

Adjusting Second Moment Bias in Eigenspace Using Bayesian Empirical Estimators, Dirichlet Tessellations and Worldview I Data for Predicting *Culex quinquefasciatus* Habitats in Trinidad

Models for West Nile Vectors

Benjamin G. Jacob¹, Dave D. Chadee², Robert J. Novak¹

¹School of Medicine, Division of Infectious Diseases, University of Alabama at Birmingham, Birmingham, Alabama, USA

²Department of Life Sciences, University of the West Indies, St. Augustine, Trinidad and Tobago, West Indies
E-mail: bjacob@uab.edu, dave.chadee@sta.uwi.edu, rjnovak@uab.edu

Received October 24, 2010; revised November 12, 2010; accepted December 23, 2010

Abstract

Temporally weighted regression models with a spatial autoregressive component may estimate nonlinearities in spatiotemporal-sampled data of *Culex quinquefasciatus*, a major vector of West Nile Virus (WNV) which can help implement control strategies by determining optimal predictors associated to prolific habitats. The design of this kind of mixed model can specifically incorporate spatial autocorrelation whilst including the influence of other aspatial predictor variables. Currently, the lack of an estimation theory that allows for heteroscedasticity and corresponding joint hypothesis testing in the presence of spatial dependence in georeferenced *Cx. quinquefasciatus* habitat data is a serious shortcoming in WNV research. In this paper we used spatially lagged and simultaneous autoregressive models based on multiple predictor variables of immature *Cx. quinquefasciatus* and Worldview 1 (WV-1) data to help implant a remote habitat-based surveillance system in Trinidad. Initially, we used Geomatica Ortho Engine[®] v. 10.2 for extracting a Digital Elevation Model (DEM) from the WV-1 raw imagery. Results of the DEM analyses indicated a statistically significant inverse linear relationship between total sampled *Cx. quinquefasciatus* data and elevation (m) ($R^2 = -0.439$; $p < 0.0001$), with a standard deviation of 10.41. Additional field-sampled information was derived using data from an orthogonal grid-matrix constructed in an ArcInfo 9.3[®] and overlaid onto the WV-1 data. A unique identifier was placed in the centroid of each grid cell. Univariate statistics and Poisson regression models were then generated using the georeferenced covariates in SAS/GIS[®]. Coefficient estimates were also used to define expectations for prior distributions in a Bayesian estimation matrix using Markov Chain Monte Carlo (MCMC) specifications. A spatial residual trend analyses was then performed using autocorrelation indices which linked tabular data in SAS PROCLMIXED[®] with the egg-raft count data in ArcInfo[®]. The estimation matrix identified prolific habitats based on the covariate distance to the nearest house. An Ordinary kriged-based interpolator was then constructed in Geostatistical Analyst Extension of ArcGIS 9.3[®] based on the adjusted Bayesian estimates. For total *Cx. quinquefasciatus* egg-raft count, first order trend was fitted to the semivariogram at a partial sill of 5.931 km, nugget of 6.374 km, lag size of 7.184 km, and a range of 31.02 km using 12 lags. We assessed the performance accuracy of the interpolation procedures based on the magnitude and distribution of errors between observed and model-predicted values using Voroni tessellations. These residuals divided the space between the individual georeferenced *Cx. quinquefasciatus* habitats by XY coordinates in 2-dimensional space which revealed that the geophysical parameter error residuals in the interpolation model were within normal statistical limitations. Newer GIS software and WV-1 data can generate highly accurate predictive *Cx. quinquefasciatus* habitat distribution models which can target prolific habitats of based on field-sampled count data. Our results suggest it may be unnecessary to manage all *Cx.*

quinquefasciatus habitats to obtain significant reductions in incidence and prevalence of WNV in Trinidad.

Keywords: *Culex Quinquefasciatus*, Trinidad, West Nile Virus, Worldview 1, ArcGIS®

1. Introduction

West Nile virus (WNV) is among the *Culex*-borne encephalitides in the *Flavivirus* genus. The virus cycles between birds and mosquitoes, while horses, humans, and a number of other vertebrates are considered incidental hosts. Since WNV first appeared in the Western Hemisphere in New York in 1999, it has spread rapidly across the North American continent, causing large numbers of human cases with neurologic disease and death and even greater amounts of milder disease characterized principally by fever and rash [1]. West Nile Virus has been spreading southward into the Caribbean Basin and Latin America as well, where its public health impact remains poorly understood [2]. The virus was first detected in 2001, in Jamaica and the Cayman Islands. Thereafter, serological evidence of WNV transmission has been accumulating. Cross-reactive WNV antibodies in humans have been detected from Mexico, the Bahamas, and Cuba, in horses from Guadalupe, Mexico, Central America, Puerto Rico, Colombia, and also in resident birds from the Dominican Republic and Venezuela. In November 2006, 4 serologically confirmed human WNV encephalitis cases were reported in Argentina. West Nile Virus may be following the same patterns of southward dissemination as other arboviruses into the Caribbean and Central and South America via migratory birds.

There have been attempts to develop and implement WNV surveillance in the Caribbean but, unfortunately present systems in the region are unprepared to track WNV spread. For example, in Guadeloupe, extensive surveillance for WNV infection demonstrated a high seroprevalence in equines in 2003, despite no human or equine case being reported. A protocol for epidemiological surveillance for WNV has been developed through a collaborative electronic working tool posted on the Caribbean Animal Health Network (CaribVET) website (www.caribvet.net). CaribVET is a collaboration of veterinary services, diagnostic laboratories, research institutes, universities, and regional/international organization to improve animal health in the Caribbean through exchange of information and data management. Improving collaborations and information dissemination for understanding WNV epidemiology is vital; however, WNV surveillance and diagnostics in the Caribbean requires tools for regional data analyses. Unfortunately, to date there have been virtually no significant findings in any established WNV surveillance programs in tropical

America. In september 2007, WNV infection detected by polymerase chain reaction (PCR) in post-mortem brain tissue taken from an encephalitic horse and by virus isolation from a dead bird confirmed WNV transmission in southwest Puerto Rico but, since then there have been no isolates reported from other areas in Central and South America or the Caribbean.

The explanation for the paucity in WNV detection in tropical America may be the lack of active WNV surveillance. One would hypothesize that the likelihood of identifying WNV cases would be maximized in the Caribbean Basin, given its location in the flight paths of infected migratory birds. However, due to cross-immunity of WNV with other tropically endemic *Flaviruses* such as St. Louis encephalitis, it is possible that WNV has not been able to establish persistent enzootic foci in the region, so that reported serological evidence may be caused by frequent but unsuccessful virus reintroductions. Regardless, presently a WNV mosquito habitat-based intervention surveillance system using a vigorous quantitative assessment of field and remote-sampled data is warranted for countries of the Caribbean.

A main challenge in devising an effective West Nile surveillance program in regions of the Caribbean; however, is the lack of sufficient empirical knowledge on the spatiotemporal patterns of *Culex quinquefasciatus* larval habitats and their underlying contribution to the adult population. The main tropical vector for WNV in birds and people is *Cx. quinquefasciatus* in the Southeastern United States and the Caribbean [3]. Previous research has revealed that *Cx. quinquefasciatus* abundance will vary significantly in different areas due to hydrologic regime and the proximity to habitats and blood-meal hosts [4-7]. These factors are not evenly distributed spatiotemporally, so it is logical to expect a variation in abundance and distribution of virus-positive mosquitoes within different ecosystems in the Caribbean.

Recently, the use of remote sensing (RS) and GIS for the quantitative prediction of geographical distributions of agent, reservoir, and vector species has been advocated to augment traditional WNV mapping methods [4,5,8]. For example, the United States Geological Survey (USGS) and Center of Disease Control (CDC) have employed GIS and RS to prepare interpretive maps showing WNV activity in many regions of North America (<http://www.cdc.gov/ncidod/dvbid/westnile/resources/wnguidelines1999.pdf>). Common GIS/RS technology for WNV data analyses include: 1) time series overlay analyses of thematic data and spatial intersection, 2)

buffer generation and neighborhood analysis, 3) vector-based grid generation and network analysis; and, 4) surface modeling [9]. Once an area has been imaged, associations between risk variables and environmental-sampled covariates can be quantified using the spatial data analysis capabilities of GIS for implementing a WNV mosquito surveillance system [9-11].

Quantification of vector-host interactions incorporated on Land Use Land Cover (LULC) maps generated in GIS time series overlay analyses suggest that *Culex* habitat productivity vary greatly over short distances, depending on the space and time at which field measurements are made [5,12-15]. For example, Jacob *et al.* (2009a) generated a land use land cover (LULC) classification in Erdas *Imagine* based on Landsat-7 ETM+ data acquired July 2003 and Landsat-5 TM data acquired July 1991 for examining two ornithophilic mosquito vector species *Culex pipiens* and *Culex restuans* in residential areas of Urbana-Champaign, Illinois [4]. A maximum likelihood unsupervised classification and LULC detection was performed using a cross-tabulation detection method and a grid-based algorithm stratified by drainage for detection of environmental predictors associated to field-sampled egg-raft abundance count data. The resulting LULC change matrix revealed that between 1991 and 2003 there was a total of 12.1% LULC change in the study site. Of 20,166 egg-rafts sampled, 76.7% was in land cover classified as maintained urban followed by 14.8% in maintained non-urban. The well-drained stratum harbored 56.4% of the total sampled data. Additionally, a remote stratification of the urban land cover site based on QuickBird visible and near infrared (NIR) data revealed that 83.3% of the egg-raft distribution was in LULC classified as residential high canopy coverage. In another study, Jacob *et al.* (2009b) constructed multiple geospatial models based on LULC and meteorological variables to predict adult abundance of the same *Culex* species in Des Plaines mosquito abatement district in northern Cook County, Illinois [5]. An ecological database of the study site was constructed in ArcGIS 9.2[®] also using multitemporal QuickBird visible and NIR data overlaid with total adult *Culex* abundance and WNV infection rates from 15 georeferenced gravid traps sampled weekly from May 2002 to October 2005. A regression analyses of LULC parameters revealed that adult *Culex* abundance in the georeferenced gravid traps was positively associated with temperature and negatively associated with precipitation. In 2002, the overall accuracy of the model was 73.3%, 65.1% in 2003, 47.2% in 2004 and 60.0% in 2005. Temperature and precipitation explained > 80.1% of the variation in each of the models. Estimates from the multivariate analysis also revealed that land cover that was classified as forest LULC along with middle range built

environment (< 40.1%) predicted heightened infected vector activity. Because abiotic factors such as LULC are measurable in GIS using submeter resolution data, these pre-epidemic abiotic signatures may be quantified for predicting arboviral epidemic transmission in isolated areas of the Caribbean. These epidemic signatures may be tracked, measured, quantified, and presented visually to help forecast human arboviral epidemics. While remote models of the observed spread of WNV have been used to evaluate potential vector and host mechanisms of dispersal and to define the suitability of the various landscape features and meteorological variables for mosquito transmission of WNV in many portions of the US, there is no literature using LULC and high resolution satellite data for evaluating ecologically sampled *Cx. quinquefasciatus* habitats in any region of the Caribbean.

Recently, digital elevation models (DEMs) have become popular with the use of GIS for surface modeling flood and swamp water mosquito abundance in many different geographic regions [16-18]. A digital elevation model (DEM) is a digital file consisting of terrain elevations for ground positions at regularly spaced horizontal intervals. It is a digital representation of ground surface topography which is usually represented as a raster or as a triangular irregular network. DEMs include a number of production strategies for determining terrain-related parameters associated to prolific *Cx. quinquefasciatus* habitats such as manual profiling from photogrammetric stereo models, digitizing of contours, digitizing topographic map contour plates, converting hypsographic and hydrographic tagged vector files and also for performing autocorrelation via automated photogrammetric systems. To obtain DEMs from satellite images for determining vector amplification and host population dynamics for mapping the spatiotemporal distribution of WNV from high resolution data two methods are possible: along-track stereoscopy from the same orbit using fore and aft images and across-track stereoscopy from two adjacent orbits [16]. The four possibilities of stereo image processing for DEM generation are Fore-Aft, Fore-Nadir, Aft-Nadir and Fore-Aft-Nadir images as a pair [19]. The simultaneous acquisition of along-track stereo data has a strong advantage in terms of radiometric variation versus the multi-date acquisition of across-track stereo data which has been traditionally employed in past research for identifying terrain-related explanatory covariates of floodwater mosquito habitats [20]. The across-track approach has been applied frequently since 1980, first with Landsat from two adjacent orbits, then with SPOT using across-track steering capabilities and finally with IRS-I C/D. But these satellite data were limited due to spatial resolution. Fortunately, along-track stereoscopy

has recently gained renewed popularity using higher resolution satellite data systems including JERS-1's optical sensor (OPS), German Modular Opto-electronic, Multi-Spectral Stereo-Scanner (MOMS), ASTER, IKONOS, QuickBird, Orthoview, SPOT-5, FORMOSAT II, Cantosat and the latest addition of Digitalglobe's Worldview (WV)-1. These data can define georeferenced terrain covariates associated with prolific *Cx. quinquefasciatus* habitats. For example, Jacob *et al.* (2010) used multispectral QuickBird data and DEM-based GIS methods to evaluate stream flow direction for determining satellite-derived characteristics of drainage basin physiographic predictors associated with *Culex erraticus*, a mosquito vector of eastern equine encephalitis virus (EEEV) and Northern Cardinals, an avian host associated with WNV and EEEV in Tuskegee, Alabama [16]. Eastern equine encephalitis virus is a virulent arbovirus maintained in an enzootic cycle between local avian fauna and ornithophilic mosquitoes in coastal regions of the eastern United States [1,14,16]. Euclidian distance-to-nearest hydrological body was calculated as the distance in geo-space from a grid cell to a stream grid cell defined by a Stream Raster Grid. A Terrain Analysis using a DEM (*i.e.*, TauDEM) was then performed in ArcGIS 9.2[®] to retrieve multiple geomorphological parameters. Additionally, a three-dimensional model of the study area was constructed based on the DEM using ArcScene extension of ArcGIS[®]. A Pearson correlation analyses computed the ratio of covariance between the sampled variables to the product of their standard deviations. Substituting estimates of the covariances and variances based on the *Cx. erraticus* samples provided the sample correlation coefficient which revealed significant inverse linear relationships between the total field-sampled adult count data and elevation ($R^2 = -0.4711$; $p < .0001$) with a standard deviation (SD) of 11.16. The analyses of the Northern Cardinal data and elevation also revealed a similar statistical relationship ($R^2 = -0.5831$; $p < .0001$) with a SD of 11.42. Even though the usefulness of selected static catchment predictor covariates (e.g., stream, density, and slope) in assessing *Culex* habitats has been verified in previous research, characteristics of drainage networks and basin physiographic parameters have not been applied in modeling WNV mosquito abundance and distribution in any country of the Caribbean.

Predictive interpolation algorithms have also not been used for forecasting productive *Cx. quinquefasciatus* habitats based on field and remote-sampled data in any country of the Caribbean. One of the most powerful and potentially valuable GIS/RS vector mosquito data modelling mechanisms is that of automated prediction. Generally, given relevant spatiotemporal field and re-

mote-sampled explanatory predictor variables with appropriate annotation, a WNV mosquito model can be constructed for making accurate forecasts of one or more dependent "target" quantities of interest. These quantities of interest may be continuous real values (a scenario often referred to as regression" or "interpolation"), in vector mosquito data analyses or, it can be in the format of discrete labels ("aquatic habitat classification" or "habitat distribution pattern recognition"). Both these cases are referred to as "supervised learning" in the GIS/RS vector mosquito data learning vernacular. Predictive interpolation algorithms can also map urban variation in vectors based on differentials in environmental-sampled parameters which can identify and validate models linking human ecological factors, while controlling for potential spatial autocorrelation in model residuals [21-25]. Spatial autocorrelation, the correlation among values of a single variable strictly attributable to their relatively close positions on a two-dimensional surface is frequently encountered in the analyses of spatiotemporal georeferenced vector mosquito data [4,8,21, 26,27]. Autocorrelation violates the ordinary least squares (OLS) assumption that the error terms are uncorrelated. While it does not bias the OLS coefficient estimates in a predictive vector mosquito habitat distribution model, the standard errors tend to be underestimated and the *t*-scores overestimated when the autocorrelations of the errors at low lags are positive [16].

Kriged-based interpolation may analyze spatiotemporal-sampled WNV mosquito data in the presence of fine-scale autocorrelation. Kriging is mathematically related to regression analyses. Both theories derive a best linear unbiased estimator based on assumptions of covariance and make use of the Gaussian-Markov theorem to prove independence amongst georeferenced vector mosquito data exploratory parameters [26]. Kriging vector mosquito data is commonly made by interpolation of a single realization of a randomized study area based on field and remote observations in multivariate environmental-sampled datasets. Under the regionalized assumption, local habitat variation is not generally unstructured but, it is spatially dependent at some scale as georeferenced points within a given habitat distance apart depend on one another in a statistical sense (16). Any variable is stationary when its distribution is invariant under translation, that is, the stationary region is the location to make estimates (26). Stationary variables provide a preliminary model decision for uncertainty about the sample value at a specific sampled habitat location $z(x)$ (16). For any space lag (h) increment in a particular direction, the distribution of $z(x_1), z(x_2) \dots z(x_n)$ habitat based observations is the same as that of $z(x_1 + h), z(x_2 + h) \dots z(x_n + h)$, where $z(x)$ is regarded as the sum of

three terms: (1) mean, drift or trend, (2) random dependent variable plus (3) error (16) This, however, requires the first two moments (*i.e.*, mean and variance) to be constant. This spatial homogeneous region in the distribution of the georeferenced *Cx. quinquefasciatus* habitats then will make random functions homogeneous and self-repeating in space. The variance of the random spatial variable will equal covariance at zero lag habitat distance ($\sigma^2 = C(0)$), while the covariance between $z(x)$ and $z(x+h)$ becomes $\sigma_{x,x+h} = \sigma^2 f(h) = C(h)$ (16). This means that the stationary covariance only depends on the sampled habitat distance and direction between two sampled habitats and not on their georeferenced locations, or the second order stationary assumption. Hence, the expected difference between sampled values at two georeferenced *Cx. quinquefasciatus* habitat locations separated by distance h is zero. If $E(\mu(x)) = m$ then $z(x) = m + e(x)$ and $E(z(x) - z(x+h)) = 0$ (23). Covariance at h lag habitat distance then will be $E\{(z(x) - m)(z(x+h) - m)\} = E(z(x) \times z(x+h)) - m^2 = \text{Cov}(h)$, being that $z(x)$ the value at x sampled habitat site, $z(x+h)$ the value at x sampled habitat site plus h lag habitat distance is based on m representing the overall mean and $E()$ symbolizing the expected value (16). Although the covariance does not exist theoretically in a predictive vector mosquito habitat distribution model, the variance of the difference between two sampled habitat locations will be assumed to be present, as it does not depend on the georeferenced location and $\text{Var}(z(x+h) - z(x)) = E\{(z(x+h) - z(x))\} = 2\gamma(h)$, being $E(z(x) - z(x+h)) = 0$. That is, differences between sampled habitat sites are merely a displacement function between them (*i.e.*, the intrinsic hypothesis). The $\gamma(h)$ (*i.e.*, dissimilarity average function) will be the key tool for the structural interpretation of the predicted *Cx. quinquefasciatus* habitats as well as for uncertainty estimation parameters. Mathematically, the variogram is defined as $\gamma(h) = 0.5 \times \text{Var}(z(x+h) - z(x))$ (26). Since, the mean of $z(x+h) - z(x)$ is zero for intrinsic variables then the variance is just the mean square difference. Consequently, the variogram for a predictive vector habitat distribution model is as follows: $\gamma(h) = 0.5 \times E\{(z(x+h) - z(x))^2\}$ (16). These predictive interpolation vector mosquito habitat distribution models can measure unknown local mean habitat values as having a local linear or quadratic trend by combining regression of the dependent variable (*e.g.*, larval/pupal count data) based on auxiliary variables, such as terrain-related parameters (*i.e.*, slope), RS imagery, and thematic maps, with kriged residuals where auxiliary predictors are used directly to solve the weights. For instance, in our previous example of research conducted in Tuskegee, Alabama for examining vector-host activities of *Culex erraticus*, a kriged-based interpolator was also constructed in Geo-statistical

Analyst Extension of ArcGIS 9.2[®] for predicting prolific habitats based on adult abundance count data [16]. For total adult *Cx. erraticus* count, a first-order trend ordinary kriging process was fitted to the semivariogram at a partial sill of 5.764 km, nugget of 6.114 km, lag size of 7.472 km, with a range of 32.62 km using 12 lags, suggesting that this species forage several kilometers beyond their habitats. The values that the semivariogram model attains at the range (*i.e.*, the value on the y-axis) is called the sill, while the nugget or nugget variance, C_0 , occurs when a curve is fitted to the set of points that lie within the range and sill and the model used has a non-zero intercept [28].

Furthermore, to improve interpolation accuracy in a kriged-based algorithm, predictive mean standard error distributions can be quantified for attaining asymptotically optimal predictors. Traditional kriging is motivated by an expected squared prediction error from a stochastic model [28]. Presently many algorithmic paradigms of computational geometry have been applied efficiently in Geomatics and GIS using Dirichlet/Voronoi/Thiessen tessellation techniques for assessing interpolation models for quantifying the maximum error variance and the extended prediction variance [28]. A Voronoi diagram is a special kind of decomposition of a metric space determined by distances to a specified discrete set of objects in the space, (*e.g.*, dataset of *Culex* habitat ground reference coordinates). Generalization towards dynamic operations and kinematic motion have been investigated for about a decade using Dirichlet neighborhood analyses. Voronoi residuals are amenable to assessment of goodness-of-fit tests in a predictive vector mosquito habitat distribution models as they are approximately Gamma distributed and tend to be far less skewed than Pearson's residuals and other residual diagnostic tests [16]. Such tessellations have been used in grid generation in volumes and surfaces, data compression, image segmentation and edge detection, optimal allocation of resources, optimal quadrature rules, optimal representation, quantization, finite difference tests and finite volume schemes. Unfortunately, there is no published literature using interpolation algorithms or Dirichlet tessellations for quantifying spatial error and predicting habitat locations using georeferenced spatiotemporal-sampled predictor variables in WNV mosquito research for any country of the Caribbean.

In this research, we generated multiple predictive geospatial models using field and remote-sampled explanatory variables of *Cx. quinquefasciatus* habitats in Trinidad. Elucidation of variability in mosquito productivity requires spatial accounts of oviposition processes [29]. Indeed, this is important because of the recent identification of sero-positive horses and domestic birds in

Trinidad [2]. Presently there is no remote habitat-based intervention for early detection of WNV activity in any regions of the Caribbean or Central and South America. A remote habitat-based intervention may provide early detection of arboviral activity by providing useful information for projection of the potential course of WNV transmission and also other arthropod-borne viral diseases (e.g., St Louis encephalitis, Venezuelan equine encephalitis) later in the season while unraveling the environmental factors associated with the maintenance and establishment of enzootic cycles in Trinidad and in other countries of the neotropics.

For mosquito larval control to be effective, however, it may be crucial to quantify individual sampled habitat variation to maximize control efforts. For example, the importance of variation in mosquito production among breeding sites in design of control programs is now broadly accepted in suppressing domestic container-breeding *Aedes aegypti*, a mosquito vector of Dengue in Trinidad. This research involved quantifying population densities of *A. aegypti* in four towns using standard house-to-house inspections of all water-holding containers to determine whether persistently positive containers and premises existed over a three-month period in the wet season, from May to July 2002. From a total of 1,503 houses inspected, 15% were positive with 3% persistently positive over the three month period and classified as 'key premises'. From a total of 24,439 containers inspected, 1.3% were positive for *A. aegypti* larvae and pupae. Of 17 container types inspected, only water drums (54%), buckets (22%), tubs and basins (8%), water tanks (5%), brick holes (4%), and tires (2%) were significant larval habitats ($P < 0.001$) producers. The role that key premises play in the introduction and reinfestation of *A. aegypti*-free communities was determined. The results suggested that *A. aegypti* control programs could be more cost effective and sustainable by targeting efforts on key premises and key containers to control mosquito densities and dengue transmission while reducing manpower needs and insecticide use.

In this research, we constructed an autocovariate error matrix in SAS/GIS[®] for spatially targeting estimates generated from individual sampled *Cx. quinquefasciatus* habitat data. In our Bayesian model we did not assume that the georeferenced predictor variables were conditionally independent, as is commonly the case in hierarchical generalized linear model (HGLM) construction practices for predicting WNV mosquito habitats. In our Bayesian network we assumed that the variables were conditionally independent. In practice, this assumption may not hold for WNV mosquito habitat models and may give rise to incorrect inferences. Violation of the assumption of statistical independence will give rise to

incorrect model specifications [28]. Thus, we proposed the creation of a hidden node for modeling the dependency in the spatiotemporal-sampled parameters. In order to determine the conditional probability matrices for the hidden node, we used a gradient descent method. The objective function to be minimized was the squared-error between the measured and computed values of the instantiated nodes. Both forward and backward propagation were used to compute the node probabilities for quantifying the dependence in the data. We also restricted our attention to the simultaneous autoregressive Gaussian spatial process and the autoregressive Gaussian response model (*i.e.*, the spatial lag model), for quantifying latent autocorrelation error components in the Bayesian uncertainty estimates. We assumed that observed parameter error estimation patterns in the response variable decomposed into three statistically independent components: 1) a systematic spatial trend component that could be specified by a parsimonious set of exogenous variables; 2) a stochastic signal that reflects either an underlying spatial process and/or a set of missing exogenous factors with an inherent spatial pattern; and, 3) independent white-noise disturbances. A specific subset of eigenvectors was then used to determine local and regional variation in the field and remote-sampled covariates by capturing dependencies among disturbances generated from the Bayesian probabilistic model. Eigenvectors are a special set of vectors associated with a linear system of equations (*i.e.*, a matrix equation) [4,5,8,10].

We constructed our remote habitat-based intervention by adoption of a landscape approach to elucidate mechanisms underlying *Cx. quinquefasciatus* productivity. Our landscape perspective was based on multiple georeferenced explanatory variables of individual sampled habitats to determine their role in WNV disease transmission in Trinidad. Arguably, our framework evolved from the traditional concept of regression models generated from LULC covariates for identification of hot zones (*i.e.*, cluster), but our remote habitat-based intervention also focused on critical spatial autoregressive modifications of the georeferenced predictors. We developed our models to emphasize the role of oviposition foraging in *Cx. quinquefasciatus* in Trinidad. Habitat-based interventions should emphasize the link between foraging behaviors of egg-laying mosquitoes and the availability of breeding sites in evaluation of environmental management programs [29]. The basic assumption in our models was that heterogeneity in immature productivity was substantially large among *Cx. quinquefasciatus* habitats and that the number of productive habitats was relatively fewer than that of unproductive ones in focal areas. Untargeted interventions are inefficient because habitats are randomly

selected without consideration of the distribution of production (29). Spatially targeting prolific breeding habitats using a remote habitat-based intervention may reduce *Cx. quinquefasciatus* populations in Trinidad and help implement a cost-effective surveillance mechanism. Therefore, our objectives of this research were to: 1) construct a DEM to identify terrain-related explanatory covariates associated with spatiotemporal-sampled mosquito data; 2) generate regression models to determine statistical significance associated with the sampled covariates, 3) quantify latent autocorrelation components in model residual variance estimates generated from an eigenfunction decomposition spatial filter algorithm; and, 4) develop spatial interpolation models of potential prolific habitats site based on field-sampled count data. Although the discussion is centered on WNV vectors specifically in Trinidad, the framework and derived guidelines in this research may be applicable to integrated control programs for other mosquito species and insect-borne diseases in other regions.

2. Material and Methods

2.1. Study Site

The island of Trinidad is the most southerly part of the Lesser Antilles located at approximately 10° north latitude and 15 km from the eastern coast of Venezuela. The geology, flora and fauna are typical of the adjacent South American mainland [31,32]. Trinidad is approximately 80 km from north to south and is on average 60 km wide, roughly rectangular in shape with large promontories on the northwest and southwest of the country. The climate of Trinidad is tropical with a “rainy” season between May and November and a “dry” season from December to May. Details about climate, physical features, vegetation, and population size have been previously described [33,34].

2.2. Mosquito Egg Sampling

The *Cx. quinquefasciatus* habitat population was surveyed by counting the number of egg-rafts deposited in oviposition traps. Field sampling was conducted from July 2008 to July 2009. One hundred and fifty two temporary, permanent, and semi-permanent mosquito larval habitats in the study site were mapped and classified, and recalibrated using a CSI-Wireless differential corrected global positioning systems (DGPS) Max receiver which yields a positional error of .179 m (+/- 392 m) [35]. Water bodies were inspected for *Cx. quinquefasciatus* egg-rafts. Oviposition traps were located in shady areas around residences and municipal buildings, as well as the

edge of parks and woodlots.

2.3. Habitat Base Mapping

Initially, Landsat images were collected using SLC-off mode (http://edc.usgs.gov/#Find_Data/Products_and_Data_Available/ETM). We used the USGS Global Visualization Viewer (GloVis) system (<http://glovis.usgs.gov/>) to search for images for performing a land cover classification. Path 233/row 53 covered the Trinidad study area. Unfortunately, the images acquired from GloVis, were all cloudy, due to the climate conditions of the island. Cloudy image is an issue for remote classification and change detection for identification of mosquito habitats [24]. We did find three Landsat 7 passes of path 233/row 53 which were less than 10% cloud free. These collection periods were April 30, May 16, and June 1, 2005. These data; however, were not able to discriminate the land cover due to poor resolution (*i.e.*, 30m resolution) associated with the overlaid georeferenced *Cx. quinquefasciatus* habitats coordinates.

We acquired WV-1 data (www.digitalglobe.com) for the study site on July 2010. Operating at an altitude of 496 kilometers, WV-1 has an average revisit time of 1.7 days and is capable of collecting up to 750,000 km² per day of half-meter resolution imagery. The satellite is equipped with state-of-the-art geolocation accuracy capabilities and exhibits stunning agility with rapid targeting and efficient in-track stereo collection. Various collection sizes of surveying are available: frame mode, route survey (along coastline, roads and linear objects), areal survey (60 × 60 km), and stereoscopic areas on a single pass as well (www.digitalglobe.com).

In this research, the WV-1 imagery was radiometrically and sensor corrected but was not projected to a plane using a map projection or datum. The sensor correction blended the pixels from all the detectors into synthetic arrays to form the image of the study site. The geolocation accuracy specification of the data was 4.0 to 5.5 m CE 90 at nadir excluding terrain and off-nadir effects. The imagery was delivered at full-swath cut into 14 km lengths. The WV-1 imagery was classified using the Iterative Self-Organizing Data Analysis Technique (ISODATA) unsupervised routine in ERDAS *Imagine* v.8.7™ (Atlanta, USA) which is commonly used to classify land covers associated with disease foci [18,26,29]. The images were co-registered manually using ground control points (GCPs) and by applying a first order polynomial algorithm with a nearest neighbor re-sampling method. The Universal Transverse Mercator (UTM) Zone 20P datum WGS-84 projection was used for all of the spatial datasets.

Base maps were prepared from the field-sampled *Cx.*

quinquefasciatus habitat data and WV-1 imagery for the study site using the DGPS ground coordinates. We generated GIS data layers in the vector shapefile format, ArcInfo® Coverage format and raster GRID format for use in the GIS software package ArcGIS v 9.3®. We constructed a discrete tessellation of the region using a digitized grid-based algorithm (**Figure 1**). For remote identification of vector mosquito habitats the first step is often to construct a discrete tessellation of the region [36]. A unique identifier was placed in each grid cell. Each *Cx. quinquefasciatus* habitat/polygon was assigned a unique identifier. Field attribute tables were then associated to the polygons.

2.4. Environmental Data Analyses

Base maps were also generated using the spatiotemporal-sampled *Cx. quinquefasciatus* habitat data which was stored in a Vector Control Management System® (VCMS) (Clarke Mosquito Control Products, Inc. 159 N. Garden Avenue, Roselle, IL 60172) database. The VCMS supported the mobile field data acquisition in the study site through a PocketPC™. All two-way, remote synchronization of data, geocoding and spatial display were processed using the embedded GIS Interface Kit™ that was built using ESRI's MapObjects™ 2 technology. The VCMS database plotted and updated the DGPS ground coordinates of the *Cx. quinquefasciatus* habitats and supported exporting data in spatial format in which any combination of the sampled habitats and explanatory covariates were described in ESRI shapefile format for use in GIS. The database displayed this information onto a user-defined field base map.

2.4.1. Spatial Hydrological Model

The latest version of PCI Geomatics Orthoengine® software was used to generate a DEM using the WV-1 data and the georeferenced *Cx. quinquefasciatus* habitat covariates (**Figure 2**). OrthoEngine offers an industry-leading variety of control sources including, manual entry, geocoded imagery, geocoded vectors, chip database, digitizing tablet, or a text file (www.pcigeomatics.com). We used the DGPS coordinates to orient the images to our map coordinate system. We used PCI Geomatics OrthoEngine to extract the DEM. PCI software supports automatic GCP/tie point (TP) collection, using Toutins rigorous models, Rational Polynomial Coefficient (RPC) model construction, automatic DEM generation, orthorectification and automatic mosciaking (www.pcigeomatics.com).

WV-1 stereo pairs were supplied by the full scene of the Trinidad study site using Basic 1B level data de-

signed for the creation of the DEM. A set of WV-1 Basic 1B stereo images were provided by Digitalglobe. The set of stereo data was in Basic 1B OR₂A product format. Due to the 2 gigabyte limit of the TIFF data, WV-1 data was distributed in tile format. Each tile came with its own RPC file. Because our OR₂A format was map projected, the effects of any high frequency movements had already been removed from the scene resulting in a good fit for the RPCs.

In order to leverage the WV-1 image for attaining topographic indices from the *Cx. quinquefasciatus* data, a geometric model was required. We generated an RPC model. The RPC model has proven to be the most popular geometric model for high resolution images (www.digitalglobe.com). Since bias or errors may still have been present in the RPCs, our results were preprocessed with a polynomial adjustment using several accurate GCPs (**Table 1**). A zero order RPC adjustment is adequate to obtain an RPC model accuracy within 1m (www.digitalglobe.com).

To generate a DEM for determining the terrain-related *Cx. quinquefasciatus* habitat parameters in the Trinidad study site, a pair of quasi-epipolar images were generated from the fore and aft stereo images to retain elevation parallax in the X-direction only. An automated image match procedure was then used to produce the DEM through a comparison of the respective gray values of these images. To find the corresponding pixels in the left and right quasi-epipolar images, we used a hierarchical sub-pixel mean normalized cross-correction matching method. The actual matching method employed generated correlation coefficients between 0 and 1 for each matched pixel, where 0 represented a total mismatch and 1 represented a perfect match. A second order surface was then fitted around the maximum correlation coefficients. The difference in the georeferenced *Cx. quinquefasciatus* habitats between the images gave the disparity

Table 1. Rational polynomial coefficient results for the WV-1 digital elevation model.

Product	Number of GCPs	Number of CPs	CP RMS error		CP RMS error	
			X	Y	X	Y
Fore	0	19	0.4	.1	0.3	2.5
	1	17	0.4	0.5	0.6	0.7
Aft	0	16	0.4	1.3	1.0	2.2
	1	17	0.5	1.1	1.3	2.7

GCP: Ground Control Points, CP: Control Point, RMS: Root Mean Square

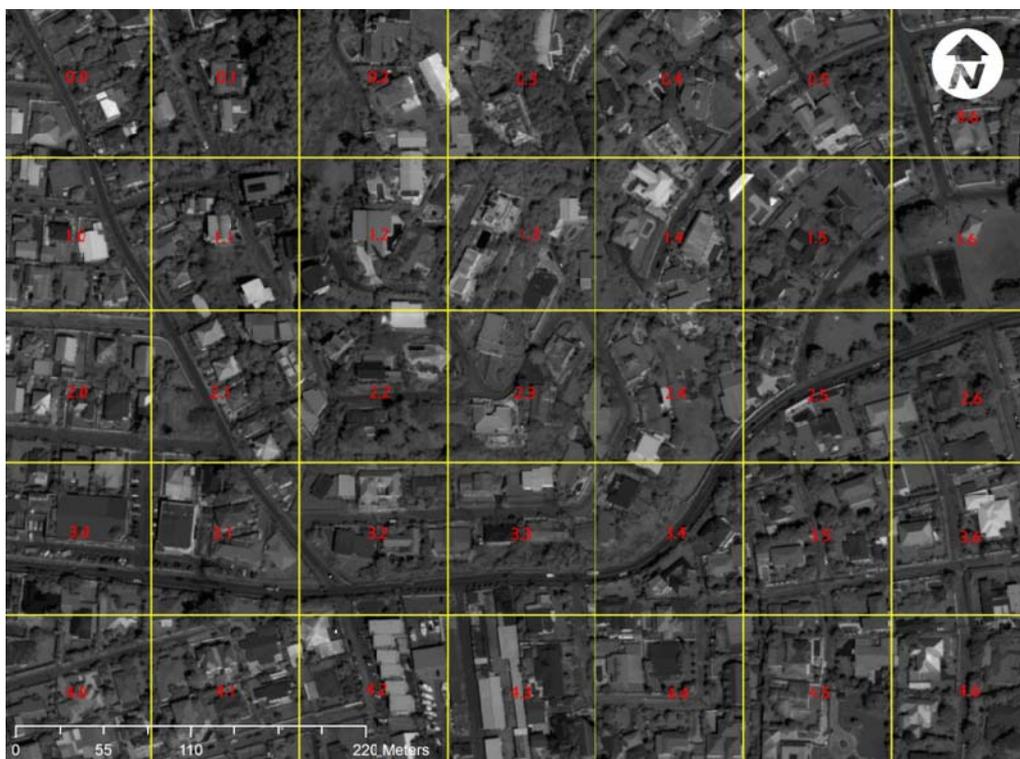


Figure 1. Gridded area of urban land cover in the Trinidad study site using WV-1 with georeferenced *Cx. quinquefasciatus* habitats.

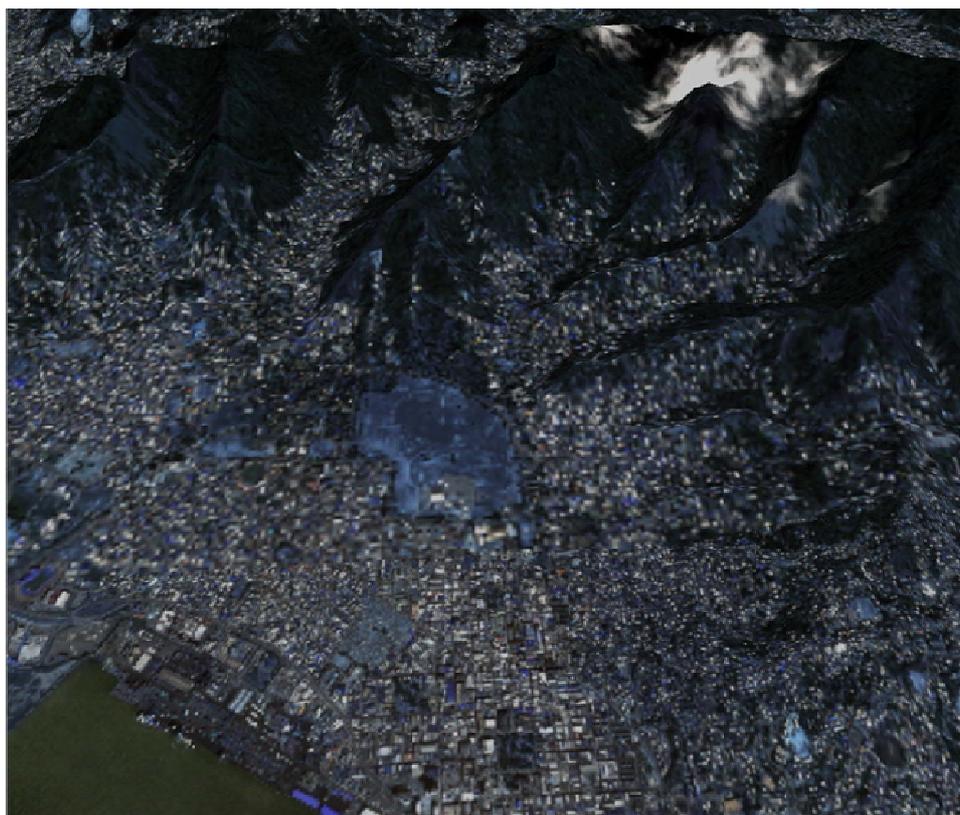


Figure 2. A digital elevation model (DEM) based on georeferenced *Cx. quinquefasciatus* habitats for the Trinidad study site.

or parrallex arising from terrain relief which was then converted to absolute elevation values above the WGS-84 absolute ellipsoid using a 3-d space interaction solution. Digital elevation model statistics were then generated employing the Pearson's correlation to find if an association existed between the georeferenced *Cx. quinquefasciatus* habitats and any of the sampled covariates (Table 2).

2.4.2. Regression Analyses

Variable selection for the multiple regression models was carried out by a combination of automatic (stepwise) procedures and goodness-of-fit criteria, and by selecting cluster covariate measurements that explained WNV prevalence. Routinely, the Analysis of Variance (ANOVA) are used to provide calculations about spatio-temporal sampled WNV habitat data regarding levels of variability within a regression model while forming a basis for tests of significance for the parameter estimates [18]. Analysis of variance is a collection of statistical models and their associated procedures in which the observed variance is partitioned into components due to different sources of variation.

In this research, the regression line concept we used was: $(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$, where the first term was the total variation in the response y (egg-raft count of *Cx. quinquefasciatus* habitats), the second term was the variation in mean response based on the sampled parameters and the third term was the residual value in the model estimates. Squaring each of these terms and adding over all of the sampled observations gave 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 then written as $SST = SSM + SSE$, where SS was notation for sum of squares and T, M, and E were the notation for total, model and error, respectively. The square of the sample correlation was equal to the ratio of the estimates while the sum of squares was related to the total sum of squares: $r^2 = SSM/SST$. This formalized the interpretation of r^2 as explaining the fraction of variability in the sampled egg-raft *Cx. quinquefasciatus* habitat population in the Trinidad study site explained by the regression-based parameters. The sample variance s_y^2 was equal to $\sum (y_i - \bar{y})^2 / (n - 1)$, which in turn was equal to the SST/DF (degrees of freedom), the total sum of squares divided by the total DF. A linear regression equation was constructed using the mean square model (MSM) = $\sum (\hat{y}_i - \bar{y})^2 / (l)$, which was equal to the SSM/DF . The corresponding mean square error (MSE) was $\sum (y_i - \hat{y}_i)^2 / (n - 2)$, which was also equal to SSE/DF and also the estimate of the variance about the regression line (i.e., σ^2). The MSE is an esti-

Table 2. Pearson correlation for *Cx. quinquefasciatus* aquatic habitat egg-raft count data and the sampled predictor variable elevation in the Trinidad study site.

Predictor variables	Statistical tests	Significance level	Elevation (m)
Total egg-raft count data	Pearson Correlation	1	-0.438
	Sig. (2-tailed)	<0.0001	<0.0001
	N	152	114

mate of σ^2 for determining whether or not the null hypothesis is true [25]. The null hypothesis for the ANOVA analyses was based on the average value of the dependent variable (i.e., egg-raft count of *Cx. quinquefasciatus*). The ANOVA calculations for multiple regression are nearly identical to the calculations for simple linear regression, except that the degrees of freedom are adjusted to reflect the number of explanatory variables included in the model [25].

For the spatiotemporal-sampled explanatory variables, (p) the DFM was equal to p and the error degrees of freedom (DFE), which was equal to $(n - p - 1)$ and the total degrees of freedom (DFT) which was equal to $(n - 1)$, the sum of DFM and DFE. The corresponding ANOVA table generated from the field and remote-sampled *Cx. quinquefasciatus* habitat parameters is shown below:

Source	Degrees of Freedom	Sum of squares	Mean Square
Model	p MSM/MSE	$\sum (\hat{y}_i - \bar{y})^2$	SSM/DFM
Error	$n - p - 1$	$\sum (y_i - \hat{y}_i)^2$	SSE/DFE
Total	$n - 1$	$\sum (y_i - \bar{y})^2$	SST/DFT

In the multiple regression analyses, the test statistic MSM/MSE had an F ($p, n - p - 1$) distribution. In this research, the null hypothesis stated that

$\beta_1 = \beta_2 = \dots = \beta_p = 0$, and the alternative hypothesis simply stated is that at least one of the sampled *Cx. quinquefasciatus* habitat parameters $\beta_j \neq 0, j = 1, 2, \dots, p$. Large values of the test statistic can provide evidence against the null hypothesis [25]. The F test did not indicate which of the parameters $\beta_j \neq 0$ was not equal to zero, only that at least one of them was linearly related to the response variable (i.e., total egg-raft count of *Cx. quinquefasciatus* habitat parameters). The ratio $SSM/SST = R^2$

(i.e., squared multiple correlation coefficient) was the proportion of the variation in the response variable that was explained by the field and remote-sampled *Cx. quinquefasciatus* habitat data. The square root of R^2 (i.e., the multiple correlation coefficient) was the correlation between the sampled observations y_i and the fitted values \hat{y}_i .

The relationship between WNV prevalence and each individual potential predictor variable sampled in the Trinidad study site was also investigated by single variable regression analysis using PROC MIXED. Since parasite prevalence data are binomial fractions, a regression model was used, as is standard practice for the analysis of the spatiotemporal-sampled WNV mosquito data [4]. Poisson regression analyses were also constructed in PROC MIXED to determine the relationship between *Cx. quinquefasciatus* habitat egg-raft count data and the sampled habitat characteristics. These regression models were generalized linear models with the logarithm as the (canonical) link function, and a Poisson distribution function.

The regression analyses assumed independent counts (i.e., n_i), taken at sampled habitats $I = 1, 2, \dots, n$. The field-sampled count data were described by a set of explanatory variables denoted by matrix \mathbf{X}_i , a $1 \times p$ vector of covariate measurements for a sampled habitat i . The expected value of these data was given by $\mu_i(\mathbf{X}_i) = n_i(\mathbf{X}_i) \exp(\mathbf{X}_i \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ was the vector of non-redundant parameters and the Poisson rates parameter was given by $\lambda_i(\mathbf{X}_i) = \mu_i(\mathbf{X}_i)/n_i(\mathbf{X}_i)$ [37]. The rates parameter $\lambda_i(\mathbf{X}_i)$ was both the mean and the variance of the Poisson distribution for each sampled *Cx. quinquefasciatus* habitat location i . The dependent variable was total field-sampled egg-raft count data. The Poisson regression model assumed that the data was equally dispersed, meaning that the conditional variance equaled the condition mean. The procedure used maximum likelihood estimation to find the regression coefficients. The data was log-transformed before analyses to normalize the distribution and minimize standard error. The predictor variables used in the regression analyses are listed in **Table 3**.

2.4.3. Bayesian Analyses

In the Bayes formulation, the specification of the *Cx. quinquefasciatus* habitat model was completed by assigning priors to all unknown parameters. We used our dataset of spatiotemporal-sampled *Cx. quinquefasciatus* observations $\mathbf{X} = [x_1, \dots, x_n]$ where each x_i for $i = 1 \dots n$ was assumed to be distributed according to some distribution $p(x_i|\theta)$. In this research, θ was a parameter that was unknown and had to be inferred from the data. Our Bayesian procedure began by assuming that θ was distributed according to some prior distribution $p(\theta|\alpha)$, where the parameter α was a hyperparameter.

Table 3. Ecological variables collected in the Trinidad study site.

Variable	Description	Units
EGG-RAFT COUNT	count data (dependent variable)	Number collected
STUDY SITE	Trinidad	10 × 10 m grid cells
DISHOUSE	Distance to house from sampled aquatic habitats	Meters
DHABITAT	Distance to house from sampled aquatic habitats	Meters
SHADE	Amount of shade	Percent of grid cell that is shaded (nearest 10%)
DRAINAGE	Drainage status of soils	0 = poorly drained, 1 = well drained
CANOPY	Canopy cover above larval habitat	Percent of grid cell with canopy cover (nearest 10%)
Elevation	The geographic height above a habitat reference point,	Meters

The joint of the *Cx. quinquefasciatus* data was then generated using:

$$p(\mathbf{X}|\theta) = p(x_1, \dots, x_n | \theta) = \prod_{i=1}^n p(x_i | \theta); \text{ whereby,}$$

$p(\mathbf{X}|\theta, a) = p(\mathbf{X}|\theta)$ and $p(x_i|\theta, a) = p(x_i|\theta)$ were conditionally independent of the hyperparameter. Bayesian inference then determined the posterior distribution of the parameter $p(\mathbf{X}|\theta, a)$ using :

$$\begin{aligned} p(\theta|\mathbf{X}, \alpha) &= \frac{p(\theta, \mathbf{X}, \alpha)}{p(\mathbf{X}, \alpha)} = \frac{p(\theta, \mathbf{X}, \alpha)}{\int_{\theta} p(\theta, \mathbf{X}, \alpha) d\theta} \\ &= \frac{p(\mathbf{X}|\theta, \alpha) p(\theta|\alpha)}{\int_{\theta} p(\mathbf{X}|\theta, \alpha) p(\theta|\alpha) d\theta} = \frac{p(\mathbf{X}|\theta) p(\theta|\alpha)}{\int_{\theta} p(\mathbf{X}|\theta) p(\theta|\alpha) d\theta} \\ &= \frac{\left[\prod_{i=1}^n p(x_i|\theta) \right] p(\theta|\alpha)}{\int_{\theta} \left[\prod_{i=1}^n p(x_i|\theta) \right] p(\theta|\alpha) d\theta} \end{aligned}$$

For the fixed regression parameters, a suitable choice was the diffuse prior, i.e., $p(y) \propto \text{const}$, but a weakly informative Gaussian prior was also possible. A second-order Gaussian random walk prior was used to allow enough flexibility while penalizing abrupt changes in the function [17]. The prior was expressed as:

$$p(f|\tau_f^2) \propto \exp\left(-\frac{\tau_f^2}{2} \sum_{t=3}^t (f_t - 2f_{t-1} + f_{t-2})^2\right) \text{ where}$$

$f = (f_1, \dots, f_p)$ and τ_f^2 was the variance, with diffuse priors $f_1 \propto \text{const}$, $f_2 \propto \text{const}$.

For additional spatial effects in the sampled *Cx. quinquefasciatus* data, smoothness priors were assigned whose joint distribution, s , was given by:

$$p(s|\tau_f^2)\alpha \exp\left(-\frac{\tau_f^2}{2}\sum_{t=12}^r (s_{t-1} + \dots + s_t)^2\right).$$

Again, assuming diffuses priors, the variance that controlled the degree of smoothness in the model was s_1, \dots, s_{11} , and τ_f^2 . The unstructured spatial heterogeneity term, u_i , was assumed to follow an exchangeable Gaussian prior with zero mean and variance, τ_u^2 , i.e., $u_i \sim N(0, \tau_u^2)$. A similar prior was then assigned to the heterogeneity term for the sampled predictor variables, i.e., $h_i \sim N(0, \tau_u^2)$.

Finally, for the spatial components, v_i , a Markov random field (MRF) prior was assigned. The conditional distribution of v_i , given an adjacent covariate, v_j , was a univariate normal distribution with mean equal the average v_j coefficient values of v_i 's neighboring sampled site and variance equal to τ_v^2 divided by the number of total adjacent sampled *Cx. quinquefasciatus* predictors. This lead to a joint density of the form:

$$p(v|\tau_v^2)\alpha \exp\left(-\frac{\tau_v^2}{2}\sum_{i \sim j} (v_i - v_j)^2\right) \text{ where } i \sim j \text{ denoted}$$

a sampled site i adjacent to j , and where the parameter estimates v_i and v_j in the adjacent sampled sites were similar. The degree of uncertainty in the spatiotemporal-sampled data was determined by the unknown precision parameter τ_v^2 .

By writing $f_j = z_j \beta_j$, $h = z_k \beta_k$, $u = z_l \beta_l$, and $v = z_m \beta_m$ for a well defined design matrix Z , a vector of regression parameters β , with all different priors, was expressed in a general Gaussian form using the expression:

$$p(\beta_j|\tau_j^2)\alpha \exp\left(\frac{1}{2\tau_j^2}\beta_j' K_j \beta_j\right) \text{ with an appropriate}$$

penalty matrix K_j . The model structure was dependent on the sampled explanatory variables and smoothness of the function. In most cases, K_j is rank deficient and, hence, the prior for β_j is improper [18]. For the variances τ_j^2 inverse Gamma priors $IG(a_j, b_j)$ was assumed, with hyperparameters a_j, b_j chosen such that this prior was weakly informative.

The Bayesian framework in this research was defined by conditional probabilities constructed from environmental-sampled *Cx. quinquefasciatus* habitat data. The observation nodes in the Bayesian estimation model were denoted by a vector $x = (x_1, x_2, \dots, x_N)$, and the set of states of the observation node x_j generated from the sampled data that was represented by $x_j \in \{1, 2, \dots, Y_j\}$. In this research the hidden nodes were denoted by

$z_k \in \{1, 2, \dots, T_k\}$. The probability that the state of the hidden node z_k was l , $1 \leq i \leq T_k$, was expressed as $\alpha_{(k,i)} := P(z_k = i)$. Because $\{\alpha_{(k,i)}, i = 1, 2, \dots, T_k\}$ is a probability distribution, $\sum_{i=1}^{T_k} \alpha_{(k,i)} = 1$, holds for $k = 1, 2, \dots, K$ [10]. The conditional probability in this research was that the j th observation node x_j was l , $1 \leq i \leq Y_j$, which was based on the condition that the states of hidden nodes were $z = (z_1, z_2, \dots, z_N)$, generated by $b_{(j,l|z)} = P(x_j = l)$. We defined $\alpha := \{\alpha_{(k,i)}\}$, $\beta := \{\beta_{(j,l|z)}\}$, and let $\omega = \{\alpha, \beta\}$ be the set of all the spatiotemporal-sampled *Cx. quinquefasciatus* habitat parameters in the study site. Then the joint probability that the states of observation nodes were

$x = (x_1, x_2, \dots, x_N)$ and the states of hidden nodes were $z = (z_1, z_2, \dots, z_N)$ based on

$$P(x, z|\omega) = \prod_{k=1}^K \alpha(k, z_k) \prod_{j=1}^N \beta(j, x_j | z).$$

The marginal probability that the states of observation nodes were x was generated using

$$P(x, z|\omega) = \left\{ \prod_{k=1}^K \sum_{z_k=1}^{T_k} \right\} \prod_{k=1}^K \alpha(k, z_k) \prod_{j=1}^N \beta(j, x_j | z),$$

and we used the notation

$$\sum_z = \left\{ \left\{ \prod_{k=1}^K \sum_{z_k=1}^{T_k} \right\} := \sum_{z_1=1}^{T_1} \sum_{z_2=1}^{T_2} \dots \sum_{z_K=1}^{T_K} \right\}$$

for the summation over all states of hidden nodes. We assumed the sampled *Cx. quinquefasciatus* habitat predictor variables $X_n = \{X_1, X_2, \dots, X_n\}$ were independently and identically taken from the true distribution $p_o(x)$. In Bayesian learning, the prior distribution $\varphi\omega$ on the parameter ω is set [18]. As such, the posterior distribution $p(\omega|X^n)$ was computed from the spatiotemporal-sampled *Cx. quinquefasciatus* habitat dataset and the prior by

$$p(\omega|X^n) = \frac{1}{Z(X^n)} \exp(-nH_n(\omega)) \varphi(\omega)$$

which was generated using the expression

$$H_1 n(\omega) = 1/n \sum_{i=1}^{\uparrow} (i=1) \uparrow n \equiv \log(p_o\{X_{\downarrow i} | \omega\}),$$

and $Z(X^n)$ (i.e., the normalization constant). The Bayesian predictive distribution $p(x|X^n)$ was provided by averaging the model over the posterior distribution as follows,

$$p\{x|X^n\} = \int \{p(x|\omega) p(\omega|x)\} X^n d\omega.$$

The Bayesian stochastic complexity $F(X^n)$ was defined by $F(X^n) = -\log Z(X^n)$ which was used as a criterion by which the model was selected and the hyperparameters in the prior were optimized. We let $E_{x^n[\square]}$ be the expectation over all the sampled *Cx. quinquefasciatus* parameters. The Bayesian stochastic complexity had the following asymptotic form using

$E_{\downarrow}(X^{\uparrow n})[F(X^{\uparrow n}) \approx \lambda \log n - (m-1) \log \log \log n + O(1)]$ where λ and m were the rational number and the natural number respectively, and the two were determined by the singularities of sampled predictor variables. In regular models, 2λ is equal to the number of parameters and $m = 1$, while in non-identifiable models, 2λ is not larger than the number of parameters and $m \geq 1$ [38]. However, Bayesian frameworks using field and remote-sampled vector mosquito data requires integration over the posterior distribution, which typically cannot be performed analytically [4].

In this research, we let $\{X^n, Z^n\}$ be the sampled *Cx. quinquefasciatus* parameter estimates corresponding to the hidden error variables in the equation $Z^n = \{Z_1, Z_2, \dots, Z_n\}$. The variational Bayesian framework approximated the Bayesian posterior $p(Z^n, \omega | X^n)$ of the hidden variables and the *Cx. quinquefasciatus* parameters using the variational posterior $q(Z^n, \omega | X^n)$, which was factorized using

$$q(Z^n, \omega | X^n) = Q(Z^n | X^n) r(\omega | X^n),$$

where $Q(Z^n | X^n)$ and $r(\omega | X^n)$ were posteriors based on the inconspicuous error variables and the sampled data respectively. The variational posterior $q(Z^n, \omega | X^n)$ was chosen to minimize the functional $\bar{F}[q]$ defined by

$$\bar{F}[q] = \sum_{Z^n} \int \frac{q(Z^n, \omega | X^n) \log(q(Z^n, \omega | X^n) p(\omega | X^n))}{p(X^n, Z^n, \omega)} d\omega,$$

which was then further defined by the equation $s = F(X^{\uparrow n}) + K(q(Z^{\uparrow n}, \omega | X^{\uparrow n}) \| p(Z^{\uparrow n}, \omega | X^{\uparrow n}))$ using the sampled parameters, where $K(q(Z^{\uparrow n}, \omega | X^{\uparrow n}) \| p(Z^{\uparrow n}, \omega | X^{\uparrow n}))$ and the true Bayesian posterior was $p(Z^n, \omega | X^n)$ and the variational posterior was $q(Z^n, \omega | X^n)$. This led to the functional $\bar{F}[q]$ being minimized under the constraint and then the variation posteriors, $r(\omega | X^n)$, and $Q(Z^n | X^n)$ which was computed using the equations

$$r(\omega | X^n) = \frac{1}{C_r} \varphi(\omega) \exp \langle \log p(X^n | Z^n | \omega) \rangle_Q \text{ and}$$

$$Q(Z^n | X^n) = \frac{1}{C_Q} \exp \langle \log p(X^n, Z^n | \omega) \rangle_r$$

where C_r and C_Q were the normalization constants. It is important to note that these equations gave only necessary conditions for the functional $\bar{F}[q]$ to be minimized in the *Cx. quinquefasciatus* habitat model. The variational posteriors were computed by an iterative algorithm. We defined the variational stochastic complexity $\bar{F}(X^n)$ by the minimum value of the functional

$\bar{F}[q]$ which was $\bar{F}(X^n) = \min_{r, Q} \bar{F}[q]$, based on the difference between $\bar{F}(X^n)$ and the Bayesian stochastic complexity $F(X^n)$.

We then generated variational posterior for the estimation matrix. We assumed that the prior distribution $\varphi(\omega)$ of the *Cx. quinquefasciatus* habitat parameters $\omega = \{a, b\}$ was the conjugate prior distribution. Then $\varphi(\omega)$ was generated by the equation

$$\varphi(a_k) = \frac{\Gamma(T_k \phi_0)}{\Gamma(\phi_0)^{T_k}} \prod_{z_k=1}^{T_k} a_{(k, z_k)}^{\phi_0 - 1}, k = 1, 2, \dots, K,$$

and covariates estimates were provided by

$$\varphi(b_{(j_n|z)}) = \frac{\Gamma(Y_j \xi_0)}{\Gamma(\xi_0)^{Y_j}} \prod_{x_j=1}^{Y_j} b_{(j_n|z)}^{\xi_0 - 1}, j = 1, 2, \dots, N,$$

which were Dirichlet distributions with hyperparameters generated using $\phi_0 > 0$ and $\xi_0 > 0$. The Dirichlet distribution [*i.e.*, $\text{Dir}(\alpha)$] is a family of continuous multivariate probability distributions parameterized by the vector α of positive reals which can generate the multivariate generalization of the beta distribution and conjugate prior of the categorical distribution and multinomial distribution in Bayesian statistics for predictive vector mosquito habitat models [4]. The Dirichlet distribution is the multinomial extension to the beta distribution for a binomial process which can also be used in quantifying probabilities in predictive vector mosquito habitat probability models [39].

We then let $\delta(n)$ be 1 when $n = 0$ and 1 otherwise, and then defined the sampled parameter uncertainty estimates using $n_{(k, z_k)}^z := \sum_{i=1}^n (\delta(Z_i^{(k)} - z_k))_Q$. In this research we also used

$$n_{(j, x_j|z)}^x := \sum_{i=1}^n (\delta(X_i^{(j)} - x_j)) \left(\prod_{k=1}^K \delta(Z_i^{(k)} - z_k) \right)_Q.$$

In these equations, $X_i^{(j)}$ was the state of the j th observation node and $Z_i^{(k)}$ was the state of the k th hidden node. The variational posterior distribution of parameters $\omega = \{a, b\}$ was given by using the equation

$$r(a_k | X^n) = \frac{\Gamma(n + T_k \phi_0)}{\prod_{z_k=1}^{T_k} \Gamma(n_{(k, z_k)}^z + \phi_0)} \prod_{z_k=1}^{T_k} a_{(k, z_k)}^{n_{(k, z_k)}^z + \phi_0 - 1}$$

where each of the spatiotemporal-sampled *Cx. Quinquefasciatus* habitat covariates were generated using the equation

$$n_{\downarrow} z^{\uparrow} x := \sum_{\downarrow} (i=1)^{\uparrow} n \equiv$$

$$\left(\prod_{\downarrow} (k=1) K \equiv (Z_{\downarrow} i^{\uparrow} ((k)) - Z_{\downarrow} k) \right)_{\downarrow} Q$$

It then followed that $\bar{n}_z^x = \sum_{x_j=1}^{y_j} \bar{n}_{(j,x_j|z)}^x$ for $j=1, \dots, N$ and $\bar{n}_{(k,z_k)}^z = \sum_{z=k} \bar{n}_z^x$ denoted the sum over $z_i (i \neq k)$.

In this research we evaluated the statistical efficiency of the MCMC sequence in WinBUGS[®] using the following steps:

- Step1: Sample $S^{(m)}$ from $f(S|X, Z^{(m-1)})$
- Step2: Sample $P^{(m)}$ from $f(P|X, S^{(m)}, Z^{(m-1)})$
- Step3: Sample $Z^{(m)}$ from $f(Z|X, P^{(m)}, S^{(m)}, \Phi^{(m-1)})$

where X was the field and remote-sampled *Cx. quinquefasciatus* habitat data and Z was randomly assigned a starting value $Z^{(1)}$ using a uniform prior distribution.

Step 1 was performed by drawing from an inverse Wishart distribution. In Bayesian statistics inverse Wishart distribution is used as the conjugate prior for the covariance matrix of a multivariate normal distribution [9]. The probability density function of the inverse Wishart was:

$$\frac{|\psi|^{m/2} |B|^{-(m+p+1)/2} e^{-\text{trace}(\psi B^{-1})/2}}{2^{\frac{mp}{2}} \tau_p\left(\frac{m}{2}\right)}$$

where B and ψ were positive definite matrices, and $\tau_p(\cdot)$ was the multivariate gamma function. In this research the distribution of the inverse of a Wishart-distributed matrix was generated from $A \sim W(\Sigma, m)$ and Σ which was of size of the matrix constructed from the field and remote-sampled *Cx. quinquefasciatus* habitat parameters. Under these circumstances $B = A^{-1}$ had an inverse Wishart distribution, $B \sim W^{-1}(\Sigma^{-1}, m)$. We calculated the probability density function:

$$p(B|\psi, m) = \frac{|\psi|^{m/2} |B|^{-\frac{m+p+1}{2}} \exp(-\text{tr} \psi B^{-1})/2}{2^{\frac{mp}{2}} \tau_p\left(\frac{m}{2}\right)}$$

where $\psi = \Sigma^{-1}$ and $\tau_p(\cdot)$ was the multivariate gamma function of the sampled data. A probability density function or density of a continuous random variable is a function that describes the relative likelihood for this random variable to occur at a given point [25].

The marginal and conditional distributions from the inverse Wishart-distributed matrix was quantified using $A \sim W^{-1}(\psi, m)$. We then partitioned the matrices for determining if ψ was conformable with each other using:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad \psi = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}$$

where A_{ij} and ψ_{ij} were $P_i \times P_j$ matrices. We then determined if:

- 1) A_{11} was independent of $A_{11}^{-1}A_{12}$ and $A_{22,1}$, when

$A_{22,1} = A_{22} - A_{21}A_{11}^{-1}A_{12}$ which was the Schur complement (*i.e.*, a submatrix within a larger matrix) of A_{11} in A ;

2) $A_{11} \sim W^{-1}(\psi_{11}, m - p_2)$;

3) $A_{11}^{-1}A_{12} | A_{22,1} \sim MN_{p_1 \times p_2}(\psi_{11}^{-1}\psi_{12}, A_{22,1} \otimes \psi_{11}^{-1})$,

where $MN_{pq}(\cdot, \cdot)$ was a matrix normal distribution generated from the spatiotemporal-sampled *Cx. quinquefasciatus* habitat parameters;

4) $A_{22,1} \sim W^{-1}(\psi_{11}, m)$.

We used the Conjugate distribution to make inference about a covariance matrix Σ whose prior $p(\Sigma)$ had a $W^{-1}(\psi_{11}, m)$ distribution. If the observations $X = x_1, \dots, x_n$ are independent p -variate Gaussian variables drawn from a $N(0, \Sigma)$ distribution, then the conditional distribution $p(\Sigma|X)$ has a $W^{-1}(A + \psi, n + m)$ distribution, where $A = XX^T$ is n times the sample covariance matrix [3]. Because the prior and posterior distributions are the same family, the inverse Wishart distribution was the conjugate to the multivariate Gaussian generated from the georeferenced *Cx. quinquefasciatus* habitat parameters.

Due to its conjugacy to the multivariate Gaussian it was possible to “integrate out” the Gaussian’s parameter Σ using:

$$P(X|\psi, m) = \int P(X|\Sigma)P(\Sigma|\psi, m)d\Sigma = \frac{|\psi|^{m/2} \tau_p\left(\frac{m+n}{2}\right)}{\pi^{\frac{np}{2}} |\psi + A|^{m+n} \tau_p\left(\frac{m}{2}\right)}$$

This task was simplified by assuming that the spatiotemporal-sampled *Cx. quinquefasciatus* habitat data analyses had covariance matrices with common eigenvectors. If covariance’s differ among sources, the inverse Wishart draws often produce invalid, especially for sources that are small components of the mixture [9]. In this research we developed a different approach, noting that a covariance matrix S can be decomposed into a vector of standard deviations, V , and a correlation matrix, R using: $S = \text{diag}(V)R\text{diag}(V)$ where $\text{diag}(V)$ was a matrix with diagonal elements V and zeros elsewhere. This decomposition permitted the independent sampling of V and R . We simulated the standard deviations, V , from an inverse gamma distribution:

$$p(V_{k,l}^2 | X, Z^{(m-1)}) \sim IG\left(\alpha + \frac{n_k}{2}, \beta + \frac{S_{k,l}^2 n_k}{2}\right), \text{ where } n_k$$

was the number of individual habitats assigned to cluster k by $Z^{(m-1)}$ (*i.e.*, an estimate of the *Cx. quinquefasciatus* habitat sample size), $s_{k,l}^2$ was the sample variance of

element l in cluster k as assigned by $Z^{(m-1)}$, and α and β were constants, both set to the non-informative prior value of 1. Quasi-likelihood techniques in a logistic regression equation and Bayesian prior distributions can enumerate intra-cluster correlations in urban vector mosquito aquatic habitats [4].

We simulated the elements of the correlation matrices, R , from a hyperbolic-tangent transformed distribution using :

$$p\left(\tanh\left(R_{k,i,j}\right)|X, Z^{(m-1)}\right) \sim N\left(\tanh\left(\hat{R}_{k,i,j}\right)\right), \frac{1}{n_k}$$

where $\hat{R}_{k,i,j}$ was the current estimate of the correlation between the i^{th} and j^{th} elements ($i \neq j$) in a cluster k given $Z^{(m-1)}$, and n_k was the same. After sampling values for both V and R , we reassembled the covariance matrix, S_k , for each *Cx. quinquefasciatus* habitat cluster, thus, completing Step 1.

Step 2 required drawing values for the elemental means, P . The field and remote-sampled data X had an approximate multivariate normal distribution as such Step 2 was performed using the vector of mean concentrations for cluster k . The multivariate normal distribution using the *Cx. quinquefasciatus* habitat parameters was given by the sample means calculated from X , generated from the cluster assignments, $Z^{(m-1)}$ and the covariance S_k from Step 1. If cluster k was empty at $m-1$ (that was, no individual sampled habitat parameters were assigned to k by $Z^{(m-1)}$), then the grand mean and covariance matrix of X were used as the sample mean and covariance matrix of k .

Step 3 involved drawing new cluster assignments using each individual sampled *Cx. quinquefasciatus* habitat in the Trinidad study site. To do so, it was necessary to calculate $\Pr(z_i = k)$ for every i, k combination (z_i was the i^{th} element of Z). Again we assumed multivariate normal distributions where $Z^{(m)}$ was simulated from:

$$\Pr\left(z^{i,(m)} = k | X, P^{(m)}, S^{(m)}, \phi\right) = \frac{\phi_k f\left(X^i | P_k^{(m)}, S_k^{(m)}, z^i = k\right)}{\sum_{k'=1}^K \phi_{k'} f\left(X^i | P_{k'}^{(m)}, S_{k'}^{(m)}, z^i = k'\right)}$$

and where the $f(\cdot)$ terms on the right-hand side were multivariate normal likelihoods. Thus, the likelihood that the i^{th} element of X was present in the *Cx. Quinquefasciatus* habitat population k , sampled in the study site normalized by the sum of likelihoods was quantified for all potential sources influencing presence of immatures. Finally, the sample $\Phi^{(m)}$ from $f(Z|X, P^{(m)}, S^{(m)}, Z^{(m)})$ was used to generate asymptotically efficient estimates from the spatiotemporal-sampled data.

2.6. Eigenvector Mapping

We conducted a residual analyses using a misspecification perspective for the error estimation models generated in SAS/GIS[®] assuming that the basic regression model $y = X\beta + \varepsilon^*$ had autocorrelated disturbances ε^* , which was decomposed into a white-noise component, ε , and a set of unspecified and/or misspecified models that had the structure $y = XB + \underbrace{E\gamma + \varepsilon}_{=\varepsilon^*}$. In this research,

the misspecification term was $E\gamma$. Quantification of topographic patterns generated from the distribution of the georeferenced *Cx. quinquefasciatus* habitat covariates was required to describe independent key dimensions and underlying spatial processes for defining a pattern in the misspecification term. This was accomplished by expanding the matrix term:

$(I - \rho V)^{-1} = \sum_{k=0}^{\infty} \rho^k V^k$, as an infinite power series, which was feasible under the assumption that the underlying spatial process in the data was stationary; note that in this research $\rho^0 V^0$ gave the identity matrix I . The simultaneous autoregressive error model was then rewritten as $y - \rho V y = X\beta - \rho V X\beta + \varepsilon$. Substituting these transformation gave:

$$\begin{aligned} y &= (I - \rho V)^{-1} [X\beta - \rho V(X\beta) + \varepsilon] \\ y &= \sum_{k=0}^{\infty} \rho^k V^k (X\beta - \rho V X\beta + \varepsilon) \\ y &= \sum_{k=0}^{\infty} \rho^k V^k X\beta - \sum_{k=0}^{\infty} \rho^{k+1} V^{k+1} (X\beta) + \sum_{k=0}^{\infty} \rho^k V^k \varepsilon \\ y &= X\beta + \underbrace{\sum_{k=1}^{\infty} \rho^k V^k X\beta - \sum_{k=1}^{\infty} \rho^k V^k (X\beta)}_{=0} + \sum_{k=0}^{\infty} \rho^k V^k \varepsilon \\ y &= X\beta + \underbrace{\sum_{k=1}^{\infty} \rho^k V^k \varepsilon}_{\text{misspecification-term}} + \varepsilon \end{aligned}$$

The misspecification term $\sum \rho^k V^k \varepsilon (k=1, \dots, \infty)$ remained uncorrelated with the exogenous variable, X , as the standard OLS assumption of the disturbances, ε , were uncorrelated with the predictor variables generated from the model. The spatial lag model was expressed as: $(I - \rho V)y = X\beta + \varepsilon$. Substituting the transformation gave:

$$\begin{aligned} y &= \sum_{k=0}^{\infty} \rho^k V^k (X\beta + \varepsilon) \\ y &= X\beta + \underbrace{\sum_{k=1}^{\infty} \rho^k V^k (X\beta + \varepsilon)}_{\text{misspecification-term}} + \varepsilon \end{aligned}$$

The misspecification term $\sum \rho^k V^k (X\beta + \varepsilon) (k=1, \dots, \infty)$ included the exogenous

variables X . Consequently, the exogenous variables were correlated with the misspecification term. Under this condition, standard OLS results for the basic regression model $y = X\beta + \varepsilon^*$, generated from the spatiotemporal-sampled covariates provided biased estimates $\hat{\beta}$ of the underlying regression parameters β .

The correlation, or lack thereof, between the exogenous variables and the misspecification terms of both habitat models were used to design spatial proxy variables, so that the properties of either model could be satisfied. We considered two different projection matrices. Different sets of eigenvectors were established using a spatial regression model. This model feature enabled us to also use the eigenvector spatial filtering approach for predictions of the endogenous variable y .

We selected an evaluation criterion using the standardized Moran's statistic. The standardized Moran's coefficient (MC) for regression residuals $\hat{\varepsilon}^*$ was defined as

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}, \text{ where } N$$

was the number of sampled habitats and their respective parameter estimates indexed by i and j respectively; X was the spatiotemporal-sampled egg-raft count data; \bar{X} was the mean of X ; and w_{ij} was a matrix of spatial weights. The expected value of Moran's I under hypothesis of no spatial autocorrelation was:

$$E(I) = \frac{-1}{N-1}. \text{ Its variance equaled:}$$

$$Var(I) = \frac{NS_4 - S_3S_5}{(N-1)(N-2)(N-3)(\sum_i \sum_j w_{ij})^2} \text{ where}$$

$$S_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2$$

$$S_2 = \frac{\sum_i (\sum_j w_{ij} + \sum_j w_{ji})^2}{1}$$

$$S_3 = \frac{N^{-1} \sum_i (x_i - \bar{x})^4}{(N^{-1} \sum_i (x_i - \bar{x})^2)^2}$$

$$S_4 = \frac{(N^2 - 3N + 3)S_1 - NS_2 + 3(\sum_i \sum_j w_{ij})^2}{1}$$

$$S_5 = S_1 - 2NS_1 + \frac{6(\sum_i \sum_j w_{ij})^2}{1}$$

In this research, negative (positive) values indicated negative (positive) spatial autocorrelation. Values ranged from -1 (indicating perfect dispersion) to +1 (perfect correlation). A zero values indicated a random spatial

pattern. For statistical hypothesis testing predictive WNV mosquito habitat modeling, MC values can be transformed to z -scores in which values greater than 1.96 or smaller than -1.96 indicate spatial autocorrelation that is significant at the 5% level [8,29].

Geary's Contiguity Ratio (GC) another measure of spatial autocorrelation was also generated. GC is inversely related to the Moran's statistic but it is not identical. Moran's indices are measure of global spatial autocorrelation, while Geary's C is more sensitive to local spatial autocorrelation in vector mosquito data analyses [16]. In this research, the sampled *Cx. quinquefasciatus* habitat data was quantified by GC which was defined as

$$C = \frac{(N-1) \sum_i \sum_j w_{ij} (X_i - X_j)^2}{2W \sum_i (X_i - \bar{X})^2} \text{ where } N \text{ was the}$$

number of the sampled habitats and explanatory parameters indexed by i and j respectively; X was the egg-raft count data; \bar{X} was the mean of X ; w_{ij} was a matrix of spatial weights; and W was the sum of all w_{ij} . The value of GC lies between 0 and 2. 1 means no spatial autocorrelation while 0 refers to a chaotic random distribution. Smaller (larger) than 1 means positive (negative) spatial autocorrelation [8].

In this research, distance between sampled habitats was defined in terms of an n -by- n geographic weights matrix, C , whose c_{ij} values were; 1 if the sampled *Cx. quinquefasciatus* egg-raft count locations i and j in the Trinidad study site were deemed nearby, and 0 otherwise. Adjusting this matrix by dividing each row entry by its row sum gave $C\mathbf{1}$, where $\mathbf{1}$ was an n -by- $\mathbf{1}$ vector of one's which converted this matrix to matrix \mathbf{W} . The resulting autoregressive model specification with no sampled covariates present (*i.e.*, the pure spatial autoregression specification) took on the following form: $\mathbf{Y} = \mu(1-\rho)\mathbf{1} + \rho\mathbf{W}\mathbf{Y} + \boldsymbol{\varepsilon}$ (2.1), where μ was the scalar conditional mean of \mathbf{Y} and $\boldsymbol{\varepsilon}$ was an n -by- $\mathbf{1}$ error vector residuals of the sampled data which in this research was represented as statistically independent and identically distributed (iid) normally random variates. The spatial covariance matrix for equation (2.1), using the sampled covariates, was

$$E[(\mathbf{Y}-\mu\mathbf{1})'(\mathbf{Y}-\mu\mathbf{1})] = \boldsymbol{\Sigma} = [(\mathbf{I}-\rho\mathbf{W}')(\mathbf{I}-\rho\mathbf{W})]^{-1} \sigma^2,$$

where $E(\cdot)$ indicated the calculus of expectations, \mathbf{I} was the n -by- n identity matrix denoting the matrix transpose operation, and σ^2 was the error variance.

Two different autoregressive parameters appeared in the spatial covariance matrix. When autocorrelation is present in residual data, a more explicit representation of both effects leads to a more accurate interpretation of empirical results [38]. Consequently, we used an SAR model specification

$$\Sigma = \left[(\mathbf{I} - \langle \rho \rangle_{\text{diag}} \mathbf{W}^n) (\mathbf{I} - \langle \rho \rangle_{\text{diag}} \mathbf{W}) \right]^{-1} \sigma^2, \quad (2.2)$$

where the diagonal matrix of autoregressive parameters, $\langle \rho \rangle_{\text{diag}}$ contained two sampled habitat parameters: ρ_+ displaying positive spatial dependency, and ρ_- for those sampled habitat pairs displaying negative spatial dependency based on the aggregation of the *Cx. quinquefasciatus* egg-raft count data. For example, by letting $\sigma^2 = 1$ and employing a 2-by-2 regular square tessellation,

$$\Sigma = \left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \rho_+ & 0 & 0 & 0 \\ 0 & \rho_+ & 0 & 0 \\ 0 & 0 & \rho_- & 0 \\ 0 & 0 & 0 & \rho_- \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \right]^{-2}$$

for the vector $\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}$ enabled positing a positive rela-

tionship between the sampled egg-raft count data and the explanatory covariates, y_1 and y_2 , a negative relationship between covariates, y_3 and y_4 , and, no relationship between covariates y_1 and y_3 and between y_2 and y_4 . This covariance yielded:

$$\mathbf{Y} = \mu (\mathbf{I} - \rho_+ \langle \mathbf{I}_+ \rangle_{\text{diag}} - \rho_- \langle \mathbf{I}_- \rangle_{\text{diag}}) \mathbf{I} + (\rho_+ \langle \mathbf{I}_+ \rangle_{\text{diag}} + \rho_- \langle \mathbf{I}_- \rangle_{\text{diag}}) \mathbf{W} \mathbf{Y} + \boldsymbol{\varepsilon}$$

where \mathbf{I}_+ was a binary 0-1 indicator variable which denoted those parameter estimates which displayed positive spatial dependency, and \mathbf{I}_- was a binary 0-1 indicator variable denoting those sampled covariates displaying negative spatial dependency, where $\mathbf{I}_+ + \mathbf{I}_- = \mathbf{1}$. Expressing the preceding 2-by-2 example in terms of an equation yielded:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \mu \left[\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \rho_+ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} - \rho_- \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \rho_+ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \rho_- \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}$$

if either $\rho_+ = 0$ then \mathbf{I}_+ was $\mathbf{0}$ and $\mathbf{I}_- = \mathbf{I}$ or if $\rho_- = 0$

then $\mathbf{I}_- = \mathbf{0}$ and $\mathbf{I}_+ = \mathbf{I}$.

In this research, this spatial structuring was achieved by constructing a linear combination of a subset of the eigenvectors of a modified geographic weights matrix, using $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ that appeared in the numerator of the MC. Spatial filtering furnishes an approach that incorporates spatial structuring into an intercept term that accounts for latent autocorrelation [38]. A subset of eigenvectors was selected with a stepwise regression procedure. Because $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n) = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}'$, where \mathbf{E} was an n -by- n matrix of eigenvectors and $\boldsymbol{\Lambda}$ was an n -by- n diagonal matrix of the corresponding eigenvalues, the resulting egg-raft count model specification was given by: $\mathbf{Y} = \mu\mathbf{1} + \mathbf{E}_k\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where μ was the scalar mean of \mathbf{Y} , \mathbf{E}_k was an n -by- k matrix containing the subset of $k \ll n$ eigenvectors selected with a stepwise regression technique and $\boldsymbol{\beta}$ was a k -by-1 vector of regression coefficients.

A number of the eigenvectors were extracted from the matrix $(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)\mathbf{C}(\mathbf{I} - \mathbf{1}\mathbf{1}'/n)$ which were affiliated with geographic patterns of the sampled *Cx. quinquefasciatus* habitat covariates portraying a negligible degree of spatial autocorrelation. Consequently, only k of the n eigenvectors were of interest for generating a candidate set for a stepwise regression procedure. One way of constructing a candidate set is to include an eigenvector in the set only when the vector's corresponding MC represents a minimum degree of spatial autocorrelation [38]. In this research, we used a spatial statistical threshold of $|\text{MC}/\text{MC}_{\text{max}}| > 0.25$ where MC_{max} was the maximum MC for a given surface partitioning, as then each candidate eigenvector represented a level of autocorrelation tending to account for the redundant information in the dataset. As redundant information spatial autocorrelation represents pseudo-replicated data linking it to missing values estimation and interpolation, as well as to notions of effective sample size and degrees of freedom [38].

2.6.1. Spatial Analyses

In this research, spatial linear prediction of the prolific *Cx. quinquefasciatus* habitat and its spatiotemporal sampled covariates was performed using an Ordinary kriged-based interpolator in ArcGIS 9.3[®] using Geostatistical Analyst. The Semivariogram Properties dialog box has several models to choose from. We set the Kriging method to Ordinary. A default value for lag size was initially set to the default output cell size. For Major range, Partial sill, and Nugget, a default value was calculated internally. The length of the longer axis to reach the sill is called the major range, and the length of the shorter axis to reach the sill is called the minor range; the angle of rotation of the line formed the major range [28]. The optional output variance of prediction raster contained the kriging variance

at each output raster cell. Assuming the kriging errors are normally distributed in a predictive vector mosquito habitat distribution model there is a 95.5 percent probability that the actual z-value at the cell is the predicted raster value, plus or minus two times the square root of the value in the prediction raster [16]. Low values within the output variance of prediction raster indicated a high degree of confidence in the predicted estimates. High values in interpolation based algorithms may indicate a need for more data points (26).

Geostatistical techniques were used to interpolate the value $Z(x_0)$ of a prolific *Cx. quinquefasciatus* habitat site, $Z(x)$ at an unobserved sampling site, x_0 , where z_i was $Z(x_i)$, $i = 1 \dots n$ based on the dimensions of the sampled habitat locations [*i.e.*, (x_1, \dots, x_n)]. The Ordinary interpolator also computed the best linear unbiased estimator, $\hat{Z}(x_0)$ of $Z(x_0)$, for the *Cx. quinquefasciatus* data based on a stochastic model quantified by the variogram $\gamma(x, y)$, by expectation $\mu(x) = E[Z(x)]$, and by the covariance function $c(x,y)$ of the random field. In this research, the kriging estimator was given by a linear combination of the sampled field and remote parameters using $\hat{Z}(x_0) = \sum_{i=1}^n w_i(x_0) Z(x_i)$; whereby, $z_i = Z(x_i)$ was the weights $w_i(x_0)$, and $i = 1 \dots n$ was the variance used to minimize any biased condition. The dependent variable was the spatiotemporal-sampled *Cx. quinquefasciatus* egg-raft count data, which were transformed to fulfill the diagnostic normality test prior to performing the kriging. The kriging weights of the algorithm were used to fulfill the unbiasedness condition in the spatial interpolation of the ecological-dependent variables using: $\sum_{i=1}^n \lambda_i =$ which was given by the equation system:

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ 1 \end{pmatrix} = \begin{pmatrix} \gamma(x_1, x_1) & \dots & \gamma(x_1, x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n, x_1) & \dots & \gamma(x_n, x_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1, x^*) \\ \vdots \\ \gamma(x_n, x^*) \\ 1 \end{pmatrix}$$

The Ordinary kriging was given by:

$$\text{var}(\hat{Z}(x^*) - Z(x^*)) =$$

$$\begin{pmatrix} \gamma(x_1, x^*) \\ \vdots \\ \gamma(x_n, x^*) \\ 1 \end{pmatrix} \begin{pmatrix} \gamma(x_1, x_1) & \dots & \gamma(x_1, x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n, x_1) & \dots & \gamma(x_n, x_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma(x_1, x^*) \\ \vdots \\ \gamma(x_n, x^*) \\ 1 \end{pmatrix}$$

The semivariogram generated described the dependence, between the *Cx. quinquefasciatus* parameters as a function of the distance between the sample sites. This allowed for abundance estimations at any point in the Trinidad study site. The value of prevalence, Z , at the coordinate (x_0, y_0) was estimated from the n nearest habitat sampling values using $Z_{obs}(x_1, y_1)$, $Z_{obs}(x_2, y_2)$.

$Z_{obs}(x_n, y_n)$ and the linear formula which generated:

$$\hat{Z}(x_0, y_0) = \sum_{i=1}^n a_i Z_{obs}(x_i, y_i)$$

The a_i was found by the Lagrange multiplier λ and solving the system using:

$$\sum_{i=1}^n a_i \gamma(h_{i,j}) + \lambda = \gamma(h_{i,0}), j = 1 \dots n \text{ under the constraint.}$$

$\sum_{i=1}^n a_i = 1$, where $h_{i,j}$ denoted the distance between any two sampled *Cx. quinquefasciatus* habitat locations in the Trinidad study site at (x_i, y_i) and (x_j, y_j) , and $h_{i,0}$ was the distance between the two habitats (x_0, y_0) . In the field of calculus of variations in mathematics, the method of Lagrange multipliers can be used to solve certain infinite-dimensional constrained optimization problems [26]. In this research the semivariance was defined as $\gamma(h)$. The magnitude of the semivariance was dependent on the distance between sampling sites. Semivariance of the deviance residuals of the *Cx. quinquefasciatus* data was calculated and a variogram was constructed to determine if there was evidence of residual spatial correlation in the data. The plot of the semivariances as a function of distance from a point is referred to as a semivariogram [26].

In this research, parameters of a fitted mathematical function (*i.e.*, the semiovariogram model) included generating a range, a nugget and a sill. Since our data was spatiotemporal, the square of the difference between expected values at habitat points had to be added: $2\gamma(x,y) = C(x,x) + C(y,y) - 2C(x,y) - (E(Z(x)) - E(Z(y)))^2$. If the covariance function of a stationary process exists, it is related to variogram by $2\gamma(x,y) = C(x,x) + C(y,y) - 2C(x,y)$ [26]. We checked for spatial dependence also in the sampled *Cx. quinquefasciatus* habitat data. We then determined the best fit variogram equation based on the number of lags (*i.e.*, bins) and the lag distance (the minimum value of h) used in the calculation of experimental variogram equation. Lag size is the width (distance) of the bins into which these vectors are grouped (www.esri.com). In this research the lag size was dependent upon the minimum and maximum distances between sampled *Cx. quinquefasciatus* habitats in the study site. In an effort to uncover our variogram's structure, similar lags were grouped together (*i.e.*, pairs of habitat that points aligned in roughly the same direction and roughly the same distance from each other) into the bins.

The semivariogram nugget coefficient allowed an interpretation of the required scale for defining spatial variability of the *Cx. quinquefasciatus* habitat data for providing minimized unbiased prediction error for residuals on the logit scale. Standard error in the models was calculated. Variations of kriging can produce noise-free predictions (www.esri.com). The kriging interpolator produced a map that was smooth and free of "jumps" at the sampled *Cx. quinquefasciatus* habitat locations. The algorithms incorporated in Geostatistical Analyst

provided exact filtered value predictions at the sampled habitat locations. This prevented discontinuities in the habitat predictions and their associated standard errors. A simple quantitative measure of the interpolation was determined by using a root mean square error (RMSE) for the models.

The kriging procedure returned the observed values at the sampled *Cx. quinquefasciatus* habitat locations and their interpolated values using the intensity and shape of the variograms. Using a neighborhood and/or distance search radius provided the standard errors of the interpolated values. Prediction errors have often been used to optimize sampling design by identifying areas where sampling effort should be increased or decreased [26]. In this research, prediction errors were generated as a function of the variogram models. A simple quantitative measure of the uncertainty in the interpolation model was determined by generating the RMSE values for the models.

Because we were working in two-dimensional space, the semivariogram and covariance functions changed not only with distance but with direction (*i.e.*, anisotropy). We quantified sampled habitats points and the vector that separated them. This vector had a distance on the x -coordinate as well as the y -coordinate. Alternatively, the vector had a distance and an angle in polar coordinates. Anisotropy was then described for the semivariogram. The isotropic model was the same in all directions; whereas, the anisotropic model reached the sill more rapidly in some directions than others.

In Geostatistical Analyst, an outline of the range was given over the empirical semivariogram surface. Having created a covariance function with which to form the kriging weights matrix, the next task was to tessellate the region into patches. Voronoi maps were used to explore the distribution of the sampled *Cx. quinquefasciatus* habitat data for global and local outliers and to quantify the covariation among the spatiotemporal-sampled data. Voronoi diagram is a special kind of decomposition of a metric space determined by distances to a specified discrete set of objects in the space [28]. Voronoi polygons were also used for interpolating neighborhoods and patches.

A Voronoi diagram was constructed with the sample points as the centers of the polygons using the Weighted Voronoi Diagram Extension in ArcGIS 9.3[®]. Using the Generate tab, we generated a weighted Voronoi diagram from the geo-sampled point features. The Graphical User Interface (GUI) has two tabs: Generate and Update (www.esri.com). The Voronoi Diagram was generated whereby a sampled *Cx. quinquefasciatus* habitat was associated with p_i (spatiotemporal-sampled covariate); whereby, $P = \{p_1, \dots, p_n\}$, where $2 \leq n \leq \infty$ and $x_i \neq x_j$ for $i \neq j, i, j \in I_n$. The region was given by $V(p_i) = \{x: \|$

$x - x_i\| \leq \|x - x_j\|$ for $j \neq i, i \in I_n\}$ was the ordinary Voronoi polygon associated with p_i , or the Voronoi polygon of p_i and the set given by $V = \{V(p_1), \dots, V(p_n)\}$ was the planar ordinary Voronoi diagram generated by P or simply Voronoi diagram of P . We also defined a planar ordinary Voronoi diagram with half planes as follows where we let $P = \{p_1, \dots, p_n\} \subseteq R^2$, where $2 \leq n \leq \infty$ and $x_i \neq x_j$ for $i \neq j, i, j \in I_n$. We called the region $V(p_i) = \bigcap_{j \in I_n(i)} H(p_i, p_j)$ which was the ordinary Voronoi polygon associated with p_i and set $V(P) = \{V(p_1), \dots, V(p_n)\}$ the planar ordinary Voronoi diagram generated by P .

A raster image showing normal Euclidean distance and adjusted Euclidean distance was generated, as well as a Voronoi polygon shapefile. The *Cx. quinquefasciatus* habitat point attributes were then transferred to Voronoi polygons automatically by appending the spatial attributes of one layer to another. Once a distance raster is generated, the user can use new points to update the distance raster and create updated Voronoi polygons [28].

The Voronoi tessellation produced a set of polygons V_i with area f_i ($i=1, \dots, n$). The (unknown) global mean z_D was estimated by a weighted mean of the spatiotemporal-sampled *Cx. quinquefasciatus* habitat values using: $m_g = \sum_{i=1}^n f_i z_i / \sum_{i=1}^n f_i$. The extension error of each polygon was calculated by using a discrete version of $\sigma_E^2 = 2\bar{\gamma}(V, v) - \bar{\gamma}(V, V) - \bar{\gamma}(v, v)$. If the elementary error terms are uncorrelated, errors can be combined to an estimate of the global error: $\sigma_g^2 = \frac{\sum_{i=1}^n \sigma_{E_i}^2 f_i^2}{\sum_{i=1}^n f_i}$ where

the summation is over all polygons V_i [28]. By convention we assumed a normal probability distribution for the global error $z_D - m_g$ and achieved a 90% confidence limits for $Z_D, m_g - 1.65 * \sigma_g \leq z_D \leq m_g + 1.65 * \sigma_g$. The global estimation error decreased with an increasing number of samples although some local deviation from this tendency did occur due to large polygons. As the number of samples increased the global estimation error converged to zero.

3. Results

The ANOVA used in data analyses tested the null hypothesis that the sampled parameters of the immature *Cx. quinquefasciatus* population in the Trinidad study site means were equal (*i.e.*, $H_0: \mu_1 = \mu_2 = \dots = \mu_a$), by comparing two estimates of variance. If the null hypothesis is false, then Mean Square Between (MSB) estimates generated from a predictive vector mosquito habitat distribution model is something larger than σ^2 [25]. In this research the MSE was an estimate of variance for determining whether or not the null hypothesis was true,

while the second estimate MSB was based on the variance of the sample means. We tested the null hypothesis as follows: if the null hypothesis was true, then MSE and MSB would be about the same since they are both estimates of the same quantity (*i.e.*, σ^2); however, if the null hypothesis was false then MSB would have been expected to be larger than MSE, since MSB is estimating a quantity larger than σ^2 . The ANOVA model encompassed all possible sources of variation in sampled *Cx. quinquefasciatus* habitat data which allowed further testing of our research hypotheses. In this research the significance test of spatiotemporal-sampled data involved the ratio of MSB to MSE: $F = \text{MSB}/\text{MSE}$. The F-statistic was used to calculate the *p*-value to determine deviations from normality. If the kurtosis is greater than 0, then the *F* tends to be too small and we cannot reject the null hypothesis even though it is incorrect; the opposite is the case when the kurtosis is less than 3 [25]. The skewness of the distribution in the *Cx. quinquefasciatus* data did not have a sizable effect on the *F* statistic. If the *n* per cell is fairly large, then deviations from normality do not matter much at all because of the central limit theorem, which implicitly states the sampling distribution of the mean approximates the normal distribution, regardless of the distribution of the variable in the population [25]. The *P* value was the estimated probability of rejecting the null hypothesis (H_0). Traditionally for WNV mosquito models, if $p < .05$ the null hypothesis is rejected [18]. If the null hypothesis is true in the estimation of a model then the *F* ratio should be approximately one, since MSB and MSE should be about the same [25]. In order to conduct a significance test using the prolific spatiotemporal-sampled *Cx. quinquefasciatus* habitats predictor variables, it was thus necessary to know the sampling distribution of *F*.

Additionally, from the sampling distribution, generated from the sampled *t* parameters, the probability of obtaining an *F* as large or larger than the one was also calculated. When there are only two means to compare, the *t*-test and the *F*-test are equivalent; the relation between ANOVA and *t* is given by $F = t^2$ [25]. If the probability values were lower than the significance level, then the null hypothesis was rejected. Significant differences by ANOVA were noted for mean numbers of *Cx. quinquefasciatus* captured throughout the sampling frame ($F = 42.2$, degrees of freedom [df] = 1).

The Poisson regression modeled the spatiotemporal-sampled *Cx. quinquefasciatus* habitat count data and contingency tables. Poisson regression assumed the response variable *Y* had a Poisson distribution and assumed the logarithm of its expected value can be modeled by a linear combination of the spatiotemporal-sampled *Cx. quinquefasciatus* parameters. A Poisson

regression model is sometimes known as a log-linear model, especially when used to model contingency tables [25]. Our model with a single independent variable *x*, took on the form: $\log(E(Y|x)) = \alpha + bx$. If Y_i are independent observations with corresponding values x_i of the sampled predictor variable, then *a* and *b* can be estimated by maximum likelihood if the number of distinct *x* values is at least 2 [25]. In this research the maximum-likelihood estimates lacked a closed-form expression and had to be quantified using numerical methods. The probability surface for maximum-likelihood Poisson regression was convex, making Newton-Raphson or other gradient-based methods appropriate estimation techniques for data analyses of the spatiotemporal-sampled *Cx. quinquefasciatus* habitat data. Linearized iteration scheme converges quickly and accurately using inversion equations solved through the Newton-Raphson method as each iteration, requires solving the finite difference equations and the linear adjoint equations only once, respectively [25].

OrthoEngine, PCI Geomatics' Geocorrection and Extraction tools generated a DEM of the Trinidad study site. OrthoEngine included a number of automation tools to improve the efficiency of georeferencing the sampled *Cx. quinquefasciatus* data including Automatic GCP collection from Chip Databases, and Automatic TP measurement. We stitched the image files. Stitching operation merged all the raster tiles and recalculated the image RPC from individual RPCs. Stitching automatically revealed the final RPCs as a binary segment in an output PIX file. Coefficients were reviewed using 'Project Report Option' in 'Reports' procedure step. We then selected geometric Model under the image information panel. The final step was the 'Schedule Ortho Generation.' We proceeded to the 'Orthogeneration' processing step and selected the files to be processed. We selected an appropriate DEM file and set other processing options before generating the final orthorectified image. This allowed for the precise orthorectification of the WV-1 imagery. The WV-1 over the Trinidad study site had look angles of approximately south at 30 degrees-off nadir and vertical. For simple corrections, polynomial, thin plate spline and rational functions were used. This allowed using image correlation from the stereo images. Once the images were geometrically corrected, they were mosaicked into a seamless image database, to form a conventional image map product and a GIS base layer for derivation of vector hypsographic and hydrographic covariates associated to the sampled *Cx. quinquefasciatus* habitat data. The range of the elevation in the DEM had a minimum value of 0 m with a maximum value of 425 m.

The spatial autocorrelation analysis results are re-

ported in **Table 4**. Eigenvectors were extracted from the matrix $(I - 11'/n) C (I - 11'/n)$ using the ecological-sampled predictor variables. Estimation results for these models appear in **Table 5**. Positive and negative spatial autocorrelation spatial filter component pseudo- R^2 values are reported. Positive and negative spatial autocorrelation spatial filter component pseudo- R^2 values are reported. These values did not exactly sum for the complete spatial filter; however, they are very close to their corresponding totals, suggesting that any induced multicollinearity was quite small.

To summarize **Table 5** results, the Trinidad study site had a random effects intercept with a near-zero mean and only trace spatial autocorrelation. The random effects term accounted for two-thirds or more of the variation in the spatiotemporal-sampled *Cx. quinquefasciatus* habitat count data. In addition, the random effects term resulted in the loss of a counterbalancing spatial autocorrelation component in the spatial filters. The spatial autocorrelation components revealed 17% redundant information in the *Cx. quinquefasciatus* habitat datasets.

Table 6 lists the improvements of fit for all model specifications and random error in the spatial analyses. The unadjusted model compared the univariate model to a model containing only the intercept term. Improvement of fit was also calculated for the first-order interaction models to determine whether including significant interactions improved fit compared to the full main effects model.

Table 7 presents the results of the regression for the interactions model. These results provided information for estimates of the prior distribution of main effect coefficients in the Bayesian analysis. The values for parameter estimates and standard were used as mean values and standard errors to parameterized prior expected values for quantifying the field and remote-sampled *Cx. quinquefasciatus* habitat explanatory variables. The prior expected mean value for the error term was assumed to be zero ('0'), with a standard deviation of 0.01. Initial values for the MCMC chains were generated.

The difference in the deviances between a simple model and the more complex model provided the improvement χ^2 values listed in **Table 8**. We examined all interaction between the sampled predictor variables and found that an interaction model did not improve the fit; therefore, no interaction terms were included in the final model. We could not examine the improvement of fit between a saturated model and the full effects model, as the number of the sampled parameters that needed to be estimated exceeded the maximum number that could estimate.

To derive the improvement of fit values listed in **Table 9**, the posterior mean deviance values were obtained with

Deviance Information Criterion (DIC) spatial analytical tools. Factors that did not improve fit were omitted from the final model. DIC tools were used to obtain mean posterior deviance values to construct improvement of fit tables and DIC statistics for identification of the best fitting model. In this research, this deviance was defined

Table 4. Global spatial analyses of *Cx. quinquefasciatus* egg-raft count data in the Trinidad study site.

Study Site	n	Transformation	MI	GR
Trinidad	152	LN(count + 1)	0.069	0.898

Table 5. Poisson spatial filtering model results for the sampled *Cx. quinquefasciatus* egg-raft count data in the Trinidad study site.

Spatial Statistics	Trinidad
SF: # of eigenvectors	5
SF: MC	0.654
SF: GR	0.402
SF pseudo- R^2	0.198
Positive SA SF: # of eigenvectors	4
Positive SA SF: MC	0.567
Positive SA SF: GR	0.484
Positive SA SF pseudo- R^2	0.150
Negative SA SF: # of eigenvectors	1
Negative SA SF: MC	-0.369
Negative SA SF: GR	1.291
Negative SA SF pseudo- R^2	0.048
Deviance statistic	1.168

Table 6. Poisson SF GLMM random effects results for the *Cx. quinquefasciatus* egg-raft count data in the Trinidad study site.

Statistics	Trinidad
Mean	0.056
Standard deviation	0.492
P(S-W)	0.0002
MC	-0.054
GR	1.028
Pseudo- R^2	0.924

Table 7. Comparison of improvement of fit measured by likelihood ratio between unadjusted and adjusted effects models.

Variable	Unadjusted effects			Adjusted effects		
	Deviance	Improvement χ^2	df	Deviance	Improvement χ^2	df
Intercept	996.9673					
SHADE	981.9554	15.0119	1	901.4757	20.0341	1
DISHOUSE	983.6985	13.2688	1	885.147	3.7054	1
DHABITATS	981.6438	14.9832	1	897.139	19.4364	1
CANOPY	988.6662	8.3011	1	890.101	8.6594	1
DRAIN	986.8716	10.0957	1	901.9639	20.5223	1
Full Main Effects						
1 st Degree Interactions				844.8677	38.9132	5

Table 8. Results of poisson regression used to estimate prior distribution of coefficients for the MCMC analysis

Variable	df	Coefficient	SE	P
Intercept	1	1.4020	0.1053	< 0.0001
SHADE	1	0.0357	0.0057	< 0.0001
DISHOUSE	1	0.0052	0.0066	0.4297
DHABITATS	1	0.02991	0.0048	< 0.0001
CANOPY	1	0.0172	0.0044	< 0.0001
DRAIN	1	0.0521	0.1702	0.7596
STBSITE	1	-0.8438	0.1067	< 0.0001
DRAIN*STBSITE	1	0.3731	0.1539	0.0154
SHADE*DRAIN	1	-0.0402	0.0102	< 0.0001
DISHOUSE*STBSITE	1	0.0594	0.0186	0.0014
DHABITATS*STBSITE	1	-0.0386	0.0163	0.0011

Table 9. Improvement of fit of the WinBUGS Hierarchical Bayesian Model (HBM) model.

Variable	Unadjusted effects		Adjusted effects	
	df	Improvement χ^2	Improvement χ^2	df
SHADE	1	-1.368	-0.353	1
DISHOUSE	1	6.089	3.242	1
CANOPY	1	1.187	1.432	1
DRAIN	1	0.722	1.548	1

as $-2 * \log(\text{likelihood})$, where ‘likelihood’ was defined as $p(y | \text{and } \theta)$, including all the normalizing constants: y comprised all stochastic node values and θ comprised the immediate stochastic parents of y . The expectation $\bar{D} = E^\theta [D(\theta)]$ was used as a measure of

model fitness based on the values of the predictor variables of the habitats sampled. The effective number of parameters included in the *Cx. quinquefasciatus* habitat model was computed as $p_D = \bar{D} - D(\bar{\theta})$, where $\bar{\theta}$ was the expectation of θ . The DIC was calculated as: $DIC = p_D + D$. The DIC value for the final model was 930.3. The DIC value for the model was 933.4.

Median parameter values, as well as the 95% credibility intervals (2.5 percentile and 97.5 percentile values), are listed in **Table 10**. As the sampling sites increased based on the covariate distance from the nearest house, the median log-egg-raft count of count data changed. The adjusted model assumed independence among the sampled predictor variables of *Cx. quinquefasciatus* aquatic habitat egg-raft count data fit better than the model that adjusted for correlation within the study site based on the RMSE (**Figure 3**).

The kriged-based model resulted in two images, a surface of estimates and a surface of estimated variances.

The latter image was used to identify problems with the fit of the model to the sampled data by revealing relative differences in the model fit across the Trinidad study site. A measure of the degree of spatial dependence between samples was also generated (*i.e.*, semivariance). The magnitude of the semivariance between any two sampled *Cx. quinquefasciatus* habitats was dependent on the distance between the georeferenced habitats. Thus, given two habitat locations x and $(x + h)$, in the study site, a measure of one half of the mean square difference (*i.e.*,

Table 10. Coefficient parameters estimates for WinBUGS Bayesian model.

Variable	Mean	SD	MC error	2.5%	50%	97.5%
Intercept	1.427	0.0804	0.0013	1.267	1.427	1.581
DISHOUSE	0.019	0.0093	0.0001	0.001	0.019	0.037

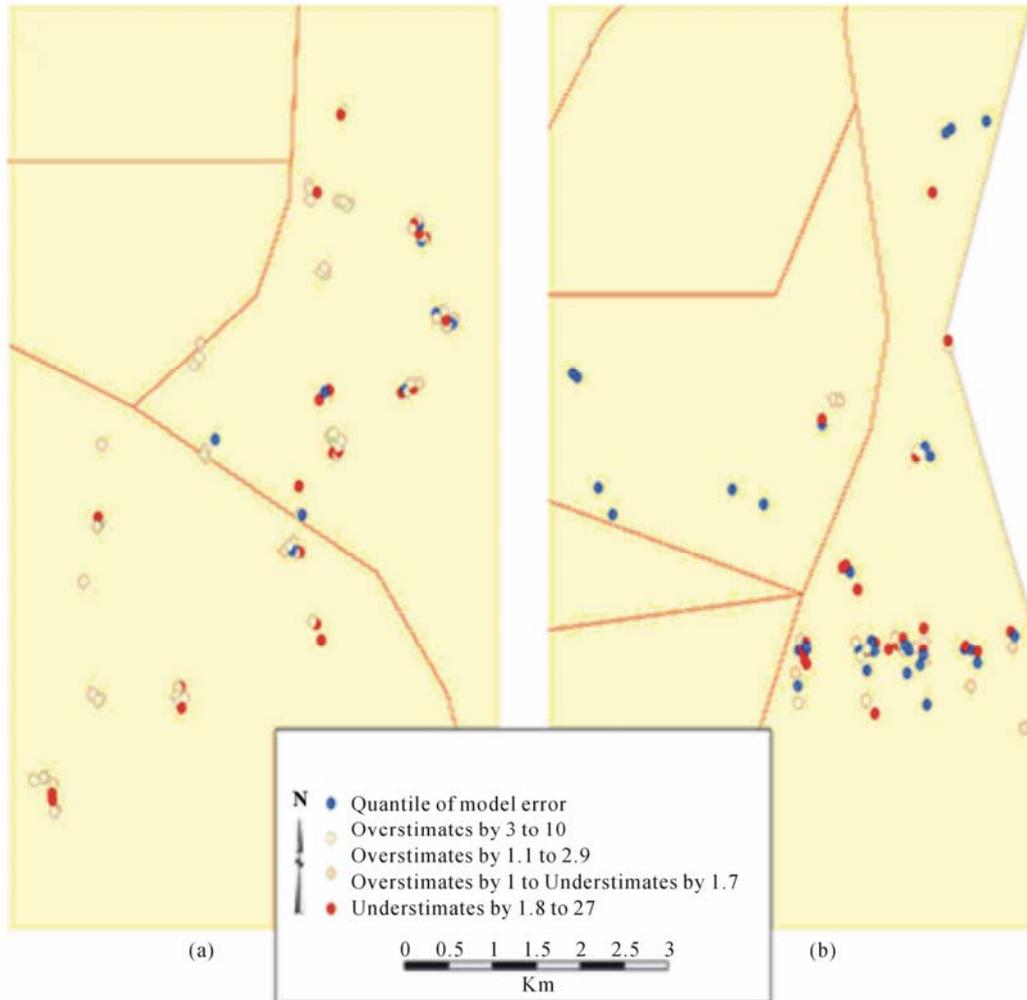


Figure 3. Observed *Cx. quinquefasciatus* habitats egg-raft count data in the Trinidad study site using a a) non-spatially adjusted and b) in a spatially adjusted model.

the semivariance) was produced by assigning the values $z(x + h)$ to the value $z(x)$, where h (*i.e.*, the lag) was the inter-sample distance between the georeferenced *Culex* habitats. Semivariance is an average of the squared deviations of values that are less than the mean [25]. In this research, plotting the semivariance versus the lag h gave a semivariogram. The summation was carried out over all distances h . The semivariance increased as the distance increases until it reached a maximum at a certain distant away from a sampled *Cx. quinquefasciatus* habitat. When the semivariance equals the variance around the averaged values, and no longer increases, this causes a flat region to occur on the semivariogram, (*i.e.*, the sill) [28]. The sill indicated that the semivariance values had been reached (*i.e.*, the value of maximum variance was equivalent to the variance of the WV-1 0.5 m pixel value). A non-zero intercept value (*i.e.*, nugget variance) of the variogram model was also generated which indi-

cated the variability of the sampled *Cx. quinquefasciatus* data. We optimized the RMSE by minimizing the spatial structure in a *Cx. quinquefasciatus* habitat model which generated a pure nugget variogram. The level of nugget variance represented noise characteristics in the sampled explanatory variables. A neighborhood distance search radius provided the mean standard errors of the interpolated values.

A kriged map of deviance residuals was then calculated, which was added to the predicted values on the logit scale before transforming the result back to proportions. Spatial dependence displayed by these plots was modeled using the constructed semivariogram. The maps developed from the kriged uncertainty residuals allowed deviation from the model and incorporated the field sampled egg raft count values. The final maps were greatly improved in sensitivity and did not deviate too severely from the field-sampled *Cx. quinquefasciatus*

habitat data. The exponential model fitted to the semi-variogram at a partial sill 0 and nugget of 5.8 km, lag size 6.5 km number of lags 12 with a range of 73.6 km which formed the basis of the model to map predictions of residuals in an 73.6 km radius around each sampled *Cx. quinquefasciatus* habitat (**Figure 4**).

Voronoi polygons were then modeled. An approximation, based on the midpoints of the legs and center of the triangle were used to evaluate the area covered by each Voronoi polygon. A new sample point was calculated using the importance metric (Gamma) of each pixel in the Voronoi polygons generated according to

$$C_p = \frac{\sum_{u,v} \Gamma_{u,v}(\vec{u}, \vec{v})}{\sum_{u,v} \Gamma_{u,v}}$$

where (u, v) were the coordinates of a WV-I pixel in the polygon, $\Gamma(u, v)$ was the importance metric of that pixel, and $C(p)$ was the coordinates of the new polygon center (*i.e.* the sampled *Cx. quinquefasciatus* habitat). To do this calculation on each

polygon we used a scan conversion. We ran a scan conversion and examined each pixel in the frame buffer for each polygon. The Voronoi tessellation provided a spatial trend analyses of the error in the model which revealed that all coefficients were within normal statistical limitations (**Figure 5**).

4. Discussion

The ANOVA tested the significant differences between means in the *Cx. quinquefasciatus* habitat data. In repeated measures ANOVA; however, containing factors such as spatiotemporal-sampled *Cx. quinquefasciatus* with more than two levels, additional special assumptions may be required for quantification of sampled predictor variables; specifically, the compound symmetry assumption and the assumption of sphericity. The compound symmetry assumption requires that the variances derived from the data (*i.e.*, pooled within-group) and

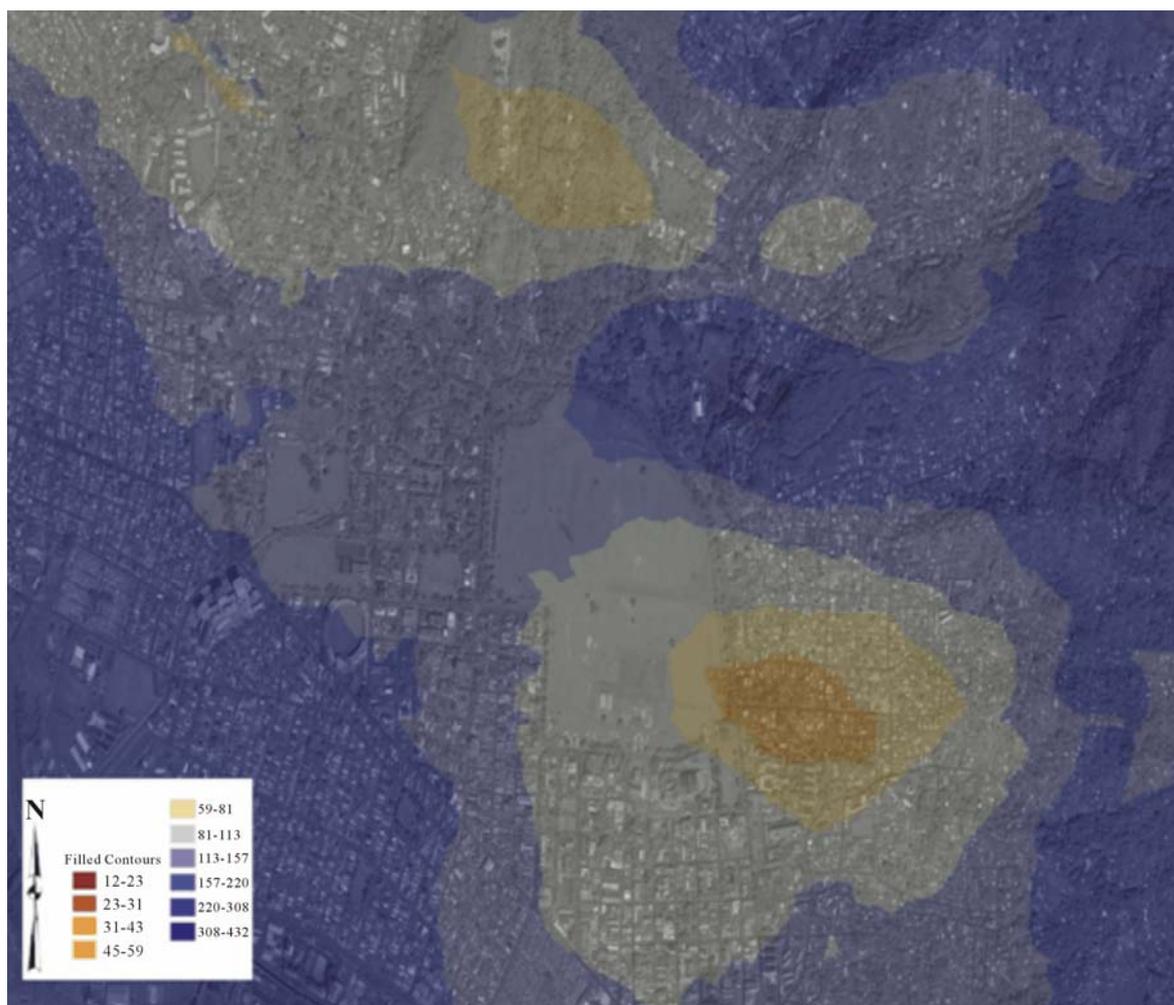


Figure 4. An ordinary kriged-based interpolation of egg-raft abundance of *Cx. quinquefasciatus* habitats in the trinidad study site.



Figure 5. Voroni tessellations generated from the predicted spatiotemporal-sampled *Cx. quinquefasciatus* habitats in the Trinidad study site.

covariances of the different repeated measures be homogeneous (25). This is a sufficient condition for the univariate F test for repeated measures to be valid (*i.e.*, for the reported F values to actually follow the F distribution); however, it is not a necessary condition in spatio-temporal *Cx. quinquefasciatus* habitat data analyses. When the compound symmetry or sphericity assumptions have been violated, the univariate ANOVA table will give erroneous results (23). A remedy may be to stress the necessity of independent hypotheses testing in spatio-temporal *Cx. quinquefasciatus* data analyses. A general algorithm implemented can attempt to generate, for each effect, a set of independent (orthogonal) contrasts. In repeated measures ANOVA, these contrasts specify a set of hypotheses about differences between the levels of the repeated measures factor. However, if these differences are correlated across subjects, then the resulting contrasts are no longer independent. In fact, in most instances where a repeated measures ANOVA is used, we would probably suspect that the changes across levels are correlated across the sampled *Cx. quinquefasciatus* habitat parameters. When this happens, the compound symmetry and sphericity assumptions have been violated, and independent contrasts cannot be computed.

Thus, the problem of compound symmetry and sphericity in *Cx. quinquefasciatus* habitat data analyses pertains to the fact that multiple contrasts involved in testing repeated measures effects with more than two levels are not independent of each other. However, they do not need to be independent of each other if we use multivariate criteria to simultaneously test the statistical significance of the two or more repeated measures contrasts. This insight is the reason why Multivariate Analysis of Variance (MANOVA) methods are increasingly applied to test the significance of univariate repeated measures factors with more than two levels. Multivariate analysis of variance is a generalized form of univariate ANOVA which is used in cases where there are two or more dependent variables (23). Analogous to ANOVA, MANOVA is based on the product of model variance matrix and error variance matrix inverse; thus, invariance considerations generated from spatio-temporal-sampled *Cx. quinquefasciatus* habitat predictors will imply the statistic is a suitable measure of magnitude of a singular value decomposition from this matrix product. Thus, instead of a univariate F value, we would obtain a multivariate F value (Wilks' λ) based on a comparison of the error variance/covariance matrix and the effect

variance/covariance matrix. The problem of compound symmetry and sphericity pertains to the fact that multiple contrasts involved in testing repeated measures of *Cx. quinquefasciatus* habitat effects with more than two levels are not independent of each other but, fortunately a multivariate criteria can simultaneously test the statistical significance of the two or more repeated measures contrasts. We thus endorse a MANOVA approach for spatiotemporal data analyses of *Cx. quinquefasciatus* habitat parameters as it simply bypasses the assumption of compound symmetry and sphericity altogether.

The spatial hydrological model generated from the environmental-sampled *Cx. quinquefasciatus* covariates revealed that elevation was an important variable in the DEM model. Anderson *et al.* (2004) captured significantly more *Culex quinquefasciatus* in a variety of traps (bird-baited, MMX mosquito magnet, CO₂-baited CDC trap) in the canopy (7.6 m) than at near ground level (1.5 m) [40]. In our research local differences among WV-1 DEM grid cells were also analyzed to calculate slope and other land surface parameters using relative and absolute accuracy estimates for determining the statistical significance of terrain-related parameters associated to the *Cx. quinquefasciatus* habitats. In this research the slope gradient was defined as the maximum rate of change in altitude (*i.e.*, $\tan(\Theta)$), aspect (ψ) as the compass direction of this maximum rate of change). Slope is defined by a plane tangent to a topographic surface, as modeled by the DEM at a point, whereby, it is classified as a vector; as such it has a quantity (gradient) and a direction (aspect). More analytically, the slope gradient at a sampled *Cx. quinquefasciatus* habitat in the Trinidad study site was the first derivative of elevation (Z) with respect of the slope (S), where S was in the aspect direction (ψ). We calculated the first derivative of a function at a sampled habitat as the angular coefficient of the tangent to the function of a sampled habitat. The resolution of our WV-1 DEM quantified the georeferenced explanatory covariates at each 0.5m pixel (*i.e.*, absolute accuracy). At the same time the relative accuracy of the sampled data generated a geometrically correct reference frame for validating the covariate elevation associated with the sampled *Cx. quinquefasciatus* habitats. WV-1 DEMs can allow for hydrologic modeling, view-shed determination, slope/aspect analyses, and 3-d surface visualization of georeferenced terrain-related parameters associated to spatiotemporal-sampled *Cx. quinquefasciatus* habitat data.

In the regression model, PROC MIXED quantified correlations among error measurements using random effects estimates and random regression coefficients. PROC MIXED used three options for the method of estimation including: 1) Maximum Likelihood (ML); 2)

Restricted or Residual maximum likelihood (REML), which was the default method; and, 3) Minimum Variance Quadratic Unbiased Estimation (MIVQUE). In this research, the ML was the regular maximum likelihood method as the *Cx. quinquefasciatus* parameter estimates maximized the likelihood function, while the REML was a variant of maximum likelihood. The sample mean was the maximum likelihood estimator of the sampled habitat population mean, based on the egg-raft count data. The sample variance was also a close approximation to the maximum likelihood estimator of the immature sampled variance. REML estimated the variance parameters using a marginal likelihood which produced unbiased estimates. REML estimators generated estimates from the spatiotemporal-sampled parameters by maximizing only that part of the likelihood function that was invariant to the fixed effects part of the linear model. The variance of a quadratic function of the random variables in the linear model was then minimized to obtain locally best unbiased estimators of the variance components. The matrices of these quadratics were computed from relationship matrices and from prior estimates of variances used in the model equations.

In the Bayesian matrix we were able to instantiate the nodes directly which was equivalent to supplying, a measurement for a radius based on the distribution of the georeferenced *Cx. quinquefasciatus* habitats. We were able to quantify an uncertainty estimate in our likelihood information by supplying a probability distribution over a set of radii generated from the sampled data. We used a backward propagation to compute the node probabilities. The error gradients were then propagated in any direction throughout the Bayesian framework. This type of probability propagation allowed choosing between the sampled predictor variables using the computed probability of each sampled covariate value based on one or more of the nodes.

Our Bayesian approach allowed flexible model fitting and estimation and geomapping of all “high risk” *Cx. quinquefasciatus* habitats based on sampled egg-raft count data in the study site. The MCMC algorithm for the *Cx. quinquefasciatus* egg-raft count model produced a sequence of parameter vectors that represented random draws from the posterior distribution. The DIC comprised two goodness-of-fit measures and the posterior distribution of the deviance, which was the number of effective sampled habitat covariates for measuring complexity in our *Cx. quinquefasciatus* model. Habitats with high egg-raft counts were compared with the results of a Monte Carlo simulation, which established the probabilities and occurrences of these habitats. Our results indicated that likelihood weights influenced the resulting posterior distributions of the field and remote-sampled

Cx. quinquefasciatus egg-raft count parameters, which, in turn influenced the spatial trends in the variance uncertainty estimates for model prediction. Distance from the nearest house was a significant predictor associated with prolific habitats in the Trinidad study site.

Environmental factors commonly associated with *Cx. quinquefasciatus* habitats include the distance from the nearest human habitation. The location of WNV cases at the place of residence can be based on the fact that *Culex* mosquitoes are crepuscular feeders, and are most likely to come in contact with people while they are at home during the evening hours. Because egg-laying is a spatial process dependent upon the location of the focal habitat relative to sources of gravid mosquitoes, habitats closer to human inhabitations tend to receive more eggs and thus are more productive if conspecific competition is negligible [29]. Furthermore, housing age can also play a big role in the explanation of the focal areas. This is expected since older houses have poorer drainage systems and open foundations which may provide larval and adult habitats. For example, houses built in the post World War II era in the study site may suffer from a general inattention. These houses may contain catch basins, rich with organic material which are breeding grounds for *Culex* mosquitoes. Most of the *Culex* species require organic-rich water for larval development and are commonly found in high numbers in artificial containers, unattended pools, retention ponds, storm drains, catch basins, sewage systems and treatment plants [41,42]. Furthermore, combined sewage and street runoff discharges into natural waterways are considered the main sources of urban stream pollution, which are associated to WNV as *Culex* mosquitoes proliferate in eutrophic hydrological networks. The identification of social factors that characterize these focal areas in the Trinidad study site may also provide insight into human risk and can help to target control and prevention strategies. Future analyses should explicitly test whether enzootic and bridge transmission are related to human population factors including household income, population age, age of housing, housing density, and population density. Since it can be assumed from this analysis that the WNV is transmitted near a person's residence in the Trinidad study site, extensive engineering to improve the runoff of water in residential areas should be considered.

A framework for assessing autocorrelation in residual error from the spatial analyses was constructed in SAS/GIS[®]. We generated a semiparametric filtering model using proxy variables constructed from the spatiotemporal-sampled georeferenced explanatory variables. In our *Cx. quinquefasciatus* habitat model, the misspecification terms was replaced by the proxy variables for implementing a conditional covariance matrix employing the

autoregressive specification. An unknown misspecification term can be approximated by a set of spatial proxy variables [38]. After modeling the misspecification terms, the remaining residuals $\hat{\varepsilon}$ become white noise. This result implied that the estimated regression parameters $\hat{\beta}$ were unbiased for the basic regression model, $y = X\beta + \varepsilon^*$, where ε^* incorporated the misspecification term and the white-noise disturbances. This allowed us to calibrate the autoregressive models. All of these eigenvectors exhibited weak positively autocorrelated spatial patterns in the habitat parameter estimates. Local weather patterns, mosquito control programs and movement of hosts can have spatial expressions, which cause immature *Culex* mosquitoes to aggregate in geographic space rendering positive autocorrelation in sampled predictor variables [8,26].

In this research, we used an Ordinary interpolation in ArcGIS[®] for identifying initial spatial structures in the sampled *Cx. quinquefasciatus* habitats and for quantifying the variance and autocorrelation in the data. The estimation method determined the best linear unbiased estimator. It was linear since the estimated values were weighted linear combinations of the spatiotemporal-sampled data and it was unbiased because the mean of the error was zero. For determining the optimal predictors, a semivariogram was constructed which expressed the spatial variation in the sampled data. In this research the variogram [*i.e.*, $2\gamma(x,y)$] was a function describing the degree of spatial dependence between the sampled *Cx. quinquefasciatus* habitats [*i.e.*, $Z(x)$]. This was defined as the expected squared increment of the values between the sampled habitat locations x and y . Our semivariogram was nonnegative since it was the expectation of a square. The covariance function was related to variogram by $2\gamma(x,y) = C(x,x) + C(y,y) - 2C(x,y)$. In this research, the $\gamma(x,y) = E(|Z(x) - Z(y)|^2) = \gamma(y,x)$ was a symmetric function, consequently, $\gamma_s(h) = \gamma_s(-h)$ was an even function. A function is a semivariogram if, and only if, it is a conditionally negative definite function, *i.e.* for all weights w_1, \dots, w_N subject to $\sum_{i=1}^N w_i = 0$ and locations x_1, \dots, x_N , thus, it holds:

$$\sum_{i=1}^N \sum_{j=1}^N w_i \gamma(x_i, x_j) w_j \leq 0$$
 [28]. The function corresponded to the fact that the variance [*i.e.*, $var(X)$ of $X = \sum_{j=1}^N w_j Z(x_j)$] generated from the sampled *Cx. quinquefasciatus* data was given by the negative of this double sum, which in this research was nonnegative. All spatial dependence was quantified in the model. If a stationary random field has no spatial dependence (*i.e.* $C(h) = 0$ if $h \neq 0$) the semivariogram is the constant $var(Z(x))$ everywhere except at the origin, where it is zero [28].

Theoretically, at zero separation distance (lag = 0), the semivariogram value is zero in a predictive vector mos-

quito habitat distribution model; however, at an infinitesimally small sampled habitat distances, the semi-variogram often exhibits a nugget effect (*i.e.*, measurement errors or spatial sources of variation at distances smaller than the sampling interval), which is some value greater than zero [16]. Additionally, since the random field was stationary and ergodic, the $\lim_{h \rightarrow \infty} \gamma_s(h) = \text{var}(Z(x))$ corresponded to the variance generated by the *Cx. quinquefasciatus* habitat data. We defined a practical range and defined the distance at which 95% of the sill was reached for the asymptotic variogram model. The semi-variogram calculated estimates of the surface at the different habitat locations. The increase in the semi-variogram values with increasing lags diminished with distance and leveled off, at around 70 km (*i.e.*, the range or active lag distance). This was the approximate distance at which spatial autocorrelation between the *Cx. quinquefasciatus* habitat data point pairs ceased or became much more variable. In the future sampling variograms should be constructed using spatiotemporal-sampled *Cx. quinquefasciatus* habitat data. The sampling variogram, unlike the semi-variogram, shows where a significant degree of spatial dependence in the sample space or sampling unit dissipates into randomness when the variance terms of a temporally or *in-situ* ordered set are plotted against the variance of the set and the lower limits of its 99% and 95% confidence ranges [28]. Sampling variograms may quantify extraneous measurement variances before spatial dependence is verified in a predictive *Cx. quinquefasciatus* habitat distribution model.

In this research interpolation accuracy was also measured by the natural logarithm of the mean squared interpolation error and Voroni polygons which revealed all uncertainty effects while quantifying several covariate interaction terms in Euclidean space. The Voroni cells generated were a d -dimensional polytope (*i.e.*, a convex hull of a finite set of georeferenced points). A set of hyperplanes were generated using the polygonal shapes around each sampled *Cx. quinquefasciatus* habitat in the Trinidad study site. If S contains only two points, a and b , then the set of all points equidistant from a and b is a hyperplane—an affine subspace of codimension 1 [28]. The set of the polytopes covered the whole space and was the Voroni tessellation corresponding to the georeferenced *Cx. quinquefasciatus* habitats. The convex Voroni polygons were visualized as space-filling polygons constructed around the dataset of the sampled covariates (*i.e.*, Voroni centers), such that each polygon contained all points closer to its Voroni center than to the center of any other polygon. The relationship of the Voroni centers to edges of polygons were used to test whether any convex tessellation existed in the polygons. This test amounted to finding Voroni centers that best fit the given tessellation. In this research, Voroni centers were found

by solving two systems of linear equations. These equations represented 1) conditions on the slope of polygon edges relative to the slope of lines through Voroni centers, and 2) conditions on the distance from edges to Voroni centers. Least squares and constrained least-squares solutions were used to solve the two systems. In order to quantitatively characterize the error in the interpolation model, we examined normalized cell size distribution related to the homogenous and inhomogeneous fixation densities (*i.e.*, *Cx. quinquefasciatus* habitat egg-raft count), which provided additional insight as to how the convex tessellations varied between the Voroni polygons. A goodness-of-fit statistic was derived using the Voroni tessellations which determined the error propagation in the multivariate model. The Voroni residuals reflected sampling errors and analytical errors based on the sampled *Cx. quinquefasciatus* habitat egg-raft count density.

In the future new network nodes using Dirichlet tessellations in spatiotemporal-sampled *Cx. quinquefasciatus* habitats should be analyzed. For example, Okabe and Satoh, (2008) have developed a software package, SANET, which can be installed as an ArcGIS add-in for “Spatial Analysis on a Network” for determining new network nodes [43]. The SANET software includes a network-based Voroni tool, very similar to some of the network analysis tools provided in ArcGIS itself and in network-oriented packages such as TransCAD and Cube. The point set (*i.e.*, off-network sites) can be linked to the nearest network edge, and then non-directed network Voroni regions can be generated from georeferenced *Cx. quinquefasciatus* habitat locations using shortest path distances. Separate inward and outward-directed Voroni regions can then be computed and each habitat can be derived using unweighted calculations.

For future research, the risk of WNV transmission in the Trinidad study site should also be investigated using wild bird roosting behavior and seasonal migratory pattern data. Quantifying WNV amplification at a local scale, with explicit attention to distances measurements from mosquito and hatch-year bird sites and human habitation areas may affect timing of infection in birds, mosquitoes, and humans. Seropositive resident and migratory birds have been documented along the Atlantic and Mississippi flyways into the Caribbean [2]. While transient birds may suggest a means by which WNV has rapidly spread across North and Central America, resident breeding birds may be more critical for zoonotic transmission in Trinidad [44-47]. In North America, high seroprevalence is associated with avian species that breed year-round and live close to humans (e.g. American robin, House sparrow, Mourning dove, Northern

cardinal, Northern mockingbird), though other wild birds may be important [47-49]. In contrast to the urban avian species commonly associated as WNV hosts in North America, avian WNV hosts that occur in Trinidad are often non-urban, wild bird species [2]. Caribbean birds that have tested seropositive for WNV and are resident to Trinidad or breed seasonally in Trinidad include: Bananaquits (*Coereba flaveola*), Palm-tanagers (*Phaenicophilus palmarum*), and Grossquit species [2]. Since the significance of avian species in WNV transmission may relate to their host competency, natural exposure to mosquito vectors and spatiotemporal availability, an avian GIS-based model using spatial filter eigenvectors and Bayesian estimates might be able to generate globally asymptotically stable parameters while quantifying the amplification cycle prior to bridge vector involvement. Bridge vectors are mosquito species that readily feed on humans and mammals [2].

In conclusion, correlation estimates generated from the WV-1 DEM revealed that the terrain covariate elevation was significantly associated with the *Cx. quinquefasciatus* habitats sampling sites. The spatial hydrological model allowed for hydrologic modeling, view-shed determination, slope/aspect analyses, and 3-d surface visualization of the georeferenced parameters associated to *Cx. quinquefasciatus* habitat data. Elevation data was represented in the WV-1 DEM by including gridded data where the terrain covariate was estimated for each cell in a regular grid, a triangular irregular network, and contours. The coefficients from the decomposition of the sampled predictor variables were then entered into a Bayesian matrix. This framework allowed flexible model fitting, estimation, and mapping of all high risk habitats based on field-sampled egg-raft count data using a MCMC specification. Residual estimates were then analyzed with an eigenvector spatial filtering algorithm which transformed all the georeferenced predictor variables containing spatial dependence into covariates free of spatial dependence by partitioning the original sampled habitat data and the sampled egg-raft abundance count data into two synthetic variates: 1) a spatial filter variate capturing latent spatial dependency and 2) a non-spatial variate that was free of spatial dependence. The geographic distribution of the sampled habitats based on the egg-raft counts exhibited positive autocorrelation in all models tested: log-egg-raft counts aggregated in geo-space. The procedure accounted for 17 % pseudo-replicated information in the datasets. All models were then adjusted based on the residual estimates from the eigenfunction decomposition algorithm and then kriged. The sill indicated that the semivariance values had been reached (*i.e.*, the value of maximum variance was equivalent to the variance of a 0.5m pixel value), while a non-zero intercept value (*i.e.*,

nugget variance) of the variogram model was indicative of the variability of the field and remotesampled *Cx. quinquefasciatus* data. A simple quantitative measure of the interpolation performed was determined by generating RMSE values for the models and by constructing Voroni polygons. Optimizing the RMSE by minimizing the spatial structure in a *Cx. quinquefasciatus* habitat model generated a pure nugget variogram, of which the level of nugget variance represented noise characteristics in the predictor variables. Interpolation accuracy tests measured by the natural logarithm of the mean squared interpolation error and the Voroni polygons revealed all error effects of the parameter estimates in the kriged-based model and quantified the covariate interaction terms. Newer GIS software, WV-I data and spatial statistics can implement larval control strategies by targeting prolific habitats of *Cx. quinquefasciatus* based on field-sampled count data in Trinidad.

In the future high quality spatial autoregressive and surface models should be constructed using spatiotemporal-sampled *Cx. quinquefasciatus* and other mosquito habitat data and Worldview-2 (WV-2) imagery for improving sensitivity in LULC, elevation and kriged-based interpolation parameters in Trinidad. On October 8, 2009 Digitalglobe launched WV-2 which is the first high resolution commercial satellite to offer 8-band capability with high accuracy, agility and spectral diversity (www.digitalglobe.com). This allows the creation of accurate *Cx. quinquefasciatus* habitat maps without the need of GCPs. A wide range of features can be extracted using the traditional four spectral bands but WV-2 offers the addition of the Coastal Yellow, Red edge and shorter wave NIR bands which may provide a wider range of potential derivatives for targeting prolific *Cx. quinquefasciatus* and other mosquito habitats. For example, the Coastal-blue bands can be used for bathymetry, benthic, wetland and atmospheric modeling which can assist in detecting subtle differences in land cover and coastal features associated with prolific mosquito habitats. Probably the best feature of WV-2 for geomapping *Cx. quinquefasciatus* and other mosquito habitats, however, is the ability to derive algorithms and functions from new band combinations for feature extraction and improved classification. For example, the Red-edge NIR band can identify plant health and age through chlorophyll production while the yellow band in conjunction with the Red Edge band can provide agricultural applications and iron oxide mineralogy which can model moisture content. Thus, WV-2 can provide detailed imagery for precise map generation, LULC change detection, construction of DEMs and autoregressive predictive models. WV-2 data can further help implement a remote-based habitat surveillance in Trinidad and in other regions.

5. Acknowledgements

We acknowledge the data collection provided by the Ministry of Health in Trinidad and Tobago and the Department of Life Sciences, University of the West Indies, St. Augustine, and Trinidad. This research was funded by the National Institute of Health Grant U01A154889 (Novak Robert) University of Alabama at Birmingham and from the Government of Trinidad and Tobago, GTTGF-6 (Chadee Dave).

6. References

- [1] D. Nash, F. Mostashari, A. Fine, J. Miller, D. O'Leary, K. Murray, A. Huang, A. Rosenberg, A. Greenberg, M. Sherman, S. Wong, G. L. Campbell, J. T. Roehrig, D. J. Gubler, W. J. Shieh, S. Zaki, P. Smith and M. Layton, "The Outbreak of West Nile Virus Infection in the New York City Area in 1999," *The New England Journal of Medicine*, Vol. 344, 2001, pp. 1807-1814. [doi:10.1056/NEJM200106143442401](https://doi.org/10.1056/NEJM200106143442401)
- [2] N. Komar and G. C. Clark, "West Nile Virus Activity in Latin America and the Caribbean," *Revista Panamericana de Salud Pública*, Vol. 19, 2006, pp. 112-117. [doi:10.1590/S1020-49892006000200006](https://doi.org/10.1590/S1020-49892006000200006)
- [3] M. J. Turell, M. R. Sardelis, M. L. O'Guinn and D. J. Dohm, "Potential Vectors of West Nile Virus in North America. Curr," *Top Microbiol. Immunol*, Vol. 267, 2002, pp. 241-252.
- [4] B. G. Jacob, R. L. Lampman, M. P. Ward, E. J. Muturi, J. A. Morris, E. X. Caamano and R. J. Novak, "Geospatial Variability in the Egg Raft Distribution and Abundance of *Culex Pipiens* and *Culex Restuans* in Urbana-Champaign, Illinois," *International Journal of Remote Sensing*, Vol. 30, 2009, pp. 2005-2019. [doi:10.1080/01431160802549195](https://doi.org/10.1080/01431160802549195)
- [5] B. G. Jacob, W. Gu, E. J. Muturi, E. X. Caamano, J. M. Morris, R. Lampman, R. J. Novak, "Developing Operational Algorithms Using Linear and Non-Linear Least Squares Estimation in Python® for Identification of *Culex Pipiens* and *Culex Restuans* Aquatic Habitats in a Mosquito Abatement District (Cook County, Illinois)," *Geospat Health*, Vol. 3, 2009, pp. 23-31.
- [6] J. K. Meece, J. S. Henkel, L. Glaser, K. D. Reed, "Mosquito Surveillance for West Nile Virus in Southeastern Wisconsin-2002," *Clinical Medicine & Research* Vol. 1, 2003, pp. 37-42. [doi:10.3121/cmr.1.1.37](https://doi.org/10.3121/cmr.1.1.37)
- [7] F. W. Kutz, T. G. Wade and B. B. Pagac, "A Geospatial Study of the Potential of Two Exotic Species of Mosquitoes to Impact the Epidemiology of West Nile Virus in Maryland," *Journal of the American Mosquito Control Association*, Vol. 19, 2003, pp. 190-198.
- [8] D. A. Griffith, "A Comparison of Six Analytical Disease Mapping Techniques as Applied to West Nile Virus in the Coterminous United States," *International Journal of Health Geographics*, Vol. 4, 2005, pp. 18-26. [doi:10.1186/1476-072X-4-18](https://doi.org/10.1186/1476-072X-4-18)
- [9] U. Kitron, "Risk Maps: Transmission and Burden of Vector-Borne Diseases," *Parasitol Today*, Vol. 16, 2000, pp. 324-325. [doi:10.1016/S0169-4758\(00\)01708-7](https://doi.org/10.1016/S0169-4758(00)01708-7)
- [10] D. J. Rogers and S. E. Randolph, "Studying the Global Distribution of Infectious Diseases Using GIS and RS," *Nature Reviews Microbiology*, Vol. 1, 2003, pp. 231-237.
- [11] S. I. Hay, J. A. Omumbo, M. H. Craig and R. W. Snow, "Earth Observation, Geographic Information Systems and Plasmodium Falciparum Malaria in Sub-Saharan Africa," *Advances in Parasitol*, Vol. 47, 2000.
- [12] E. P. Pegoraro, M. J. Monson, R. K. Rey, A. Barron-Gafford and C. B. Osmond, "The Effect of Elevated CO₂, Soil and Atmospheric Water Deficit and Seasonal Phenology on Leaf and Ecosystem Isoprene Emission," *Functional Plant Biology*, Vol. 34, 2007, p. 774. [doi:10.1071/FP07021](https://doi.org/10.1071/FP07021)
- [13] J. S. Brownstein, T. R. Holford, D. Fish, "Enhancing West Nile Virus Surveillance, United States," *Emerging Infectious Diseases*, Vol. 10, 2004, pp. 1129-1133.
- [14] N. Komar, S. Langevin, S. Hinten, N. Nemeth, E. Edwards, D. Hettler, B. Davis, R. Bowen, M. Bunning, "Experimental Infection of North American Birds with the New York 1999 Strain of West Nile Virus," *Emerging Infectious Diseases*, Vol. 9, 2003, pp. 311-322.
- [15] L. R. Petersen, A. A. Marfin, D. J. Gubler, "West Nile Virus," *JAMA, the Journal of the American Medical Association*, Vol. 4, No. 290, 2003, pp. 524-528.
- [16] B. G. Jacob, N. D. Burkett-Cadena, J. C. Luvall, S. H. Parcak, C. J. W. McClure, L. K. Estep, G. E. Hill, E. W. Cupp and R. J. Novak, "Developing GIS-Based Eastern Equine Encephalitis Vector-Host Models in Tuskegee, Alabama," *International Journal of Health Geographics*, Vol. 9, 2010, p. 12. [doi:10.1186/1476-072X-9-12](https://doi.org/10.1186/1476-072X-9-12)
- [17] E. Mushinzimana, S. Munga, N. Minakawa, L. Li, C. C. Feng, L. Bian, U. Kitron, C. Schmidt, L. Beck, G. Zhou and A. K. Githeko, "Landscape Determinants and Remote Sensing of Anopheline Mosquito Larval Habitats in the Western Kenya Highlands," *Malaria Journal*, Vol. 5, 2006, p. 13. [doi:10.1186/1475-2875-5-13](https://doi.org/10.1186/1475-2875-5-13)
- [18] J. Shaman, J. F. Day and M. Stieglitz, "Drought-Induced Amplification and Epidemic Transmission of West Nile Virus in Southern Florida," *Journal of Medical Entomology*, Vol. 42, No. 2, 2005, pp. 134-141. [doi:10.1603/0022-2585\(2005\)042\[0134:DAAETO\]2.0.CO;2](https://doi.org/10.1603/0022-2585(2005)042[0134:DAAETO]2.0.CO;2)
- [19] J. R. Jensen, "Remote Sensing of Urban Suburban Infrastructure and Socio-Economic Attributes," *Photogrammetric Engineering and Remote Sensing*, Vol. 5, No. 65, 2005, pp. 611-622.
- [20] B. G. Jacob, E. J. Muturi, E. X. Caamano, J. T. Gunter, E. Mpanza, R. Ayine, J. Okelloonen, J. P. M. Nyeko, J. I. Shililu, J. I. Githure, J. L. Regens and R. J. Novak, "Hydrological Modeling of Geophysical Parameters of Arboviral and Protozoan Disease Vectors in Internally Displaced People Camps in Gulu, Uganda," *International Journal of Health Geographics*, Vol. 7, 2008, pp. 1-11. [doi:10.1186/1476-072X-7-11](https://doi.org/10.1186/1476-072X-7-11)

- [21] B. G. Jacob, D. A. Griffith and R. J. Novak, "Decomposing Malaria Mosquito Aquatic Habitat Data into Spatial Autocorrelation Eigenvectors in a SAS/GIS® Module," *Transactions in GIS* Vol. 12, 2008, pp. 341-364. [doi:10.1111/j.1467-9671.2008.01104.x](https://doi.org/10.1111/j.1467-9671.2008.01104.x)
- [22] C. N. Theophilides, S. C. Ahearn, E. S. Binkowski, W. S. Paul and K. Gibbs, "First Evidence of West Nile Virus Amplification and Relationship to Human Infections," *International Journal of Geographical Information Science*, Vol. 20, 2006, pp. 103-115. [doi:10.1080/13658810500286968](https://doi.org/10.1080/13658810500286968)
- [23] T. G. Andreadis, J. F. Anderson and C. R. Vossbrink, "Mosquito Surveillance for West Nile Virus in Connecticut, 2000: Isolation from *Culex Pipiens*, *Cx. Restuans*, *Cx. Salinarius*, and *Culiseta Melanura*," *Emerging Infectious Diseases*, Vol. 7, 2001, pp. 670-674. [doi:10.3201/eid0704.010413](https://doi.org/10.3201/eid0704.010413)
- [24] M. Eidson, J. Miller, L. Kramer, B. Cherry and Y. Hagiwara, "Dead Crow Densities and Human Cases of West Nile Virus New York State 2000," *Emerging Infectious Diseases*, Vol. 7, 2001, pp. 662-664. [doi:10.3201/eid0704.010411](https://doi.org/10.3201/eid0704.010411)
- [25] M. Kulldroff, "A Spatial Scan Statistic," *Community Statistical Theory Methods*, Vol. 26, 1997, pp. 1481-1496. [doi:10.1080/03610929708831995](https://doi.org/10.1080/03610929708831995)
- [26] B. G. Jacob, E. Muturi, J. Mwangangi, J. Funes, J. Shililu, J. Githure and R. J. Novak, "Remote and Field Level Quantification of Vegetation Covariates for Malaria Mapping in Three Rice Agro-Village Complexes in Central Kenya," *International Journal of Health Geographics*, Vol. 6, 2007, pp. 21-28. [doi:10.1186/1476-072X-6-21](https://doi.org/10.1186/1476-072X-6-21)
- [27] I. Kleinschmidt, M. Bagayoko, G. P. Y. Clarke, M. Craig and D. L. Le Sueur, "A Spatial Statistical Approach to Malaria Mapping," *International Epidemiology Association*, Vol. 29, No. 2, 2000, pp. 355-361.
- [28] N. Cressie, "Aggregation in Geostatistical Problems," *Geostatistics troia*, 1992.
- [29] W. Gu and R. J. Novak, "Habitat-Based Modeling of Impacts of Mosquito Larval Interventions on Entomological Inoculation Rates, Incidence, and Prevalence of Malaria," *American Journal of Tropical Medicine and Hygiene*, Vol. 73, 2005, pp. 546-552.
- [30] B. G. Jacob, D. A. Griffith, E. J. Muturi, E. X. Caamano, J. I. Githure, J. T. Gunter, R. J. Novak, "A Heteroskedastic Error Covariance Matrix Estimator Using a First-Order Conditional Autoregressive Markov Simulation for Deriving Asymptotical Efficient Estimates from Ecological Sampled *Anopheles Arabiensis* Aquatic Habitat Covariates," *Malaria Journal*, Vol. 8, 2009, pp. 216-224. [doi:10.1186/1475-2875-8-216](https://doi.org/10.1186/1475-2875-8-216)
- [31] D. Chadee, "Key Premises, a Guide to *Aedes Aegypti* (Diptera: Culicidae) Surveillance and Control," *Bulletin of Entomological Research*, Vol. 94, 2004, pp. 201-207. [doi:10.1079/BER2004297](https://doi.org/10.1079/BER2004297)
- [32] J. P. Bradbury, B. Leyden, M. Salgado-Labouriau, Lewis, W. M. Jr., C. Schubert, M. W. Binford, D. G. Frey, D. R. Whitehead and F. W. Weibezahn, "Late Quaternary Environmental History of Lake Valencia," *Ven. Science*, Vol. 4, 1981, pp. 18-26.
- [33] J. R. Flenley, "The Equatorial Rain Forest: A Geological History," Butterworth, London, Vol. 9, 1979, p. 162.
- [34] J. S. Beard, "The Mora Forests of Auburn, British West Indies," *Journal of Ecology*, Vol. 33, 1946, pp. 173-192. [doi:10.2307/2256464](https://doi.org/10.2307/2256464)
- [35] D. D. Chadee, "Indoor and Outdoor Host-Seeking Rhythms of *Anopheles Albitarsis* (Diptera: Culicidae) in Trinidad, West Indies," *Journal of Medical Entomology*, Vol. 29, 1992, pp. 567-569.
- [36] B. G. Jacob, P. G. Nelson, R. Lampman, J. A. Morris, A. Raims, J. Funes, C. LaPonte and R. J. Novak, "Comparing Global Positioning System (GPS) Technology for Identifying Spatial Ecological Variation for Urban Mosquito Abatement," *Wing Beats*, Vol. 16, 2005, pp. 30-33.
- [37] B. G. Jacob, J. Shililu, E. J. Muturi, J. M. Mwangangi, S. M. Muriu, J. Funes, J. Githure, J. L. Regens and R. J. Novak, "Spatially Targeting *Culex Quinquefasciatus* Aquatic Habitats on Modified Land Cover for Implementing an Integrated Vector Management (IVM) Program in Three Villages Within the Mwea Rice Scheme, Kenya," *International Journal of Health Geographics*, Vol. 5, 2006, pp. 18-27. [doi:10.1186/1476-072X-5-18](https://doi.org/10.1186/1476-072X-5-18)
- [38] F. A. Haight, "Handbook of the Poisson Distribution," New York, Wiley, 1967.
- [39] S. Lang and A. Brezger, "Bayesian P-Splines," *Journal of Computational and Graphical Statistics* Vol. 13, 2004, pp. 183-212. [doi:10.1198/1061860043010](https://doi.org/10.1198/1061860043010)
- [40] D. A. Griffith, "Spatial Autocorrelation on Spatial Filtering," Springer, 2003.
- [41] J. F. Anderson, T. G. Andreadis, A. J. Main and D. L. Kline, "Prevalence of West Nile Virus in Tree Canopy-Inhabiting *Culex Pipiens* and Associated Mosquitoes," *American Journal of Tropical Medicine and Hygiene*, Vol. 71, 2004, pp. 112-119.
- [42] W. K. Reisen, J. O. Lundstrom, T. W. Scott, B. F. Eldridge, R. E. Chiles, R. Cusack, V. M. Martinez, H. D. Lothrop, D. Gutierrez, S. E. Wright, K. Boyce and B. R. Hill, "Patterns of Avian Seroprevalence to Western Equine Encephalomyelitis and St. Louis Encephalitis Viruses in California, USA," *Journal of Medical Entomology*, Vol. 37, 2000, pp. 507-527. [doi:10.1603/0022-2585-37.4.507](https://doi.org/10.1603/0022-2585-37.4.507)
- [43] L. M. Calhoan, M. Avery, L. Jones, K. Gunarto, R. King, J. Roberts and T. R. Burkot, "Combined Sewage Overflows (CSO) Are Major Urban Breeding Sites for *Culex Quinquefasciatus* in Atlanta, Georgia," *American Journal of Tropical Medicine and Hygiene*, Vol. 77, 2007, pp. 478-484.
- [44] K. Sugihara, A. Okabe and T. Satoh, "Computation Method for the Point Cluster Analysis on Networks," *Geoinformatica*, Vol. 15, No. 1, 2008, pp. 9-92.
- [45] J. H. Rappole, S. R. Derrickson and Z. Hubálek, "Migratory Birds and Spread of West Nile Virus in the Western Hemisphere," *Emerging Infectious Diseases*, Vol. 6, 2000,

- pp. 11-19. [doi:10.3201/eid0604.000401](https://doi.org/10.3201/eid0604.000401)
- [46] A. T. Peterson, D. A. Vieglaiss and J. K. Andreasen, "Migratory Birds Modeled as Critical Transport Agents for West Nile Virus in North America," *Vector-Borne and Zoonotic Diseases*, Vol. 3, 2003, pp. 27-37. [doi:10.1089/153036603765627433](https://doi.org/10.1089/153036603765627433)
- [47] R. B Tesh, R. Parson, M. Siirin, Y. Randle, C. Sargent, H. Guzman, T. Wuithiranyagool, S. Higgs, D. L. Vanlandingham, A. A. Bala, K. Haas and B. Zerinque, "Year-Round West Nile Virus Activity, Gulf Coast Region, Texas and Louisiana," *Emerging Infectious Diseases*, Vol. 10, 2004, pp. 1649-1652.
- [48] T. A. Beveroth, M. P. Ward, R. L. Lampman, A. M. Ringia and R. J. Novak, "Changes in Seroprevalence of West Nile Virus Across Illinois in Free-Ranging Birds from 2001 through 2004," *American Journal of Tropical Medicine and Hygiene*, Vol. 74, 2006, pp. 174-179.
- [49] R. M. Gleiser, A. J. MacKay, A. Roy, M. M. Yates, R. H. Vaeth, G. M. Faget, A. E. Folsom, W. F. Augustine, R. A. Jr. Wells and M. J. Perich, "West Nile Virus Surveillance in East Baton Rouge Parish, Louisiana," *Journal of the American Mosquito Control Association*, Vol. 23, 2007, pp. 29-36. [doi:10.2987/8756-971X\(2007\)23\[29:WNVSIE\]2.0.CO;2](https://doi.org/10.2987/8756-971X(2007)23[29:WNVSIE]2.0.CO;2)
- [50] P. P. Marra, S. Griffing, C. Caffrey, A. M. Kilpatrick, R. McLean, C. Brand, E. Saito, A. P. Dupuis, L. Kramer and R. J. Novak, "West Nile Virus and Wildlife," *BioScience*, Vol. 54, 2004, pp. 393-402. [doi:10.1641/0006-3568\(2004\)054\[0393:WNVAV\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2004)054[0393:WNVAV]2.0.CO;2)