Journal of Remote Sensing & GIS

Research Article Open Access

A Negative Binomial with a Non-Homogenous Gamma Distributed Mean for Robustifying Pseudo R^2 Regression Values of Immature Vector and Nuisance Mosquito Count Data for Optimally Discerning Un-Geosampled Waste Tires in a Subtropical Oviposition Site in SAS®/GIS employing Worldview-3 Visible and Near Infra-Red Data in Hillsborough County, Florida

Dinh ETN* and Jacob Benjamin

Department of Global Health, College of Public Health, University of South Florida, Tampa, FL 33612, USA

Abstract

Refuse vehicle tires on undeveloped land plots near human dwellings may be a public health threat, as they can provide a suitable habitat for vector and nuisance mosquito (Diptera: Culicidae) population growth. These tires are currently found only through ground-based searches, so interpolated spectral signature of a geo-referenceable, known, positive tire may help expedite discriminating unknown waste tire geolocations. However, frequentistic and non-frequentistic quantification of bioenvironmental explanatorial time series covariates statistically significant to mosquito hyperproductivitt in waste tire habitats is needed to limit the search criteria of the signature. This study aimed to develop an iteratively interpolative geo-spectral biosignature for detecting unknown, un-geosampled waste tires conducive to mosquito propagation. After constructing various regression models, we found that the field geo-sampled mosquito count data featured deviations from the assumptions of regression modeling (i.e., collinear and heteroskedastic parameters). Thus, a negative binomial paradigm was utilized to assuage the violations of regression analysis and to robustify the model's R^2 value. Based on the results of the linear analyses, a spectral signature of a productive habitat was created from multispectral band imagery from WorldView-3 satellite sensor data. The signature was then applied in Hillsborough County, FL to remotely determine the eco-geographical geo-locations of anthropogenic waste tire habitats. The signature model exhibited a sensitivity of 83% and a specificity of 87%. In conclusion, the regression and signature models constructed here provided a parsimonious yet accurate estimation of undiscovered waste tire habitats that may yield many mosquitoes.

Keywords: Statistics; Spectral signature; Diptera; Culicidae; Spatial interpolation; Mixels; Overdispersion

Introduction

Automobile tires discarded in undeveloped land plots near anthropogenic and animal habitations are a health hazard because they can support an immature population of vector mosquitoes. These mosquitoes may transmit several zoonotic arboviral diseases. Hillsborough County, Florida has recorded anthropogenic cases of locally-acquired West Nile virus, Eastern equine encephalitis, and La Crosse encephalitis in the past few years [1]. Unfortunately, there are no means to locate waste tires dumped near georeferenced human dwellings besides ground-based searches. Thus, location techniques for waste tires that conserve limited funds and human resources are needed for arboviral disease prevention via tire removal in Hillsborough County.

Many of the entomological studies that have regressed eco-epidemiological time series dependent co-factors influencing mosquito productivity in waste tires are limited to evaluating habitat productivity in urban land use land cover (LULC) zones using binomialized logistic regression derivatives [2-4]. Literature has not seen a parsimonious, geo-spatiotemporal, multivariate, regression-oriented, epidemiological, forecasting model for evaluating the extent to which individual parameterized covariate estimators acquired from empirical geo-sampled data map high georferenceable mosquito habitat productivity counts anywhere in Florida. Elucidating parameters statistically significant ($p<0.05$) to high mosquito count data would help geo-locate waste tire habitats favorable to vector and nuisance mosquito multiplication in unbuilt landscapes in subtropical central west Florida.

Jacob et al. [5] composed geo-optical algorithms for decomposing sub-meter spatial resolution (i.e., panchromatic Quickbird 0.61m IFOV data) imagery of rice field environments in order to geo-locate undiscovered productive aquatic larval habitats of malaria mosquito vector of *Anopheles arabiensis* (Diptera: Culicidae). The reference bio-signatures of these habitats generated from the unmixing algorithmic geomteri-optical models were then used to perform an ordinary krig-based interpolation in ArcGIS® [5]. Likewise, Jacob et al. [6] geo-predicted seasonal trailing vegation, discontinuous, infrequently canopied, turbid water, seasonal black –fly vector of onchocercisasis, *Simulium damnosum* s. l. (Diptera: Simuliidae) by extracting spectral end members of canopy shaded riverine sites featuring black Precambrian rock from QuickBird imagery [6]. The end members were decomposed to orthogonal eigenevctors render a specified graphical

***Corresponding author:** Dinh ETN, Department of Global Health, College of Public Health, University of South Florida, 13201 Bruce B. Downs Blvd., MDC 56, Tampa, FL 33612, USA, Tel: (813)9749784; Fax: (813)9740992; E-mail: emilydinh@health.usf.edu

Received April 25, 2016; **Accepted** June 17, 2016; **Published** June 24, 2016

Citation: Dinh ETN, Jacob BG (2016) A Negative Binomial with a Non-Homogenous Gamma Distributed Mean for Robustifying Pseudo R^2 Regression Values of Immature Vector and Nuisance Mosquito Count Data for Optimally Discerning Un-Geosampled Waste Tires in a Subtropical Oviposition Site in SAS®/GIS employing Worldview-3 Visible and Near Infra-Red Data in Hillsborough County, Florida. J Remote Sensing & GIS 5: 167. doi:10.4172/2469-4134.1000167

indicator of black fly larval proliferation sites in seasonal rmeandering riverine tribuataries from mixed sub-pixels. Linear spectral mixture analysis is a common acceptable approach for conducting optimizable, hierarchically-oriented, frequentistic, or non-frequentistic, geo classification cartographic routines which often involves defining unique illuminative signatures of pure ground components (i.e., end members) and linear combinations of end member materials (i.e., eigenvectors). Given a set of mixed, multispectral or hyperspectral vectors, spectral end member eigenvectors aims at estimating the number of reference substances (i.e., end members), their spectral signatures, and their abundance fractions. The purified end members were then kriged to identify similar yet unknown locations.

End member unmixing has never been used for cartographically forecasting hyper-prolific arboviral mosquito discarded tire habitats. We aimed to develop an interpolative geo-spectral proxy bio-signature for detecting unknown, un-geosampled waste tires in an area with a specified vegetation index (VI) that was shown to be conducive to mosquito propagation. We pursued four objectives to accomplish this goal:

- Construct logistic, Poisson, and negative binomial with a non-homogenous gamma distributed mean regression models to determine which field or remote geo-sampled explanatorial characteristic(s) were statistically significant ($p<0.05$) in affecting geo-spatiotemporal, field-sampled immature mosquito count data. immature mosquito count data. Then, the robustness (quantified by the pseudo R2 value) of the regression frameworks were compared to determine the optimal probabilistic explanative model for forecasting geo-spatiotemporal mosquito production.
- Quantitate levels of vegetation land cover with the normalized difference vegetation index (NDVI) and soil-adjusted vegetation index (SAVI) to generate proxy covariates. The statistical significance of the NDVI was compared to that of the SAVI to determine which eco-geographic geoclassified LULC characterization of the vegetation canopied landscape would have more forecasting power in the statistical models constructed in first objective.
- Quantitate elevation surface slope coefficients that contribute to geo-spatiotemporal idiosyncrasies in mosquito immature abundance and distribution by forming three-dimensional digital elevation models (DEM).
- Attain the spectral signature of known mosquito vector eco-epidemiological capture points for interpolating spectrally uncoalesced sub-meter resolution (i.e., WorldView-3) unknown unknown, unsampled fecund waste tire habitats in a specific landscape classification via an stochastic interpolator (i.e., Ordinary kriging algorithm).

Methods

Study site description

Field-derived mosquito immatures were geo-sampled from the University of South Florida's Forest Preserve in Tampa, FL. The Preserve is a 500-acre plot of wetland and sand hill habitat. The site has a humid subtropical climate with a distinct rainy season from June to September, peaking in August. The average minimum and maximum daily temperatures during this season are 24°C and 32°C, respectively. Average monthly rainfall during this time ranges from approximately 160 mm in September to 200 mm in August. The rest of the year receives an average of nearly 60 mm of rainfall per month and the average minimum and maximum daily temperatures are 15°C and 25°C, respectively [7].

The five georeferenced areas within the preserve each had 4 automobile tires propped upright against standing vegetation. Site A was located at 28.070900° N, 82.397600° W in a low-lying area with an elevation of 11 ft above sea level. Site B was at the highest elevation of the 5 sites at 82 ft and was located at 28.071267 ° N, 82.389650° W. Site C was located at 28.070433° N, 82.388367° W and was set 60 ft above sea level. Site D was at 28.070517° N, 82.387717° W and 24 ft. Site E was at 19 ft and 28.074733° N, 82.388950° W. All GPS coordinates and elevation readings were taken from a Garmin eTrex® H handheld unit. These locations are depicted at the top of Figure 1.

Most of the sites featured large *Quercus* oak and *Pinus* pine trees mixed in with saw palmetto *Serenoa* sp. (Figure 1, bottom left). Site B was unique from the others in that it was a scrub area predominately composed of saw palmetto, grass, and small oak and pine trees (Figure 1, bottom right). Spanish moss *Tillandsia usneoides* was often found on oak and bald cypress *Taxodium distichum* trees from nearby cypress domes.

Immature mosquito sampling

Fourth instar and pupal mosquitoes were collected from every tire approximately every 2 weeks from 27 September 2014 to 19 September 2015. The facing direction (°) of the tire and quantity of water (mm) in the tire was recorded before each collection.

Daily weather data

Daily minimum and maximum temperature (°C), minimum and maximum humidity (%), and precipitation (cm) were obtained from Tampa Executive Airport via wunderground.com, except for October 25-27, 2014, which were recorded from Tampa International Airport. The averages of each weather regressor were taken in between collection events, inclusive of the day of collection.

Multivariate regression modeling

Variable selection for the multiple regression models was carried out in SAS® 9.4 (SAS® Institute, Cary, NC) by a combination of multicollinearity diagnostics and an automatic forward stepwise procedure.

First, the presence of multicollinearity was diagnosed through a correlation matrix constructed in the CORR procedure. Additionally, the variance inflation factor (VIF) and tolerance options were specified in the REG procedure to quantify the severity of collinearity in the ordinary least squares regression analysis. The VIF measures how inflated the standard errors of the estimated regression coefficients become due to multicollinearity [8]. Tolerance is the reciprocal of VIF. Table 1 lists the predictor variables analyzed. The five study sites were categorized 0-4. Variables were eliminated from further regression analyses if their pairwise coefficient of correlation (r) ≥ 0.80, which indicated a high level of co-dependency on each another, and if their VIF was ≥ 10.0 and tolerance value was ≤ 0.10.

After non-independent regressors were removed from analysis, the residuals of the dependent variable total immature count were examined to assess their adherence to the normality assumption of regression modeling. The check for normality was completed by generating in the univariate procedure a quantile-quantile (QQ) plot of the normal quantiles against the residuals of the dependent variable. The plot (Figure 2) revealed that the errors in the dependent variable were not normally distributed. Moreover, the null hypothesis of the Shapiro-Wilk W test that the residuals were normally distributed was rejected

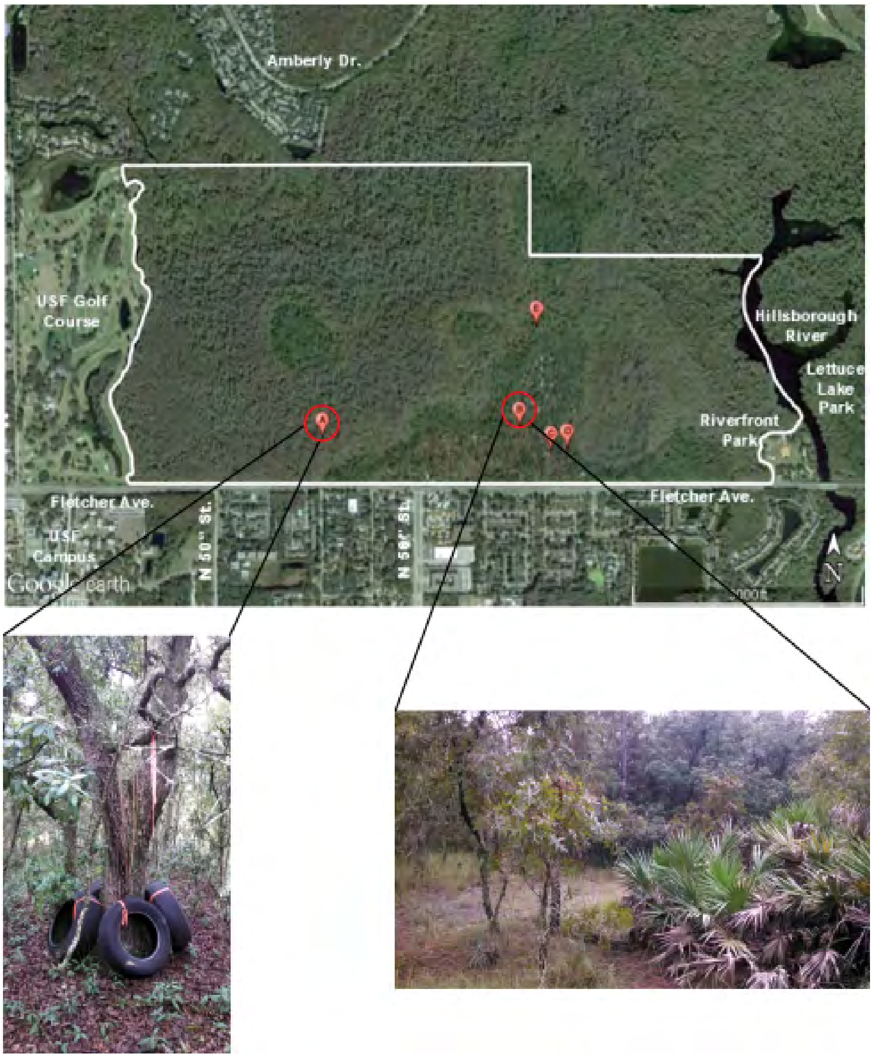

Figure 1: Top: Aerial view of the five sampling sites within the USF Forest Preserve. The Preserve's boundaries are outlined in white. Imagery source and date: modified from Google Earth imagery captured on January 17, 2015. Bottom left: Site A, exemplifying typical environment and tire set up. Bottom right: Environment surrounding Tire B.

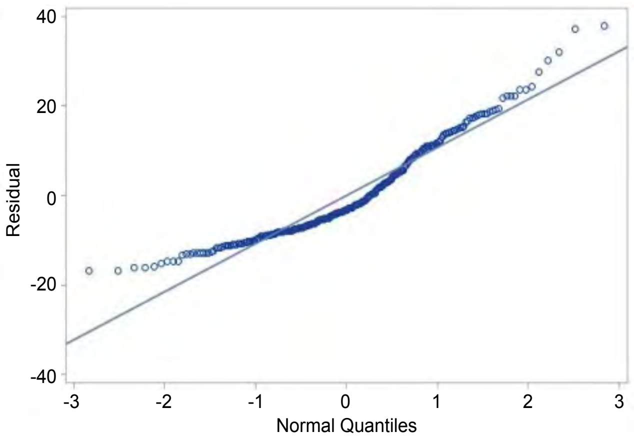

Figure 2: Quantile-quantile (QQ) plot of the normal quantiles against the residuals of the untransformed dependent variable to check for compliance with normality assumption of regression modeling.

(p-value<0.0001). Thus, the response variable was log-transformed to normalize the distribution and minimize standard error.

Finally, the log-transformed response variable was binarized, with amounts greater than 1.75 coded as a 1. Both the dichotomized and continuous log-transformed response variable were subjected to the forward stepwise procedure to determine the covariates to be included in the following regression analyses. The optimal model for forecasting immature mosquito count was decided through an automatic forward stepwise procedure that utilized the default p-value entry and exit values (α=0.15 each). p-values more liberal than α=0.05 were specified to prevent the selection procedure from ceasing prematurely, which could have reduced the amount of variation in the dependent variable explained by the explanatory predictors in the model. This variation was represented by the coefficient of determination R^2.

Logistic: A logistic regression paradigm attempts to describe the relationship of several predictor covariates to a dichotomous response variable, usually coded as 0 or 1 [9]. The model was generated with the logistic procedure in SAS$^®$ 9.4. The log-transformed response variable

Variable	Description	Units
TOTAL	Immature count data (dependent variable)	Number collected
DATE	Collection date	None
SITE_CAT	Site	A-E, categorized 0-4
TIRE	Tire	Numbered 1-4
ORIEN	Tire facing orientation	Degrees (°)
DEPTHMM	Amount of water in a tire	mm
NDVI	Normalized difference vegetation index	None; range -1 to +1 inclusive
SAVI	Soil adjusted vegetation index of site	None; range -1 to +1 inclusive
ELEVM	Elevation of site	M
MAXTC	Average† maximum daily temperature	°C
MINTC	Average† minimum daily temperature	°C
MEANTC	Average† mean daily temperature	°C
MAXH	Average† maximum daily humidity	Percent (%)
MINH	Average† minimum daily humidity	Percent (%)
PRECIP_CM	Average† daily precipitation	cm

Table 1: Ecological variables sampled of the immature mosquito capture point waste tire habitats in the Tampa, FL subtropical forest study site. †"Average" refers to the amount between each collection event averaged together, including the day of collection.

immature mosquito count was binarized with amounts greater than 1.75 coded as 1. The independent variables significant at α ≤ 0.05 level were derived from the forward stepwise selection procedure employing the dichotomized mosquito immature count were employed in the logistic forecasting model.

Poisson: Unlike a logistic response variable, a Poisson random variable can take any nonnegative integer value [10]. Hence, Poisson probability regression model using the GENMOD procedure was created to robustify model fit by employing the actual count data as the dependent variable rather than the bilateralized data. The procedure used maximum likelihood estimation to find the regression coefficients. Furthermore, the Poisson regression paradigm assumed that the data were equally dispersed (i.e., that the conditional variance equaled the conditional mean). The log-transformed count data was the dependent variable and the forward stepwise selection procedure decided the predictors included in the Poisson regression analysis.

Negative binomial: After assessing the data's adherence to the assumptions of regression model construction, a negative binomial model was composed in the GENMOD procedure to compensate for overdispersion (i.e., over-Poissonian) in geo-sampled field and remote explicative characteristic(s) eco-geographically representing immature mosquito count data geosampled from waste tires in the USF Forest Preserve in Hillsborough County, FL.

GIS: Normalized difference vegetation indices (NDVI) of each study site, disregarding weather conditions, land cover classification, plant physiognomy, and soil type, were calculated in ArcMap 10.3.1* utilizing the equation:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} \times \rho_{RED}} \tag{1}$$

NDVI values fall within a range of -1.0 and 1.0. New Para VI in arboviral mosquito epidemiology is the Normalized Difference vegetation index (NDVI) [11-17]. For example, Brown et al. [11]. used canonical correlation analyses to determine if a significant relationship existed between NDVI, disease/water stress index and distance to water and four local West Nile virus vectors [11]. (*Cx. pipiens, Cx. restuans, Cx. salinarius, and Ae. vexans*). Their model determined a significant relationship existed between the sampled explanatory predictor covariates and the sampled mosquito habitats (0.93, P=0.03).

Data originated from Landsat 8 data (www.nasa.gov) of Hillsborough County, FL acquired on 13 and 20 February 2015, downloaded from the USGS Earth Explorer website [8]. Band 4 corresponded to the red band whereas band 5 was near infrared.

Soil adjusted vegetation indices (SAVI) were similarly computed employing the equation:

$$SAVI = \frac{NIR - RED}{(NIR + RED + L)} \times (1 + L) \tag{2}$$

In areas where vegetative cover is low (i.e., <40%) and the soil surface is exposed, the reflectance of light in the red and near-infrared spectra can influence vegetation index values [6]. In previous research, Jacob et al. [12]. constructed multiple NDVI and NDVI variant geographic maps using QuickBird visible and NIR data and georeferenced *Cx. pipiens/restuans* explanatory predictor covariates sampled in a mosquito abatement district in northern Illinois [12]. Their models revealed that NDVI and soil adjusted vegetation index (SAVI) parameters can quantify prolific habitats based on spatiotemporal field-sampled count data. based on spatiotemporal field-sampled count data. Therefore, SAVI were computed employing the equation below for this investigation. The adjustment factor L = 0.5, as it was shown to reduce soil-induced noise throughout a range of vegetation densities [18].

Spectral biosignature construction: WorldView-3 satellite sensor data was used to image land cover types in Hillsborough County and remotely determine the geographical locations of anthropogenic waste tire habitats. Radiometrically corrected, mixelated multispectral band imagery was employed to create a spectral signature of a unit productive habitat. The WorldView-3 data had 0.31 m panchromatic 450-800 nm band resolution and 1.24 meter 8 band (red, red edge, coastal, blue, green, yellow, near-infrared 1, and near-inrared 2 400-1040 nm) multispectral resolution. The images were remotely taken over Hills borough County on February 8, 2015 and contained 3279 Km^2 of land cover. The satellite operated at an altitude of 617 Km and collected 680,000 Km^2 of ground data per day with an average revisit time of <1 day. The sensor images were delivered in a processed GeoTIFF file format, and with an Orthorectified Map Scale of 1:12,000 from Digital Globe Inc. (Longmont, CO, USA). There were 4 composite bands (Red, Blue, Green, and Infrared), with 16 bit value pixel collection depth, and were without any cloud cover (0% cloud cover).

Preliminary validation on the remote sensing images was conducted by converting WorldView-3's Digital Number (DN) to Top of Atmosphere Reflectance (TOA), which is the spectral radiance entering the telescope aperture at the altitude of 617 Km, in ArcGIS. A manual pertaining to WorldView-2 imagery was utilized in this exercise since during this research; a calibration methodology for WorldView-3 imagery had not yet been implemented. However, the same equation used for converting DN values to TOA reflectance on WorldView-2 images applied to WorldView-3. The conversion from radiometrically corrected image pixels to spectral radiance uses the following general equation for each band of a WorldView-3 product:

$$L_{\text{ëPixel,Band}} = \frac{K_{Band} \times q_{Pixel,Band}}{\Delta \text{ë}_{Band}} \tag{3}$$

where $L_{\lambda\text{Pixel, Band}}$ are TOA spectral radiance image pixels (Wm^{-2}sr^{-1}µm^{-1}), K_{Band} is the absolute radiometric calibration factor (Wm^{-2}sr^{-}

^{1}count^{-1}) for a given band, $q_{Pixel,Band}$ are radiometrically corrected image pixels (counts), and $\Delta\lambda_{Band}$ is the effective bandwidth (μm) for a given band. Conversion to TOA spectral radiance involved two major steps: multiplying radiometrically corrected image pixels by the appropriate absolute radiometric calibration factor (K) to obtain a band-integrated radiance (Wm^{-2}sr^{-1}) and then dividing the result by the appropriate effective bandwidth to get spectral radiance (Wm^{-2}sr^{-1}μm^{-1}) [19].

The remote sensing imagery that was analyzed in the initial preliminary validation step was employed to obtain a spectral biosignature representing waste tire habitats conducive to mosquito production in Hillsborough County. All tire habitats were analyzed in ArcGIS and the most prolific in terms of availability of larvae throughout the sampling period was identified for end member signature extraction. End member signature estimates, which are sub-pixel spectral surface radiance generated from a georeferenced unit habitat [5], were geosampled following interactive supervised image classification in ArcGIS, whereby homogenous waste tire spectral training samples were polygonised, merged, and their Red, Green and Blue band wavelength estimates generated (Figure 5). Prior to this probabilistic outcome, the remote sensor data was subdivided into two major phases: calibration, in which the algorithm identified a classification scheme based on signatures of different bands obtained from known training sites with known class labels; and prediction, in which the classification algorithm based on *a* priori probability file in ASCII format or training samples was applied to find other imaged sites with unknown signature classification membership based on known, sampled signatures [6].

In the ArcGIS cyber-environment, geostatistical Kriging/CoKriging interpolation was employed to validate the ability of the signature model in forecasting the presence of unknown, prolific waste tire sites in Hillsborough County. The spatial interpolation technique transformed irregular sampled known waste tire habitats to raster representation and resampled between multiband raster resolutions. In doing so, unknown, unsampled anthropogenic waste tire habitats at the study site were stochastically identified. Indicator Kriging in geostatistical analyst was further employed to obtain an epidemiological probabilistic, surface based maps based on primary threshold values obtained from the positively geosampled biosignature.

Results

Multivariate regression modeling

The multicollinearity diagnostic correlation matrix revealed that the independence assumption of regression modeling was violated when all of the predictors listed in Table 2 (except MEANTC) were included. NDVI and SITE_CAT, NDVI and SAVI, SAVI and SITE_CAT, and MAXTC and MINTC had *r* values ≥ 0.80, indicating that they were strongly correlated to each other: 0.97567, 0.97194, 0.90651, and 0.89865, respectively. These covariates also possessed high VIF values (≥ 10.0): NDVI 237.43157, SITE_CAT 75.71868, SAVI 62.82885, MINTC 16.62231, and MAXTC 13.64677. These same variables had low tolerance values (≤ 0.10): NDVI 0.00421, SITE_CAT 0.01321, SAVI 0.01592, MINTC 0.05370, and MAXTC 0.07328.

NDVI was removed from further statistical analyses due to its high level of correlation with multiple predictors, high VIF, and low tolerance value. Likewise, the temperature measurements were replaced by a one collective quantity, MEANTC. As a result of these adjustments, multicollinearity was removed from the data and compliance with the independence assumption of regression analysis was ensured.

Normal quantile plots were constructed with the UNIVARIATE procedure to test for the normality of the residuals. The residuals for the dependent variable total immature mosquito count were discovered to lack a normal distribution via visual inspection (Figure 3). This finding was quantitated and verified through the Shapiro-Wilk W test for normality conducted in the UNIVARIATE procedure. The test had a p-value<0.0001, which led to the conclusion that the null hypothesis declaring normality of the residuals should be rejected in favor of the alternative, which states non-normality. The response variable was then log-transformed to Gaussianize the distribution of its residuals and minimize standard error. The log-transformed count data was the response variable in subsequent model generation.

Logistic: The bilateralized log-transformed response variable total immature mosquito count underwent an automatic forward stepwise selection procedure to specify regressors to be included into the logistic regression paradigm. Log-transformed count numbers with a value greater than 1.75 were coded as 1. Six explanative factors had p-values under the α=0.15 model entry and exit levels, but only three of these six were statistically significant at the α=0.05 level. These three predictors were MINH with a p-value of 0.0002, DATE<0.0001, and SAVI 0.0049. Just those three were incorporated into the logistic regression model forecasting immature mosquito count. The explanatory output from the selection process is summarized in Table 2.

The least squares regressive model under the logistic paradigm was found to be Y=2.2634+[-1.79 × 10^{-7} (DATE)]+(-8.9439 × SAVI)+(0.0759 × MINH), with Y symbolizing the log-transformed mosquito immature count. Model R^2 was 0.1307.

Poisson: The kurtosis in logistic analysis (Figure 2) justified

Step	Variable Entered	Partial R^2	Model R^2	Mallows' C_p	p-value >F
1	MINH	0.0494	0.0494	33.0945	0.0002
2	DATE	0.0588	0.1082	16.5175	<0.0001
3	SAVI	0.0261	0.1343	10.293	0.0049
4	DEPTHMM	0.0081	0.1424	9.7379	0.1143
5	ORIEN	0.0111	0.1535	8.2199	0.0629
6	ELEVM	0.0118	0.1653	6.484	0.0541

Table 2: Explantory forecast of the automatic forward stepwise procedure utilizing the binomialized dependent variable to select variables for regression model construction.

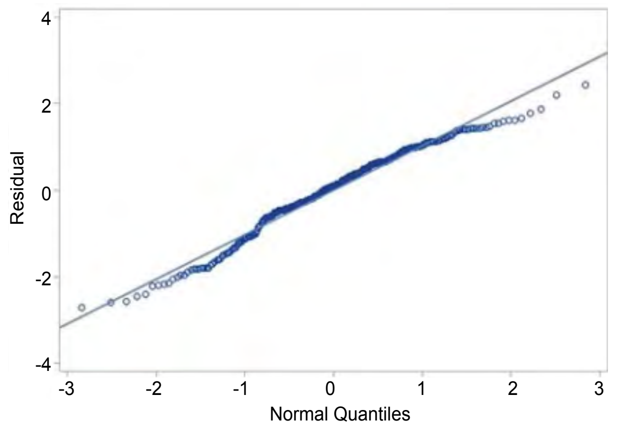

Figure 3: Quantile-quantile (QQ) plot of the normal quantiles against the residuals of the log-transformed dependent variable to check for compliance with normality assumption of regression modeling.

the creation of a Poisson model with a gamma-distributed mean to compensate for light tails on both the positive and negative ends and to robustify the model. Similar to the assembly of the logistic estimating standard, the forward stepwise selection process was applied to the continuous logarithmic transformation of the dependent variable to select expository factors for regression modeling. Like in the logistic analysis, only three covariates were chosen for inclusion in the Poissonian model: MINH with a p-value of 0.0040, SAVI 0.0070, and DATE 0.0078. The output from the selection process utilizing the continuous dependent variable is summarized in Table 3.

The least squares regression model under the Poissonian distribution was found to be Y=3.0498+DATE+(-4.5092 × SAVI)+(0.0281 × MINH), with Y symbolizing the log-transformed mosquito immature count. The model's R^2 of 0.0811 was determined by a regression procedure that included just the three cofactors used in the Poisson regression model.

Negative binomial: Next, heteroskedascity was evaluated in the REG procedure by plotting the residuals of the response against its predicted values. Additionally, the White test, which confirms whether the variance residuals are constant and homogenous (i.e., homoskedastic), was included via the SPEC option. Finally, the presence of outliers was detected by calculating the studentized residuals (r_i), leverage, Cook's distance, and DFFITS values in the REG procedure, then locating values that exceeded cutoffs specified for each outlier diagnostic.

Since the Poisson model's R^2 decreased relative to that of the logistic model, compliance of the data with the assumptions of regression analysis was ensured in several ways. First, normality of the residuals of the log-transformed continuous response variable was ensured by creating in the UNIVARIATE procedure quantile-quantile (QQ) plots of the normal quantiles against the residuals of the dependent variable. The plot, shown in Figure 3, indicates that the non-Gaussianism of the residual error distribution still held true despite log-transformation. The Shapiro-Wilk W test for the normality of the error had a p-value of 0.0001. This lower level of significance in comparison to that from the test of the untransformed dependent variable (p-value of <0.0001) confirmed that the log-transformation did succeed in fitting the residuals to a more normal distribution and in reducing standard error.

Next, linearity between the response and predictor variables was tested by constructing partial regression plots in the REG procedure. This output is depicted in Figure 4. The plots of each regressor involved in the Poisson analysis against the log-transformed response variable displayed non-linear relationships.

Then, a plot of the residuals versus the fitted (predicted) values was created to reveal the possible presence of heteroskedascity in the model. The SPEC option was attached after the model statement while constructing the plot in the REG procedure in order to obtain a White test. The null hypothesis of the White test asserts that the variance of the residuals is homogeneous. When the model is correctly specified and the errors are independent of the regressors, the rejection of this null hypothesis is evidence of heteroskedasticity [20]. The p-value of the White test was 0.0830, so the data was deemed to be non-homoskedastic.

Step	Variable Entered	Partial R^2	Model R^2	Mallows' C_p	*p*-value >F
1	MINH	0.0303	0.0303	15.1058	0.004
2	SAVI	0.026	0.0564	9.53	0.007
3	DATE	0.0248	0.0811	4.3234	0.0078

Table 3: Explantory forecast of the automatic forward stepwise procedure utilizing the discrete dependent variable to select variables for Poissonian and negative binomial regression model construction.

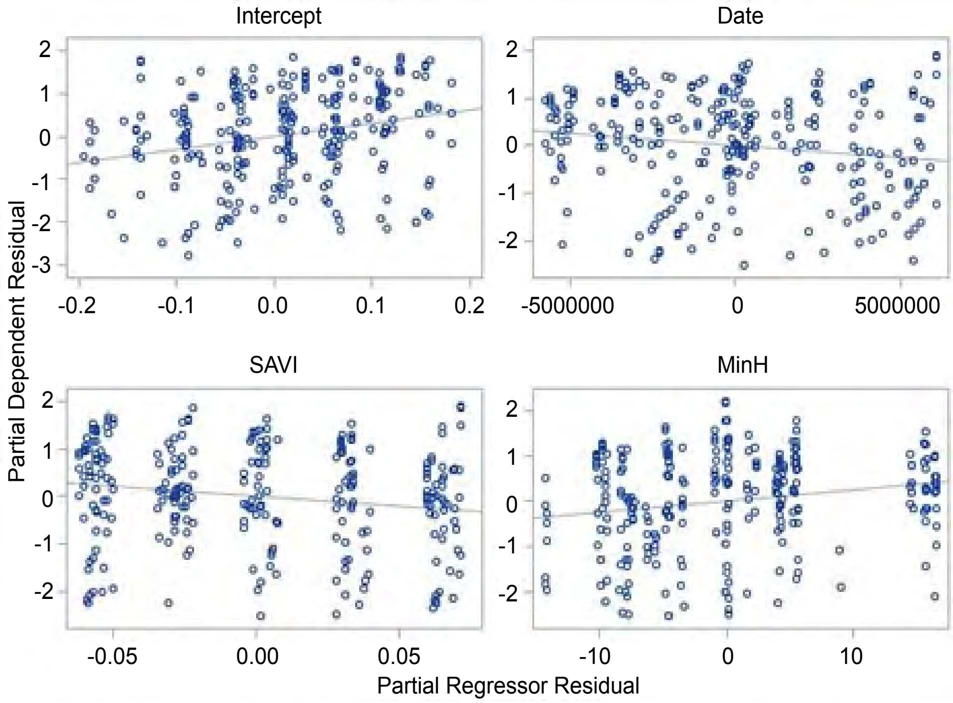

Figure 4: Non-linear relationship of the between the predictors and outcome variable exhibited by partial regression leverage plots of the residuals of the log-transformed dependent variable.

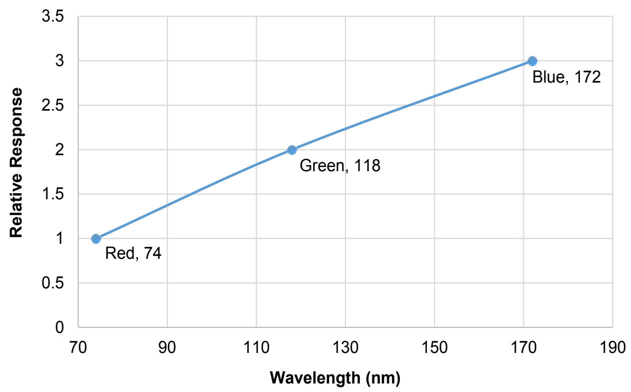

Figure 5: Spectral radiance band emissivities for an anthropogenic vector and nuisance mosquito waste tire habitat.

Finally, studentized residuals (r_i), difference in fits (DFFITS), Cook's distance, leverage, and difference in beta (DFBETA) were calculated to identify any influential cases amongst the data used for the Poisson regression. None of the measurements indicated any influential outliers.

In summary, the data featured deviations from the assumptions of regression modeling, which may have contributed to a low R^2 value and increased error in the Poisson model. The source of the error was likely due to *overdispersion,* a phenomenon that may occur with binomial and Poisson data. For Poisson data, overdispersion arises when the variance of the response Y exceeds the Poisson variance [21]. Recall that the Poisson variance equals the response mean ($Var(y)=\mu$).

Since the error variance was revealed to be inconstant (i.e., non-homoskedastic), we created a multivariate negative binomial regression model to compensate for the overdispersion present in the mosquito count data [10].

The negative binomial paradigm resulted in a least squares regressive model identical to that of the Poisson: Y=3.0498+DATE+(-4.5092 × SAVI)+(0.0281 × MINH). The *p*-values of each variable were the same as well: DATE 0.0326, 0.0215 SAVI, and 0.0124 MINH.

GIS: The generality of the biosignature model was tested in Hillsborough County. To control for data redundancy, a unit sample point employing independently distributed mean values per pixel was obtained. Of the 12 geosampled potential vector and nuisance mosquito waste tire sites predicted to be conducive habitats (Figure 6), 10 (83%) were found to meet or exceed the threshold value contained in the spectral signature. The model thus exhibited a sensitivity of 83% and a specificity of 87% when applied to find unknown, hyper-productive anthropogenic waste tire habitats in Hillsborough County. Since some waste tire habitats in the study area were within short spatial distances to each other, their presence at a similar geospatial location could have resulted in the 10 coincidental sample points seen in Figure 7.

Discussion

Previous work studying the extent to which field-sampled environmental parameters influenced mosquito count focused on to predicting prolific *Aedes albopictus* and *Culex quinquefasciatus* habitats identifying *Anopheles* habitats in tropical Africa [22-26]. The few studies that have attempted this in North America were in suburban Alabama [23,27]. and Illinois [12]. This work is the first attempt to determine positive predictors of associate with prolific mosquito habitats in an undeveloped subtropical North American forest environment.

The significance of the collection date in determining immature mosquito count from tires confirmed that there was seasonal variation to production in the subtropical study area. However, against the established knowledge that culicid abundance is sensitive to climatic changes, the minimum humidity level (%) was the only meteorological variable deemed statistically significant to mosquito production in waste tires in this study. Also, the SAVI as a more significant predictor than the NDVI indicated that the SAVI is a better proxy covariate of vegetation greenness and therefore should be favoured over the NDVI in upcoming GIS/public health research that necessitates usage of a eco-geographic LULC characterization of the vegetation canopied landscape in order to increase disease forecasting power.

Although the negative binomial paradigm assuaged heteroskedascity within the data, its application here resulted in a regressive model identical to that of the Poisson. Hence, it could not be confirmed that the negative binomial model was more or less robust than the Poisson. Since deviance from normality could not be entirely fixed by log-transformation of the response variable, error within the data itself may have contributed to these findings. Sources of model uncertainty include inaccuracies in recording elevation and weather data. The Garmin eTrex® H GPS unit altimeter may have been off by ± 3 m, which may have made enough of a statistically significant difference for ELEVM to become a predictor in this study's regression models. Weather data may have added error into the statistical analyses due to the distance of Tampa Executive Airport from the study area. Since rainfall in Florida can be localized, more precise weather information could have helped the violations of regression assumptions seen in this study.

Therefore, future studies should use a lag or ARIMA model to assess site-specific weather data prior to regressing with it to reveal individual site-level time-series dependent trends [30].

Despite the shortcomings seen here, the statistical and signature models constructed here provide a parsimonious yet accurate estimator

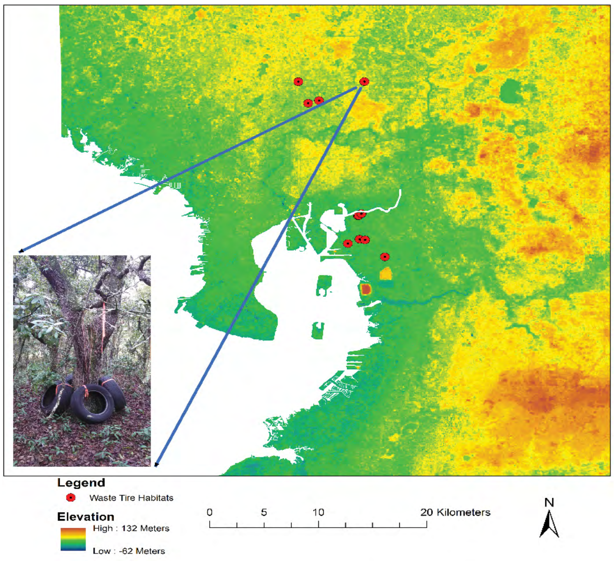

Figure 6: Geosampled waste tire habitats overlaid on to a 3-D slope coefficient model of Hillsborough County, FL.

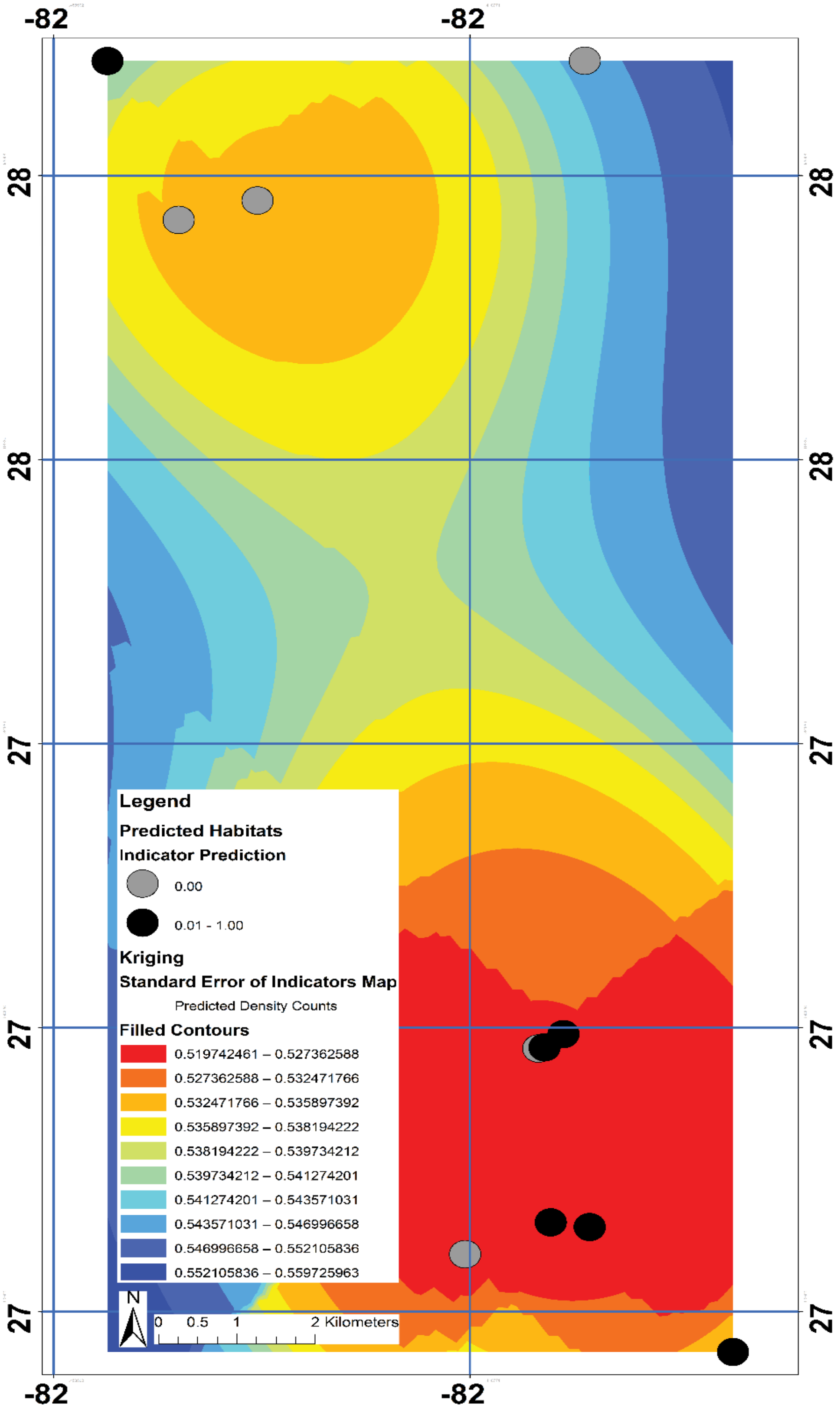

Figure 7: Probability prediction model outcome from validated anthropogenic waste tire habitats in Hillsborough County, FL.

of undiscovered waste tires near human dwellings in a subtropical, undeveloped zone that may yield many mosquitoes. Mosquito abatement managers could use the methods proposed here to model environment sampled explanatory covariate coefficients in their proposed area, then strategically implement control efforts as time and budgetary restraints allow.

In this research the NDVI and SAVI were very similar in their optimally forecasted estimates. Another approach for regressively quantitating geomorphological, soil-related explanatorial covariate coefficinets is by quantitating land cover information of a georeferenced hyperproductive waste tire habitat ArcGIS-based Land Information Surface (LIS) model. These models are designed explicitly for soil moisture estimation. LIS is very customizable with the ability to choose many different inputs for geo-spatiotemporal geosampled parameters (e.g., elevation, soil types, land use classification) and other related data (i.e., radiation and meteorological fields including precipitation updated hourly or 3-hourly). LIS has the ability to run several "tiles" within a sub-meter resolution digitized grid cell that has different land use classifications (www.nasa.gov) so even if a digitized cell classified at the Hillsborough study sites was for example 5% urban, that portion could still be remotelyhmonitored. For example, in previous research Jacob et al. [19] generated an LIS map using field and QuickBird-geosampled explanatory predictors of *Culex. pipiens/restuans* for 15 larval habitats in Urbana/Champaign, Illinois USA. The LIS framework may be used to provide information on surface soil moisture conditions related to geosampled georferenced waste tire habitats. In these models the configuration, will be based on a Land Data Assimilation System (NLDAS) forcing data (1/8- degree, hourly) up to 3 days before the sampling day and a Global Land Data Assimilation System (GLDAS) forcing data (1/2 degree, 3-hourly) up to 12 hours previous to the time of sampling.

Additionally, the model could be augmented with extra georeferenced explanatory predictor variables at each grid point to quantify other data related to soil variable including:

- water depth in an open container such as a bucket (taking into account precipitation and evaporation influenced by temperature, humidity, winds, and radiation),
- water depth in a infrequently shaded container such as a tire (similar to the first, but with little or no solar radiation),
- potential standing water on the ground, assuming an area with suitable topography exists inside a stratified grid cell with no drainage by runoff. This would provide monitoring of potential conditions favorable to arboviral mosquito outbreaks. Thereafter, all generated model covariates may be analyzed using various statistical algorithms (e.g., linear, exponential, logarithmic, power or polynomial).

References

1. USGS Disease Maps (2014) January 13, 2015 (Cited 2015 April 13) Available from: http://diseasemaps.usgs.gov/
2. Kling LJ, Juliano SA, Yee DA (2007) Larval mosquito communities in discarded vehicle tires in a forested and unforested site: detritus type, amount, and water nutrient differences. Journal of Vector Ecology 32: 207-217.
3. Beier JC (1983) Influence of water chemical and environmental parameters on larval mosquito dynamics in tires. Environmental Entomology 12: 434-438.
4. Morris CD, Robinson JW (1994) Distribution of mosquito larvae in a waste tire pile in Florida - an initial study. Journal of the American Mosquito Control Association 10: 174-180.
5. Jacob BG (2011) A taxonomy of unmixing algorithms using Li-Strahler geometric optical model and other spectral end member extraction techniques for decomposing a QuickBird visible and near infra-red pixel of an Anopheles arabiensis habitat. Open Remote Sensing Journal 4: 1-25.
6. Jacob B (2013) Unbiasing a stochastic end member interpolator using ENVI object-based classifiers, a Farquhar's single voxel leaf photosynthetic response explanatory model and Boolean time series statistics for forecasting shade-canopied Simulium damnosum s.l. larval habitats in Burkina Faso. Journal of Geophysics and Remote Sensing 2: 109.
7. NOAA (1981-2010) Station normals of temperature, precipitation, and heating and cooling degree days - Tampa international airport. NOAA, Editor. 2015, NOAA: Asheville, NC, p: 4.
8. USGS Earth Explorer (2015) (Cited 2015 July 11); Available from: http://earthexplorer.usgs.gov/
9. Huete AR (1995) A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment 25: 295-309.
10. Updike T, Comp C (2010) Radiometric use of WorldView-2 Imagery Digital Globe®, Inc.
11. White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica: Journal of the Econometric Society 48: 817-838.
12. Rao JN, Scott AJ (1999) A simple method for analysing overdispersion in clustered Poisson data. Stat Med 18: 1373-1385.
13. Haight FA (1967) Handbook of the poison distribution, New York: Wiley Press.
14. Jacob BG, Arheart KL, Griffith DA, Mbogo CM, Githeko AK, et al. (2005) Evaluation of environmental data for identification of Anopheles (Diptera: Culicidae) aquatic larval habitats in Kisumu and Malindi, Kenya. J Med Entomol 42: 751-755.
15. Jacob BG, Morris JA, Caamano EX, Griffith DA, Novak RJ (2011) Geomapping generalized eigenvalue frequency distributions for predicting prolific Aedes albopictus and Culex quinquefasciatus habitats based on spatiotemporal field-sampled count data. Acta Trop 117: 61-68.
16. Jacob BG, Griffith DA, Muturi EJ, Caamano EX, Githure JI, et al. (2009) A heteroskedastic error covariance matrix estimator using a first-order conditional autoregressive Markov simulation for deriving asympotical efficient estimates from ecological sampled Anopheles arabiensis aquatic habitat covariates. Malar J 8: 216.
17. Jacob BG, Griffith DA, Novak RJ (2008) Decomposing Malaria Mosquito Aquatic Habitat Data into Spatial Autocorrelation Eigenvectors in a SAS/GIS® Module. Transactions in GIS 12: 341-364.
18. Hay SI, Snow RW, Rogers DJ (1998) Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. Trans R Soc Trop Med Hyg 92: 12-20.
19. Jacob BG (2010) Developing GIS-based eastern equine encephalitis vector-host models in Tuskegee, Alabama. International Journal of Health Geographics 9: 12-27.
20. Jacob BG (2009) Developing operational algorithms using linear and non-linear squares estimation in Python® for the identification of Culex pipiens and Culex restuans in a mosquito abatement district (Cook County, Illinois, USA). Geospat Health 3: 157-176.
21. Jacob BG (2010) A Random-effects Regression Specification Using a Local Intercept Term and a Global Mean for Forecasting Malarial Prevalance. American Journal of Computational and Applied Mathematics 3: 49-67.
22. Marniemi J, Parkki MG (1975) Radiochemical assay of glutathione S-epoxide transferase and its enhancement by phenobarbital in rat liver in vivo. Biochem Pharmacol 24: 1569-1572.