ISSN: 2641-6336
DOI: 10.33552/ABBA.2024.06.000631
Annals of
Biostatistics & Biometric Applications
Iris Publishers

Research Article

Employing Second Order Geospatial Autocorrelation Statistics and an Eigen-Bayesian Semi-Parametric Markovian Non-Gaussian Model for Interpolating Culex quinquefasciatus Storm Sewer Habitats in Bexar and Dallas Counties Texas, U.S.A

Alexander M. Gambrell[1]*, Namit Choudhari[1], Saurav Chakraborty[1], Jing Liu[2], Heather McDonald[2], Sasha Mosich[2], Benjamin G. Jacob[2]

[1]*School of Geosciences, University of South Florida, United States of America*

[2]*College of Public Health, University of South Florida, United States of America*

***Corresponding author:** Alexander M. Gambrell, School of Geosciences University of South Florida, Cutler Bay, Florida, United States of America.

Received Date: June 19, 2024

Published Date: July 01, 2024

Abstract

Second-order eigenfunction eigen-decomposition eigen-algorithms have been used to determine hyper/hypo-endemic, aggregation/non-aggregation-oriented, aquatic larval habitat foci of multiple nuisance and vector arthropods. Unfortunately, a major West Nile Virus (WNV) mosquito vector, *Culex quinquefasciatus*, has not been geospatially forecasted for determining geolocations of unknown foci for implementing county level larval control. In this article, we would like to employ geospatial autocorrelation eigen-algorithms to determine georeferenceable hot and cold spots of storm sewer *Cx. quinquefasciatus* habitats in Dallas County and Bexar County in Texas. We aimed to predictively map unknown storm sewer foci of *Cx. quinquefasciatus* by employing interpolated capture point sentinel sites. Initially, we employed Google Earth Pro tools to identify sentinel site storm sewers in Dallas and Bexar. We then generated a Red, Green, and Blue (RGB) wavelength reflectance signature using Sentinel-2 visible and near Infrared, 10-meter spatial resolution data. We interpolated previously unknown storm sewer habitats in both entomological intervention counties. Remote validation models revealed sensitivity and specificity approaching 100 %. Currently we are implementing Larval Source Management (LSM) tactics from "Seek and Destroy" (e.g., targeted habitat modification/destruction) for treating these important foci.

Keywords: West nile virus; *culex quinquefasciatus*; storm sewers; seek and destroy; dallas and bexar; texas; second order autocorrelation

List of Abbreviations: WNV: West Nile Virus; RGB: Red, Green, and Blue; LSM: Larval Source Management; S&D: Seek and Destroy; U. S. A: United States of America; ITN : Insecticide-Treated Mosquito Nets; IRS: Indoor Residual Spraying; ySALA: Python Spatial Analysis Library; RE: Random Effect; SSRE: Spatially Structured Random Effect; SURE: Spatially Unstructured Random Effect; MCMC: Markov Chain Monte Carlo; NIR: Near Infrared; ODS: Output Delivery System; SAR: Spatial Autoregressive; UAV: Unmanned Aerial Vehicles; SWIR: Short-Wave Infrared; LULC : Land Use Land Cover

Introduction

West Nile Virus (WNV) is an arthropod vector-borne virus that belongs to the genus *Flavivirus* in the family *Flaviviridae*. WNV can be traced back to 1937 where it was isolated from the blood of a pregnant female patient in the West Nile District of Northwest Uganda [1]. The first observed occurrence of WNV in North America was in 1999 in New York City [2]. From 1999 and 2001, 149 cases were reported in humans across 10 states in the United States (U.S) [3,4]. The first widespread occurrence of WNV across the U.S. was documented in 2002, with 4,156 human cases reported across 40 states. Although WNV occurrences in Texas have only been observed since 2002, between 2002–2011 1,514 cases were reported to the Texas Department of State Health Services [5]. Caseload peaks occurring in 2006 and 2009 respectively [6]. Endemicity of WNV in Texas constitutes a significant threat to vulnerable human populations with Texas itself having the third largest elderly population in the United States [7]. Alleviation of these areas requires treatment of populations consisting of significant "bridge-vectors" such as *Cx. quinquefasciatus* and surveillance of reservoir species.

The primary vectors of WNV are primarily varied species of *Culex* mosquitos which includes Common House Mosquito (*Culex pipiens*), Eastern Encephalitis Mosquito (*Culex tarsalis*), and Southern House Mosquito (*Culex quinquefasciatus*). However, Asian Tiger Mosquito (Aedes albopictus), White Dotted Mosquito (*Culex restuans*), Unbanded Saltmarsh Mosquito (Culex salinarius), and Northern "Tree Hole Mosquito" (*Ochleratatus triceriatus*) are also known to carry WNV [8]. WNV is commonly transmitted by inoculation from common arthropod vectors in the southeastern U.S.A such as *Cx. quinquefasciatus* and other species of mosquitos that include the *Cx. Erraticus*, a prime vector for eastern equine encephalitis virus, as well as *Ae. Albopictus* [9]. Inoculation begins when an infected avian host, usually from the *Corvidae* family is bitten by an uninfected mosquito for a blood meal leading to the mosquito now being infected with WNV. This mosquito will then move towards either inoculation to transmit WNV to another avian host or become a "bridge-vector" [10]. Infected mosquitos subsequently transmit WNV to human hosts through the collection of blood meals for healthy embryonic development and release of ovum [11].

The salivary glands of mosquitos release the virus into the bloodstream of an uninfected host. 2 to 14 days after the incubation period symptoms will begin to manifest in the host in humans. 75%-80% of those infected with WNV will be asymptomatic or possess a mild fever [12]. In addition, 20% of individuals with WNV will develop flu like symptoms and less <1% of individuals will develop infections of the nervous system [13]. When severe WNV is present about 50% will be neuroinvasive while 50% will be non-neuroinvasive. *Cx. quinquefasciatus*, is enclosed within the *Cx. pipiens* species complex. This complex is known for the primary role it plays in the transmission of WNV in the southeastern United States [14]. *Cx. quinquefasciatus* resides in sub-tropical temperatures and has a broad thermal range from 14.1-32.2°C [15]. *Cx. quinquefasciatus* thrives at a temperature of 25.2°C and survivorship rates of mature *Cx. quinquefasciatus* is extremely high at temperatures around 25.2°C [16]. Subsequently, this temperature range increases infection rates of WNV in *Cx. quinquefasciatus* [17].

The life cycle of *Cx. quinquefasciatus* begins with the ovipositing of an egg raft on the top of the surface of a body of water. Temperatures between 24 – 29°C Care preferential for embryonic development of *Cx. quinquefasciatus*. Oviposited egg rafts are oval shaped in nature with about 100 to 300 larvae will be present after oviposition depending on climatic conditions. On average a single *Cx. quinquefasciatus* female will oviposit 155 ova [18]. In addition, a *Cx. quinquefasciatus* females can lay 3-5 egg rafts in a single lifetime [19]. The ideal water surface for oviposition is primarily a standing body of water with an abundance of organic material from which the ovum and subsequent larvae can obtain nutrients. These larvae in normal climatic conditions will hatch in an estimated 24-30 hours [20]. *Cx. quinquefasciatus* larvae prefer temperatures between 20-24 °C and experiences increased survivorship rates [21]. After completing successful larvae development *Cx. quinquefasciatus* briefly enters the pupa stage where emergence occurs in a few hours [22].

This process takes about 5-8 days in total with optimal climatic conditions. Climate can have profound consequences for mosquito arthropod diseases by altering the distribution and spread of vector-borne diseases. These include malaria, dengue and West Nile Virus by changing the ecological response, distribution, seasonality, and range [23-24]. It is of paramount importance to understand the interplay of socioeconomic and climatological conditions and how they influence mosquito habitats. Current literature lacks a clear understanding of the impact of varied species of mosquitoes on WNV outbreaks in Texas, particularly the analysis of combined influence of socioeconomic factors and climatological conditions on mosquito habitats. This article addresses the question of whether unknown foci of *Cx. quinquefasciatus* habitats can be mapped using interpolated sentinel signatures. The objective of this research is to implement Larval Source Management (LSM) real-time tactics associated with the "Seek and Destroy" (S&D) approach which can provide significant insights into new, real-time control measures for reducing larval vector density [Macro S&D] and blood parasite levels [Micro S&D]. "S&D" has already been employed in the treatment of malaria within suspected populations [25].

Utilization of these tactics involved a low-cost (< $1000) drone (DJI Phantom) for eco-geographically locating water bodies, including natural water bodies, irrigated rice paddies, cultivated swamps, ditches, ponds, and other common breeding sites for *Anopheles* mosquitoes in the Gulu district of Northern Uganda. The first hypothesis was the feasibility of integrating real-time, scaled-up sentinel site spectral signatures with UAV (Unmanned Aerial Vehicles) or drone imagery and satellite WorldView-2 (0.46 meter) data using geospatial artificial intelligence (geo-AI), infused into an iOS application. This would allow a local vector control officer to retrieve a ranked list of visually similar breeding sites for similar species of mosquito such as *Cx. quinquefasciatus* and their respective county-level capture points with GPS-indexed centroid coordinates. A smartphone app can map each georeferenced UAV-

sensed capture point by a mobile field team, consisting of trained local village residents led by a vector control officer. That same day habitats are signature mapped, spatially forecasted, and treated with larvicide.

Implementing a real-time Macro and Micro "S&D" intervention tool, along with other existing tools (insecticide-treated mosquito nets (ITNs) and indoor residual spraying (IRS), at a county-level intervention site can lower seasonal West Nile Virus prevalence through timely modification of aquatic *Cx. quinquefasciatus* larval habitats through precise targeted larvicide interventions. The hypothesis in this research is to utilize non-asymptotic, multicollinear, independent covariates and skewed heteroscedastic (unequal variance), non-Gaussian, county-level data (i.e., Dallas and Bexar) and interpolated Sentinel-2 visible and near-infrared (NIR) 10-meter resolution indexable RGB (Red, Green, and Blue) signature estimators to address violations of regression assumptions in eigenvector Eigen-Bayesian eigen-geospace. In so doing we would be able to establish precise georeferenced locations of unknown foci of *Cx. quinquefasciatus* habitats in the entomological intervention sites. We will also construct multiple vegetation and elevation, catchment watershed proxy satellite models using visible infrared and near-infrared bands to visualize values of co-variances.

In literature, extensive use of co-variances such as Normalized Difference Vegetation Index (NDVI), Digital Elevation Model (DEM), Modified Normalized Difference Water Index (MNDWI), and Land Use Land Cover (LULC) have all been utilized for interpolating various mosquito vector arthropod signatures of *A. arabiensis* [26], *Ae. Albopictus* [27], and *Ae. Aegypti* [28]. Normalized Difference Vegetation Index (NDVI), Modified Normalized Difference Water Index (MNDWI), Soil-Adjusted Vegetation Index (SAVI), Digital Elevation Model (DEM), and Land Use and Land Cover (LULC) are essential remote sensing indices and classifications used in environmental monitoring and analysis. NDVI, developed to quantify vegetation density and health by comparing the near-infrared and red-light reflectance from vegetation, helps in identifying and analyzing green biomass [29]. MNDWI is used for enhancing the presence of open water bodies while suppressing the noise from built-up land areas, vegetation, and soil, which improves water feature extraction in various landscapes [30]. SAVI, on the other hand, is an adaptation of NDVI designed to minimize the influence of soil brightness in areas where vegetation cover is low, by incorporating a soil brightness correction factor, making it particularly useful in arid and semi-arid regions [31].

DEMs provide 3D representations of terrain elevations, offering crucial data for topographic analysis and watershed management [32]. LULC classifications provide comprehensive information on the distribution and changes in land use and cover types, which is crucial for understanding human-environment interactions and managing natural resources [33]. Together, these indices and classifications play a crucial role in eco-hydrological studies, providing vital information for sustainable environmental management. Specifically, in this experiment we are employing NDVI for storm sewer *Cx. quinquefasciatus* mosquitoes due to the correlation between vegetation abundance and the propensity of *Culex* mosquitoes to thrive in vegetative environments. The importance of vegetation is underscored by studies showing that ecotones, areas where different types of vegetation meet, can influence mosquito populations. For example, *Cx. tarsalis* mosquitoes were found to be more abundant in ecotones with high vegetation, as they provide suitable habitats for these mosquitoes and their hosts. In this research, community-level spatial models were developed for determining mosquito abundance and environmental factors that could aid in the risk prediction of WNV outbreaks [34]. We used various covariates like DEM, which provides crucial data on the topography of the study area.

This information helps in understanding how elevation and terrain can affect mosquito habitat breeding sites. NDVI was employed to assess the density and health of vegetation, which is a critical factor since *Culex* mosquitoes are often found in vegetative areas where they can find both breeding sites and hosts. LULC data was used to classify different land types and their usage, revealing how built environments, forests, and water bodies contribute to mosquito habitat suitability. Additionally, the MNDWI was used to identify water bodies, which are potential breeding sites for immature *Cx. quinquefasciatus* mosquitoes. This index is particularly useful in distinguishing water features from built-up areas and vegetation. By integrating these indices—NDVI, DEM, LULC, and MNDWI—into our models, we can accurately predict mosquito abundance and potential hotspots for WNV outbreaks, leveraging the strengths of each type of data to provide a comprehensive understanding of the environmental factors influencing mosquito populations. Furthermore, an eigen-Bayesian second-order Markovian, semi-parametric autocorrelation model was tested using eigen-spatial filter and eigen algorithm to identify outliers in the interpolated signature data in Python.

Given that *Cx. quinquefasciatus* is a primary bridge vector for WNV, treating areas with high densities of *Cx. quinquefasciatus* larvae is advisable, with the only restriction being the identification of habitats. *Cx. quinquefasciatus* has already been observed to oviposit inside storm sewer habitats. For example, immatures of multiple vector-competent mosquito species present in these habitats, totaling over 64,560 larvae in 1,761 storm sewer drains in one study site [35]. Identifying autocorrelation would allow the forecast of potential georeferenceable hot and cold spots of *Cx. quinquefasciatus* storm sewer larval habitats. Therefore, the objective in this article is 1. To identify storm sewer *Cx. quinquefasciatus* habitats using google earth mapping tools. 2. Generate a spectral signature of a sentinel site. 3. To interpolate the signature. 4. To develop satellite proxy maps. 5. To develop a customized "S&D" for the intervention counties.

Materials and Methods

Study Sites

Dallas County, Texas, is the second most populous county in the state, with an approximate population of 2,613,539 people as of 2020 [36]. Covering an area of 2,350 km^2, Dallas County is characterized by extensive urban land use due to the land being

predominantly zoned for metropolitan purposes [37]. The Dallas/Fort Worth Metroplex, located in North Central Texas, features a mixture of prairie and oak woodlands, though much of the natural landscape has been overtaken by urbanization and agriculture. The region's climate is humid subtropical with hot summers and considerable annual temperature variation. Precipitation varies from less than 20 inches to more than 50 inches annually, typically favoring wet periods in spring and fall and dry periods in summer and winter. The area experiences mild winters with periodic cold fronts, and the urban heat island effect reduces the number of freezes compared to surrounding rural areas. Summers are characterized by high temperatures often exceeding 100 degrees, high humidity, and brief afternoon thunderstorms. Spring and fall are the wettest periods, with frequent thunderstorms that can bring severe weather such as hail, damaging winds, flooding, and tornadoes.

The county's urbanization, combined with its climate, creates numerous artificial habitats for *Cx. quinquefasciatus*, the primary vector for WNV in Texas [38,39]. Bexar County, Texas, located in the south-central part of the state, is home to a population of approximately 2,009,324 people as of 2020. Spanning an area of 3,250 km^2, Bexar County features significant urban land use concentrated in the city center of San Antonio [40]. The outer areas of the county are primarily rural, suburban, or in transition between rural and urban, creating a diverse range of land covers. This variation significantly affects the patterns of *Cx. quinquefasciatus*, the primary vector for WNV in Texas. The climate of Bexar County has seen a 2.4°F increase in the 12-month average temperature from May 1900 to April 2023, with the average temperature over this period being 68.7°F [41]. This warming trend, coupled with the diverse land cover, impacts the distribution and abundance of *Cx. quinquefasciatus* across the county, making Bexar County a critical area for studying the environmental factors influencing mosquito populations and the spread of WNV.

Spatial Analysis

This article presents two space-time model specifications, one based upon the generalizable, linearized, mixed model and the other upon Moran's eigenvector space-time filters. Moran's coefficient is an index of spatial autocorrelation, involving the computation of cross products of mean-adjusted values that are geographic neighbors (i.e., covariations). These ranges are roughly (-1, -0.5, 0) to nearly 0 for negative, and nearly 0 to approximately 1 for positive, spatial autocorrelation, with an expected value of -1/(n-1) for zero spatial autocorrelation, where n denotes the number of areal units [42]. We identify signature algorithms to fit an RGB satellite signature of a georeferenced storm sewer *Cx. quinquefasciatus* aquatic larval habitat, for a training dataset. We did so to learn more about how regression functions characterized spilled over hierarchical diffusion of interpolated, non-Gaussian, *Cx. quinquefasciatus* storm sewer capture point habitats. Our objective was to predictively prioritize and target, georeferenecable, hyper/hypo endemic aggregation/non-aggregation oriented, potential transmission sites based on the occurrence, abundance, and distribution of the interpolated *Cx. quinquefasciatus* storm sewer for implementing "S&D".

A Moran spatial filtering technique was generated by employing Python code using GeoPandas and LibPySAL. The code performed an eigenfunction, second order eigen-spatial filter eigen-decomposition of the random effects (REs) on the varying potential, capture point, *Cx. Quinquefasciatus*, storm sewer, larval habitats. Spectral emissivity determinants rendered (SSRE) and spatially unstructured random effect (SURE) components. The RE model incorporated synthetic eigen-spatial filtered signature capture points at the county level. The eigen-orthogonalized, eigenfunction eigenvectors derived from a geographic connectivity matrix were used to account for SSRE and SURE using standardized z scores stratified by zip codes due to geo-spatiotemporal spill-over from the interpolated Sentinel-2, 10-meter *Cx. quinquefasciatus* RGB signatures. We then calculated the conditional probabilities for each county zip code using the conditional distribution functions derived from the regressed diagnostic spectral interpolated RGB signature habitat determinants.

This included the probability density function, cumulative density function, and quantile function. Subsequently, a random variable mean response specification was written in scikit-learn as follows: esitk and eHith were the ith elements of the K < NT and H < NT selected eigenvectors, respectively. Estk and EHth were parsimoniously extracted from the doubly centered space-time signature capture points, which included sentinel site, zip code stratified areas, and oviposition-related storm sewer habitats. The expectation attached to the equation, (i.e., RE ≡ SURE) was satisfiable both having trivial SSRE components. For the eigen-Bayesian post-treatment context, the SSRE component was also modeled in scikit-learn with a conditional autoregressive specification which revealed residual non-zero autocorrelation (i.e., geographic non-chaos). This component also revealed non-homoscedasticity, non-asymptotic behavior, and Gaussian non-multicollinearity in the zip code stratified georeferenced clustering propensities of *Cx. quinquefasciatus* storm sewer larval habitats.

The model's variance implied a substantial variability in the prevalence of storm sewer habitat occurrence, abundance, and distribution across each entomological intervention site due to the hierarchical diffusion of the interpolated capture point sentinel sites. We surmised that using an interpolated RGB signature eigen-Bayesian, semi-parametric Markovian, non-frequentist model would reveal non-Gaussian residual error amplification due to violations of regression assumptions [i.e., geospatial heteroscedasticity, multicollinearity, and or zero autocorrelation]. Hence, we assumed that regardless of whether the true processes in our storm sewer signature model were explicitly eco-geographical, the spatial relationships among our georeferenced, *Cx. Quinquefasciatus*, storm sewers, larval habitats would be informative. We assumed these relationships along with aggregation/non-aggregation-oriented entomological foci and satellite-sensed, georeferenced signature RGB interpolation would reveal more information about nearby potential storm sewer breeding sites.

In this experiment we built space into the traditional entomological prognosticative model for Dallas and Bexar counties for the intended usage of constructing a robust georeferencable, zip code stratified, scalable, *Cx. quinquefasciatus* larval habitat regression model framework using Python Spatial Analysis Library (ySALA). The construction began with a standard linear regression model devoid of any geographical reference, after which we formalized spatial relationships in three main ways. First, we encoded an empirical dataset of synthesized Sentinel-2 visible and NIR data, as well as signaturized exogenous explanatory capture points. Secondly, we considered spatial heterogeneity as the systematic variation of scalable, georeferencable, autoregressive, eigen-spatial filter eigen-algorithmic outcomes across eigenvector eigen-Bayesian eigen-geospace. Third, we examined dependence and the effects associated with the characteristics of neighboring storm sewer habitats in both intervention counties. Throughout the analysis, we focused on the conceptual differences on how each approach entails rather than on the technical details.

RGB interpolation of *Cx. quinquefasciatus* storm sewer habitats

Maximum likelihood classification, a remote sensing technique, employs Bayesian methods and multivariate Gaussian distribution [43-44]. It is a widely used and popular method in remote sensing, which has been in place for many decades [45-46]. To determine if a real time, sensed, sentinel site, storm sewer, capture point, georeferenced, *Cx. quinquefasciatus* habitat breeding sites pixels in the satellite data was part of a given RGB-LULC. The probability of this occurring was definable by the posterior probability $P(P_i \mid x) = \frac{P(P_i)P(x \mid P_i)}{P(x)}$. Where $P(P_i)$ represented the probability of the i-th class and $P(x)$ represented the probability of a sensed pixel. $P(x \mid P_i)$ which in this experiment was defined by where x was the feature vector of the georeferenced, oviposition capture point.

In the prognosticative geospatial analysis, pixel μ_i was the mean vector of the i-th class features Σ_i was the covariance matrix of the i-th signature, RGB-LULC, class features, Σ_i^{-1} was the inverse covariance matrix of the -th class features and was the determinant of the covariance matrix of the i-th class attributes. This process has already been used to evaluate land cover feature attributes in Dacca, Bangladesh for identifying unknown georeferenceable *Ae. aegypti*, mosquito vector of dengue, habitats signatures and implementation of "S&D" at the state level, using the real time sensed Sentinel-2 visible and NIR satellite imagery.

Eigenvector Analysis

Initially, we developed a misspecification perspective for the model incorporating potentially eigenizable capture points of georeferenceable storm sewers *Cx. quinquefasciatus* habitats. This interpolation model was constructed using the acf () function from the stats model's library in Python. We specified the number of lags to use with the `nlags` argument, assuming the georeferenced risk-based storm sewer larval habitat-related parameter fit was $y = X\beta + \varepsilon^*$ (i.e., a standard regression equation). The primary function of our RGB interpolation model generation was to identify agenized hot/cold spots at the county level in georeferenceable locations, and to detect noise (atmospheric, sensor, geometric) caused by violations of regression assumptions (i.e., non-Gaussian zero autocorrelation, often termed 'geographic chaos'). This analysis focused on explanatory factors such as hyper/hypo endemic conditions, aggregation/non-aggregation-oriented patterns, and georeferenced determinants derived from capture points and sentinel sites. We forecasted RGB signature georeferenced scalable capture points for storm sewer *Cx. quinquefasciatus* larval habitats with Matplotlib, Pandas, Geopandas and Basemap in Python.

The mapping functionalities in Python allowed us to create choropleth, prism, block, and surface maps along with the RGB signature interpolation maps. Three key techniques were emphasized to generate regressed county-level maps related to storm sewers. These techniques included Map Data, Annotation, and the Output Delivery System (ODS) in Python. The stratified latent coefficients were optimally eigendecomposed into a white-noise component, ε, and a set of unspecified model outputs that had the structure $y = XB + \underbrace{E_\gamma + \varepsilon}$. White noise is a univariate or multivariate discrete-time stochastic process, whose terms are independently distributed with a zero mean [47]. The annotate facility enabled us to generate a special dataset of graphics commands from which we created the potential interpolated, RGB signature, and model annotate output. The annotate output created multiple customizable, signature topographic maps of capture points, sentinel sites, georeferenced county-level areas, and zip code stratified LULCs. The misspecification term was $S(T) = S(t)\exp$ the interpretive land cover patterns were quantified based on the distribution of various determinants derived from the regressed, non-time-series, dependent variables.

These determinants, which included hyper/hypo endemic, transmission-related, stratified, and aggregation/non-aggregation-oriented factors were georeferenced and related to storm sewers. In this experiment, independent key dimensions of the underlying uncertainty processes required analyzing empirical Geosampled data at the sentinel site and county level; as well as interpolated larval habitat data capture points. We were able to define a pattern in the misspecification term. Python provided an efficient interactive tool for organizing and analyzing the georeferenced stratified, aggregation/non-aggregation-oriented, hot/cold spots. Python is efficient at finding sentinel site entomological county-level capture points that were georeferenced non-geo-spatiotemporally [48]. Next, we generated an autoregressive model for *Cx. quinquefasciatus* at the county level, incorporating spatialized hyper/hypo endemic and aggregation/non-aggregation-oriented feature attributes. This model utilized RGB signature, capture points, and sentinel sites all scaled up to cover a larger geographical area.

We employed an asymptotical, stratified variable Y as a function of a nearby regressor Y in the epidemiologic hot/cold spot model. A capture point at a sentinel site, representing storm sewer habitats for *Cx. quinquefasciatus*, was stratified residually and subjected to eigen spatial filtering to determine

estimator determinants. The explanatory indicator value I (an autoregressive response) and the residual of Y were treated as a function of a nearby Geosampled Y residual. This situation follows a spatial error specification or SAR (Spatial Autoregressive) model for autoregressive, dependent, hot/cold spot modeling. The autocorrelation model furnishes an alternative specification that frequently is written in terms of matrix W. Here, the spatial covariance of the potential georeferenceable data was a function of the matrix $(I-\rho CD^{-1})(I-\rho D^{-1}C)=(I-\rho W^{T}W)(I-\rho W)$, where T denoted the matrix transpose. The resulting matrix was symmetric and was considered a second-order specification as it included the product of two spatial structure matrices $(i.e.,W^{T}W)$. This matrix restricted positive v remote sample discrete integer values of the autoregressive parameter to the more intuitively interpretable range of $0\le\hat{\rho}\le 1$.

Euclidean distances between the georeferenced sentinel site capture point, scaled-up to county-level, *Cx. quinquefasciatus* storm sewer larval habitat estimator determinants were definable in terms of an n-by-n geographic weights matrix C. These c_{ij} values were, 1 if the geosampled hold/cold spot geolocations i and j were deemed nearby; and 0 otherwise. Adjusting this matrix by dividing each row entry by its row sum subsequently rendered C1, where 1 was an n-by-1 vector of ones which converted the regression-based matrix to matrix W. The resulting Python model specification with no georeferenced scaled-up sentinel site capture point, estimator determinants (i.e., the pure spatial autoregression specification) subsequently took on the following form: $Y=\mu(1-\rho)1+\rho WY+\varepsilon$, where μ was the scalar conditional mean of Y and ε was an n-by-1 error vector. The parameters were statistically independently, identifiable, distributable, and normalized random variates. Spatial autoregressive models are fitted using datasets that contained observations on geographical areas or on any units with a spatial representation [49].

Approximate standard errors were calculated for the eigenizable TB estimator determinant in the county-level. Stratified vulnerability RGB signature model which is geo-spatially specifiable was computable as the square roots of the diagonal elements of the estimated covariance matrix. The covariance matrix for analyzing the capture point, sentinel site, *Cx. quinquefasciatus* storm sewer estimator determinants were expressible employing $=\Sigma=\left[(I-\rho W')(I-\rho W)\right]^{-1}\sigma^{2}$, where $E(\bullet)$ designated the calculus of expectations. It was the n-by-n identity matrix denoting the matrix transpose operation and σ^{2} was the asymptotical error variance [i.e., the variance of how far the geosampled, non-homogenous, hyper/hypoendemic, aggregation/non-aggregation-oriented, interpolated storm sewer habitat estimator determinants were spread out, eco-geographically]. Subsequently, we employed a Hessian matrix in Python; we quantified an empirical, georeferenced dataset at the county level. This dataset included stratified, hyper/hypoendemic, hot spot areas, focusing on entomological observations related to *Cx. quinquefasciatus* storm sewer larval habitats. As in Newton's method, we employed a second order approximation to find the minimum of a function $f(x)$ in the geosampled, estimator, determinant dataset.

We employed a Taylor series of $f(x)$ to generate $f(x_k+\Delta x)\approx f(x_k)+\nabla f(x_k)^{T}\,\Delta x+\frac{1}{2}\Delta x^{T}B\Delta x,$ where (∇f) the gradient was B, which in this experiment approximated the Hessian matrix. In mathematics, the Hessian matrix or Hessian is a square matrix of second-order partial derivatives of a function [50]. This matrix described the local curvature of a function of the hyper-hypoendemic, *Cx. quinquefasciatus* storm sewer larval habitat estimator determinants. Given the function, $f(x_1,x_2,...,x_n)$, if all second partial derivatives of existed and were continuous over the domain of the function, then the Hessian matrix off was derivable from $H(f)_{ij}(x)=D_iD_jf(x)$, where $x=(x_1,x_2,...,x_n)$ and D_i were the differentiation operator with respect to the i-th argument. The matrix rendered in Python was

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

We noted the determinant of the sensitive matrix in our

scaled-up, epidemiologic prognosticative model was Hessian. In image analysis, the Hessian matrix can describe the second-order variations of local image intensity [i.e., hyper/hypoendemic, aggregation/non-aggregation-oriented, *Cx. quinquefasciatus* stratified geolocation around a scaled-up, capture point, sentinel site, classified, LULC, stratified, reflectance pixel] thereby encoding the shape information.

In practice, it is computable by convolving an image with second derivatives of the Gaussian kernel in the x- and y-directions in the literature. The matrix described the local curvature forecasted and spatial structures over a whole county. The Hessian matrix is suitable for detecting the shape of hot/cold spots related to vector arthropods at the epidemiologic, county-level, capture points, and sentinel sites [51]. We assumed the Gaussian kernel to have a standard deviation of 1, which was applied to each capture point pixel in a georeferenced intervention county satellite, Sentinel-2, 10-meter spatial resolution image so that the Hessian matrix was expressible as $Hf=HxxHxyHyy$. In this experiment, the Hessian matrix was relatable to the Jacobian matrix by $H(f)(x)=J(\nabla f)(x)$. In vector calculus, the Jacobian matrix of a vector-valued function of several variables is the matrix of all its first-order partial derivatives [52]. When this matrix is square, that is, when the function takes the same number of variables as input as the number of vector components of its output, its determinant is referred to as the Jacobian determinant.

Both the matrix and the determinant are often referred to simply as the Jacobian in literature. The Jacobian of a vector-valued function in the georeferenced storm sewer *Cx. quinquefasciatus*

related estimator determinants generalized the gradient of a scalar-valued function. This gradient, in turn, generalized the derivative of every entomological estimator in the geosampled discrete integer values of the empirical dataset. In other words, the Jacobian matrix of a scalar-valued function in the epidemiologic risk model, when transposed, represents the gradient. This gradient, in turn, reflects the derivative of a scalar-valued function for a single, georeferenceable, aggregation/non-aggregation-oriented explanatory variable at capture point, sentinel sites, at the county-level. Furthermore, differentiation was possible at each sentinel site. The Jacobian described the amount of "stretching", "rotating", or "transforming" that the function imposed locally near that capture points.

For example, here $(x', y') = f(x, y)$ was employed to smoothly transform a evidential prognosticator in the Jacobian matrix $Jf(x, y)$ which in this case described how the geosampled, georeferenced, county-level, estimator, determinant in the model performed. In our semi-parametric, Markovian, eigen-Bayesian, county-level, prognosticative, epidemiological, storm sewer *Cx. quinquefasciatus* habitat model θ was a parameter that was unknown and had to be inferred from the, geosampled, capture point epidemiologic empirical data. The joint probability of the data was optimally quantifiable by employing:

$$p(X \mid \theta) = p(x_1, \ldots x_n \mid \theta) = \prod_{i=1}^{n} p(x_i \mid \theta);$$

whereby, $p(X \mid \theta, \alpha) = p(X \mid \theta)$ and $p(x_i \mid \theta, \alpha) = p(x_i \mid \theta)$ was conditionally independent of the hyperparameters. Eigen-Bayesian semiparametric, Markovian, non-Gaussian inference was determined by the posterior distribution of the parameter $p(\theta \mid X, \alpha)$ by employing:

$$p(\theta \mid X, \alpha) = \frac{p(\theta, X, \alpha)}{p(X, \alpha)} = \frac{p(\theta, X, \alpha)}{\int_\theta p(\theta, X, \alpha) d\theta} = \frac{p(X \mid \theta, \alpha) p(\theta \mid \alpha)}{\int_\theta p(X \mid \theta, \alpha) p(\theta \mid \alpha) d\theta} = \frac{p(X \mid \theta) p(\theta \mid \alpha)}{\int_\theta p(X \mid \theta) p(\theta \mid \alpha) d\theta} = \frac{\left[\prod_{i=1}^{n} p(x_i \mid \theta)\right] p(\theta \mid \alpha)}{\int_\theta \left[\prod_{i=1}^{n} p(x_i \mid \theta)\right] p(\theta \mid \alpha) d\theta}.$$

A second-order Gaussian random walk prior was employed to allow enough flexibility while penalizing abrupt changes in the function. The second-order random walk model is commonly used for smoothing data and for quantifying response functions as it is computationally efficient due to the Markov properties of the joint (intrinsic) Gaussian density [53].

Results and Discussion

RGB Signature Results

The RGB signature of a storm sewer habitat for *Cx. quinquefasciatus* was captured, providing a detailed spectral profile across various Sentinel-2 bands (Figure 1). The spectral profile shows the following RGB values: Coastal Aerosol (Band 1) at 0.2, Blue (Band 2) at 0.3, Green (Band 3) at 0.25, Red (Band 4) at 0.35, Vegetation Red Edge 1 (Band 5) at 0.32, Vegetation Red Edge 2 (Band 6) at 0.28, Vegetation Red Edge 3 (Band 7) at 0.3, NIR (Band 8) at 0.1, Narrow NIR (Band 8) at 0.05, Water Vapor (Band 9) at 0.08, SWIR-Cirrus (Band 10) at 0.12, and SWIR (Bands 11 and 12) at 0.05 (Figure 2). The spectral profile indicates a peak reflectance in the red band (Band 4) with an RGB value of 0.35, while the lowest reflectance is observed in Narrow NIR (Band 8) and SWIR (Band 12) with RGB values of 0.05.

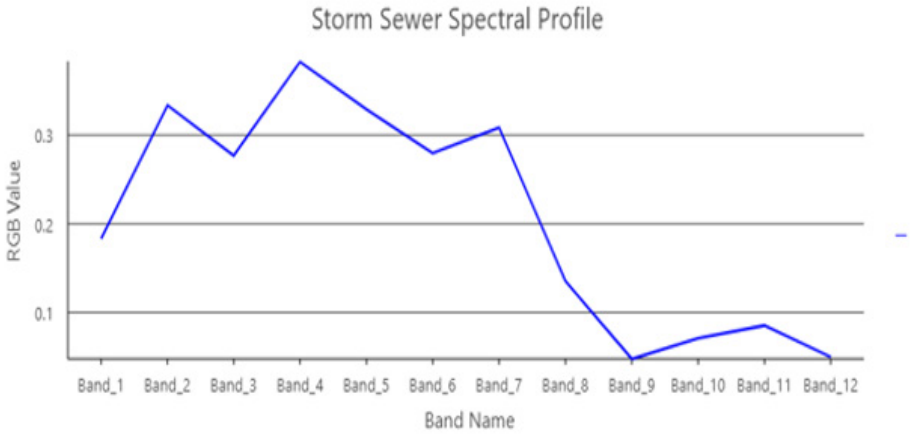

Figure 1: Spectral profile for Storm Sewers in Dallas Texas.

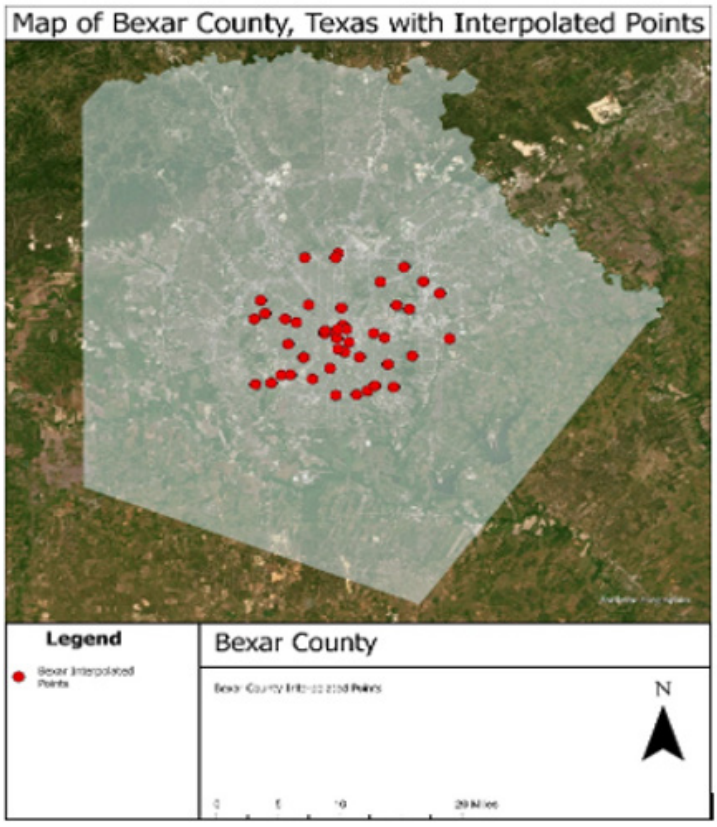

Figure 2: Interpolation Results of Bexar County, Texas.

Stochastic Interpolation

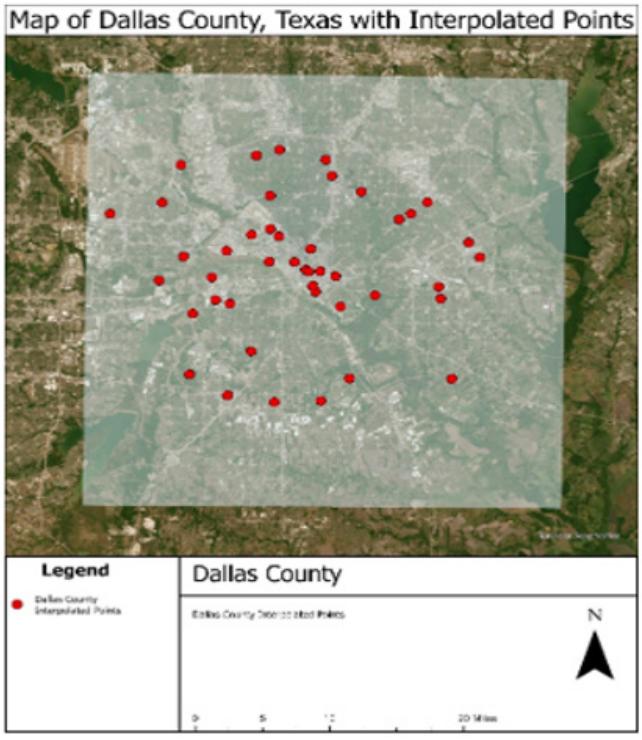

Figure 3: Interpolation Results Dallas County, Texas.

The interpolation algorithm was successful providing results for Dallas and Bexar counties. Bexar had 25 original points and 20 newly interpolated points. These newly interpolated points are potential habitats for *Cx. quinquefasciatus* larvae. Dallas County had 25 original points too with 19 newly interpolated points after running the eigen-spatial filter eigen algorithm (Figure 3).

Eigen-Bayesian Hot/Cold Spots Results

The analysis of hot and cold spots in Bexar and Dallas Counties revealed that in Bexar County, hot spots with 95% and 99% confidence are located in central and northeastern areas, while cold spots with similar confidence levels are in the southern area. Bexar County also displayed potential aggregation of storm sewer habitats in areas of high elevation in the northwest region of San Antonio. In Dallas County, the NDVI-based map showed hot spots concentrated in the central region, particularly west of the city center, with cold spots scattered in the eastern parts. The DEM-based analysis for Dallas County indicated a central clustering of cold spots with high confidence, while hot spots were found in the southern and northeastern regions. Non-zero autocorrelation was examined and identified after employing eigen-Bayesian semiparametric, Markovian, Hot/Cold spots analysis for the interpolated points. Dependent variables had significant correlations between *Cx. quinquefasciatus* storm sewer habitat aggregation/non-aggregation. Due to the findings of non-zero autocorrelation, we can assume heteroscedasticity and multicollinearity are both absent (Figures 4-7).

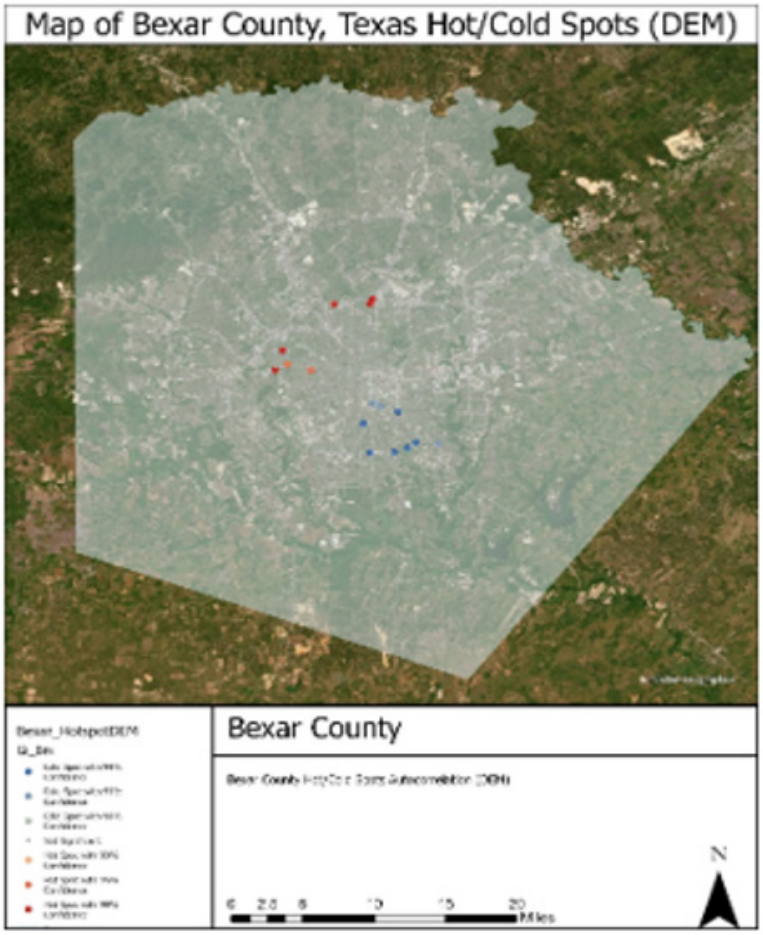

Figure 4: Eigen-Bayesian Fixed (DEM) Results for Bexar County, Texas.

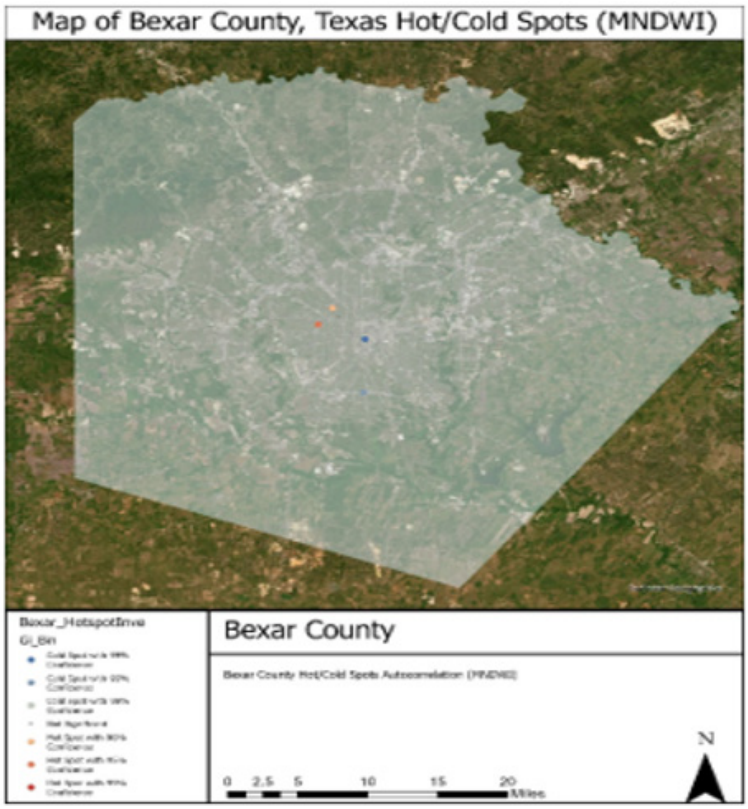

Figure 5: Eigen-Bayesian Inverse Distance Squared (MNDWI) Results for Bexar County, Texas.

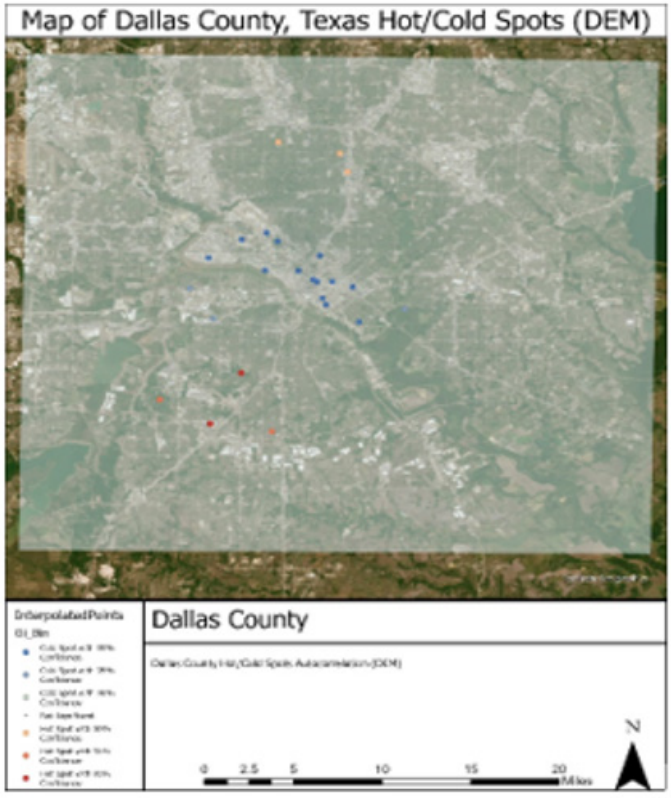

Figure 6: Eigen-Bayesian Fixed (DEM) Results Dallas County, Texas.

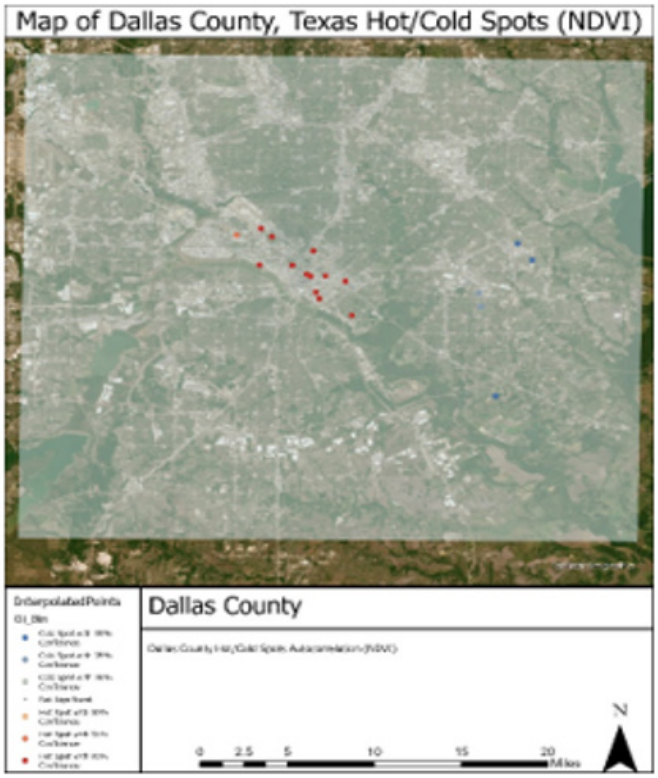

Figure 7: Eigen-Bayesian Fixed (NDVI) Results Dallas County, Texas.

Results of Moran's I Autocorrelation

Second-order spatial autocorrelation analysis were conducted for all five variables, DEM, NDVI, MNDWI, LULC, and SAVI, for Bexar and Dallas counties to detect the clustering propensities and hot/cold spots across geospatial locations. All three types, inverse distance, inverse distance squared, and fixed distance, were tested to reveal the model robustness and the most optimal technique for spatial autocorrelation. Moran's Index for DEM shows a strong positive spatial autocorrelation across all three methods (Tables 1&2), with values of 0.69 (inverse distance), 0.78 (inverse distance squared), and 0.82 (fixed distance). The inverse distance method revealed the highest Moran's index with a Z-score of 7.35 with a P-value of less than 0.05, indicating Moran's index is significantly different than the expected randomness (Griffith, 2003), thereby revealing statistically significant spatial autocorrelation. The results of inverse distance and inverse distance squared showed a statistically significant spatial autocorrelation (P-value<0.05). However, the fixed distance method showed the highest Moran's index (I=0.82) due to less non-homogeneity in the elevation values across Bexar County. Comparable results were detected for Dallas County with a Moran's index of 0.43 with a P-value of less than 0.05, indicating statistically significant spatial autocorrelation.

Table 1: Moran's, I value, z-score, and p-value coefficients for Bexar County, Texas.

s	DEM	NDVI	MNDWI	LULC	SAVI
Inverse Distance Values	Moran's Index 0.690739	Moran's Index 0.253519	Moran's Index 0.388133	Moran's Index 0.201381	Moran's Index 0.272512
	z-score 8.122145	z-score 2.925978	z-score 4.244192	z-score 2.306759	z-score 3.210838
	p-value 0.000000	p-value 0.003434	p-value 0.000022	p-value 0.021068	p-value 0.001323
Inverse Distance Squared Values	Moran's Index 0.781775	Moran's Index 0.327540	Moran's Index 0.438034	Moran's Index 0.239313	Moran's Index 0.352232
	z-score 8.400416	z-score 3.074162	z-score 3.944057	z-score 2.235016	z-score 3.378827
	p-value 0.000000	p-value 0.002111	p-value 0.000080	p-value 0.025416	p-value 0.000728

Fixed Distance Values	Moran's Index 0.827579	Moran's Index 0.186630	Moran's Index 0.324006	Moran's Index 0.143842	Moran's Index 0.202450
	z-score 7.357146	z-score 2.417691	z-score 3.904900	z-score 1.869173	z-score 2.670127
	p-value 0.000000	p-value 0.015619	p-value 0.000094	p-value 0.061599	p-value 0.007582

List of Abbreviations: DEM: Digital Elevation Model; NDVI: Normalized Difference Vegetation Index; MNDWI: Modified Normalized Difference Water Index; LULC: Land Use Land Cover; SAVI: Soil Adjusted Vegetation Index.

Table 2: Moran's, I value, z-score, and p-value coefficients for Dallas County, Texas.

	DEM	NDVI	MNDWI	LULC	SAVI
Inverse Distance Values	Moran's Index 0.506616	Moran's Index 0.520591	Moran's Index 0.492859	Moran's Index 0.155518	Moran's Index 0.196176
	z-score 6.313416	z-score 6.435024	z-score 6.10355	z-score 2.083073	z-score 2.984458
	p-value 0	p-value 0	p-value 0	p-value 0.037245	p-value 0.002841
Inverse Distance Squared Values	Moran's Index 0.571401	Moran's Index 0.611794	Moran's Index 0.566116	Moran's Index 0.331703	Moran's Index 0.199066
	z-score 5.270352	z-score 5.589394	z-score 5.184529	z-score 3.076581	z-score 1.962689
	p-value 0	p-value 0	p-value 0	p-value 0.002094	p-value 0.049682
Fixed Distance Values	Moran's Index 0.437698	Moran's Index 0.413956	Moran's Index 0.406534	Moran's Index 0.000571	Moran's Index 0.196176
	z-score 6.293987	z-score 5.928479	z-score 5.824667	z-score 0.318164	z-score 2.984458
	p-value 0	p-value 0	p-value 0	p-value 0.750361	p-value 0.002841

List of Abbreviations: DEM: Digital Elevation Model; NDVI : Normalized Difference Vegetation Index; MNDWI: Modified Normalized Difference Water Index; LULC: Land Use Land Cover; SAVI: Soil Adjusted Vegetation Index.

However, inverse distance and inverse distance squared methods showed the highest Moran's index; only the fixed distance method revealed the hot and cold spots of similar elevation. Inverse distance methods were not able to reveal significant hot/cold spots due to the absence of homogeneity in the elevation values across geospatial locations. Moran's Index for NDVI demonstrates positive spatial autocorrelation across all three analytical methods (Table 1), with values of 0.25 (inverse distance), 0.32 (inverse distance squared), and 0.18 (fixed distance) for Bexar County. The highest Moran's Index was observed with the inverse distance square method, yielding a Z score of 3.07 and a P-value less than 0.05, suggesting a statistically significant deviation from expected randomness, thus confirming the geospatial clusters in NDVI patterns. This inverse distance squared method assigns the highest weights to the nearest points, thereby having more chances of revealing a high Moran's index than the fixed distance where the weight of all the NDVI values across geospatial locations is equal [54]. Similarly, the highest Moran's index (0.61) was revealed for Dallas County for the inverse distance squared method with a P-value of less than 0.05. However, the inverse distance squared method failed to detect statistically significant hot/cold spots.

The inverse distance squared method assumes that nearby ground-referenced points have a more considerable influence (ability to detect hot spots) on each other than distant points. However, the impact of revealing hot/cold spots diminishes when comparing geospatial coordinates of longer Euclidean distances in inverse distance and inverse distance square method due to the assumption of detecting clusters of similar values and hot spots for the least Euclidean distance between geospatial coordinates. However, fixed distance revealed statistically significant hot/cold spots at a 95% confidence interval due to the assumption of equal probability of detecting hot/cold spots, unlike in the Inverse Distance Method. Moran's Index for MNDWI shows a strong positive spatial autocorrelation across all three methods for both Bexar and Dallas counties. In Bexar County, the inverse distance method

resulted in a Moran's Index of 0.38 with a Z-score of 4.24 and a P-value of less than 0.05 at a 95% confidence interval. The inverse distance squared method showed a higher Moran's Index of 0.43, with a Z-score of 3.94 and a significant P-vale (<0.05). The fixed distance method revealed a Moran's Index of 0.32, with a Z-score of 3.90. For Dallas County, the inverse distance method yielded a Moran's Index of 0.492859, a Z-score of 6.10355, and a P-value of 0.

The inverse distance squared method showed the highest Moran's Index of 0.566116, with a Z-score of 5.184529 and a P-value of 0. The fixed distance method resulted in a Moran's Index of 0.406534, a Z-score of 5.824667, and a P-value of 0. These results indicated statistically significant spatial autocorrelation for all methods in both counties. Moran's Index for LULC indicates varying degrees of positive spatial autocorrelation across different methods for both Bexar and Dallas counties. In Bexar County, the inverse distance method produced a Moran's Index of 0.201381, with a Z-score of 2.306759 and a P-value of 0.021068. The inverse distance squared method showed a slightly higher Moran's Index of 0.239313, a Z-score of 2.235016, and a P-value of 0.025416. The fixed distance method had the lowest Moran's Index of 0.143842, with a Z-score of 1.869173 and a P-value of 0.061599. For Dallas County, the inverse distance method resulted in a Moran's Index of 0.155518, a Z-score of 2.083073, and a P-value of 0.037245. The inverse distance squared method showed a higher Moran's Index of 0.331703, with a Z-score of 3.076581 and a P-value of 0.002094. The fixed distance method yielded an incredibly low Moran's Index of 0.000571, with a Z-score of 0.318164 and a P-value of 0.750361.

These results revealed statistically significant spatial autocorrelation for most methods, except for the fixed distance method in both counties regarding LULC. Moran's Index for SAVI demonstrated positive spatial autocorrelation across different methods for both Bexar and Dallas counties. In Bexar County, the inverse distance method yielded a Moran's Index of 0.272512, with a Z-score of 3.210838 and a P-value of 0.001323. The inverse distance squared method showed a higher Moran's Index of 0.352232, a Z-score of 3.378827, and a P-value of 0.000728. The fixed distance method resulted in a Moran's Index of 0.202450, a Z-score of 2.670127, and a P-value of 0.007582. For Dallas County, the inverse distance method had a Moran's Index of 0.196176, a Z-score of 2.984458, and a P-value of 0.002841. The inverse distance squared method produced a Moran's Index of 0.199066, a Z-score of 1.962689, and a P-value of 0.049682. The fixed distance method, matching the inverse distance method, had a Moran's Index of 0.196176, a Z-score of 2.984458, and a P-value of 0.002841. These results indicated statistically significant spatial autocorrelation for most methods, except for the inverse distance squared method in Dallas County which barely meets the threshold for significance.

Results of Linear Regression

These results pertain to the DEM analysis where sampling across four chains with 10,000 tuning iterations and 20,000 draw iterations (a total of 40,000 tuning and 80,000 draw iterations) took 833 seconds. The rhat statistic exceeded 1.01 for some parameters, indicating convergence problems, and the effective sample size (ESS) per chain was less than 100 for certain parameters, suggesting insufficient sampling for reliable rhat and ESS computations. Chain 0 reached the maximum tree depth, indicating a need to increase the `max_treedepth` parameter or raise the `target_accept` rate. Summary statistics for the parameters showed the intercept had a mean of -15.250, an SD of 9.938, an HDI (3%) of -29.035, an HDI (97%) of -0.680, an MCSE (mean) of 4.263, an MCSE (SD) of 3.193, and an ESS (bulk) of 7.0. The beta parameter had a mean of 0.201, an SD of 0.166, an HDI (3%) of 0.059, an HDI (97%) of 0.483, an MCSE (mean) of 0.082, an MCSE (SD) of 0.063, and an ESS (bulk) of 7.0 (Figures 8-12).

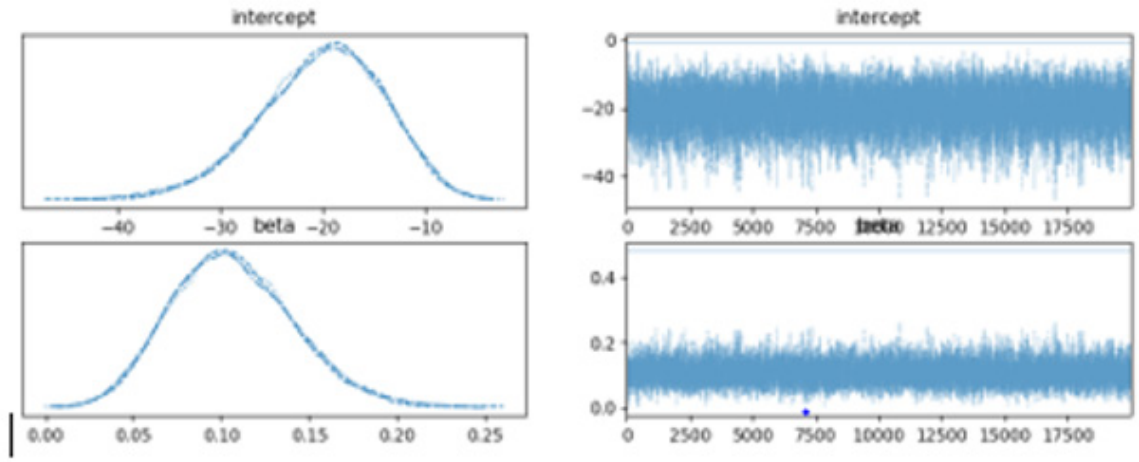

Figure 8: Eigen-Bayesian logistic Regression analysis (DEM).

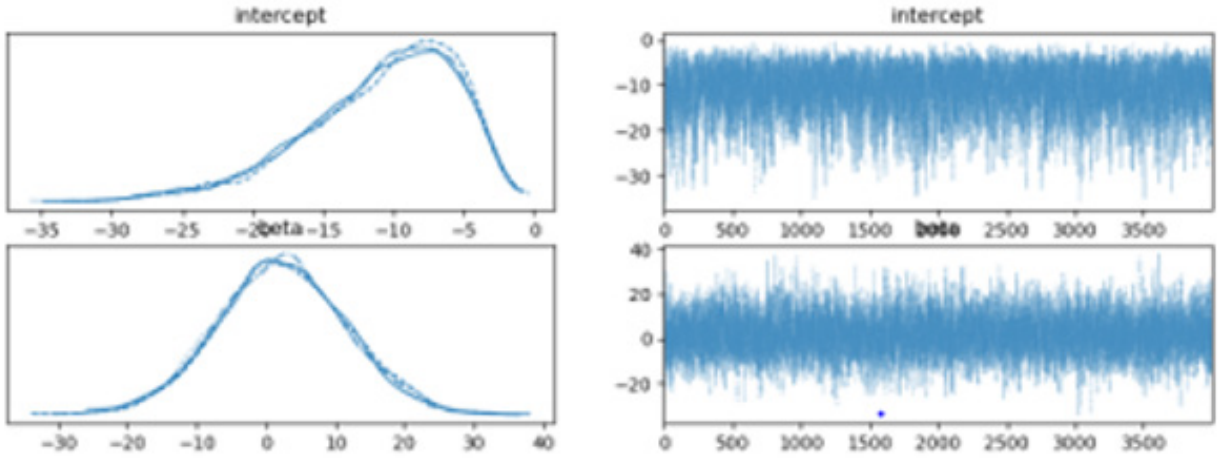

Figure 9: Eigen-Bayesian logistic Regression analysis (MNDWI).

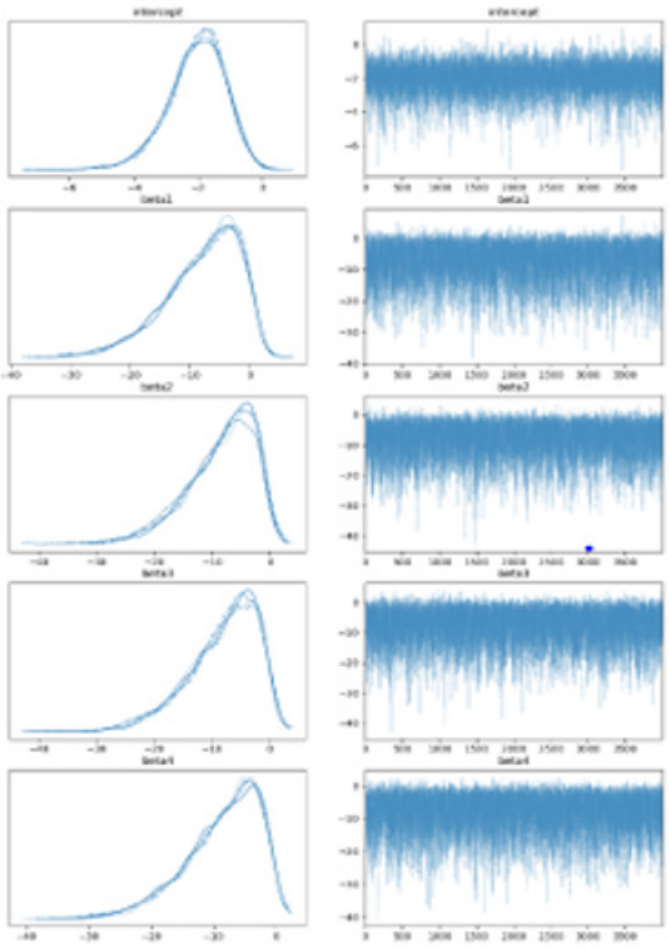

Figure 10: Eigen-Bayesian logistic Regression analysis (LULC).

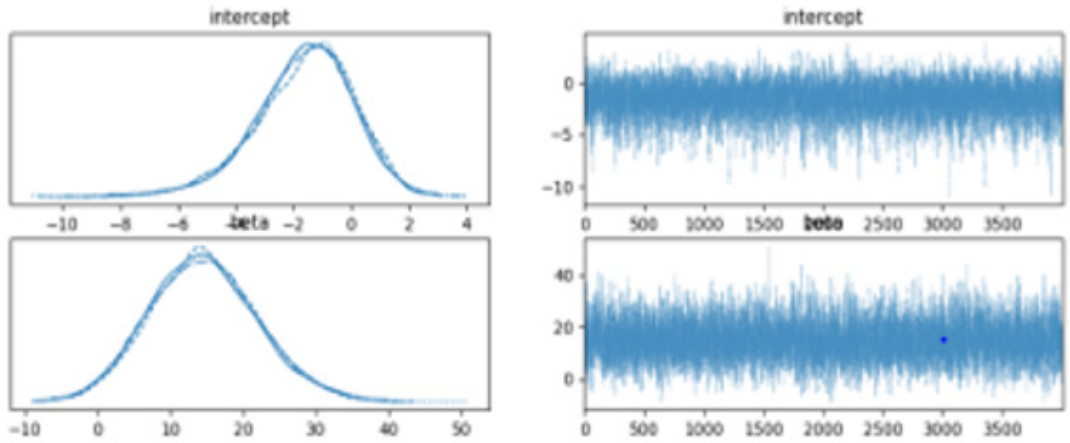

Figure 11: Eigen-Bayesian logistic Regression analysis (NDVI).

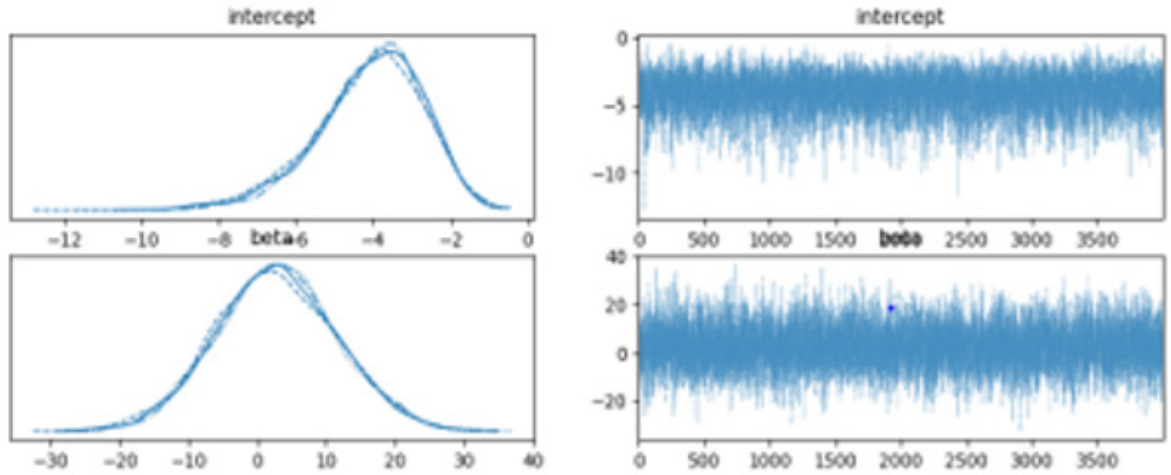

Figure 12: Eigen-Bayesian logistic Regression analysis (SAVI).

These results pertain to the MNDWI analysis where sampling across four chains with 2,000 tuning iterations and 4,000 draw iterations (a total of 8,000 tuning and 16,000 draw iterations) took 4 seconds. Summary statistics for the parameters indicated the intercept had a mean of -10.915, an SD of 5.496, an HDI (3%) of -21.170, an HDI (97%) of -2.109, an MCSE (mean) of 0.089, an MCSE (SD) of 0.065, an ESS (bulk) of 4115.0, and an ESS (tail) of 4846.0, with a rhat of 1.0. The beta parameter had a mean of 2.249, an SD of 9.330, an HDI (3%) of -15.175, an HDI (97%) of 20.135, an MCSE (mean) of 0.124, an MCSE (SD) of 0.093, an ESS (bulk) of 5701.0, and an ESS (tail) of 6086.0, with a rhat of 1.0. These results pertain to the LULC model analysis where sampling across four chains with 2,000 tuning iterations and 4,000 draw iterations (a total of 8,000 tuning and 16,000 draw iterations) took 4 seconds. Summary statistics for the parameters indicated the intercept had a mean of -2.063, an SD of 0.918, an HDI (3%) of -3.737, an HDI (97%) of -0.395, an MCSE (mean) of 0.012, an MCSE (SD) of 0.009, an ESS (bulk) of 6848.0, and an ESS (tail) of 5698.0, with a rhat of 1.0.

The beta1 parameter had a mean of -7.854, an SD of 6.275, an HDI (3%) of -19.740, an HDI (97%) of 1.794, an MCSE (mean) of 0.097, an MCSE (SD) of 0.072, an ESS (bulk) of 4782.0, and an ESS (tail) of 4963.0, with a rhat of 1.0. The beta2 parameter had a mean of -8.491, an SD of 6.014, an HDI (3%) of -19.993, an HDI (97%) of 0.644, an MCSE (mean) of 0.086, an MCSE (SD) of 0.066, an ESS (bulk) of 5799.0, and an ESS (tail) of 5734.0, with a rhat of

1.0. The beta3 parameter had a mean of -8.058, an SD of 6.117, an HDI (3%) of -19.679, an HDI (97%) of 1.246, an MCSE (mean) of 0.088, an MCSE (SD) of 0.070, an ESS (bulk) of 5852.0, and an ESS (tail) of 5093.0, with a rhat of 1.0. The beta4 parameter had a mean of -8.479, an SD of 6.020, an HDI (3%) of -19.836, an HDI (97%) of 0.520, an MCSE (mean) of 0.089, an MCSE (SD) of 0.068, an ESS (bulk) of 5556.0, and an ESS (tail) of 5440.0, with a rhat of 1.0.

These results pertain to the NDVI analysis where sampling across four chains with 2,000 tuning iterations and 4,000 draw iterations (a total of 8,000 tuning and 16,000 draw iterations) took 4 seconds. Summary statistics for the parameters indicated the intercept had a mean of -2.063, an SD of 0.918, an HDI (3%) of -3.737, an HDI (97%) of -0.395, an MCSE (mean) of 0.012, an MCSE (SD) of 0.009, an ESS (bulk) of 6848.0, and an ESS (tail) of 5698.0, with a rhat of 1.0. The beta1 parameter had a mean of -7.854, an SD of 6.275, an HDI (3%) of -19.740, an HDI (97%) of 1.794, an MCSE (mean) of 0.097, an MCSE (SD) of 0.072, an ESS (bulk) of 4782.0, and an ESS (tail) of 4963.0, with a rhat of 1.0. The beta2 parameter had a mean of -8.491, an SD of 6.014, an HDI (3%) of -19.993, an HDI (97%) of 0.644, an MCSE (mean) of 0.086, an MCSE (SD) of 0.066, an ESS (bulk) of 5799.0, and an ESS (tail) of 5734.0, with a rhat of 1.0. The beta3 parameter had a mean of -8.058, an SD of 6.117, an HDI (3%) of -19.679, an HDI (97%) of 1.246, an MCSE (mean) of 0.088, an MCSE (SD) of 0.070, an ESS (bulk) of 5852.0, and an ESS (tail) of 5093.0, with a rhat of 1.0.

The beta4 parameter had a mean of -8.479, an SD of 6.020, an HDI (3%) of -19.836, an HDI (97%) of 0.520, an MCSE (mean) of 0.089, an MCSE (SD) of 0.068, an ESS (bulk) of 5556.0, and an ESS (tail) of 5440.0, with a rhat of 1.0. These results pertain to the SAVI model analysis where sampling across four chains with 2,000 tuning iterations and 4,000 draw iterations (a total of 8,000 tuning and 16,000 draw iterations) took 1 second. Summary statistics for the parameters indicated the intercept had a mean of -4.104, an SD of 1.415, an HDI (3%) of -6.788, an HDI (97%) of -1.606, an MCSE (mean) of 0.020, an MCSE (SD) of 0.014, an ESS (bulk) of 5457.0, and an ESS (tail) of 6538.0, with a rhat of 1.0. The beta parameter had a mean of 2.706, an SD of 8.819, an HDI (3%) of -13.591, an HDI (97%) of 19.706, an MCSE (mean) of 0.112, an MCSE (SD) of 0.079, an ESS (bulk) of 6259.0, and an ESS (tail) of 7702.0, with a rhat of 1.0.

Eigen-Bayesian semiparametric Markovian non-Gaussian Model

We extended an approach for estimating a series of county-level mixed models for determining eigen-Bayesian, semiparametric, eigen algorithmic random effects parameters under a random intercept to include standard errors; thereby including confidence interval. The dependent variable was the RGB Sentinel-2 signature storm sewer capture point *Cx. quinquefasciatus* larval habitat. The independent variables were proxy satellite indicators. The procedure may entail numerical integration to yield posterior empirical eigen-Bayesian estimates of statistically significant capture point sentinel sites with interrogatable signature and parameters; as well as their corresponding posterior standard errors. In this experiment we incorporate an adjustment of the standard error to account for the non-Gaussian variability in estimating the variance component of the eigen-Bayesian, semi-parametric, Markovian, algorithmic random effects distribution in the interpolated storm sewer *Cx. quinquefasciatus* habitats. The final model revealed only non-zero autocorrelated coefficients.

Discussion

The RGB signature analysis of storm sewer habitats for *Cx. quinquefasciatus*, as illustrated in (Figure 1), provides critical insights into the environmental conditions favorable for mosquito larvae. The spectral profile shows a notable peak in reflectance at the red band (Band 4), suggesting that storm sewers with higher reflectance in this band may be more conducive to mosquito habitation. Conversely, the significantly lower reflectance values in the Narrow NIR (Band 8) and SWIR (Band 12) bands indicate that these wavelengths are less reflective in storm sewer environments. This spectral information aids in identifying and characterizing larval habitats, offering valuable data for enhancing mosquito control measures and predicting areas at risk for mosquito-borne diseases. The analysis of hot and cold spots in Bexar and Dallas Counties reveals distinct spatial patterns. In Bexar County, hot spots with 95% and 99% confidence are in central and northeastern areas, while cold spots are in the southern region. In Dallas County, the NDVI-based map shows hot spots concentrated centrally, particularly west of the city center, with cold spots in the east. The DEM-based analysis indicates central clustering of cold spots with high confidence, while hot spots are found in the southern and northeastern regions. These patterns can guide targeted environmental management and resource allocation in both counties.

Second-order spatial autocorrelation analysis was conducted using DEM, NDVI, MNDWI, LULC, and SAVI maps in Bexar and Dallas counties using inverse distance, inverse distance squared, and fixed distance methods to assess clustering propensity and hot/cold spots. For DEM, strong positive spatial autocorrelation was observed with Moran's Index values of 0.69, 0.78, and 0.82 for the three methods, respectively, in Bexar County, and 0.43 for Dallas County, indicating statistically significant spatial autocorrelation. NDVI also showed positive spatial autocorrelation, with the highest Moran's Index of 0.32 for the inverse distance squared method in Bexar County and 0.61 in Dallas County. MNDWI demonstrated significant spatial autocorrelation in both counties, with the highest Moran's Index of 0.43 in Bexar County and 0.566116 in Dallas County using the inverse distance squared method. LULC revealed varying degrees of spatial autocorrelation, with the inverse distance squared method showing the highest Moran's Index of 0.239313 in Bexar County and 0.331703 in Dallas County, while the fixed distance method showed no significant spatial autocorrelation in Dallas. SAVI results indicated positive spatial autocorrelation, with the inverse distance squared method showing the highest Moran's Index of 0.352232 in Bexar County and 0.199066 in Dallas County, although the significance was marginal in Dallas.

These findings confirm the presence of spatial autocorrelation for most methods across different variables and counties. Bayesian logistic regression analysis was performed for the binary data of the

confidence level bin [55,56] as a dependent variable against DEM, NDVI, MNWI, LULC, and SAVI as covariates. Bernoulli distribution likelihood with a Gaussian prior was used to model the dichotomous dependent variable with a draw of 20000, tune of 10000, and 4 chains. These high draws and chains were selected to maximize the convergence rate in the Markovian chains due to the presence of excess random noise [57]. However, the results indicated that MCMC was not able to converge in the model due to the presence of geospatial outliers caused by geospatial multicollinearity, non-homogeneity, and high variance, which were not able to tease out at a linear level. Furthermore, the 2-dimensional logistic regression is based on binary data, which can give a non-robust output of significant and non-significant predictors due to either inflation or deflation of pseudo-R square values. Additionally, 10-meter resolution sentinel data in this model may involve noises (geospatial multicollinearity and heteroscedasticity) and further increase the difficulties in convergence between different chains. In future research efforts, quantitative data, such as larvae counts, should be applied to investigate *Culex* habitats to provide robust and heterogeneous data.

In addition, images with higher resolution may increase heterogeneity and create robust results. The data suggest that the presupposed co-variances are correlated with position and non-random aggregation of storm sewer drains. *Cx. quinquefasciatus* indexes such as this one can prove useful in identifying potential storm sewer aquatic larval habitats of *Cx. quinquefasciatus*. Furthermore, the identification of hot and cold spots in Dallas and Bexar County can prove to be useful in implementing larval source management. Using this model, Texas abatement districts located in Dallas and Bexar can begin to treat these man-made foci using "S&D" tactics. Consequently, the risk of human infection resulting from WNV will decrease as vectors for disease transmission are lessened. The findings in this paper indicated a strong correlation between values attributable to the DEM and aggregation of storm sewers in both Dallas and Bexar counties. Moran's I coefficients as high as 0.827579 analyzed the fixed distance values in Dallas. Strong non-random association between aggregation of storm sewers can be quintessential in identifying potential hot spots and cold spots for storm sewer drains. Current efforts in Dallas and Bexar counties omit treatment of storm sewer drains from abatement practices which inhibits the elimination of *Cx. quinquefasciatus* from habitats completely [58]. One notable limitation of this study is the spatial resolution of the satellite imagery used, which may not adequately capture fine-scale variations in mosquito breeding habitats.

Man made storm sewer drains are only a few meters in distances and using 10-meter Sentinel-2 satellite imagery makes interpolation of points extremely difficult. This limits the robustness of the model to accurately predict where storm sewer habitats are forecasted due to spatial noise. The color of the storm sewers themselves proved to also generate limitations in the differentiation between storms sewers and other man-made structures of the same color such as communication wire covers and water main covers. Implementation of higher resolution Planet 3-meter satellite imagery can feasibly remove these limitations and increase robustness of the storm sewer model. Future research should aim to address existing limitations and broaden the study's scope to improve vector control strategies. Exploring other *Cx. quinquefasciatus* habitats such as wetland foci, drainage canals, and artificial water containers [59-62]. Broadening the habitats being modeled can result in more effective treatment of *Cx. quinquefasciatus* and potential reduction of cases of WNV. Subsequently, after incorporating all potential capture point habitats of *Cx. quinquefasciatus* future studies can run the model again but at a larger scale. Instead of just Dallas and Bexar County the whole state of Texas can be interpolated to find habitat points and also hot and cold spots. Creating a broad-based model for *Cx. quinquefasciatus* for Texas can prove useful in reducing endemicity of WNV in Texas.

This could reveal significant *Cx. quinquefasciatus* populations in the state and allow treatment of previously unknown habitats to begin. As of now, Dallas and Bexar County and their respective abatement districts do not account for treatment predicated on predictive modeling. Further studies should also go about obtaining satellite imagery with increased spatial resolution. Creating a comparative study between Dallas and Bexar counties interpolated points at 10-meters and 3-meters can prove useful in quantifying the robustness of the storm sewer model. The result of this study can then affect what resolution points are interpolated to reflect reality. Another potential future study approach involves creating a feedback loop using a convolutional neural network (CNN) to identify potential storm sewer habitats. This method would entail an iterative process where potential habitats identified by the CNN are validated, and the parameters of the interpolation algorithm are updated based on this validation. New points would then be interpolated and validated, with these new data points fed back into the CNN to further refine its accuracy. This feedback loop would continuously improve the model's precision in predicting storm sewer habitats, thereby enhancing vector control strategies and potentially reducing the prevalence of WNV.

Conclusion

In conclusion, the findings in this study reveal previously unknown geolocations of *Cx. quinquefasciatus* larval storm sewer oviposition habitats. These results confirm that it is possible to interpolate more predicted aquatic storm sewer habitat points using available online spatial satellite imagery. Using second-order autocorrelation allows geospatially autocorrelating storm sewer habitats in Dallas and Bexar County, highlighting the significant impact of spatial autocorrelation on the interpolation of storm sewers. Interpolated Sentinel-2 10-meter resolution and RGB signatures of *Cx. quinquefasciatus* storm sewer variables can generate erroneous noisy signatures due to violations of regression assumptions, which may be corrected in an eigen-Bayesian eigen-geospace. Eigen-orthogonalizable Eigen-spatial filters are useful in revealing the influence of non-Gaussianism, such as heterogeneity of variances. Markovian semiparametric non-frequentist uncertainty-oriented models are accurate in predicting zero-autocorrelated georeferenceable hyper/hypo endemic geolocations compared to a global model. These models, as documented in the literature, handle interpolated erroneous non-homogeneous RGB signature

non-evidential uncertainty-oriented probabilities consistently across Bayesian eigenvector eigen-geospace.

Conflicts of interest

No conflicts of interest.

References

1. Sejvar JJ (2003) West nile virus: an historical overview. Ochsner J 5(3): 6-10.
2. Lanciotti RS, Roehrig JT, Deubel V, Smith J, Parker M, et al. (1999) Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. Science 286(5448): 2333-2337.
3. Historic Data (1999-2022) West Nile Virus.
4. Karki S, Brown WM, Uelmen J, Ruiz MO, Smith RL (2020) The drivers of West Nile virus human illness in the Chicago, Illinois, USA area: Fine scale dynamic effects of weather, mosquito infection, social, and biological conditions. PLoS One 15(5): e0227160.
5. Murray KO, Ruktanonchai D, Hesalroad D, Fonken E, Nolan MS (2013) West Nile virus, Texas, USA, 2012. Emerg Infect Dis 19(11): 1836-1838.
6. Nolan MS, Schuermann J, Murray KO (2013) West Nile virus infection among humans, Texas, USA, 2002-2011. Emerg Infect Dis 19(1): 137-139.
7. Texas Demographic Center (2016) Aging in Texas: Introduction.
8. Harrison Couty Health Department (2024) Major Mosquito Carries West Nile Virus.
9. Little EAH, Harriott OT, Akaratovic KI, Kiser JP, Abadam CF, et al. (2021) Host interactions of Aedes albopictus, an invasive vector of arboviruses, in Virginia, USA. PLoS Negl Trop Dis 15(2): e0009173.
10. Habarugira G, Suen WW, Hobson-Peters J, Hall RA, Bielefeldt-Ohmann H (2020) West Nile Virus: An Update on Pathobiology, Epidemiology, Diagnostics, Control and "One Health" Implications. Pathogens Pathogens 9(7): 589.
11. Farajollahi A, Fonseca DM, Kramer LD, Kilpatrick AM (2011) "Bird biting" mosquitoes and human disease: a review of the role of Culex pipiens complex mosquitoes in epidemiology. Infect Genet Evol 11(7): 1577-1585.
12. Petersen LR, Brault AC, Nasci RS (2013) West Nile virus: review of the literature. JAMA 310(3): 308-315.
13. Chowdhury P, Khan SA (2021) Global emergence of West Nile virus: Threat & preparedness in special perspective to India. Indian J Med Res 154(1): 36-50.
14. Gorris ME, Bartlow AW, Temple SD, Romero-Alvarez D, Shutt DP, et al. (2021) Updated distribution maps of predominant Culex mosquitoes across the Americas. Parasit Vectors 14(1): 547.
15. Mordecai EA, Caldwell JM, Grossman MK, Lippi CA, Johnson LR, et al. (2019) Thermal biology of mosquito-borne disease. Ecols Lett 22(10): 1690-1708.
16. Hill S, Connelly R (2009) Featured Creatures. UF/IFAS, University of Florida Entomology & Nematology.
17. Eastwood G, Cunningham AA, Kramer LD, Goodman SJ (2019) The vector ecology of introduced Culex quinquefasciatus populations, and implications for future risk of West Nile virus emergence in the Galápagos archipelago. Med Vet Entomol 33(1): 44-55.
18. Richards SL, Anderson SL, Lord CC, Smartt CT, Tabachnick WJ (2012) Relationships Between Infection, Dissemination, and Transmission of West Nile Virus RNA in *Culex pipiens quinquefasciatus* (Diptera: Culicidae). J Med Entomol *49*(1): 132-142.
19. Gerberg EJ, Barnard DR, Ward RA (1994) Manual for mosquito rearing and experimental techniques. American Mosquito Control Association Bulletin (5): 61-62.
20. Bates M (1949) The Natural History of Mosquitoes. Macmillian Company. New York, USA pp. 379.
21. Mitchell CJ, Francy DB, Monath TP (1980) Chapter 7: Arthropod vectors. In: St. Louis Encephalitis. Monath TP, [Eds.,]. Washington, DC, USA, American Public Health Association Inc, pp. 313-373.
22. Rueda LM, Patel KJ, Axtell RC, Stinner RE (1990) Temperature-dependent development and survival rates of Culex quinquefasciatus and Aedes aegypti (Diptera: Culicidae). J Med Entomol 27(5): 892-898.
23. Paz S (2015) Climate change impacts on West Nile virus transmission in a global context. Philos Trans R Soc Lond B Biol Sci 370(1665): 20130561.
24. Texas Department of State Health Services (2014) Arbovirus Activity in Texas 2012 Surveillance Report.
25. Jacob BG, Izureta R, Bell J, Parikh JU, Loum D, et al. (2023) Approximating Non-Asymptoticalness, Skew Heteroscedascity and Geo-spatiotemporal Multicollinearity in Posterior Probabilities in Bayesian Eigenvector Eigen-Geospace for Optimizing Hierarchical Diffusion-Oriented COVID-19 Random Effect Specifications Geosampled in Uganda. American Journal of Mathematics and Statistics 13(1): 1-43.
26. Jacob BG, Novak RJ, Toe L, Sanfo MS, Tingueria R, et al. (2014) Full Length Research Paper. Public Health and Epidemiology 347.
27. Mwangangi JM, Shililu J, Muturi EJ, Muriu S, Jacob B, et al, (2010) Anopheles larval abundance and diversity in three rice agro-village complexes Mwea irrigation scheme, central Kenya. Malar J 9: 228.
28. Ritchie KK, Lzurieta R, Hoare I, Choudhari N, Murray K, et al. (2024) Mapping Ae aegypti Bird Bath Habitats for Implementing" Seek and Destroy" Larval Source Management in Hillsborough County, FL. USA. American Journal of Entomology 8(1): 1-17.
29. Tucker CJ, Newcomb WW, Los SO, Prince SD (1991) Mean and inter-year variation of growing-season normalized difference vegetation index for the Sahel 1981-1989. International Journal of Remote Sensing 12(6): 1133-1135.
30. Xu H (2006) Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International journal of remote sensing 27(14): 3025-3033.
31. Qi J, Chehbouni A, Huete AR, Kerr YH, Sorooshian S (1994) A modified soil adjusted vegetation index. Remote Sensing of Environment 48(2): 119-126.
32. Polidori L, El Hage M (2020) Digital elevation model quality assessment methods: A critical review. Remote sensing 12(21): 3522.
33. Bhattacharjee S, Ghosh SK (2015) Spatio-temporal change modeling of LULC: a semantic kriging approach. Annals of the Photogrammetry Remote Sensing and Spatial Information Sciences II-4/W2: 177-184.
34. Reisen WK (2010) Landscape Epidemiology of Vector-Borne Diseases. Annu Rev Entomol 55(1): 461-483.
35. Arana-Guardia R, Baak-Baak CM, Loroño-Pino MA, Machain-Williams C, Beaty BJ, et al. (2014) Stormwater drains and catch basins as sources for production of Aedes aegypti and Culex quinquefasciatus. Acta Trop 134: 33-42.
36. Dallas County, Home.
37. U.S. Census Bureau. Explore Census data.
38. Bexar County TX.
39. NOAA's National Weather Service. Dallas/Fort Worth climate narrative.
40. U.S. Census Bureau. (n.d.-b). Explore Census data.

41. NCEI.Monitoring.Info@noaa.gov. (n.d.). Climate at a Glance | County Time Series | National Centers for Environmental Information (NCEI).

42. Griffith DA (2005) A comparison of six analytical disease mapping techniques as applied to West Nile Virus in the coterminous United States. Int J Health Geogr 2: 4-18.

43. Stahler AH (1980) The use of prior probabilities in maximum likelihood classification of remotely sensed data. Remote Sensing of Environment 10(2): 135-163.

44. Al-Ahmadi FS, Hames AS (2009) Comparison of four classification methods to extract land use and land cover from raw satellite images for some remote arid areas, Kingdom of Saudi Arabia. Journal of King Abdulaziz University, Earth Sciences 20(1): 167-191.

45. Maselli F, Conese C, Petkov L, Resti R (1992) Inclusion of prior probabilities derived from a nonparametric process into the maximum likelihood classifier. Photogrammetric Engineering and Remote Sensing 58(2): 201-207.

46. Mingguo Z, Qianguo C, Mingzhou Q (2009) The effect of prior probabilities in the maximum likelihood classification on individual classes: A theoretical reasoning and empirical testing. Photogrammetric Engineering and Remote Sensing 75(9): 1109-1117.

47. Stein ML (2005) Statistical methods for regular monitoring data. Journal of the Royal Statistical Society Series B: Statistical Methodology 67(5): 667-687.

48. Jacob BG, Loum D, Kaddumukasa M, Kamgno J, Djeunga HCN, et al. (2021) Geospatial artificial intelligence infused into a smartphone drone application for implementing 'Seek and Destroy' in Uganda. American Journal of Entomology 5(4): 92-109.

49. Cressie N (2015) Statistics for spatial data. John Wiley & Sons.

50. Binmore K (2008) Rational Decisions in Large Worlds. Annales d Économie et de Statistique 86: 25-41.

51. Bouazizi H, Brunette I, Meunier J (2023) Comparison of polynomial and rational function cornea models for effective dimensionality reduction. Comput Biol Med 167: 107635.

52. Marsden JE, Tromba A (2003) Vector calculus, 5th edition, WH Freeman, [Eds.,]. United Kingdom. PP. 182.

53. Gelman A, Shalizi CR (2012) Philosophy and the practice of Bayesian statistics. British Journal of Mathematical and Statistical Psychology 66(1): 8-38.

54. Ord K, Getis A (2010) Local spatial autocorrelation statistics: distributional issues and an application. Geographical analysis 27(4): 286-306.

55. ESRI. How Hot Spot Analysis (Getis-Ord GI*) works—ArCGIS Pro.

56. Getis A, Ord JK (1992) The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis 24(3): 189-206.

57. Hoffman MD, Gelman A (2014) The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research 15: 1593-1623.

58. Müller GC, Junnila A, Qualls W, Revay EE, Kline DL, et al. (2010) Control of Culex quinquefasciatus in a storm drain system in Florida using attractive toxic sugar baits. Med Vet Entomol 24(4): 346-351.

59. Bhattacharya S, Basu P (2016) The southern house mosquito, Culex quinquefasciatus: profile of a smart vector. J Entomol Zool Stud 4(2): 73-81.

60. Mckeever S (2014) Differentiating Geo-Spatiotemporal Aquatic Larval Habitats of Anopheles gambiae complex in Urban Agriculture and Urban Non-Agriculture Environments in Accra, Ghana (Order No. 1570429). Available from Dissertations & Theses @ University of South Florida - FCLA; ProQuest Dissertations -& Theses A&I; ProQuest Dissertations & Theses Global. (1640735016).

61. Jacob BG, Gu W, Caamano EX, Novak RJ (2009) Developing operational algorithms using linear and non-linear squares estimation in Python for the identification of Culex pipiens and Culex restuans in a mosquito abatement district (Cook County, Illinois, USA). Geospat Health 3(2): 157-176.

62. Rothman SE, Jones JA, Ladeau L, Leisnham PT (2021) Higher West Nile Virus Infection in Aedes albopictus (Diptera: Culicidae) and Culex (Diptera: Culicidae) Mosquitoes from Lower Income Neighborhoods in Urban Baltimore, MD. J Med Entomol 58(3): 1424-1428.