

Chapter

Spatial Statistics in Vector-Borne Diseases

*Manuel Solís-Navarro, Susana G. Guzmán-Aquino,
María Guzmán-Martínez and Jazmín García-Machorro*

Abstract

Vector-borne diseases are those caused by the bite of an infected arthropod, such as the *Aedes aegypti* mosquito, which can infect humans with dengue or Zika. Spatial statistics is an interesting tool that is currently implemented to predict and analyze the behavior of biological systems or natural phenomena. In this chapter, fundamental characteristics of spatial statistics are presented and its application in epidemiology is exemplified by presenting a study on the prediction of the dispersion of dengue disease in Chiapas, Mexico. A total of 573 confirmed dengue cases (CDCs) were studied over the period of January–August 2019. As part of the spatial modeling, the existence of spatial correlation in CDCs was verified with the Moran index (MI) and subsequently the spatial correlation structure was identified with the mean squarer normalized error (MSNE) criterion. A Generalized Linear Spatial Model (GLSM) was used to model the CDCs. CDCs were found to be spatially correlated, and this can be explained by a Matérn covariance function. Finally, the explanatory variables were maximum environmental temperature, altitude, average monthly rainfall, and patient age. The prediction model shows the importance of considering these variables for the prevention of future CDCs in vulnerable areas of Chiapas.

Keywords: vector-borne diseases, Gaussian process, generalized spatial linear models, georeferenced data, spatial correlation

1. Introduction

Vector-borne diseases are infections caused by viruses, bacteria, or parasites that are transmitted to humans by the bite of infected arthropod species, these can be diseases transmitted by mosquitoes (dengue fever, West Nile fever, chikungunya, malaria, Zika, etc.), by sandflies (leishmaniasis), by ticks (encephalitis, Lyme Borreliosis, Crimean-Congo hemorrhagic fever, Human Granulocytic Anaplasmosis) by triatomines (Chagas disease), among others. These diseases account for more than 17% of all infectious diseases and cause more than 700, 000 deaths per year [1, 2].

Vectors are living organisms that can transmit infectious pathogens between humans or from animals to humans. Many of these vectors are insects that ingest disease-causing microorganisms during a blood meal from an infected host and then transmit it to a new host after the pathogen has replicated. Another characteristic of arthropod vectors is that they are cold-blooded (ectothermic) and therefore very sensitive to climatic factors, although the climate is only one of many factors that influence vector distribution, as there are also geographic and sociodemographic factors [1].

In order to interpret the behavior of vector-borne diseases in the most accurate and simplified way possible, statistical models are used. A statistical model is a simplified representation of a phenomenon of interest [3, 4]. With their help, it is possible to model, predict and make inferences about natural phenomena, biological systems, epidemiological studies, and others [5]. One of the most widely used statistical models is linear regression models, which predict a continuous target based on linear relationships between the target and one or more predictors. But there is another type of model that extends the general linear model, so that the dependent variable is linearly related to the factors and covariates by means of a certain link function, which is known as a generalized linear model [6].

Generalized Linear Models (GLMs) provide a collection of linear regression models including the exponential family, such as the Binomial and Poisson, which are distributions for counting data. The GLMs were introduced by Nelder in 1972 [7], in 1989 they were studied in greater depth by McCullagh [8] and over time more authors were integrated [9–13].

There are three components in GLMs: A response variable distribution, a linear predictor, and a link function. A response variable \mathbf{Y} is assumed (Y_1, Y_2, \dots, Y_n) , where Y_1, Y_2, \dots, Y_n are independent of each other; its expected value is related to a linear predictor $E[Y] = g^{-1}(\mathbf{d}'\boldsymbol{\beta})$, where $\boldsymbol{\beta} \in \mathcal{R}^p$ is a vector of regression parameters, \mathbf{d} are known explanatory variables and g is a known function called a link function, which allows to define the relationship between the systematic and random components [14].

GLMs can help in numerous areas such as epidemiology, mining engineering, Earth and environmental sciences, ecology, biology, geography, economics, agronomy, forestry, image processing, and more [15, 16]. For epidemiology in particular, as it is about understanding diseases that affect a population, the most usual thing is to find a binary variable that represents the presence or absence of a disease or to count the events of a disease for certain areas.

Such is the case of a study conducted by Hashizume et al. [17] in Bangladesh, 2012. They used a Generalized Linear Poisson Regression Model to examine weekly dengue hospitalizations in relation to river levels, during the years 2005 to 2009, and the climatic variables daily precipitation and average temperature. The models were adjusted according to seasonal variation and temperature. They found evidence of a 6.9% increase in dengue with high river levels, but a 29.6% increase in disease when rivers were very low.

An important extension of the GLMs is the Generalized Linear Mixed Models (GLMMs) [18]. GLMMs provide a range of analyses for those data that are correlated in space and belong to the exponential family (Gamma, Poisson, Binomial, among others) [19]. Generalized Linear Spatial Models (GLSMs) are basically GLMMs, since latent variables are derived from a spatial process. In recent years, there has been a growing interest in the analysis of spatial data in epidemiology, in order to predict the incidence of vector-borne diseases.

Using techniques available to epidemiologists and other health professionals, the potential of remote sensing, Geographic Information Systems (GIS), and spatial analysis of epidemiological data has been demonstrated by some authors such as those mentioned below; however, there are still few studies that adequately prove the potential of these tools, since they are still being exploited in the fight against diseases [20].

For instance, a Colombian paper published in 2012, Sanchez et al. [21] estimated Generalized Linear Spatial Regression Models with a Poisson response to explain the behavior of malaria and dengue in different years. Health determinants were identified in the occurrence of these diseases and risk maps were obtained. Finally, it demonstrated the need to link spatial effects in the models and the explanatory variables considered, to explain the number of reported cases of the disease in the years analyzed.

Another example is the work of Estallo et al. [22] in 2021, which evaluated the species responsible for the transmission of *Leishmaniasis* (phlebotom-*Phlebotominae*) during the period 2012 – 2014 in northern Argentina. Through Generalized Linear Mixed Models, the implications of vectors in disease transmission were evaluated, using meteorological and teledetection environmental factors. It was observed that the species *Lutzomyia longipalpis* was the most abundant in urban areas. The findings allowed detecting of high-risk areas and the developing of predictive models to optimize resources and prevent *leishmaniasis* transmission in the area.

As can be seen, spatial analysis is a powerful tool for the analysis of georeferenced data, as it can give health research a broader perspective of the occurrence of health events and diseases. Spatial statistical models are useful because they estimate the spatial variance inherent in the data, and can also be used to perform statistical inference throughout the study area. Spatial prediction can be made based entirely on a stochastic model or in combination with a deterministic trend [20, 23].

The aim of this chapter is to show an example of the application of spatial statistics, implementing a Generalized Linear Spatial Model for the prediction of dengue disease in the state of Chiapas. For this, there are considered patient age and the next information of each municipality: garbage disposal service, maximum environmental temperature, average monthly rainfall, and altitude as covariates. For the study of the disease in the 118 municipalities of Chiapas, the cases observed in 36 municipalities in the state of Chiapas and the information in the aforementioned explanatory variables were considered.

2. Spatial statistical models

Space models have a simple structure, flexible enough to handle a variety of problems. The data may be continuous or discrete, present spatial aggregations, or be point observations in space. As for the spatial locations can be regular or irregular. A spatial model is usually used to predict sites where the study phenom was not observed.

Let $x \in A \subset \mathbb{R}^d$ and $S(x)$ the data observed at the x location, this results in a stochastic process

$$S(x) : x \in A \tag{1}$$

Structure 1, allows to differentiate and talk about problems with continuous spatial indexes, lattice, and point patterns giving rise to three types of data: geospatial, lattice

data, and point patterns. In geospatial data, A is a fixed set in \mathbb{R}^d containing a d -dimensional rectangular with positive volume; $S(x)$ is a random vector in the location $x \in A$. These data arise in areas such as atmospheric sciences, mining, and public health. In point patterns A is a point process in \mathbb{R}^d or a subset of \mathbb{R}^d ; $S(x)$ is a random vector in the location $x \in A$. In its most general form, it results in a spatial point process marked when $S(x) = 1$, for all $x \in A$. Point patterns arise when the variable to be analyzed is a location of “events”.

Finally, the entangled data or also known as area data, A is a regular or irregular fixed set (with additional information from the surrounding neighborhood) of \mathbb{R}^d ; $S(x)$ is a random vector in location $x \in A$. When locations are in regular meshes it is the closest analogy to time series observed at equally spaced time points. In the entangled data, based on the general spatial process 1, it is assumed that A is an accounting collection of space sites, in which the data are observed. The most common entangled data models are the Conditional Autoregressive Model (CAR) and the Simultaneous Autoregressive Model (SAR). CAR models form the basis of Markovian Gaussian random fields and Integrated Nested Laplace Approximation (INLA) methods. SAR models are popular in geographic information systems. Other models are the spatial autoregressive moving average (ARMA) [24, 25].

2.1 Gaussian spatial processes

Knowing the type of variables with which they are working and taking into account their spatial dependence, helps to determine the regression technique that best fits the characteristics of the data [21]. For the study of spatial data Gaussian processes can be used, which are stochastic processes, a collection of variables. This allows any subset of finite random variables to have a multivariate Gaussian distribution. Gaussian processes can thus be thought of as distributions of random vectors or random functions [26]. Gaussian processes began to be studied in the 1940s, but until the 1970s they were used in geostatistics and meteorology; In the 1990s Cressie [24] began to implement them in spatial statistics. In fact, the term “model-based geostatistics” was first used to describe an approach to geostatistical problems based on formal statistical models and inference procedures [27].

Gaussian stochastic processes are widely used as models for geostatic data. If a transformation of the original response variable is used, the scope of the Gaussian models can be amplified, and so with this extra flexibility the model provides a good empirical fit to the data.

A Gaussian process, $\{S(x) : x \in \mathbb{R}^2\}$, is a stochastic process with the property that for any collection of locations $x_1, \dots, x_n, x_i \in \mathbb{R}^2$, the joint distribution of $\mathbf{S} = \{S(x_1), \dots, S(x_n)\}$ is multivariate Gaussian.

Any such process is fully specified by the average function $\mu(x) = E[S(x)]$ and the covariance function $Cov\{S(x), S(x')\}$. As given x_1, \dots, x_n an arbitrary set of locations with $\boldsymbol{\mu} = (\mu(x_1), \dots, \mu(x_n))$ and \mathbf{G} an $n \times n$ matrix with elements $G_{ij} = Cov(S(x_i), S(x_j))$; then \mathbf{S} has a multivariate normal distribution (MN).

$$\mathbf{S} \sim MN(\boldsymbol{\mu}, \mathbf{G}) \tag{2}$$

A spatial Gaussian process is stationary if $\mu(x)$ is constant, $\mu(x) = \mu$, for all x and $Cov(S(x), S(x')) = Cov(u)$; where $u = \|x - x'\|$ is the Euclidean distance. A stationary process is isotropic if the covariance between the values of $S(x)$ at any two locations

depends only on the distance between them. The term stationary is often used as the equivalence of stationary and isotropic. A process for which $S(x) - \mu(x)$ is stationary is called covariance stationary. Processes of this type are widely used in practice as models for geostatistical data [28].

Among the parametric functions for the covariance function [29] are the following:
 Exponential:

$$Cov(u) = \sigma^2 \left[\exp \left(\frac{-u}{\phi} \right) \right] \quad (3)$$

Gaussian:

$$Cov(u) = \sigma^2 \left[\exp \left(-\frac{u}{\phi} \right)^2 \right] \quad (4)$$

Matérn:

$$Cov(u) = \sigma^2 \left[\frac{2}{2^{\kappa-1} \Gamma(\kappa)} \left(\frac{u}{\phi} \right)^{\kappa} K_{\kappa} \left(\frac{u}{\phi} \right) \right] \quad (5)$$

In these covariance functions (Eqs. (3)–(5)) $u > 0$, $\phi > 0$, $\kappa > 0$; function K_{κ} denotes the modified Bessel function of order κ and $\Gamma(\cdot)$ denotes the gamma function.

2.2 Criteria for evaluating the covariance structure of the Gaussian process

There are several criteria in the literature to validate the covariance structure of a Gaussian process Eq. (2). Among the most used are: Mean Error (ME), Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Square Normalized Error (MSNE) (**Table 1**). ME and MSE should tend to zero when the covariance structure of the Gaussian process was correctly estimated. The MAE and RMSE criteria are considered as the most efficient criteria to validate the covariance structure of the Gaussian process. The RMSE is expected to be small like MAE, while the MSNE is expected to be close to 1 [29, 30].

Measurement	Definition
Mean error	$ME = \frac{1}{n} \sum_{i=1}^n (Y(s_i) - \hat{Y}(s_i))$
Mean square error	$MSE = \frac{1}{n} \sum_{i=1}^n (Y(s_i) - \hat{Y}(s_i))^2$
Mean absolute error	$MAE = \frac{1}{n} \sum_{i=1}^n (Y(s_i) - \hat{Y}(s_i))$
Root mean square error	$RMSE = \left[\frac{1}{n} \sum_{i=1}^n (Y(s_i) - \hat{Y}(s_i))^2 \right]^{\frac{1}{2}}$
Mean square normalized error	$MSNE = \frac{1}{n} \sum_{i=1}^n \frac{(Y(s_i) - \hat{Y}(s_i))^2}{\hat{\sigma}_{ok}^2(s_i)}$

$\hat{\sigma}_{ok}^2$ is a variance estimated by the ordinary kriging interpolation method [29].

Table 1.
 Criteria for evaluating the covariance structure of the Gaussian process.

2.3 Generalized linear spatial models

Spatial Generalized Linear Models were introduced by Diggle et al. in 1998 [31]; if the variable response Y has Poisson distribution, then

$$Y_i|S(\cdot) \sim \text{Poisson}(\mu_i) \quad (6)$$

Where

$$\mathbf{S} \sim MN(\mathbf{D}\beta, \mathbf{G})$$

It is assumed that $\{Y_i : i = 1, \dots, n\}$ conditioned in \mathbf{S} are independent, $E[Y_i|S(\cdot)] = \mu_i$, g a known link function such that $g(\mu_i) = \eta_i$ then $\mu_i = g^{-1}(\eta_i)$, $i = 1, \dots, n$. $\mathbf{D} = (\mathbf{1}, \mathbf{d}_1, \dots, \mathbf{d}_p)$ is a design matrix of $n \times (p + 1)$ of full range, $\mathbf{1}$ a vector $n \times 1$ of ones and $\mathbf{d}_j = (d_j(x_1), \dots, d_j(x_n))'$, where $d_j(x_i)$ is the value of the covariate j -th of the i -th location; $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ the regression parameters.

2.4 Moran's index for spatial autocorrelation

To prove the existence of spatial dependence on a variable Y , the Moran index [32, 33], given by

$$IM = \frac{n \sum_i \sum_i^n w_{ij} (Y_i - \bar{Y})(Y_i - \bar{Y})}{\sum_{i \neq j} w_{ij} \sum_i^n (Y_i - \bar{Y})^2} \quad (7)$$

Where \mathbf{W} is the weights matrix that defines the relationships between the regions of the study. In this case $w_{ij} = 1$ denotes areas with a common border and $w_{ij} = 0$ in another case. Y_i and Y_j would be the values observed in regions i and j respectively, while \bar{Y} is the average incidence of the districts studied, n is the total number of localities.

2.5 Statistical software packages R for spatial data

Several packages are available in statistical software R [34] to perform spatial modeling.

The *geoR* package is used for performing geostatistical data analysis and spatial prediction, which expands the set of methods and tools presently available for spatial data analysis in R. The package executes methods for Gaussian and Gaussian models transformed, incorporates functions and methods for reading and preparing the data, exploratory analysis, inference on model parameters and spatial interpolation, and it also contains functions for parameter estimation under Bayesian methods [35].

The *geoRglm* package is used to implement Generalized Linear Spatial Model. The subsequent and predictive inference is based on Markov Chains Monte Carlo (MCMC) methods. This package, which is an extension of the *geoR* package, help with GLSM conditional simulation and prediction, and with Bayesian inference for the models Poisson (*pois.krige*) and Binomial (*binom.krige*) [35, 36]. A Langevin-Hastings algorithm is used to obtain MCMC simulations. In the *pois.krige* and *binom.krige* functions, the user can provide a value for the variation of the proposal *S.scale*, a value

initial, $S.start$, the thinning, $thin$, the length of the burn, $burn.in$, and the number of iterations, $n.iter$ [35].

2.5.1 Inference for the generalized linear spatial model

The geostatistical model assumes the response variable to be Gaussian, which may be an unrealistic assumption for some data sets. The GLSM provides a framework for analyzing Binomial and Poisson distributed data. The likelihood for such a model, in general, cannot be represented in closed form, since it is a high-dimensional integral

$$L(\boldsymbol{\beta}, \sigma^2, \phi) = \int \prod_{i=1}^n f(y_i; g^{-1}(s_i)) p(\mathbf{s}; \boldsymbol{\beta}, \sigma^2, \phi) ds \quad (8)$$

where $f(y; \mu)$ denotes the density of the distribution with mean μ , $p(\mathbf{s}; \boldsymbol{\beta}, \sigma^2, \phi)$ is the multivariate Gaussian density for the vector \mathbf{s} of random effects at the data locations and $g(\cdot)$ is the link function. In practice, the high dimensionality of this integral precludes direct computation, so the inference is based on MCMC.

3. Description of data

This section shows the application of a spatial model taking into account the social, climatic, and geographical characteristics of the municipalities of the state of Chiapas in relation to dengue virus infections registered from January to August of the year 2019.

3.1 Study area

Dengue disease is endemic to the state of Chiapas with scattered case reports, this is due to the different geographic characteristics of the state, such as the altitude of its municipalities and its border condition with the country of Guatemala. It is known that at different altitudes, in the regions, the climatic conditions tend to vary and this can favor the reproduction of the vector. The state of Chiapas is divided into 118 municipalities, each with different sociodemographic and climatic conditions. The population density, according to the INEGI, is around 5,544 million inhabitants; being the state capital, Tuxtla Gutiérrez, the municipality with the highest population density; for the year 2019, 604,147 inhabitants were registered [37].

3.2 Data collection

The data, which were collected at the municipal level, being 36 the municipalities that registered positive cases of dengue and were considered for the analysis, were obtained from different sources that are mentioned below.

3.2.1 Dengue cases

The database with dengue cases registered in the state of Chiapas, during the period January–August 2019, was obtained from the Secretary of Health of the state of

Chiapas, in collaboration with the area of vector-borne diseases. This database is updated week by week, fulfilling 52 Epidemiological Weeks (EW) reports per year.

3.2.2 Climatic data

The climatic data were obtained from the World Meteorological Organization (WMO) [38], for each municipality of residence where the dengue cases were registered, working with the daily reports of average environmental temperature, maximum and minimum environmental temperature and average monthly rainfall. The climatic data were taken into account for the analysis, 6 days before the onset of symptoms for each case, this was done considering the intrinsic incubation period in order to obtain an approximate date of infection and capture the daily climatic data for each municipality [39]. With respect to the rainfall variable, it was decided to work with the monthly average, since there were days in which there were no records.

3.2.3 Non-climatic data

Other factors related to infection were also considered in the analysis. Data on the population density and altitude of each municipality of residence per observed case were obtained from the INEGI, the other variables such as garbage disposal, contact with the mosquito, drinking water service, patient age, and sex were obtained from the original database of registered dengue cases provided by the secretary of health [37].

3.3 Georeferencing

For the georeferencing of dengue cases registered in the period January–August 2019, the postal code and the world geographic coordinate system, WGS84, were used. With the *pois.krige* function from the *geoRglm* package, in R software version 4.0.3 [34] and the projection of the cases was carried out on a map of Chiapas.

4. Results

The database that is made up of 573 dengue cases, reported in the state of Chiapas, Mexico, during the period January–August of the year 2019; being the state capital, (Tuxtla Gutiérrez) the municipality with the highest number of CDCs, with 49.04%, the rest of the cases were scattered in other 35 municipalities of the State. The average age of the cases was 14 years, with the female sex being the most affected with 53%, in the same way, 15% indicated not having the drinking water service.

4.1 Spacial location

The spatial distribution of the 573 dengue cases is heterogeneous in 36 municipalities in the state of Chiapas (blue points in **Figure 1**).

4.2 Moran's index

The Moran's Index obtained, with the number of CDCs in the 36 municipalities of Chiapas, was 0.115, which indicates that there is a spatial relationship in the number

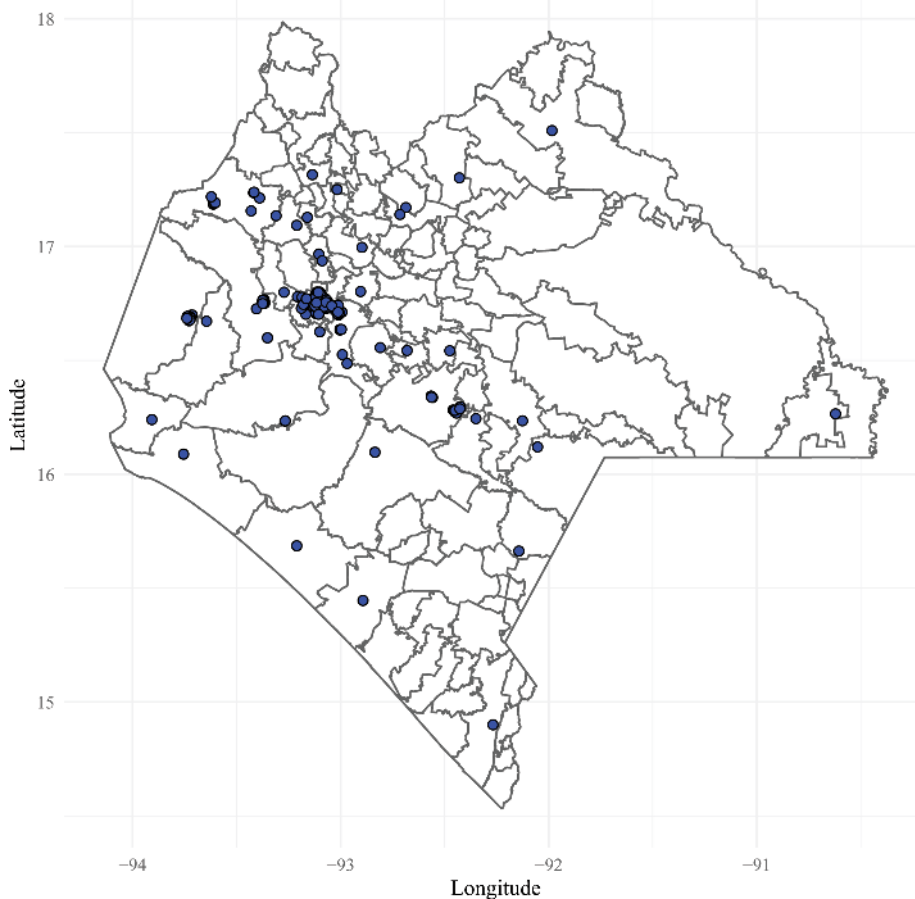


Figure 1.
Georeferencing of cases 573 DCs registered in Chiapas.

of cases observed in the municipalities ($p - value = 0.001$). Therefore, it is convenient to study the CDCs with a spatial model, since it is a counting variable, a Generalized Linear Spatial Model was used.

4.3 Evaluation of the covariance structure of georeferenced data

After selecting the spatial model and the variables, we proceeded to estimate the covariance structure of the Gaussian process. For this, the Exponential, Gaussian, and Matérn covariance functions were tested, taking CDCs as the response variable, and measures of central tendency of the explanatory variables maximum environmental temperature, altitude, patient age, and average monthly rainfall were taken. Of the three functions, the Matérn covariance function generated the best value for $ME = -1.185$ and $MSNE = 0.885$, that is, ME tends to 0 and $MSNE$ tends to 1, therefore a covariance function Matérn can be assumed for the fitted spatial model.

4.4 Parameter estimation

For the simulation and conditional prediction of the process Eq. (6) MCMC was used, since this provides a solution to the impediment of direct calculation of the

Parameter	Estimation coeff.	95% Confidence intervals
Intercept (β_0)	1.88952	(1.88178, 1.96555)
Maximum temp. (β_1)	0.00740	(0.00523, 0.00763)
Altitude (β_2)	0.00028	(0.00026, 0.00028)
Rainfall (β_3)	-0.02549	(-0.02607, -0.02423)
Age (β_4)	-0.05356	(-0.05444, -0.05275)

Table 2.

Estimation of parameters and their confidence intervals of the selected model.

predictive distribution due to the high dimensionality of the integral Eq. (8) [36]. For this, 505000 simulations were performed, with a burn-in period of 5000 data and a thinning of chains of 100 data. Ordinary kriging was used for data interpolation. The initial values for the GLSM parameters were $\sigma^2 = 3$, $\phi = 0.5$ and $\beta = (0.1, 0.1, 0.1, 0.1)$. The estimation of β was carried out under the classical approach. Confidence intervals at 95% were obtained using 1000 Monte Carlo simulated samples [40].

For modeling the number of registered dengue cases in the 36 municipalities of Chiapas, Y_i , $i = 1, \dots, 36$. As for the 13 covariates considered, only the variables maximum environmental temperature, altitude above sea level in the municipality, average monthly rainfall, and patient age showed a relationship with the number of confirmed dengue cases. It was verified that the problem of multicollinearity did not exist in those included in the model: altitude and maximum environmental temperature ($r = -0.2231$, p - value = 0.191), average monthly rainfall and maximum temperature ($r = 0.243$, p - value = 0.1534), average monthly rainfall and altitude ($r = 0.1724$, p - value = 0.3147).

In **Table 2**, it is observed that the variables that have an effect on the cases of dengue observed are maximum environmental temperature, altitude of the municipalities, average monthly rainfall, and patient age. High temperatures and altitudes favor the presence of the disease, while young people will be preferred factors by the vector, as well as low rainfall because in seasons where there is no continuous flow of water in the rivers, stagnation causes an increase in the proliferation of *Aedes* mosquitoes.

4.5 Prediction of the model to the Chiapas map

The projection of the model was carried out on a map of the state of Chiapas which was made based on the municipalities where the cases were registered, as can be seen in **Figure 2**, the prediction is divided by zones in shades of green to yellow with a contour delimited by contour lines that show the area in which the model predicts the number of cases for that area. As we can see, most of the predicted cases occur within the metropolitan area where the state capital Tuxtla Gutiérrez and the municipalities of Chiapa de Corzo, Berriozábal and Suchiapa are located, this corresponds to the observed data, since most of the cases occurred in the same area. On the other hand, it is observed that the prediction power is diminished in areas where no dengue cases were registered.

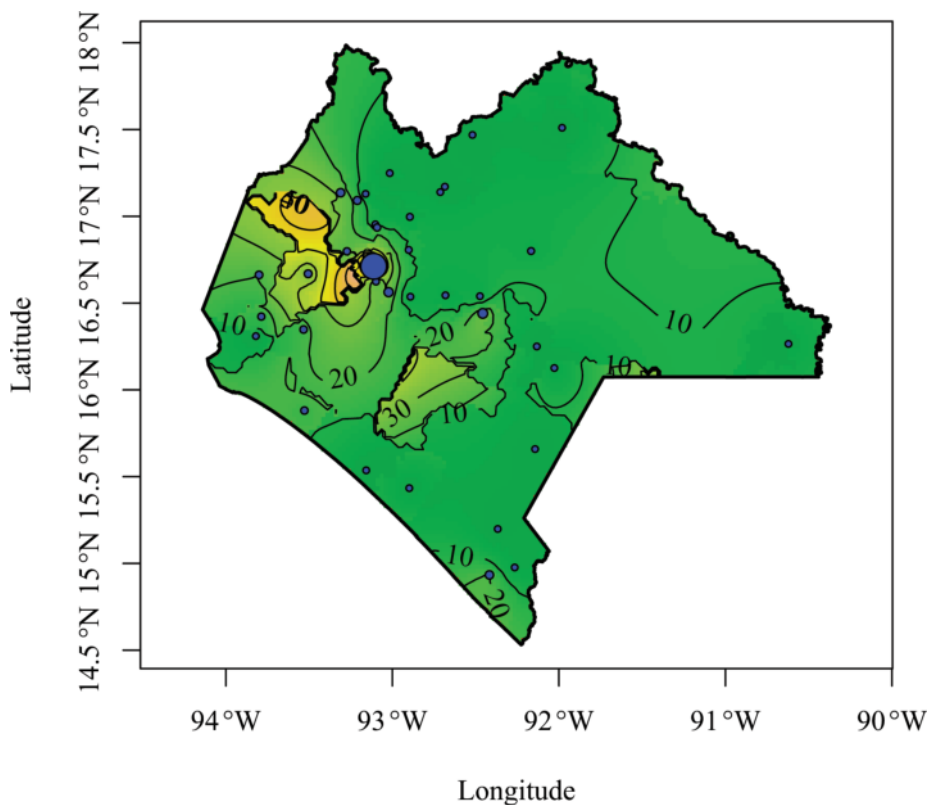


Figure 2.
Prediction of confirmed dengue cases.

5. Discussion

The purpose of this chapter is to present and expand the use of spatial statistics to contribute to public health and the epidemiology of vector-borne diseases, and for this reason, the example of the use of a GLSM was proposed to model the distribution of dengue in Chiapas, since this is one of the endemic diseases that cause numerous infections per year. Climatological, geographic, and sociodemographic variables were used for the modeling, where it was found that the maximum environmental temperature, altitude, patient age, and average monthly rainfall are the variables that best predict the spread of dengue.

Maximum environmental temperature is shown to have a significant effect on dengue cases, as it is an environmental risk factor for dengue transmission, higher temperatures increase viral replication in the vector in a shorter time and thus increase the potential for transmission of dengue viruses. This is described by a study on the extrinsic incubation period. Liu et al. [41] found that the virus remained in the midgut of the vector at 18°C, but could spread and invade the salivary glands at temperatures between 23°C and 32°C, thus demonstrating that higher temperatures create a shorter extrinsic incubation period and greater transmission potential.

The altitude above sea level of each municipality was also an important variable in the study, which is consistent with the findings of the systematic review by Aswi [42], where this variable was used in different statistical models in order to describe the behavior of the disease, since the spread of the *Aedes aegypti* mosquitoes is limited by

climatic conditions and this will be governed by the location of the geographical area and its altitude. The study of Reinhold et al. [43] alludes that *Aedes Aegypti* cannot regulate its body temperature because it is an endothermic arthropod, and that is why its temperature is defined by the climatic conditions of its environment. Thus, geographic location and altitude are important variables for dengue disease.

On the other hand, we have average monthly rainfall, where we see a negative association, since the less rainfall, the more cases of dengue. This coincides with the results of the work of Hashizume et al. [17], where they indicate that dengue cases increase by 29.6% in the months when the rivers have low flow, and this is understandable, since, in those seasons of the year when rainfall is scarce, the rivers do not have a continuous flow of water, which produces stagnation and these, in turn, become ideal breeding grounds for mosquitoes, causing an increase in the proliferation of *Aedes*.

Finally, we have the variable patient age, as can be seen in the results, the correlation was negative too, due to the young population being preferred by the vector, since there is a greater number of cases at an average age of 14 years. As demonstrated by Phanitchat et al. [44] in their work, where it was reported that the age range of dengue cases was between 5 and 14 years in northeastern Thailand.

6. Conclusions

Vector-borne diseases (VBD) are an important public health issue worldwide. The distribution of these diseases as well as their transmission and seasonality are known to be largely determined by environmental, geographic, and socio-demographic factors. GLSMs allow robust analysis of the complex and diverse factors that influence the occurrence of VBD, incorporating spatial dimensions. They can also be a valuable tool for targeting interventions in surveillance and control programs for VBD at the global or regional level. These analytical approaches have recently been used in the field of public health, but in Mexico there are still very few studies that contribute to this knowledge. For this reason, this chapter presents an example of the application of GLSM with a study of dengue, one of the most common VBD in Mexico, finding that the maximum temperature, altitude, and average monthly rainfall of each municipality, as well as patient age, are the factors that best predicted the presence of dengue cases in the state of Chiapas in the period from January to August 2019.

Conflict of interest

The authors declare no conflict of interest.

Abbreviations

ARMA	Autoregressive Moving Average
CAR	Conditional Autoregressive Model
CDCs	Confirmed Dengue Cases
DENV	Dengue Virus
EW	Epidemiological Week
GLM	Generalized Linear Model

GLMM	Generalized Linear Mixed Model
GLSM	Generalized Linear Spatial Model
INEGI	National Institute of Statistic and Geography
INLA	Integrated Nested Laplace Approximation
MAE	Mean Absolute Error
MCMC	Markov Chain Monte Carlo
ME	Mean Error
MI	Moran's Index
MSE	Mean Square Error
MSNE	Mean Square Normalized Error
RMSE	Root Mean Square Error
SAR	Simultaneous Autoregressive Model
VBD	Vector-Borne Diseases
WMO	World Meteorological Organization
ZCL	Zoonotic Cutaneous <i>Leishmaniasis</i>

Author details

Manuel Solís-Navarro^{1†}, Susana G. Guzmán-Aquino^{1†}, María Guzmán-Martínez^{2*}
and Jazmín García-Machorro¹


1 Instituto Politécnico Nacional, Escuela Superior de Medicina, Ciudad de México, México

2 Universidad Autónoma de Guerrero, Chilpancingo de los Bravo, México

*Address all correspondence to: manguzgm@gmail.com

† These authors contributed equally.

IntechOpen

© 2022 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Semenza JC, Menne B. Climate change and infectious diseases in Europe. *Lancet ID*. 2009;**9**:365-375. DOI: 10.1016/S1473-3099(09)70104-5
- [2] World Health Organization [Internet]. 2020. Available from: <https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases>. [Accessed: 2022-03-23]
- [3] Navarrete J, Vázquez J, Gómez H. Epidemiología del dengue y dengue hemorrágico en el Instituto Mexicano del Seguro Social (IMSS). *Revista Peruana de Epidemiología*. 2002; **10**(1):1-12
- [4] Gaetan C, Guyon X. Second-order spatial models and geostatistics. In: *Spatial Statistics and Modeling*. Springer Series in Statistics. New York, NY: Springer; 2010. pp. 1-52. DOI: 10.1007/978-0-387-92257-7_1
- [5] Hernández-Ávila JE, Rodríguez MH, Santos-Luna R, Sánchez-Castañeda V, Román-Pérez S, Ríos-Salgado VH, et al. Nation-wide, web-based, geographic information system for the integrated surveillance and control of dengue fever in Mexico. *PLoS One*. 2013;**8**(8):1-9. DOI: 10.1371/journal.pone.0070231
- [6] Sinharay S. An overview of statistics in education. In: *International Encyclopedia of Education*. Third ed. Princeton, NJ, USA: Elsevier Science; 2011. pp. 1-11. DOI: 10.1016/B978-0-08-044894-7.01719-X
- [7] Nelder JA, Wedderburn RWM. Generalized linear models. *Journal of the Royal Statistical Society A*. 1972;**135**: 370-384. DOI: 10.2307/2344614
- [8] McCullagh P, Nelder JA. *Generalized Linear Models*. London, UK: Chapman and Hall; 1989
- [9] Christensen OF, Waagepetersen R. Bayesian prediction of spatial count data using generalized linear mixed models. *Biometrics*. 2002;**58**(2):280-286
- [10] Diggle P, Moyeed R, Rowlingson B, Thomson M. Childhood malaria in the Gambia: A case-study in model-based geostatistics. *Journal of the Royal Statistical Society C*. 2002;**51**(4):493-506
- [11] Diggle PJ, Ribeiro PJ, Christensen OF. An introduction to model-based geostatistics. In: Møller J, editor. *Spatial Statistics and Computational Methods*. New York, NY, USA: Springer; 2003. pp. 43-86. DOI: 10.1007/978-0-387-21811-3_2
- [12] Zhang H. On estimation and prediction for spatial generalized linear mixed models. *Biometrics*. 2002;**58**(1): 129-136. DOI: 10.1111/j.0006-341X.2002.00129.x
- [13] Zhang H. Optimal interpolation and the appropriateness of cross-validating variogram in spatial generalized linear mixed models. *Journal of Computational and Graphical Statistics*. 2003;**12**(3): 698-713. DOI: 10.1198/1061860032265
- [14] Saavedra A, Taboada J, Araújo M, Giráldez E. Generalized linear spatial models to predict slate exploitability. *Journal of Applied Mathematics*. 2013;**2013**:1-7. DOI: 10.1155/2013/531062
- [15] Guyon X. Modelacion para la estadística espacial. *Revista de Investigacion Operacional*. 2010;**31**(1): 1-33. DOI: 10.0257-4306
- [16] Ver Hoef JM, Hanks EM, Hooten MB. On the relationship between conditional (CAR) and simultaneous

- (SAR) autoregressive models. *SpatStat*. 2018;**25**:68-85
- [17] Hashizume M, Dewan AM, Sunahara T, Rahman MZ, Yamamoto T. Hydroclimatological variability and dengue transmission in Dhaka, Bangladesh: A time-series study. *BMC Infectious Diseases*. 2012;**12**(98):1-9. DOI: 10.1186/1471-2334-12-98
- [18] Breslow N, Clayton D. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*. 1993;**88**:9-25. DOI: 10.2307/2290687
- [19] Bandera Fernández E, Pérez PL. Los modelos lineales generalizados mixtos. Su aplicación en el mejoramiento de plantas. *Cultivos Tropicales*. 2018;**39**(1): 127-133. ISSN: 0258-5936
- [20] Graham AJ, Atkinson PM, Danson FM. Spatial analysis for epidemiology. *Science Direct*. 2004;**91**(3):219-225. DOI: 10.1016/j.actatropica.2004.05.001
- [21] Sánchez Pérez DL, Santa F, Fuentes López HJ. Spatial regression models for the behavior of infectious diseases dengue and malaria in Colombia for the years 2000, 2005 and 2010. *Revista Geomática UD.GEO*. 2012;**6**:110-128
- [22] Estallo EL, Santana M, Martín ME, Galindo LM, Willener JA, Kuruc JA, et al. Environmental effects on phlebotominae sand flies (Diptera: Phychodidae) and implications for sand fly vector disease transmission in Corrientes city, northern Argentina. *Anais da Academia Brasileira de Ciências*. 2021;**93**(3):1-17. DOI: 10.1590/0001-3765202120191278
- [23] Valbuena-Garcia AM, Rodríguez-Villamizar LA. Spatial analysis in epidemiology: Methodological review. *Revista de la Universidad Industrial de Santander. Salud*. 2018;**50**(4):358-365. DOI: 10.18273/revsal.v50n4-2018009
- [24] Cressie N. *Statistics for Spatial Data*. Chichester-Toronto-Brisbane-Singapore. New York: John Wiley & Sons, Inc.; 1991. DOI: 10.1002/bimj.4710350210
- [25] Sherman M. *Spatial Statistics and Spatio-Temporal Data*. Chichester, England: John Wiley & Sons, Inc; 2011. ISBN: 978-0-470-69958-4
- [26] Chen Z, Wang B. How priors of initial hyperparameters affect Gaussian process regression models. *Neurocomputing*. 2018;**275**:1702-1710. DOI: 10.1016/j.neucom.2017.10.028
- [27] Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. *Journal of the Royal Statistical Society C*. 1998;**47**(3): 299-350. DOI: 10.1111/1467-9876.00113
- [28] Diggle PJ, Ribeiro PJ Jr. *Model-based Geostatistics*. In: Springer Series in Statistics. New York, NY, USA: Springer Science + Business Media, LLC; 2007. ISBN: 978-0-387-48536-2
- [29] Gaetan C, Guyon X. *Spatial Statistics and Modeling*. New York: Springer; 2010. p. 90. ISBN: 9780387922577
- [30] Kerry R, Oliver MA. Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma*. 2007;**140**(4): 383-396. DOI: 10.1016/j.geoderma.2007.04.019
- [31] Diggle PJ, Tawn JA, Moyeed RA. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1998;**47**(3):299-350. DOI: 10.1111/1467-9876.00113
- [32] Moran PA. Notes on continuous stochastic phenomena. *Biometrika*. 1950; **37**(1/2):17-23. DOI: 10.2307/2332142

- [33] Banerjee S, Carlin BP, Gelfand AE. Hierarchical Modeling and Analysis for Spatial Data. Florida, USA: CRC press; 2014. ISBN 9781439819173
- [34] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021
- [35] Ribeiro PJ Jr, Christensen OF, Diggle PJ. Geostatistical software - geoR and geoRglm. In: DSC 2003 Working Papers. Vienna, Austria: Distributed Statistical Computing; 2003
- [36] Christensen OF, Ribeiro PJ. geoRglm: A Package for Generalised Linear Spatial Models. Vienna, Austria: R news; 2002
- [37] Instituto Nacional de Estadística y Geografía. [Internet]. 2019. Available from: <https://www.inegi.org.mx/>. [Accessed: 2019-12-11]
- [38] World Meteorological Organization [Internet]. 2019. Available from: <https://public.wmo.int/es>. [Accessed: 2019-11-23]
- [39] Chan M, Johansson A. The incubation periods of dengue viruses. PLoS One. 2012;7(11):2-6. DOI: 10.1371/journal.pone.0050972
- [40] Davison AC, Hinkley DV. Bootstrap Methods and their Application. Vol. No. 1. Cambridge, England: Cambridge University Press; 1997. DOI: 10.1017/CBO9780511802843
- [41] Liu Z, Zhang Z, Lai Z, Zhou T, Jia Z, Gu J, et al. Temperature increase enhances *Aedes albopictus* competence to transmit dengue virus. *Frontiers in Microbiology*. 2017;8(2337):1-7. DOI: 10.3389/fmicb.2017.02337
- [42] Aswi A, Cramb SM, Moraga P, Mengersen K. Bayesian spatial and spatio-temporal approaches to modelling dengue fever: A systematic review. *Epidemiology and Infection*. 2018; 29(147):1-14
- [43] Reinhold JM, Lazzari CR, Lahondère C. Effects of the environmental temperature on *Aedes aegypti* and *Aedes albopictus* mosquitoes: A review. *Insects*. 2018;9(4):1-17. DOI: 10.3390/insects9040158
- [44] Phanitchat T, Zhao B, Haque U, et al. Spatial and temporal patterns of dengue incidence in northeastern Thailand 2006–2016. *BMC Infectious Diseases*. 2019;19(743):1-12. DOI: 10.1186/s12879-019-4379-3