N(h)}$ This approximation is possible by assuming $Z(s)$ to be a Gaussian SRF,in a *N. fowleri* stochastic interpolator and by further assuming the squared differences in empirical semivariances to be uncorrelated for different grid-stratifiable, Euclidean distances h. Typically, semivariance estimates are correlated because of the underlying spatial correlation amongst the observations, and also because the same observation pairs might be used for the estimation of more than one semivariogram point, as described in the following subsections(Griffith 2003). Despite these restrictive assumptions, the approximate variance provides an idea about the semivariance estimate variance and enables fitting of a theoretical *N. fowleri* eco-epidemiological, endmember signature, LULC eco-georeferenceable model to the empirical semivariance.

In conclusiona SAVI signature transformation technique minimized soil brightness using the Landsat-derived red and NIR wavelengths. Graphically, the transformation involved a shifting of the origin of reflectance, *N. fowleri*, endmember LULV spectra which was plotted in NIR-red wavelength space to account for first-order soil-vegetation interactions and differential red and NIR flux extinction through inhomogeneous vegetated canopies. Thereafter an autocorrelation model was able to forecast an eco-georeferenced sample capture point.

Appendix 1

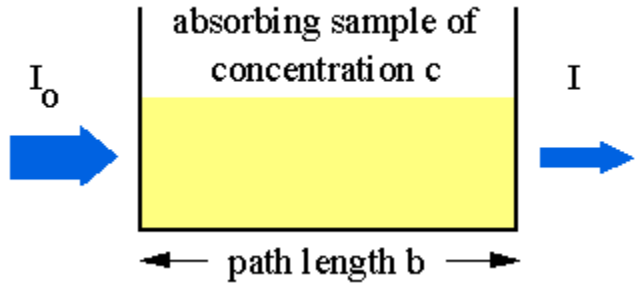
The Beer-Lambert law (or Beer's law) is the linear relationship between absorbance and concentration of an absorbing species. The general Beer-Lambert law is usually written as:A = a(λ) * b * c where A is the measured absorbance, a(λ) is a wavelength-dependent absorptivity coefficient, **b** is the path length, and **c** is the analyte concentration. When working in concentration units of molarity, the Beer-Lambert law is written as:A = ε * b * c where ε is the wavelength-dependent molar absorptivity coefficient with units of M^{-1} cm^{-1}. Data are frequently reported in percent transmission (I/I_0 * 100) or in absorbannce [A = log (I/I_0)]. The latter is particularly convenient. [common coefficients of near-ultraviolet absorption bands of some amino acids and nucleotides]Sometimes the extinction coefficient is given in other units; for example,$A = E^{1\%}$ * b * cwhere the concentration C is in gram per 100 ml of solution. This useful when the molecular weight of the solute is unknown or uncertain.

Instrumentation

Experimental measurements are usually made in terms of transmittance (T), which is defined as:

$T = I / I_o$where I is the light intensity after it passes through the sample and I_o is the initial light intensity. The relation between A and T is:
$A = -\log T = -\log (I / I_o)$.

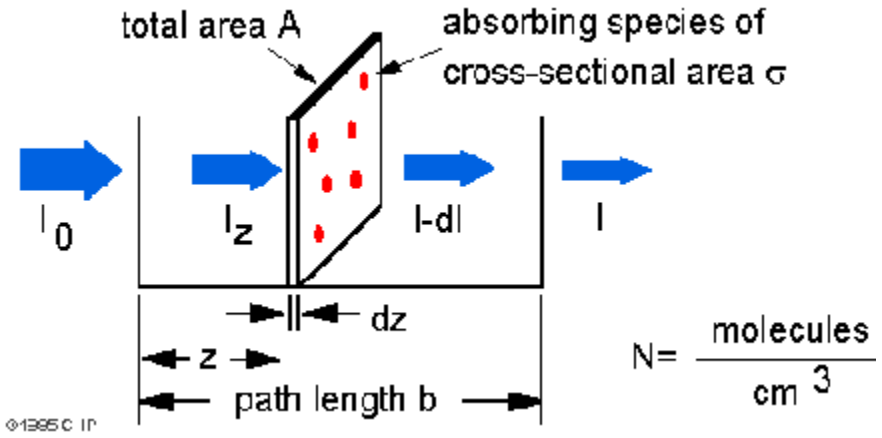
Absorption of light by a sample



Modern absorption instruments can usually display the data as either transmittance, %-transmittance, or absorbance. An unknown concentration of an analyte can be determined by measuring the amount of light that a sample absorbs and applying Beer's law. If the absorptivity coefficient is not known, the unknown concentration can be determined using a working curve of absorbance versus concentration derived from standards.

Derivation of the Beer-Lambert law

The Beer-Lambert law can be derived from an approximation for the absorption coefficient for a molecule by approximating the molecule by an opaque disk whose cross-sectional area, σ, represents the effective area seen by a photon of frequency *w*. If the frequency of the light is far from resonance, the area is approximately 0, and if *w* is close to resonance the area is a maximum. Taking an infinitesimal slab, dz, of sample:



I_o is the intensity entering the sample at z=0, I_z is the intensity entering the infinitesimal slab at z, dI is the intensity absorbed in the slab, and I is the intensity of light leaving the sample. Then, the total opaque area on the slab due to the absorbers is σ * N * A * dz. Then, the fraction of photons absorbed will be σ * N * A * dz / A so,

$dI / I_z = -$ σ * N * dz

Integrating this equation from z = 0 to z = b gives:

$\ln(I) - \ln(I_o) = -\sigma * N * b$

or $-\ln(I / I_o) = \sigma * N * b$.

Since N (molecules/cm^3) * (1 mole / $6.023x10^{23}$ molecules) * 1000 cm^3 / liter = c (moles/liter) and 2.303 * log(x) = ln(x) then

$-\log(I / I_o) = \sigma * (6.023x10^{20} / 2.303) * c * b$

$-\log(I / I_o) = A = \varepsilon * b * c$

where $\varepsilon = \sigma * (6.023x10^{20} / 2.303) = \sigma * 2.61x10^{20}$

Typical cross-sections and molar absorptivities are:

	σ (cm^2)	ε (M^{-1} cm^{-1})
absorption - atoms	10^{-12}	$3x10^8$
molecules	10^{-16}	$3x10^4$
infrared	10^{-19}	3x10
Raman scattering	10^{-29}	$3x10^{-9}$

Limitations of the Beer-Lambert law

The linearity of the Beer-Lambert law is limited by chemical and instrumental factors. Causes of nonlinearity include:

- deviations in absorptivity coefficients at high concentrations (>0.01M) due to electrostatic interactions between molecules in close proximity
- scattering of light due to particulates in the sample
- fluorescence or phosphorescence of the sample
- changes in refractive index at high analyte concentration
- shifts in chemical equilibria as a function of concentration
- non-monochromatic radiation, deviations can be minimized by using a relatively flat part of the absorption spectrum such as the maximum of an absorption band
- stray light

References:

Aitken, A.C., 1935. *On Least Squares and Linear Combinations of Observations*. Proceedings of the Royal Society of Edinburgh. 55, 42-48.

Asrar, G., et al., 1984. Estimating absorbed photosynthetic radiation and leaf area index from spectral reflectance in wheat. *Agronomy Journal* [online]. 76, 300-306. Available from: https://dl.sciencesocieties.org/publications/aj/abstracts/76/2/AJ0760020300 [Accessed 20 July 2017].

Asner, G.P., Hicke, J.A., Lobell, D.B., 2003. Per-pixel analysis of forest structure: Vegetation Indices, Spectral Mixture analysis and Canopy Reflectance Modeling. *Remote Sensing of Forest Environments* [online]. 10, 1-48. Available from: https://link.springer.com/chapter/10.1007%2F978-1-4615-0306-4_8 [Accessed 28 August 2017].

Bannari, A., Morin, D., Bonn, F., Huete, A., 1995. A review of vegetation indices. Remote Sensing Reviews [online]. 13 (1), 95–120. Available from: http://www.tandfonline.com/doi/abs/10.1080/02757259509532298 [Accessed 28 August 2017].

Baret, F., Guyot, G., 1991. Potentials and limitations of vegetation indices for LAI and APAR assessment. *Remote Sensing Of Environment* [online]. 35, 161–173. Available from:

https://www.researchgate.net/profile/Baret_Frederic/publication/222736598_Potentials_and_limits_of_vegetation_indices_of_LAI_and_APAR_assessment/links/02e7e537e0067511c1000000/Potentials-and-limits-of-vegetation-indices-of-LAI-and-APAR-assessment.pdf [Accessed 28 August 2017].

Bausch, W., 1993. Soil background effects on reflectance-based crop coefficients for corn. Remote Sensing Of Environment [online]. 46, 1–10. Available from: https://naldc.nal.usda.gov/download/37844/PDF [Accessed 28 August 2017].

Bazaar, F.A., Pickett, S.T.A., 1980. Physiological ecology of tropical succession: A comparative review. Annu. Rev. Ecol. Syst. 11, 287-310.

Brown, H., et al., 2008. Remotely-Sensed Vegetation Indices Identify Mosquito Clusters of West Nile Virus Vectors in an Urban Landscape in the Northeastern United States. *Vector-Borne and Zoonotic Diseases* [online]. 8 (2), 197-206. Available from: https://doi.org/10.1089/vbz.2007.0154 [Accessed 20 July 2017].

Buschmann, C., Nagel, E., 1993. *In vivo* spectroscopy and internal optics of leaves as basis for remote sensing of vegetation. *International Journal of Remote Sensing*. 14, 711–722.

Capewell, L.G., et al., 2014. Diagnosis, Clinical Course, and Treatment of Primary Amoebic Meningoencephalitis in the United States, 1937-2013. *Journal of the Pediatric Infectious Diseases Society* [online]. 4 (4), 68-75. Available from: https://doi.org/10.1093/jpids/piu103 [Accessed 28 May 2017]

Cameron, A.C., Trivedi, P.K., 1998. *Regression analysis of count data*. USA: Cambridge University Press.

Chalmers, R.M., et al., 2013. Naegleria. *Microbiology of Waterborne Diseases (Second Edition)* London, UK: Elsevier, 407-416.

Christakos, G., 1992. Random Field Models in Earth Sciences. 1st ed. *Elsevier*. USA: Academic Press.

Christensen, R., 1997. Log-linear models and logistic regression. Springer Texts in Statistics. 2nd ed. New York: Springer-Verlag.

Co, J., Townshend, J.R.G., Holben, B.N., Tucker, C.J., 1985. Phenology of global vegetation using meteorological satellite data. *International Journal of Remote Sensing* [online]. 6 (8), 1271–1318. Available from: http://www.tandfonline.com/doi/abs/10.1080/01431168508948281 [Accessed 28 August 2017].

Colwell, J.E., 1974. Vegetation canopy reflectance. *Remote Sensing of Environment* [online]. 3 (3), 175-183. Available from: http://www.sciencedirect.com/science/article/pii/0034425774900030p [Accessed 20 July 2017].

Cressie, N.A., 1993. Statistics for Spatial Data. New York: A Wiley- Interscience.

Cressie, N.A., 2003. Plenary: Nonparametric hypothesis testing for spatial signal. 12th ed. United States: IEEE.

De Jong, S.M., and Jetten, V.G., 2007.Estimating spatial patterns of rainfall interception from remotely sensed vegetation indices and spectral mixture analysis. *International Journal of GIS* [online]. Available from: http://www.tandfonline.com/doi/full/10.1080/13658810601064884 [Accessed 9 June 2017]. Digital Globe [online]. MacDonald, Detailer and Associates. Available from: https://www.digitalglobe.com [Accessed 20 August 2017].

Dijk-Wasser, M.A., et al., 2006. Spatial prediction of West Nile virus vector habitats in Connecticut, USA. *Vector-Borne and Zoonotic Diseases* [online]. 6 (3), 281-293. Available from: https://mariadiukwasser.files.wordpress.com/2014/11/diukwasser_2006_vbzd_wnvvectors.pdf [Accessed 20 July 2017].

Draper, N.R., and Smith, H., 1981. Series in probability and mathematical statistics. *Applied regression analysis* [online]. Available from: http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471170828.html [Accessed 9 July 2017].

Draper, N.R., and Smith, H., 1998. Fitting a straight line by least squares.*Applied regression analysis.* 3, 15-46.

Earth Explorer [online]. U.S. Geological Survey. Available from: https://earthexplorer.usgs.gov/ [Accessed 25 May 2017].

Freedman, D.A., 2005.*Statistical models: Theory and practice (first ed.) [online]*. Cambridge University Press. Available: www.cambridge.org/catalogue/catalogue.asp?isbn=9780521671057 [Accessed 28 August 2017].

Ehrlich, D., Lambin, E.F., 1996. Broad scale land-cover classification and interannual climatic variability. International Journal of Remote Sensing. 17, 845–862.

Gonsamo, A., Pellikka, P., 2008. Methodology comparison for slope correction in conopy leaf area index estimation using hemispherical photography. *Forest Ecology and Management* [online]. 256, 749-759. Available from: http://www.helsinki.fi/geography/geoinformatics/pdf/FORECO_Gonsamo%20slope_lai.pdf [Accessed 20 July 2017].

Griffith, D.A., 2003. *Spatial Autocorrelation and Spatial Filtering.* 1st ed. Berlin: Springer-Verlag.

Griffith, D.A., 2005. Effective Geographic Sample Size in the Presence of Spatial Autocorrelation. *Annals of the Association of American Geographers* [online]. 95 (4), 740-760. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8306.2005.00484.x/full [Accessed 20 July 2017].

Haight, F.A., 1967. *Handbook of the Poisson Distribution*. New York, John Wiley.

Hansen, C. B., 2007. Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics*, 140, 670-694.

Hatfield, J.L., et al., 1984. Intercepted photosynthetically active radiation estimated by spectral reflectance. *Remote Sensing of Environment* [online]. 14, 65-75. Available from: https://www.researchgate.net/publication/222437027_Intercepted_photosynthetically_active_radiation_estimated_by_spectral_reflectance [Accessed 20 July 2017].

Hay, S.I., 1997. Remote sensing and disease control: past, present and future. *Transactions of The Royal Society of Tropical Medicine and Hygiene* [online]. 91 (2). 105-106. Available from: https://academic.oup.com/trstmh/article/91/2/105/1914977/Remote-sensing-and-disease-control-past-present [Accessed 20 July 2017].

Hay, S.I., Snow, R.W., Rogers, D.J., 1998. From predicting mosquito habitat to malaria seasons using remotely sensed data: practice, problems and perspectives. *Parasiltol Today* [online]. 14 (8), 306–313. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17040796 [Accessed 28 August 2017].

Heilman, J.L., Kress, M.R., 1987. Effects of vegetation on spectral irradiance at the soil surface. *Agronomy Journal* [online]. 79, 765-768. Available from: https://dl.sciencesocieties.org/publications/aj/abstracts/79/5/AJ0790050765 [Accessed 20 July 2017].

Hosmer D.W., Hosmer T., Cessie L.E., Lemeshow S., 1980. A goodness-of-fit test for the logistic regression model. *Communications in Statistics.* 16, 965-980.

Hubbell, S.P., Foster, R.B., 1986. *Plant Ecology.* Oxford, UK: Blackwell. 77-95.

Huete, A.R., et al., 1985. Spectral response of a plant canopy with different soil backgrounds. *Remote Sensing of Environment* [online]. 17, 37-53. Available from: http://www.sciencedirect.com/science/article/pii/0034425785901117 [Accessed 20 July 2017].

Huete, A.R., 1987. Soil-Dependent Spectral Response in a Developing Plant Canopy. *Argon* [online]. 79, 61-68. Available from: https://dl.sciencesocieties.org/publications/aj/abstracts/79/1/AJ0790010061 [Accessed 20 July 2017].

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* [online]. 25 (3), 295-309. Available from: http://www.sciencedirect.com/science/article/pii/003442578890106X [Accessed 20 July 2017].

Huete, A.R., Hua, G., Qi, J., Chehbouni, van Leeuwen W.J.D., 1992. Normalization of multidirectional red and NIR reflectances with the SAVI. *Remote Sensing Of Environment* [online]. 41 (2), 143–154. Available from:https://arizona.pure.elsevier.com/en/publications/normalization-of-multidirectional-red-and-nir-reflectances-with-t [Accessed 25 May 2017].

Huete, A.R., Justice C, Liu H., 1994. Development of vegetation and soil indexes for Modis-Eos. *Remote Sensing Of Environment* [online]. 49(3), 224–234. Available from:https://opus.lib.uts.edu.au/handle/10453/13513 [Accessed 28 August 2017].

Huete, A.R., et al., 1994. Development of Vegetation and soil indexes for MODIS-EOS. *Remote Sensing of Environment* [online]. 49, 224-234. Available from: https://www.researchgate.net/publication/236770063_Development_of_vegetation_and_soil_indexes_for_MODIS-EOS [Accessed 20 July 2017].

Jackson, R.D., 1980. Hand-Held Radiometry. *Agricultural Research (Western Region), Science and Education Administration, U.S. Department of Agriculture, Oakland, California* [online]. 19, 1-61. Available from: https://naldc.nal.usda.gov/download/50523/PDF [Accessed 20 July 2017].

Jackson, R.D., 1983. Spectral indices in N-Space. *Remote Sensing of Environment* [online]. 13 (5), 409-421. Available from: http://www.sciencedirect.com/science/article/pii/003442578390010X [Accessed 20 July 2017].

Jackson, R.D., Pinter, P.J. Jr., 1986. Spectral response of architecturally different wheat canopies. *Remote Sensing of Environment* [online]. 20 (1), 43–56. Available from: https://pubag.nal.usda.gov/catalog/50603 [Accessed 28 August 2017].

Jackson, R.D., Huete, A.R., 1991. Interpreting vegetation indices. *Preventive Veterinary Medicine* [online]. 11, 185–200. Available from: http://www.uprm.edu/biology/profs/chinea/gis/lectesc/intvegindx.pdf [Accessed 28 August 2017].

Jacob, B.G., et al., 2008. Describing *Anopheles arabiensis* aquatic habitats in two riceland agro-ecosystems in Mwea, Kenya using a negative bionmial regression model with a non-homogenous mean. *Acta Tropica* [online]. 109 (2009), 17-26. Available from: http://www.sciencedirect.com/science/article/pii/S0001706X08002325 [Accessed 18 June 2017].

Jacob, B.G., et al., 2009. A heteroskedastic error covoariance matrix estimator using a first-order conditional autoregressive Markov simulation for deriving asympotical efficient esetimates from ecological sampled Anopheles arabiensis aquatic habitat covariates. *Malaria Journal* [online]. 8(1), 216-225. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2760564/ [Assessed 18 June 2017].

Jacob, B.G., et al., 2010. Developing operational algorithms using linear and non-linear square estimation in Python for the identification of *Culex pipiens* and *Culex restuans* in a mosquito abatement districts (Cook County, Illinois, USA). *Geospatial Health*. 3 (2), 157-176.

Jacob, B.G., et al., 2011. Adjusting second moment bias in eignspace using Bayesian empirical estimators, Dirichlet tessellations and WorldviewI data for predicting *Culex quinquefasciatus* in Trinidad. *Journal of Geographic Information Systems* [online]. 14 (2), 244-274. Available from: https://file.scirp.org/pdf/JGIS20110100007_97385209.pdf [Accessed 20 July 2017].

Jacob, B.G., et al., 2015. Ecogeographically and NonEcogeographically Forecasting Discontinuously Canopied Seasonally Hyperproductive Trailing Vegetation Precambrian rock Simulium damnosum s.l., Eco-epidemiological Capture Point Morphometrics by Geo-spectrotemporally Iteratively Stochastically Interpolating Metrizable Sub-Mixel Mean Solar Exoatmospheric Quantum Scalar Irradiance Wavelength Periodicities where θi is a Zenith Angle and Diatomically Etiolated Xanthophylls with Azimutually Isotropic Sources of Chloroplastic Carotenoid Zeaxanthins Stoichiometerically Extracted from a RapidEye™ Red Edge Normalized Difference Vegetation Index Reference Biosignature: A Case Study in Burkina Faso and Uganda. *Journal of Geophysics and Remote sensing* [online]. 5(1), 12-103. Available from: https://www.omicsgroup.org/journals/ecogeographically-and-nonecogeographically-forecasting-discontinuous-canopied-simulium-damnosum-sl-habitats-by-interpolating-metri-2469-4134-1000152.php?aid=64896 [Accessed 20 July 2017].

Jacob, B.G., et al. 2016. Lexicographically, cartesian-ordered, differential calculi in canonically extractable insitu near infra-red fluorescence quantum spectroscopic sub-surface continuous geodesic fluxions for metaheursitic chorophyll-a translucent emissivty mapping intermittently canopied immature narrow riverine tributary Simulium damnosum s.l. oviposition sites for bio-optically delineating multivariate normalized Gaussian processes elucidatively administrated by prior covariances and aspline within a reproducing non-frequentist simultaneous diagonalization of amalgamized positive definite kernels in Hilbert space: Implementation of a 'Slash and Clear' control intervention in two ecogeoreferenceable agro-village complexes in northern Uganda Uganda Journal of Geophysics and Remote sensing. *Journal of Geophysics and Remote sensing.*

Jensen, J.R., 2005. *Introductory Digital Image Processing*. 3rd ed. New Jersey, USA: Prentice Hall.

Karnieli, A., Kaufman, Y.J., Remer, L., Wald, A., 2001. A Modified Aerosol free vegetation index algorithm for. aerosol optic depth retrieval using GOSAT TANSO-CHI data. *Remote Sensing Of Environment* [online]. 77, 10–21. Available from: www.mdpi.com/2072-4292/8/12/998/pdf [Accessed 28 August 2017].

Kaufman, Y.J., Tanre, D., 1992. Atmospherical resistant vegetation index (ARVI) for EOSMODIS, IEEE. Trans Geoscience Remote Sensing [online]. 30, 261–270. Available from: https://modis.gsfc.nasa.gov/sci_team/pubs/abstract_new.php?id=03667 [Accessed 28 August 2017].

Kricher, J., 2011. *Tropical ecology.* Princeton, New Jersey: Princeton University Press. *188-126.*

Liu, H.Q., Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background atmospheric noise. *Geoscience and Remote Sensing, IEEE Trans Geoscience Remote Sensing* [online]. 33 (2), 457–465. Available from: http://dx.doi.org/10.1109/36.377946 [Accessed 28 August 2017].

Marciano-Cabral, F., Cabral, G.A., 2007. The immune response to *Naegleria fowleri* amebae and pathogenesis of infection. *FEMS Immunology & Medical Microbiology* [online]. 51 (2), 243-259. Available from: https://academic-oup-com.ezproxy.lib.usf.edu/femspd/article/51/2/243/article#15374625 [Accessed 18 May 2017].

Myneni, R.B., Ganapol, B.D., Asrar, G., 1992. Remote sensing of vegetation canopy photosynthetic and stomatal conductance efficiencies. *Remote Sensing Of Environment* [online]. 42 (3), 217–238. Available from:https://arizona.pure.elsevier.com/en/publications/remote-sensing-of-vegetation-canopy-photosynthetic-and-stomatal-c [Accessed 28 August 2017].

Naegleria fowleri- Primary Amebic Meningoencephalitis (PAM) - Amebic Encephalitis [online]. Centers for Disease Control and Prevention. Available from: https://www.cdc.gov/parasites/naegleria/general.html [Accessed 28 May 2017].

Olgemann, J.E., 1990. Comparison between two vegetation indices for measuring different types of forest damage in the north-eastern United States. *Intentional Journal of Remote Sensing.* 11, 2281–2297.

Richardson, A.J., Wiegand, C.L., 1977. Distinguishing Vegetation from Soil Background Information. Photogrammetric Engineering & Remote Sensing [Online]. 43 (12), 1541-1552. Available from: https://www.asprs.org/wp-content/uploads/pers/1977journal/dec/1977_dec_1541-1552.pdf [Accessed 20 July 2017].

Ronald, S., 2004. *Integers, Polynomials, and Rings: A Course in Algebra.* Springer.

Sellers, P.J., 1985. Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing* [online]. 6 (8), 1335-1372. Available from: http://www.tandfonline.com/doi/abs/10.1080/01431168508948283 [Accessed 20 July 2017].

Shamen, J., 2002. Using a Dynamic Hydrology Model To Predict Mosquito Abundances in Flood and Swamp Water. *Emerging Infectious Diseases* [online]. 8 (1), 8-13. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2730265/ [Accessed 20 July 2017].

Slater, P.N., Jackson, R.D., 1982. Atmospheric effects on radiation reflected from soil and vegetation as measured by orbital sensors using various scanning directions. *Applied Optics* [online]. 21 (21), 3923-3931. Available from: https://www.semanticscholar.org/paper/Atmospheric-effects-on-radiation-reflected-from-so-Slater-Jackson/af95d890d6c1ffb4636b4c9d32052a6ee0979435 [Accessed 20 July 2017].

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* [online]. 8 (2), 127-150. Available from: http://www.sciencedirect.com/science/article/pii/0034425779900130 [Accessed 20 July 2017].

Tucker, C.J., Fung, I.Y., Keeling, C.D., Gammon, R.H., 1986. Relationship between atmospheric CO2 variations and a satellite-derived vegetation index. Nature [online]. 319, 195-199. Available from: http://www.nature.com/nature/journal/v319/n6050/abs/319195a0.html?foxtrotcallback=true [Accessed 28 August 2017].

Tucker, C.J., 1991. Mean and inter-year variation of growing season normalized difference vegetation index for the Sahel 1981-1989. *International Journal of Remote Sensing*. 12, 1113-1115.

Yoder, J.S., et al., 2009. The epidemiology of primary amoebic meningoencephalitis in the USA, 1962–2008 [Online]. Cambridge University. Available from: https://www.cambridge.org/core/journals/epidemiology-and-infection/article/epidemiology-of-primary-amoebic-meningoencephalitis-in-the-usa-19622008/1CADA8AB942359501CCD94BA032B4DF5138 [Accessed 28 May 2017].