

RESEARCH ARTICLE

ARBOALVO: A Bayesian spatiotemporal learning and predictive model for dengue cases in the endemic Northeast city of Natal, Rio Grande do Norte, Brazil

Mariane Branco Alves^{1*}, Rafael Santos Erbisti^{2*}, Aline Araújo Nobre³, Taynãna César Simões⁴, Alessandre de Medeiros Tavares[†], Márcia Cristina Melo⁵, Rodrigo Moreira Pedreira⁵, Jan Pierre Martins de Araújo⁵, Marília Sá Carvalho³, Nildimar Alves Honório^{6,7*}

1 Departamento de Métodos Estatísticos, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, **2** Departamento de Estatística, Instituto de Matemática e Estatística, Universidade Federal Fluminense, Niterói, Brazil, **3** Programa de Computação Científica da Fiocruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, **4** Instituto René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Brazil, **5** Centro de Controle de Zoonoses, Secretaria Municipal de Saúde de Natal, Rio Grande do Norte, Brazil, **6** Laboratório das Interações Vírus-Hospedeiros, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, **7** Núcleo Operacional Sentinela de Mosquitos Vetores—Nosmove/Fiocruz, Rio de Janeiro, Brazil

† in memoriam

* erbisti@id.uff.br (RSE); nildimar.honorio@ioc.fiocruz.br (NAH); mariane@im.ufrj.br (MBA)



OPEN ACCESS

Citation: Alves MB, Erbisti RS, Nobre AA, Simões TC, Tavares (in memoriam) A de M, Melo MC, et al. (2025) ARBOALVO: A Bayesian spatiotemporal learning and predictive model for dengue cases in the endemic Northeast city of Natal, Rio Grande do Norte, Brazil. PLoS Negl Trop Dis 19(4): e0012984. <https://doi.org/10.1371/journal.pntd.0012984>

Editor: Paul O. Mireji, Kenya Agricultural and Livestock Research Organization, KENYA

Received: July 23, 2024

Accepted: March 12, 2025

Published: April 29, 2025

Copyright: © 2025 Alves et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: Epidemiological data were obtained from the Information System of Notifiable Diseases (SINAN) of the SUS Informatics Department (DATASUS) (<https://datasus.saude.gov.br/>)

Abstract

Background

Urban arbovirus transmission is spatially and temporally heterogeneous. Estimating the risk of dengue through statistical models that consider simultaneous variability in space and time provides more realistic estimates of transmission dynamics, facilitating the identification of priority areas for intervention focused on surveillance and control. These models also enable predictions to support timely interventions for arboviruses like dengue, chikungunya, and Zika.

Methodology/principal findings

We analyzed dengue case reports by epidemiological week and neighborhood in Natal, RN from 2015 to 2018. Temporal conditional autoregressive models were fitted using the Integrated Nested Laplace Approximation method. The predictors included a set of entomological, climatic and sociosanitary indicators with temporal lags, along with structures of temporal and spatial dependence. Additionally, we used an offset term to represent the expected number of dengue cases per neighborhood at each epidemiological week, under the hypothesis of homogeneity in the occurrence of cases across the municipality. We forecasted dengue case counts for the subsequent four weeks, addressing both zero occurrences and fluctuations during non-zero

[transferencia-de-arquivos/](#)) and requested from the Zoonosis Control Center of the Municipal Health Secretariat of Natal/RN (<https://natal.rn.gov.br/sms/contato>). The entomological data are confidential and can be requested from the Zoonosis Control Center of the Municipal Health Secretariat of Natal/RN (<https://natal.rn.gov.br/sms/contato>). Sociosanitary data can be obtained from the microdata of the 2010 Demographic Census – Universe Results, conducted by the Brazilian Institute of Geography and Statistics (<https://www.ibge.gov.br/estatisticas/downloads-estatisticas.html>). Climatic data were obtained from satellite images available from the International Research Institute for Climate and Society (IRI), Columbia University (<https://iridl.ldeo.columbia.edu/SOURCES/>), and data from a rainfall station located in the city of Natal-RN, provided by the National Institute of Meteorology (INMET) (<https://portal.inmet.gov.br/dadoshistoricos>).

Funding: Funded by the Brazilian Ministry of Health. Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) - grant number E-26/201.207/2022 (NAH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

periods. Weekly risk dynamics were visualized through predictive maps, enabling the timely identification of neighborhoods with high and persistent dengue risk, that is, areas consistently exhibiting a high number of dengue cases that remained concentrated in the same location for several weeks. The optimal model revealed a significant rise in dengue occurrence probability during the observation week, associated with increased cases in the previous week, the *Aedes* egg positivity index from the prior four weeks, and the mean daytime temperature 6–8 weeks earlier. Dengue risk also rose with a one-standard-deviation increase in the density of the impoverished population per occupied area and the mean *Aedes* egg density index from the preceding 3–5 weeks.

Conclusions/significance

The proposed Bayesian space-time analysis can contribute to the operational control of dengue and *Aedes aegypti* by identifying priority areas and forecasting dengue cases for the next four weeks. It also quantifies the effects of entomological, socio-sanitary, climatic and demographic indicators on both the likelihood of dengue occurrence and the intensity of outbreaks.

Author summary

Urban arbovirus transmission exhibits spatial and temporal heterogeneity. Focusing on dengue, we employed statistical models that account for spatiotemporal variability to provide a realistic assessment of transmission dynamics. By using dengue case data from Natal, RN (2015–2018) and incorporating entomological, climatic, and sociosanitary indicators, we forecasted case counts for the following four weeks. Our analysis identified significant correlations between increased dengue risk and key factors: reported cases in the previous week, the *Aedes* egg positivity index, and mean daytime temperature from preceding weeks. High-risk neighborhoods with persistent dengue transmission were pinpointed, emphasizing areas for targeted interventions. This Bayesian space-time approach supports operational control efforts by identifying priority areas and clarifying how various factors influence dengue transmission. The findings have significant public health implications, enabling more precise, proactive strategies to mitigate the incidence and impact of urban arboviruses.

Introduction

Emerging arboviruses such as dengue, Zika, and chikungunya pose one of the most significant global public health problems [1–4]. They are primarily transmitted by *Ae. aegypti* (Linnaeus, 1762), a vector well-established in tropical and subtropical regions [5,6]. Dengue has been endemic in Brazil for over 30 years, with successive

epidemics [7–10]. This ongoing challenge was compounded by the emergence of chikungunya virus in 2014 [2] and Zika virus in 2015 [4,11]. According to the data publicized by the Brazilian Health Ministry, 2023 saw 1,658,816 dengue cases (81.69 cases per 10,000 inhabitants), 154,800 probable chikungunya cases (7.62 cases per 10,000), and 7,292 probable Zika cases (0.36 cases per 10,000).

In Brazil, the co-circulation of dengue, Zika and chikungunya intensifies public health challenges, contributing to significant morbidity, mortality and economic burden [2,10]. Continuous monitoring strategies are crucial, particularly given the territorial disparities of disease transmission. While arbovirus case counts vary across time and space, patterns of occurrence may exhibit similarities in neighboring areas and during proximate time periods. Among Brazilian areas affected by dengue epidemics, Natal, the capital of Rio Grande do Norte, has faced recurrent outbreaks since the arbovirus emerged in 1996 [12], with noteworthy epidemics recorded in 2008, 2016, and 2022. Located in Brazil's Northeast region, Natal experiences high tourist activity, potentially contributing to the co-circulation of arboviruses and the dispersal of their vectors [13–19].

Current strategies for controlling these arboviruses involve targeting their primary vector, *Ae. aegypti*, but have yielded unsatisfactory results. The spatial heterogeneity of urban arbovirus transmission highlights the need for strategies that account for territorial complexity and differences, including the identification of priority areas—an approach considered promising for effectively targeted control actions [20,21]. Integrating entomological, epidemiological, and socio-environmental information specific to each territory is essential for the comprehensive management of urban arboviruses. To develop a more rational and tailored approach suitable for countries where dengue is endemic or epidemic, current control strategies should be reevaluated, beginning with an understanding of the demographic, environmental and social factors that drive dengue fever transmission dynamics [22,23].

Reliable data on intra-urban vector population estimates, trends in susceptible human populations, and virus entry timing are essential for understanding epidemic dynamics but are often lacking. In their absence, proxies such as entomological, climatic and sociosanitary indicators serve as indirect influences on urban arbovirus cases. Climatic factors, such as temperature, impact *Ae. aegypti* populations and viral dynamics [24]. Sociosanitary factors are expected to directly affect mosquito populations and indirectly increase infection risks due to urban conditions and high population densities [20,23].

This study pertains to the methodology portfolio developed by the ARBOALVO project to stratify priority intervention areas for urban arboviruses transmitted by *Ae. aegypti* [23,25]. The underlying hypothesis of this project is that the spatial distribution of cases is not homogeneous across territories and over time. To address this, Bayesian space-time models were applied, incorporating spatially and temporally structured random effects and accounting for dengue occurrences in neighboring areas and adjacent time periods. The model used was Conditional AutoRegressive (CAR), which adjusts for such effects in each neighborhood-epidemiological week combination. This approach allows assessing the impact of explanatory variables on the epidemiological risk of arbovirus transmission [26]. These effects were jointly estimated, considering various dimensions of the transmission process. The resulting relative risks of dengue occurrence were mapped, formally accounting for estimation uncertainty and identifying regions with persistent and/or heightened risks over time.

The spatial and temporal heterogeneity of transmission poses a challenge to understanding the risk of epidemics, and hinders the design of strategic control measures. Thus, the quantitative analysis of epidemiological dynamics in a spatiotemporal framework, as enabled by the model class employed, can provide valuable insights to support the development of more effective control strategies. These measures can include focused campaigns in areas and periods of persistently high risk and monitoring the emergence of new high-transmission zones. This study aimed to i) identify and quantify the effects of entomological, climatic, sociosanitary and demographic indicators as potential explanatory factors for dengue occurrence in Natal, RN; ii) predict the time series of dengue cases for each neighborhood up to four weeks ahead; and iii) forecast dengue cases stratified by the municipality's neighborhoods, generating dynamic risk maps over time.

Methods

Ethics statement

This study was conducted within the framework of ARBOALVO program, approved by the research ethics committee of the Oswaldo Cruz Institute (CEP-IOC), under Protocol No. 51057015.5.0000.5537.

Area and design of study

This study employed an ecological analysis incorporating time-series and spatial heterogeneity methods to investigate factors influencing dengue occurrence in Natal, the capital of Rio Grande do Norte (RN) in Northeast Brazil. According to the Brazilian Institute of Geography and Statistics (IBGE), Natal has a population of approximately 896,708, ranking as the 20th most populous municipality in the country, with an area of 167 km² [27]. Located at 05°47'42" S and 35°12'32" W at an altitude of 31 meters, Natal has a tropical climate with a dry summer, small temperature variation, and rainfall concentrated between March and July. The city is divided into 36 neighborhoods across four administrative regions, each characterized by distinct territorial, physical, demographic and infrastructural attributes (Fig 1). The eastern region comprises 12 neighborhoods, averaging 114,364 residents during the analysis period. The northern region includes 7 neighborhoods with a total population of 324,995. The western region consists of 10 neighborhoods with an average population of 226,734, while the southern region has 7 neighborhoods with an average of 170,364 residents. The population distribution by region and neighborhood for the analysis period is detailed in S1 Table. According to indicators such as the percentage of households with exclusive-use bathrooms and sewage disposal via rudimentary pits, trenches or natural water bodies; the percentage of households with open sewage in the vicinity; and the density of poor residents per occupied area, the highest sociosanitary deprivation was observed in neighborhoods in the northern and western regions [23,27]. Throughout this text, "occupied area" refers to the territorial space physically inhabited by people, defined as the ratio of the area occupied by anthropogenic structures to the total coverage area.

Data source

The data used in this study are part of the ARBOALVO project—Methodological Proposal for Stratification of Risk Areas for Dengue, Chikungunya, and Zika in Brazilian Endemic Cities—funded by the Ministry of Health. The epidemiological data consist of reported dengue cases from 2015 to 2018, including laboratory-confirmed and suspected cases. These data were obtained from the Ministry of Health through the Information System for Notifiable Diseases (SINAN) of the SUS Informatics Department (DATASUS) and were subsequently aggregated by epidemiological week and neighborhood in Natal. Entomological data were sourced from the Municipal Health Department of Natal, consisting of weekly counts of eggs deposited by female *Ae. aegypti* mosquitoes in 400 oviposition traps (ovitrap). The ovitrap is a black plastic container (500 mL) with a Eucatex fiberboard paddle attached to its edge by a clip (details described in [28]). These traps were distributed in 36 neighborhoods from September 2015 to April 2018. The egg density index (EDI) was calculated by dividing the total number of *Aedes* eggs found on the paddles by the number of positive ovitraps. The egg positivity index (EPI) was calculated as the percentage of ovitraps positive for *Aedes* eggs, aggregated by neighborhood and epidemiological week. Sociosanitary and demographic data were obtained from the 2010 demographic census [29]. Climatic data, including precipitation and day and night surface temperatures, were obtained from two datasets: satellite data from the International Research Institute for Climate and Society (IRI) at Columbia University, and rainfall data from a station located in Natal, provided by the National Institute of Meteorology (INMET). The temporal window of the study, encompassing all available data, was defined by the span of the entomological data, which ranged from the 37th epidemiological week of 2015 to the 16th epidemiological week of 2018. Indicators derived from the databases at the neighborhood level of Natal are presented in S2 Table.

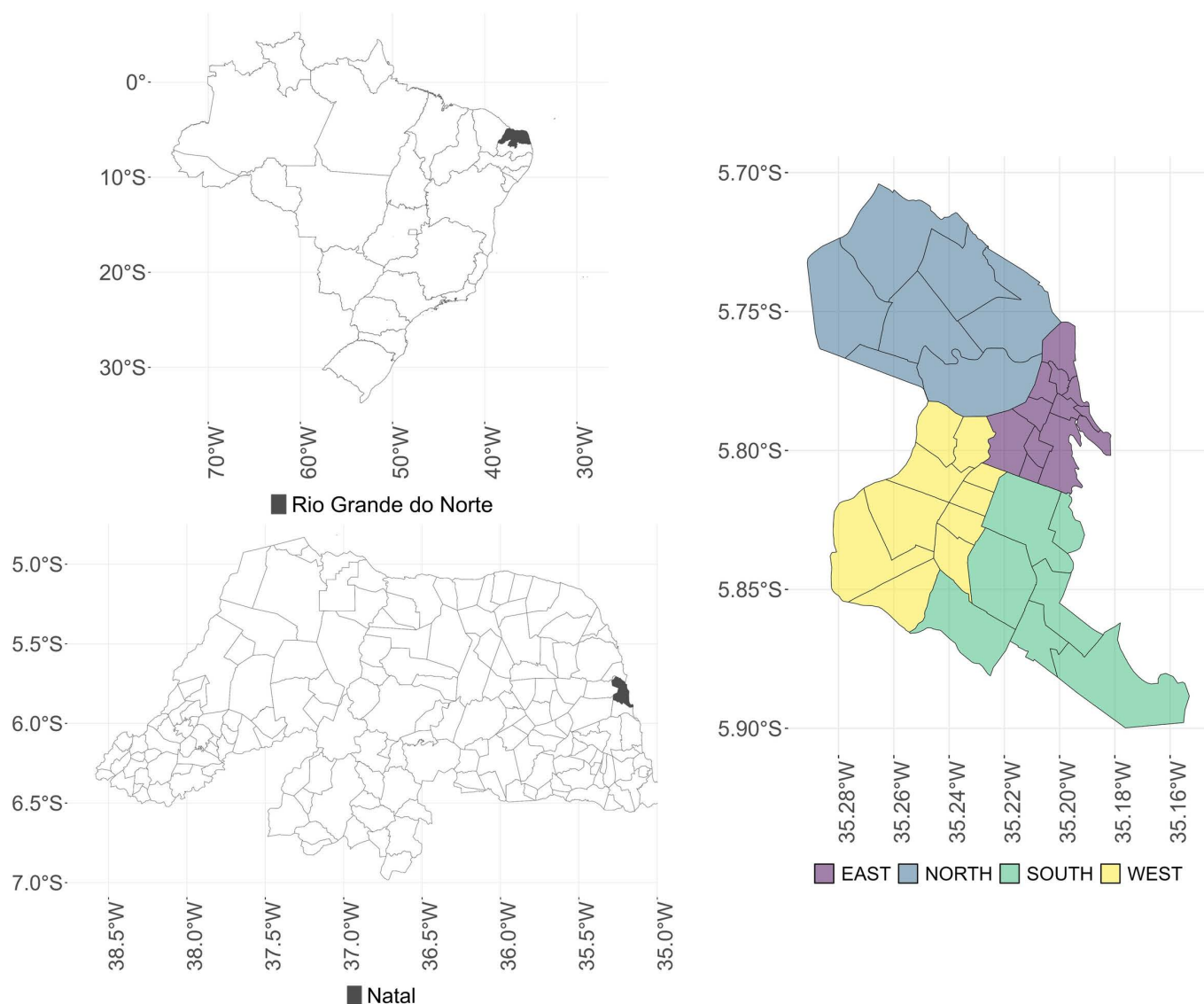


Fig 1. Map of the city of Natal by administrative regions. Source: Figure created by the authors. Shapefile of the neighborhoods in the city of Natal: <https://www.natal.rn.gov.br/semurb/geoinformacoes>. Shapefile of the municipalities in the state of Rio Grande do Norte and shapefile of the Brazilian states: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malas-territoriais/15774-malas.html>.

<https://doi.org/10.1371/journal.pntd.0012984.g001>

Statistical analysis - preliminary stages to modeling dengue cases

Data imputation and scale compatibility of climate data. Day and night temperature data gathered by the satellite were recorded every eight days, with some periods having missing data. Similarly, entomological data were incomplete in certain periods due to loss or other issues related to the ovitraps. To address these gaps, missing records of day and night temperatures, as well as *Aedes* egg density and positivity, were imputed using temporal conditional autoregressive models (temporal CAR) [30]. Climatic variables exhibit varying temporal and spatial scales across various data sources. Precipitation data from satellites were provided on a per-neighborhood and monthly basis, while daily rainfall data were obtained from a single station within the municipality. To ensure compatibility on the epidemiological weekly time scale, data from both sources were combined by multiplying the daily rainfall proportion from the municipal station by the monthly data

for each neighborhood. This approach assumes a uniform rainfall regime across all neighborhoods and depends on data availability. For temperature data, which were provided by the satellite every eight days, a weighted average was calculated for each epidemiological week. The weights were based on the number of overlapping days between the observed epidemiological week and each eight-day satellite cycle. After this process, a comprehensive and compatible database containing temperature and rainfall data for all epidemiological weeks and neighborhoods was developed.

Analysis of the lagged impact of entomological and climatic indicators. One of the initial steps in developing the spatiotemporal models proposed in this study was to evaluate the effects of entomological and climatic indicators at different time lags on dengue cases. For instance, it is unreasonable to assume that egg density observed in a specific epidemiological week would immediately influence dengue occurrence. Similarly, other climatic indicators do not immediately affect the outcomes. To address this, we adopted purely temporal models for the entire municipality to assess the lagged effects of these variables on dengue transmission. Let Y_t denote the number of dengue cases in the municipality of Natal during week t , and X_t represent a set of indicators (e.g., average rainfall volume, average egg density, proportion of positive ovitraps, and average daytime or nighttime temperature) at period t . A distributed lag structure [31,32] is assumed, which can be expressed as:

$$Y_t \sim \text{Poisson}(\mu_t)$$

$$\log(\mu_t) = \zeta + \gamma_0 X_t + \gamma_1 X_{t-1} + \dots + \gamma_q X_{t-q}. \quad (1)$$

The proper estimation of the coefficients $\gamma_0, \gamma_1, \dots, \gamma_q$, would provide insight into the temporal propagation pattern of the impact of variations in the indicator X , observed q periods ago, on the response Y at the current time. However, from a statistical perspective, the estimation of these coefficients is hindered by the association present in indicator X across successive periods. One way to address this issue is to approximate the trajectory of the coefficients γ_j (as a function of the lag order j) using a low-degree ($d \ll q$) polynomial. The following reparameterization was applied:

$$\gamma_j = \sum_{k=0}^d \eta_k j^k, \quad j = 0, 1, 2, \dots, q. \quad (2)$$

Rewriting the parameters γ_j in (1), according to the approximation (2), results in an expression for the predictor in terms of $d + 1$ parameters ($\eta_0, \eta_1, \dots, \eta_d$):

$$\log(\mu_t) = \eta_0 S_{t0} + \eta_1 S_{t1} + \dots + \eta_d S_{td},$$

$$S_{tk} = \sum_{j=0}^q j^k X_{t-j}, \quad k = 1, \dots, d. \quad (3)$$

For smooth trajectories, a low-degree ($d \ll q$) polynomial is sufficient to approximate the lag effects. Additionally, the new explanatory variables S_{tj} , obtained by aggregating the original indicators across their various lags, do not suffer from the collinearity problem typically observed in the original indicators. This approach resolves the instability in the estimation of the individual coefficients γ_j , leading to more stable and reliable parameter estimates. As a result, the estimation of the trajectory of the lagged impact curve of X on Y is improved, providing a more robust understanding of the temporal dynamics in the model [32].

Applications of the distributed-lag method in an epidemiological context have been explored in studies by Schwartz [32], Zanobetti et al. [33], and Xiang et al. [34]. In the present study, this method was employed to assess the influence horizon of indicators such as the egg density index (EDI), egg positivity index (EPI), rainfall volume, daytime temperature, and nighttime temperature on dengue case counts. The analysis was conducted using the `dlm` package [35] in the statistical software R [36] to determine the appropriate number of time lags for each indicator included in the spatiotemporal models that consider the counts of dengue cases, in addition to entomological, climatic and sociosanitary variables.

Selection of Sociosanitary Indicators. Building predictors based on combining several indicators can compromise the practical application of statistical models. Furthermore, uncertainties associated with estimated coefficients and adjusted values may be influenced by the inclusion of terms that do not effectively contribute to predictive capacity. Variable selection methods, which aim to reduce the dimensionality of the explanatory variable set, are valuable in this context. In this study, we applied the LASSO (Least Absolute Shrinkage and Selection Operator) regularization method [37], which imposes a constraint on model coefficients by shrinking them towards zero, effectively nullifying some coefficients. Therefore, this method serves as a variable selection approach. A preliminary analysis of various sociosanitary covariates was conducted using LASSO to identify those with the greatest ability to discriminate dengue cases. The sociosanitary indicators did not vary over time, since the source of the available data was the 2010 demographic census [29]. We adopted a simplifying assumption of temporal constancy of their impacts. Case counts were aggregated for the years 2016 and 2017, as well as over the entire analysis period. Thus, the LASSO procedure was applied to the accumulated dengue case counts at this stage to avoid the issue of excess zeros. The indicators identified as relevant in this preliminary analysis were then tested in the predictor of spatiotemporal CAR models in a subsequent stage. The R glmnet library [38] was used to implement the LASSO procedure, considering various neighborhoods as independent observations for each year. A Poisson distribution was assumed for the dengue occurrence counts. The initial selection of sociosanitary indicators as potential predictors was carried out according to a preliminary study using principal component analysis (PCA) [25]. At the conclusion of the LASSO procedure, the indicators marked with the symbol “*” in S2 Table were selected as candidates to compose the predictors for the models described in the following section.

Mixed Bernoulli-Poisson spatiotemporal model

Proposed model. In the more general spatiotemporal model adopted in this study, similar to the approach of Da Costa et al. [39], we assumed that the distribution of dengue cases could be represented by a mixture of two processes: a binary process (guided by a Bernoulli distribution) describing the absence or occurrence of cases, and a second process describing the counts of cases (guided by a Poisson distribution). For each combination of spatial unit i ($i = 1, \dots, N$) and temporal unit t ($t = 1, \dots, T$), Y_{it} denotes a random variable representing the counts of cases, and we assumed that:

$$Y_{it} | \mu_{it} \sim P(\mu_{it})$$

$$\log(\mu_{it}) = P_{it1} = \log(E_{it}) + \alpha_t + \sum_{l=1}^L \beta_l^y x_{i,t,l}^{(y)} + \xi_{i,t} \quad (4)$$

$$\alpha_t = \alpha_{t-1} + \omega_t, \quad \omega_t \sim N(0, W)$$

$$\xi_{i,t} = \rho \xi_{i,t-1} + v_{i,t}$$

where $P(\mu_{it})$ denotes the Poisson or zero-truncated Poisson distribution (truncated Poisson distribution at zero: $P(Y_{it} = 0 | \mu_{it}) = 0$), with intensity μ_{it} , which is linked by logarithmic transformation to a linear predictor P_{it1} described by L regressors, $\mathbf{x}_{i,t}^{(y)} = (x_{i,t,1}^{(y)}, \dots, x_{i,t,L}^{(y)})$, representing entomological, climatic and sociosanitary indicators recorded in neighborhood i and time t . The vector $\beta^y = (\beta_1^y, \dots, \beta_L^y)$ denotes the coefficients of the regressors in the predictor for counts of cases. For the relativization of case counts by the population living in the neighborhoods, *offset* terms $\log(E_{it})$ were used, with E_{it} denoting the expected number of cases per neighborhood at time t , assuming homogeneity in the distribution of cases throughout the municipality. The municipal rate of cases per inhabitant was multiplied, at each week, by the size of the resident population of each neighborhood: $E_{it} = \frac{\text{reported cases}_t}{\text{population}_t} \times \text{population}_{it} = \frac{\sum_{i=1}^N \text{reported cases}_{it}}{\sum_{i=1}^N \text{population}_{it}} \times \text{population}_{it}$. The

projection for the inhabitants in each neighborhood was estimated based on the average annual growth rate between 2000 and 2010 derived from census data. We emphasize that in this study, the offset terms are primarily important due to the spatial variation of population sizes rather than temporal variation during the analysis period.

The linear predictor is also composed of a dynamic level α_t with a random walk temporal evolution, combined with a spatiotemporal structured error $\xi_{i,t}$, guided by the autoregressive coefficient ρ ($|\rho| < 1$), allowing accommodation of possible overdispersion of the counts. The main purpose of including the effects $\xi_{i,t}$ is to formally account for temporal and spatial autocorrelation in case occurrences. These components reflect the notion that spatial and temporal dependencies within each neighborhood are linked to the space-time patterns of adjacent neighborhoods and previous weeks. This structure can be described by an intrinsic CAR specification [26] for the random effects $v_{i,t}$:

$$v_{i,t}|v_{-i,t}, \tau \sim N\left(\frac{1}{n_i} \sum_{j \sim i} v_{j,t}, \frac{1}{\tau n_i}\right) \quad (5)$$

where $i \sim j \equiv j \sim i$ indicates that the neighborhoods i and j are adjacent and n_i denotes the number of adjacent spatial units to neighborhood i . In this study, neighborhoods were considered adjacent if they shared borders. Thus, the spatial structure induced by $v_{i,t}$ considers an average of the process in neighborhood i , with increasing precision depending on the number of neighbors. Although quite intuitive, specification (5) results in an improper prior distribution for $v_{i,t}$, which can be corrected with the adoption of the specification

$$v_{i,t}|v_{-i,t}, \tau, \delta \sim N\left(\frac{1}{\delta + n_i} \sum_{j \sim i} v_{j,t}, \frac{1}{\tau(\delta + n_i)}\right) \quad (6)$$

where $\delta > 0$ is an additional parameter introduced on the main diagonal of the precision matrix. The specification (6) corresponds to the adoption of a precision matrix $\mathbf{Q} = \{\mathbf{Q}_{ij}\}$; $\mathbf{Q}_{ij} = -\tau$; $\mathbf{Q}_{ii} = \tau(\delta + n_i)$. In this work, we adopted the specification proposed by Leroux et al. [30], where the precision matrix of the intrinsic CAR process is defined as a convex linear combination of the identity matrix (representing spatially independent effects) and the precision matrix (6): $(1 - \lambda)\mathbf{I} + \lambda\mathbf{Q}$. Thus, the parameter λ ($0 < \lambda < 1$) reflects the level of spatial structure in the data, with values close to one indicating strong spatial dependence and values near zero indicating spatial independence.

As previously mentioned, the general spatiotemporal model used in this study assumes a mixture of the Poisson or zero-truncated Poisson model for counts and a binary model for distinguishing null occurrences. Let Z_{it} denote the binary variable such that

$$Z_{it}|\pi_{it} \sim \text{Bernoulli}(\pi_{it}) \quad (7)$$

$$\log\left(\frac{\pi_{it}}{1 - \pi_{it}}\right) = P_{it2} = \beta_0^z + \sum_{m=1}^M \beta_m^z \mathbf{x}_{i,t,m}^{(z)} + \varphi \xi_{i,t}$$

where $1 - \pi_{it}$ is the unobserved probability of belonging to the point mass component. Thus, the predictor P_{it2} describes the occurrence or non-occurrence of dengue cases, structured in terms of parameters $\beta^z = (\beta_0^z, \beta_1^z, \dots, \beta_M^z)$ for entomological, climatic and sociosanitary covariates, $\mathbf{x}_{i,t}^{(z)}$, which may or may not overlap with the predictor P_{it1} for case counts. The temporally and spatially structured random effect $\xi_{i,t}$ in (4) ensures smoothness in time and space for the binary process of case occurrence, scaled by a factor φ . From a Bayesian perspective, the model specification is completed with a prior distribution for the latent components. We assumed prior independence, expressed through the following distributions for the parametric vector $\Theta = (\beta, \mathbf{W}, \tau, \delta, \lambda, \varphi, \rho)$, where $\beta = (\beta^y, \beta^z)$: $\beta \sim \text{MVN}(\mathbf{0}, 100000\mathbf{I}_{(L+M)})$, where $\mathbf{I}_{(L+M)}$ denotes the identity matrix of dimension $(L + M) \times (L + M)$, $\mathbf{W}^{-1} \sim \text{Gamma}(0.0005, 0.0005)$, $\tau \sim \text{Gamma}(0.0005, 0.0005)$, $\delta \sim \text{Gamma}(1, 1)$, $\text{logit}(\lambda) \sim N(0, 0.45^{-1})$, $\varphi \sim N(1, 0.1)$, $\log\left(\frac{1+\rho}{1-\rho}\right) \sim N(0, 0.15^{-1})$.

The general spatiotemporal model is therefore composed of equations (4), (6) and (7) and can be represented by the following mixture form:

$$(1 - \pi_{it}) I_{(Y_{it}=0)} + \pi_{it} p(y_{it} | \mu_{it}) = (1 - \pi_{it}) (1 - Z_{it}) + \pi_{it} p(y_{it} | \mu_{it}), \quad (8)$$

where $I_{(Y_{it}=0)}$ is the indicator function of no occurrence of dengue cases; $p(y_{it} | \mu_{it})$ denotes the Poisson or zero-truncated Poisson probability function (for a zero-inflated or hurdle formulation, respectively); and μ_{it} and π_{it} follow the predictive structures specified in equations (4), (6) and (7). The following cases of the general mixing structure in (8) were considered for comparison:

- *Poisson model*: $\pi_{it} = 1$ and $p(y_{it} | \mu_{it})$ is the usual non-truncated Poisson probability function, with predictive structure (4);
- *Hurdle Poisson model*: $p(y_{it} | \mu_{it})$ is the zero-truncated Poisson probability function, with predictive structure (4), and $\pi_{it} = \pi$, which is equivalent to $\beta_m^Z = \varphi = 0$, $m = 1, 2, \dots, M$. This implies that no predictive structure is assumed for the occurrence of zeros;
- *Zero-inflated Poisson model*: $p(y_{it} | \mu_{it})$ is the usual non-truncated Poisson probability function, with predictive structure (4), and $\pi_{it} = \pi$, which is equivalent to $\beta_m^Z = \varphi = 0$, $m = 1, 2, \dots, M$. This implies that no predictive structure is assumed for the occurrence of zeros;
- *Hurdle Logit-Poisson model*: $p(y_{it} | \mu_{it})$ is the zero-truncated Poisson probability function, with predictive structure (4). Additionally, a predictive structure for the mass point component was assumed, as described in Equation (7).

In addition to statistical criteria for the composition of predictors, the entomological and epidemiological plausibility of potentially relevant regressors for dengue transmission were also considered. The models described in this section were adjusted in the R software [36] with functions from the INLA package [40], which applies nested Laplace approximations to the posterior distribution.

Comparison of models and significance of variables. We used the *Watanabe-Akaike information criterion* (WAIC) and *logarithm of the pseudo marginal likelihood* (LPML) [41] as statistical criteria for comparing the different models, which account for both the quality of the fit and the predictive capacity for unobserved periods and/or areas.

In classical or frequentist approaches, the significance of each regressor variable is typically assessed through a binary significance analysis to determine whether the unitary relative risk or odds ratio lies within a confidence interval. Since the models were fitted using a Bayesian approach, we instead examined the posterior estimated probability that the relative risk (or odds ratio) is greater (or smaller) than 1. If this probability significantly deviates from 0.5, we considered the effect of the corresponding regressor to be relevant.

Prediction of dengue cases over a k weeks horizon. A municipality's dengue risk levels can be evaluated through temporal maps depicting relative risks derived from spatiotemporal models. Bayesian methods enable these models to naturally generate predictive maps of the future spatial distribution of dengue cases. Such predictive distributions are essential since they can accurately represent future observations and align with the actual dynamics of the analyzed process. Let $\mathbf{y}^o = (\mathbf{y}_1, \dots, \mathbf{y}_T)$ denote the vector of observed outcomes during the analysis period, with components $\mathbf{y}_t = (y_{1t}, \dots, y_{Nt})'$, $t = 1, \dots, T$, and assume an interest in generating k week-ahead predictions for dengue occurrences $\mathbf{y}^f = (\mathbf{y}_{T+1}, \dots, \mathbf{y}_{T+k})$, where each component of the future value of observations has dimension $N \times 1$. The goal is to obtain the predictive distribution: $p(\mathbf{y}^f | \mathbf{y}^o) = \int p(\mathbf{y}^f | \boldsymbol{\Theta}) p(\boldsymbol{\Theta} | \mathbf{y}^o) d\boldsymbol{\Theta}$, where $p(\mathbf{y}^f | \boldsymbol{\Theta})$ denotes the observational model for future observations, assumed to be identical to the model used for past observations, and $p(\boldsymbol{\Theta} | \mathbf{y}^o)$ denotes the a posteriori distribution of the parametric vector, based on observations recorded up to time T . The predictive distributions were approximated using sampling methods, first by generating samples from the posterior distribution of the predictors P_{it1} and P_{it2} in (4) and (7), which were obtained using INLA for the Hurdle Poisson, Zero-inflated Poisson and Hurdle

Logit-Poisson models. The points generated from the posterior samples were used to generate null observations or positive counts from structures (4) and (7), allowing us to obtain point predictions (given by the sample mean of the predictive distribution) and predictive intervals with a 95% credibility level, given by the 2.5th and 97.5th percentiles of the predictive distribution sample.

Results

The analysis period spans from the 45th week of 2015 to the 16th week of 2018, during which 31,131 dengue cases were reported. In 2016, 14,907 cases were recorded, with an average monthly incidence rate of 14.65 cases per 10,000 inhabitants, while 5,334 cases were reported in 2017, corresponding to a monthly incidence of 5.19 per 10,000 inhabitants. Between September and December 2015, 622 cases were reported (monthly incidence of 1.85 per 10,000 inhabitants), and between January and April 2018, 10,268 cases were recorded (monthly incidence of 32.04 per 10,000 inhabitants). For the years analyzed, namely 2016 and 2017, the annual dengue incidence rates were 175.8 and 62.6 cases per 10,000 inhabitants, respectively. In April 2018, the cumulative annual incidence had already reached 128.2 cases per 10,000 inhabitants. Fig 2 illustrates the time series of dengue incidence rates per epidemiological week for each neighborhood in Natal, stratified by region. A peak in dengue cases was observed in March 2016 across all regions. The neighborhoods in the northern and western regions of the municipality had the highest number of cases, while those in the southern region did not exceed 50 weekly dengue cases throughout the study period.

Fig 3 presents the density of the *Aedes* egg index across neighborhoods, from the 45th epidemiological week of 2015 to the 16th week of 2018, with weekly boxplots grouped by administrative region. The solid lines in each panel represent

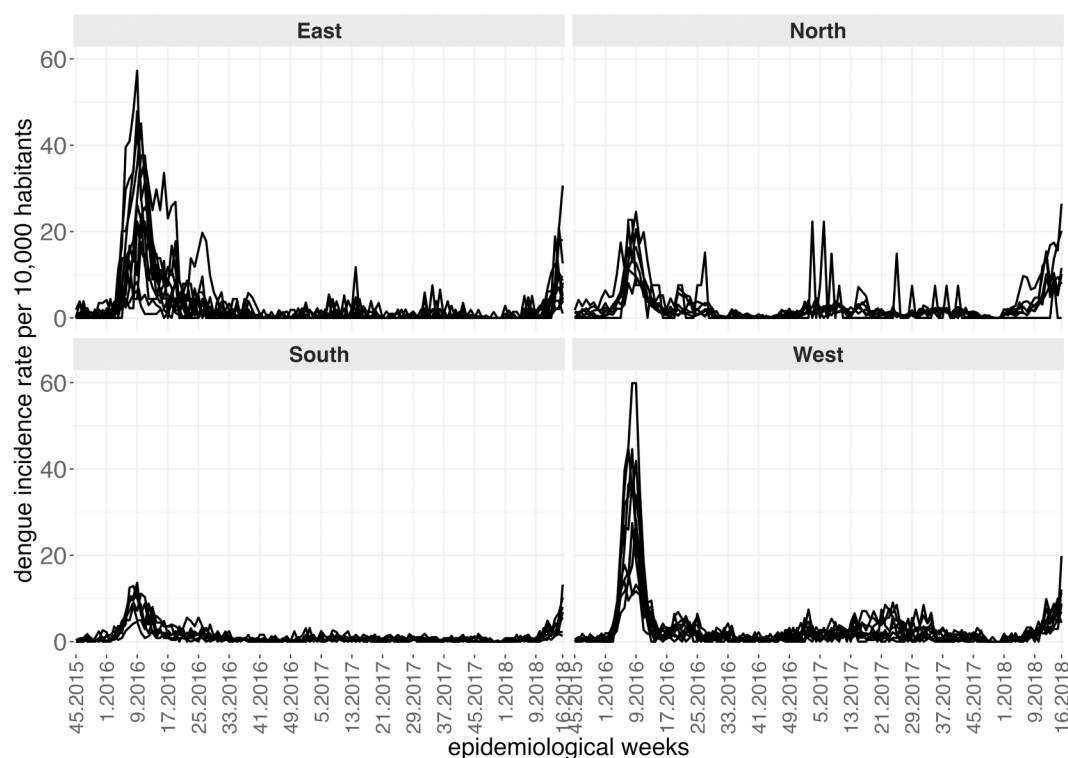


Fig 2. Time series of dengue cases, between the 45th epidemiological week of 2015 and the 16th week of 2018, for each neighborhood, according to administrative region, city of Natal, RN.

<https://doi.org/10.1371/journal.pntd.0012984.g002>

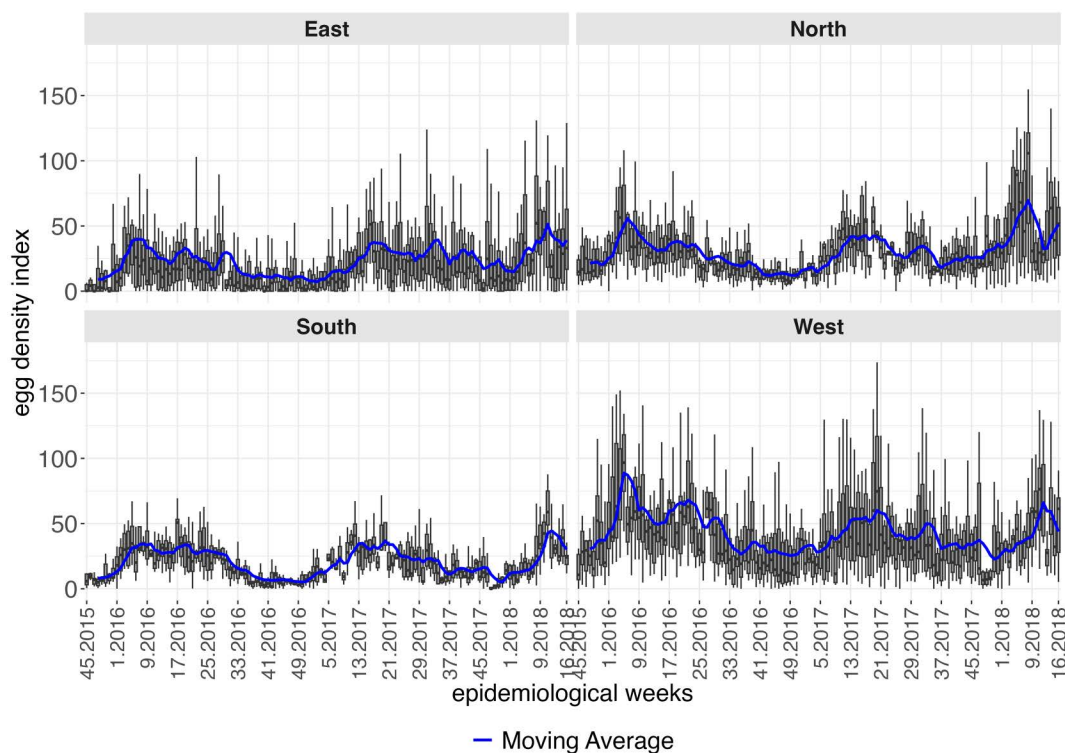


Fig 3. Boxplots of the egg density index across neighborhoods between the 45th epidemiological week of 2015 and 16th week of 2018, according to administrative region, city of Natal, RN. The solid lines represent the moving average of the egg density index in a temporal window of four epidemiological weeks.

<https://doi.org/10.1371/journal.pntd.0012984.g003>

the moving average of the indicator, calculated in a four-week temporal window. The southern region exhibited the lowest and most uniform levels of this indicator, aligned with the lowest frequency of dengue cases shown in Fig 2. Peaks in egg density were generally synchronous across all four regions and preceded the peaks in the epidemiological curves.

For the *Aedes* egg positivity index, Fig 4 presents weekly boxplots indicating frequency of ovitrap saturation in the western, and especially the eastern regions. Periods were common where over 50% of ovitraps were positive in nearly all neighborhoods in the eastern, northern and western regions.

Fig 5 illustrates the evolution of weekly boxplots of average daytime temperatures across neighborhoods throughout the analysis period. A homogeneous pattern can be observed in the territory, characterized by consistently high temperatures, with peaks and declines occurring simultaneously in all four regions.

As shown in Fig 6, the northern and western regions, which recorded the highest number of cases during the analysis period (see Fig 2), also had the highest values of the poor population density indicator per occupied area. This indicator is defined as the number of households with a per capita monthly income of up to the monthly minimum wage per square kilometer of built area in a given spatial unit of analysis.

Polynomial lag models were applied with a maximum lag of $q = 12$ weeks, using polynomials with degrees $d = 2$, $d = 3$, with the best results being obtained for $d = 3$. These preliminary models helped guide the selection of lags to be applied to entomological and climatic indicators for explaining the number of dengue cases. The Poisson, Hurdle Poisson, Zero-inflated Poisson and Hurdle Logit-Poisson models were fitted with different predictive structures specified for the binary occurrence and count of cases. Sociosanitary indicators, selected using the LASSO procedure, were included as predictors in the CAR Bernoulli-Poisson models, and their significance was assessed within the predictive frameworks

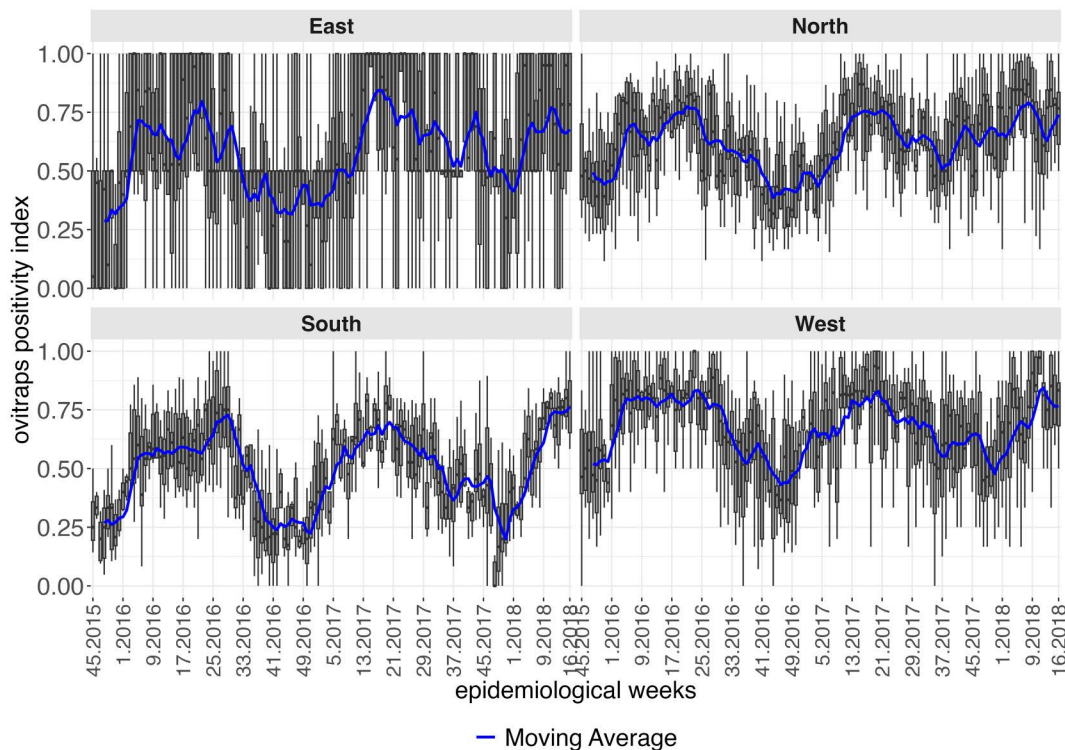


Fig 4. Boxplots of the egg positivity index across neighborhoods between the 45th epidemiological week of 2015 and 16th week of 2018, according to administrative region, city of Natal, RN. The solid lines represent the moving average of the egg positivity index in a temporal window of four epidemiological weeks.

<https://doi.org/10.1371/journal.pntd.0012984.g004>

jointly established with other entomological and climatic indicators. The optimal composition for each predictor was selected from several predictive structures tested across the four models (Poisson, Hurdle Poisson, Zero-inflated Poisson and Hurdle Logit-Poisson). Table 1 presents the values obtained from the model selection criteria for the best fit of each model category. The adoption of the predictive structure for the Bernoulli portion of the model showed advantages over both the Hurdle Poisson and Zero-inflated Poisson models. The best results were clearly achieved with the Hurdle Logit-Poisson model, which combines a Bernoulli/Zero-truncated Poisson mixture and a predictive structure for both mixture components.

The Hurdle Logit-Poisson model, presented in Table 1, includes indicators of poor population density by occupied area (*vul14*) and the mean of the *Aedes* egg density indices (EDI) over the past three, four and five weeks. The regressors that make up the binary part of the model are dengue cases lagged by one week, egg positivity index (EPI) lagged by four weeks and average daytime temperatures lagged by six, seven and eight weeks. To obtain a standardized unit of variation for the indicators, we used increments of standard deviation to assess relative risks and odds ratios, as shown in Table 2.

The results presented in Table 2 indicate that for each increment of one standard deviation in the density of poor people per occupied area (1,013 poor per km² of occupied area), the risk of dengue increased by 20%. For each standard deviation increment in the egg density index (averaging 26 *Aedes* eggs per ovitrap per neighborhood per week), the estimated risk of dengue rose by 3%. The estimated posterior probability of a significant positive impact was 90.9%. Hence, with high probability, increases in EDI levels were associated with an increase in the number of expected cases. Regarding the occurrence of dengue cases (binary part of the model), it was estimated that for each standard deviation increase (14 cases of dengue) in the number of cases from the previous week, the odds of having cases in the current week increased

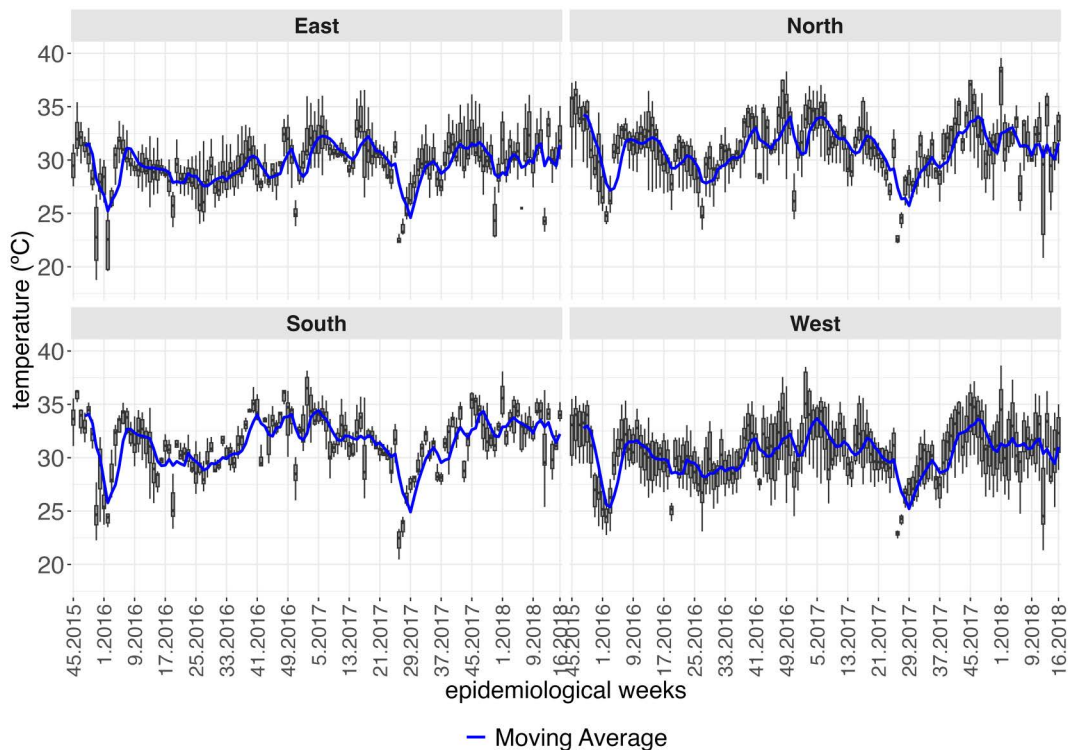


Fig 5. Boxplots of daytime temperature across neighborhoods between the 45th epidemiological week of 2015 and 16th week of 2018, according to administrative region, city of Natal, RN. The solid lines represent the moving average of daytime temperatures in a temporal window of four epidemiological weeks.

<https://doi.org/10.1371/journal.pntd.0012984.g005>

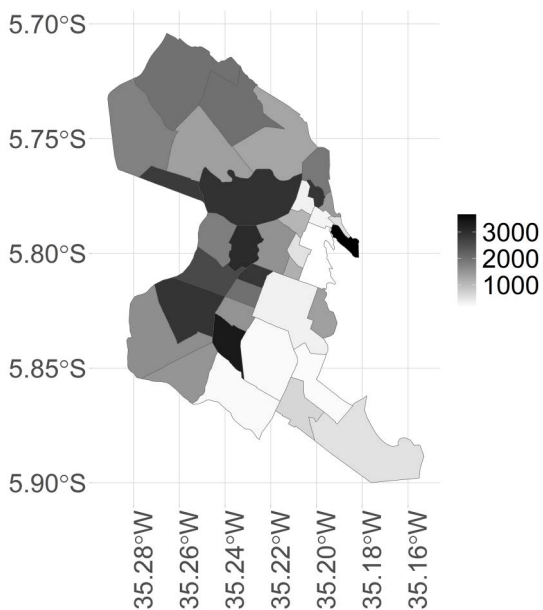


Fig 6. Stratification map by neighborhood of the poor population density indicator, by occupied area, city of Natal, RN. Source: Figure created by the authors. Shapefile of the neighborhoods in the city of Natal: <https://www.natal.rn.gov.br/semurb/geoinformacoes>.

<https://doi.org/10.1371/journal.pntd.0012984.g006>

Table 1. WAIC and LPML for models' comparison.

MODEL	WAIC	LPML
Poisson	15519.86	-7886.10
Hurdle Poisson	18048.40	-9205.85
Zero-Inflated Poisson	15517.78	-7880.08
Hurdle Logit-Poisson	14294.14	-7274.86

<https://doi.org/10.1371/journal.pntd.0012984.t001>

Table 2. Summary measures associated with the posterior distribution of parameters comprising the predictive structure of the selected Hurdle Logit-Poisson model.

Binary portion of the model	I	Odds Ratio ^a		P ^b
		mean	CI 95%	
Intercept		9.76	(6.27; 15.27)	
Number of dengue cases (lag = 1 week)	14	9.39	(5.25; 15.97)	100.0%
EPI (lag = 4 weeks)	0.29	1.20	(1.09; 1.33)	99.9%
Daytime temperature (lags = 6,7,8 weeks)	2.5	1.37	(1.20; 1.57)	100.0%
Poisson portion of the model	I	Relative Risk ^a		P ^b
		mean	CI 95%	
Intercept		0.59	(0.37; 0.94)	
EDI (lags = 3,4,5 weeks)	26	1.03	(0.99; 1.07)	90.9%
Poor population density	1013	1.20	(1.13; 1.28)	100.0%

^aAssociated with increment I.

^bP: Posterior probability of relative risk/odds ratio greater than 1, associated with increment I on the regressor.

<https://doi.org/10.1371/journal.pntd.0012984.t002>

8.39 times. Additionally, for each standard deviation increase in the proportion of positive ovitraps (29% positive ovitraps on average, per neighborhood per week) from four weeks prior, the odds of cases in the current week increased by 20%. For each standard deviation increase in the average daytime temperature (2.5 °C) from six to eight weeks ago, the odds of cases in the current week were expected to rise by 37%.

While a single spatiotemporal model provides results for all neighborhoods of the territory, to illustrate the predictive performance of the Hurdle Logit-Poisson model, we also present the results for four neighborhoods (Mãe Luíza, Pajuçara, Ponta Negra, and Potengi), exhibiting distinct epidemiological processes. To assess the model's predictive quality, we fitted four temporal windows, starting from the first data point available (the 45th epidemiological week of 2015). In each scenario, four-week forecasts were generated for periods: t_2 , the 49th to 52nd epidemiological weeks of 2017; t_2 , the 5th to 8th epidemiological weeks of 2018; t_2 , the 9th to 12th epidemiological weeks of 2018; and t_2 , the 13th to 16th epidemiological weeks of 2018. The forecasts four steps ahead are conditional on the lagged values of the entomological and climatic indicators, except for the number of cases lagged by one week. For this variable, we assumed a constant and identical value corresponding to the average of the last four weeks of observations. Figs 7–10 show the results of the Hurdle Logit-Poisson model for the four selected neighborhoods and forecasts four weeks ahead for horizons t_2 , t_2 , t_2 and t_2 , respectively. The vertical lines in Figs 7–11 mark the moment of interruption of the adjustment and start of the forecasting period.

As expected, the credibility intervals for the predictions widened with the predictive horizon and successfully captured four weeks of future dengue case counts, except during periods of sudden and rapid growth, as demonstrated by the forecast for the Pajuçara neighborhood, shown in Fig 10.

Fig 11 shows the updated revised forecasts for the Pajuçara neighborhood in the t_4 period, using a one-week predictive horizon and accommodating new case data each week to forecast the following week. The model quickly adapted to the

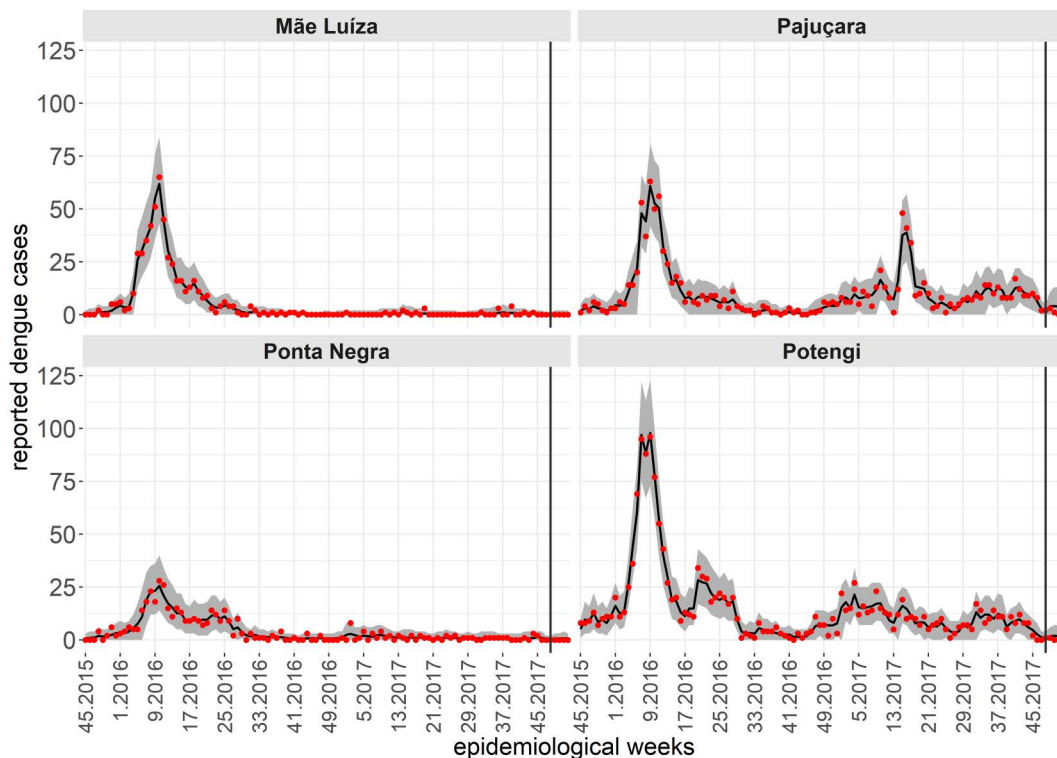


Fig 7. Fit and prediction of the average number of dengue cases for the selected neighborhoods in period t_1 , which starts at the 45th epidemiological week of 2015. Prediction for the epidemiological weeks of 2017, city of Natal, RN. Solid line: Point adjustment/forecast. Gray region: Credibility interval at the 95% credibility level for the adjustment /forecast. Red dots: Reported dengue case counts.

<https://doi.org/10.1371/journal.pntd.0012984.g007>

accelerated growth of dengue cases, providing satisfactory forecasts when updated frequently. Moreover, the predictive intervals adapted to the increasing magnitude of the number of dengue cases. Forecasting should be evaluated not only based on the point estimates, but also by considering the associated uncertainty.

Fig 12 displays maps of the stratification of dengue cases across the entire municipality of Natal based on predictions from the Hurdle Logit-Poisson model in the t_4 predictive window. These maps were generated using a one-week forecast of dengue case counts. Notably, the Hurdle Logit-Poisson model demonstrated an enhanced ability to predict areas with higher case intensity. Throughout all evaluated weeks, the high-risk areas identified by the forecasts were subsequently validated by dengue case counts one week later. Additionally, the model effectively distinguished regions within the city with low or no dengue cases.

Discussion

Dengue is the most prevalent mosquito-borne viral disease worldwide [42]. People from tropical and subtropical countries, including Brazil, can be infected with one or more of the four serotypes [3,43]. Statistical modeling estimates suggest that dengue causes approximately 390 million infections globally each year, including 96 million cases of classical symptomatic dengue and 20,000 deaths [44].

Dengue epidemics in Brazil have persisted over the years, with significant morbidity and mortality burdens on public health [10]. Therefore, it is of utmost importance to propose different strategies for the routine monitoring of this disease, considering the disparities in transmission across the country. Although dengue case counts exhibit heterogeneity in both time and space, it is reasonable to expect neighboring areas and close time points to be similar. To better understand the

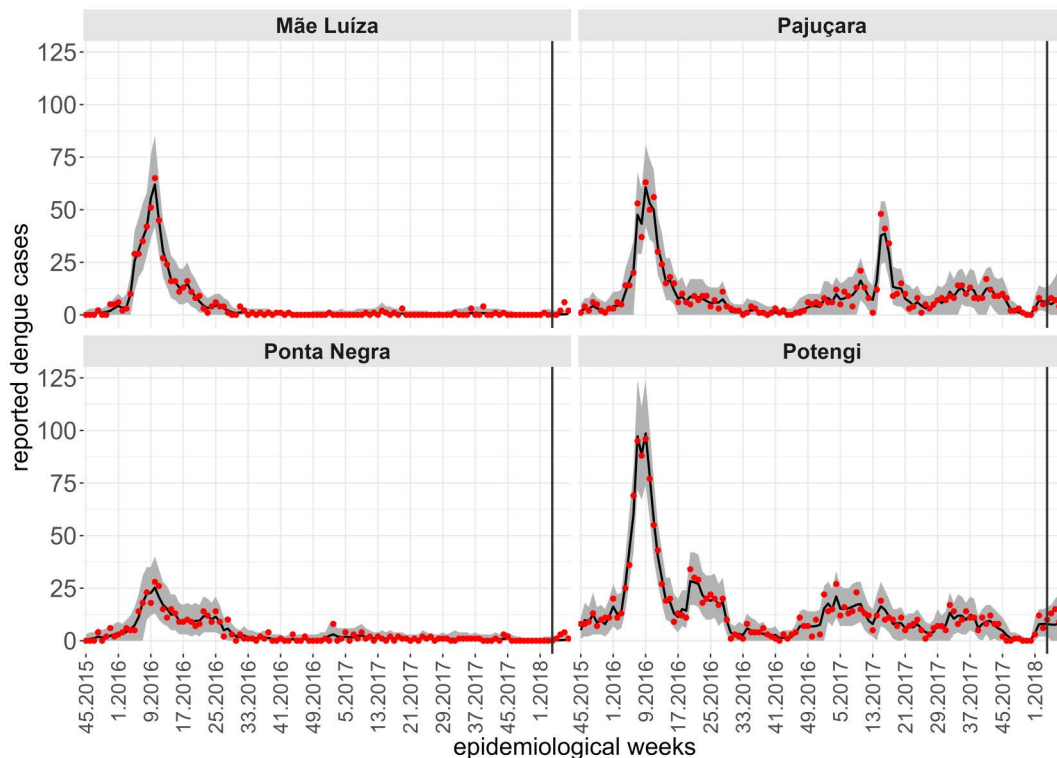


Fig 8. Fit and prediction of the average number of dengue cases for the selected neighborhoods in period t_2 , which starts in the 45th epidemiological week of 2015. Prediction for the 5th to 8th epidemiological weeks of 2018, city of Natal, RN. Solid line: Point adjustment/forecast. Gray region: Credibility interval at the 95% credibility level for the adjustment /forecast. Red dots: Reported dengue case counts.

<https://doi.org/10.1371/journal.pntd.0012984.g008>

patterns in the occurrence of dengue in the municipality of Natal, we identified and quantified the effects of entomological indicators (oviposition indices), climatic factors (daytime and nighttime temperature), and sociosanitary and demographic indicators as potential explanatory factors for the occurrence of this arbovirus.

Our findings revealed that the largest effects of each indicator on the number of reported dengue cases in the municipality of Natal were as follows: i) the egg density index had the greatest impact at a lag of four epidemiological weeks, with notable effects also observed in the third and fifth lagged weeks; ii) likewise, the egg positivity index had its strongest effect at lag four; and iii) the daytime temperature had a significant influence at lags of six, seven and eight epidemiological weeks. We conducted a single fit of a polynomial lag model [32,35,52] capable of integrating the temporal trajectory of these effects along with the number of periods until the peak of the lagged effect and its subsequent decline, until it no longer had an impact. From a temporal perspective, it is crucial to note that the impact of variations in entomological and climatic indicators on the occurrence of dengue cases is not immediate. Some studies have sought to describe the temporal lag of the effects between predictor variables and the response variable by fitting various models based on individual or grouped lags of the predictors, which naturally poses limitations, as discussed by Sanchez-Gendriz et al. [17].

The seasonal dynamics of dengue transmission has been linked to the delayed effects of climate and entomological indices on vector population dynamics. In Brazil, one study showed that mean weekly temperatures above 22–24 °C were strongly associated with high *Ae. aegypti* populations, increasing the risk of dengue transmission. Both temperature and rainfall were significantly correlated with *Ae. aegypti* indices at a short (1 week) time lag [45]. In an epidemiological study of dengue occurrence in the municipality of Cambé, in the northern part of the Brazilian state of Paraná, temperature and the number of patients positive for dengue showed a positive cross-correlation with a lag

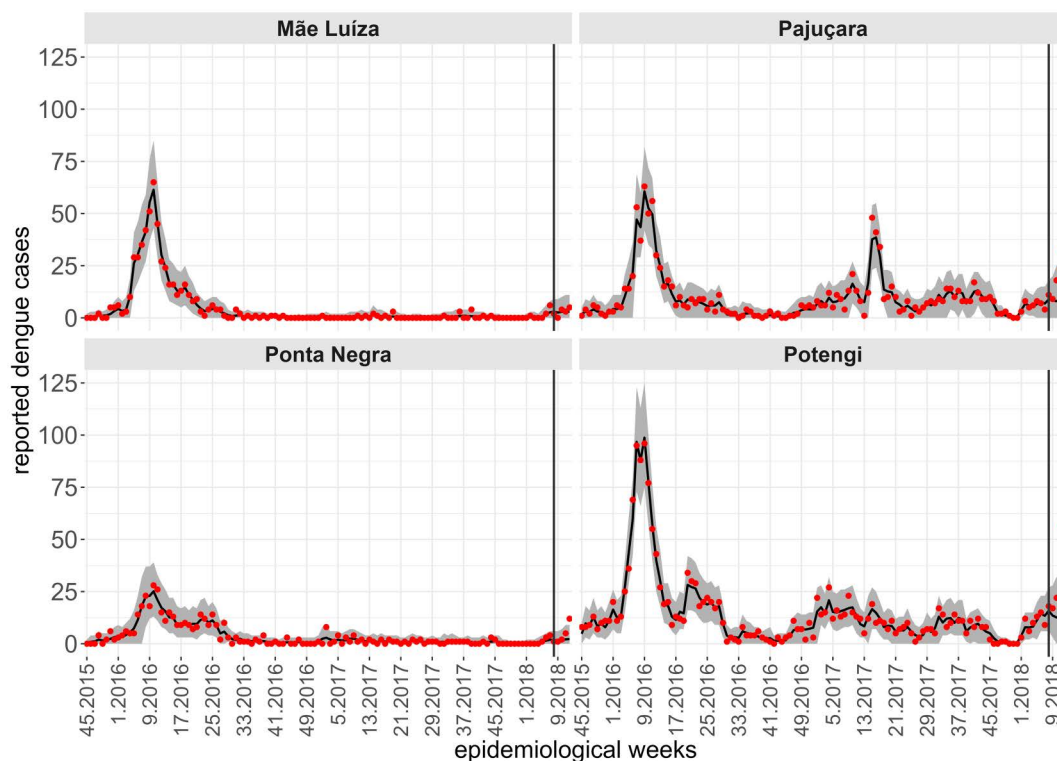


Fig 9. Fit and prediction of the average number of dengue cases for the selected neighborhoods in period t_3 , which starts in the 45th epidemiological week of 2015. Prediction for the 9th to 12th epidemiological weeks of 2018, city of Natal, RN. Solid line: Point adjustment/forecast. Gray region: Credibility interval at the 95% credibility level for the adjustment /forecast. Red dots: Reported dengue case counts.

<https://doi.org/10.1371/journal.pntd.0012984.g009>

interval of two months (eight weeks) [46]. In India, significant effects of temperatures between 26 and 28 °C on the relative risk of dengue incidence were estimated, with peak effects occurring around 5 weeks and persisting for up to 10 weeks [47].

Considering dengue cases in Natal, Brazil, Sanchez-Gendriz et al. [17] employed cross-correlation to estimate the temporal lag in the impact of the ovitrap density index on dengue incidence, finding a lag of 4 weeks. Betanzos et al. [48] observed a significant effect between ovitrap monitoring indicators and dengue incidence in Mexico, with a better fit achieved when considering the lagged effects of egg counts on case incidence. The authors obtained improved predictive results by incorporating egg counts with a lag of three weeks during the dry season and four weeks during the wet season, emphasizing that recognizing these lags enhances the predictive utility of indicators derived from ovitrap monitoring. The adoption of a single distributed lag structure for the entire territory of Natal is a limitation of our work, since these effects may vary over time and across spatial units. Estimating the spatiotemporal evolution of covariate effects propagation, as well as jointly modeling these effects, is part of our research agenda.

In our study, sociosanitary indicators did not exhibit temporal variation, since they were observed only during the census year. The application of a LASSO procedure [37] allowed the selection of a subset of sociosanitary indicators capable of explaining the spatial dynamics of dengue cases. Among the selected indicators were the density of impoverished individuals per occupied area and the percentage of households without access to urban sanitation services. Cabrera et al. [49] discussed how a low gross domestic product (GDP) per capita and high population density are potential indicators of dengue severity. Lowe et al. [50] explored the exposure-lag-response associations between hydrometeorological

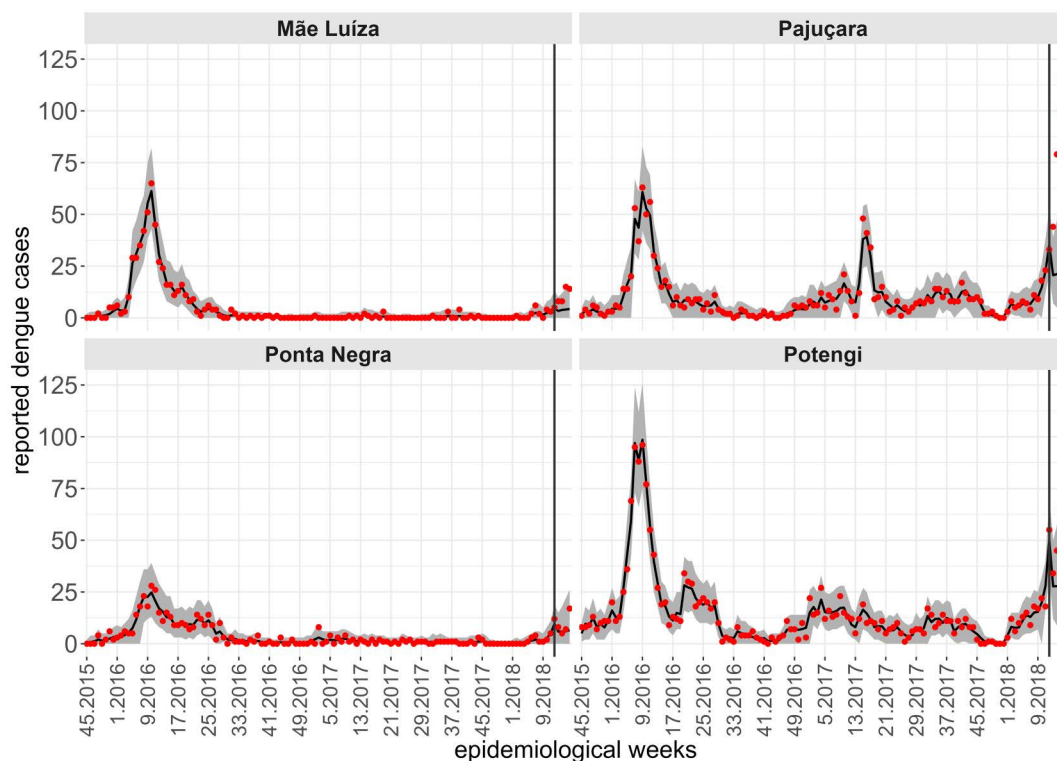


Fig 10. Fit and prediction of the average number of dengue cases for the selected neighborhoods in period t_4 , which starts in the 45th epidemiological week of 2015. Prediction for the 13th to 16th epidemiological weeks of 2018, city of Natal, RN. Solid line: Point adjustment/forecast. Gray region: Credibility interval at the 95% credibility level for the adjustment /forecast. Red dots: Reported dengue case counts.

<https://doi.org/10.1371/journal.pntd.0012984.g010>

extremes and dengue risk in the months leading up to an outbreak using a Bayesian spatiotemporal formulation. Ferreira and Schmidt [51] found that individuals residing in slums in Rio de Janeiro, Brazil, were more susceptible to dengue infection. Additionally, by categorizing patients into mean income intervals in Cambé, Paraná, Brazil, the authors observed that dengue and income were negatively correlated, indicating that poorer areas were more likely to have a higher than average risk of dengue [46]. In another study from the ARBOALVO project, a stratification methodology using the Territorial Receptivity Index (TRI) for urban arboviruses was introduced, which served as a tool for the surveillance and control of dengue, Zika, and chikungunya. The authors analyzed various sociosanitary indicators strongly correlated with areas of greater susceptibility to urban arboviruses, which were directly or indirectly associated with conditions of poverty and lack of sanitation infrastructure [25].

Our results included the prediction of dengue case occurrences for each neighborhood, four weeks ahead, based on a single fit in the municipality of Natal. The best-fit model captured the spatial heterogeneity of the dengue case series as well as the peaks and nonoccurrence periods of the disease. The selected model enabled the generation of forecasts for each neighborhood, accurately predicting case peaks, periods of no occurrences, and projecting cases even under significantly different conditions, as observed in neighborhoods like Mãe Luíza, Pajuçara, Ponta Negra, and Potengi. The credibility intervals successfully captured the future values of reported dengue cases in a four-week horizon, except during moments of sharp and rapid growth. When updated at an appropriate frequency, the model quickly adapted to the accelerated growth in dengue cases, providing satisfactory forecasts. In all evaluated weeks, the model predicted high case counts in neighborhoods with elevated dengue occurrences. Additionally, the model accurately stratified areas in Natal, by identifying regions with low or no dengue occurrences.

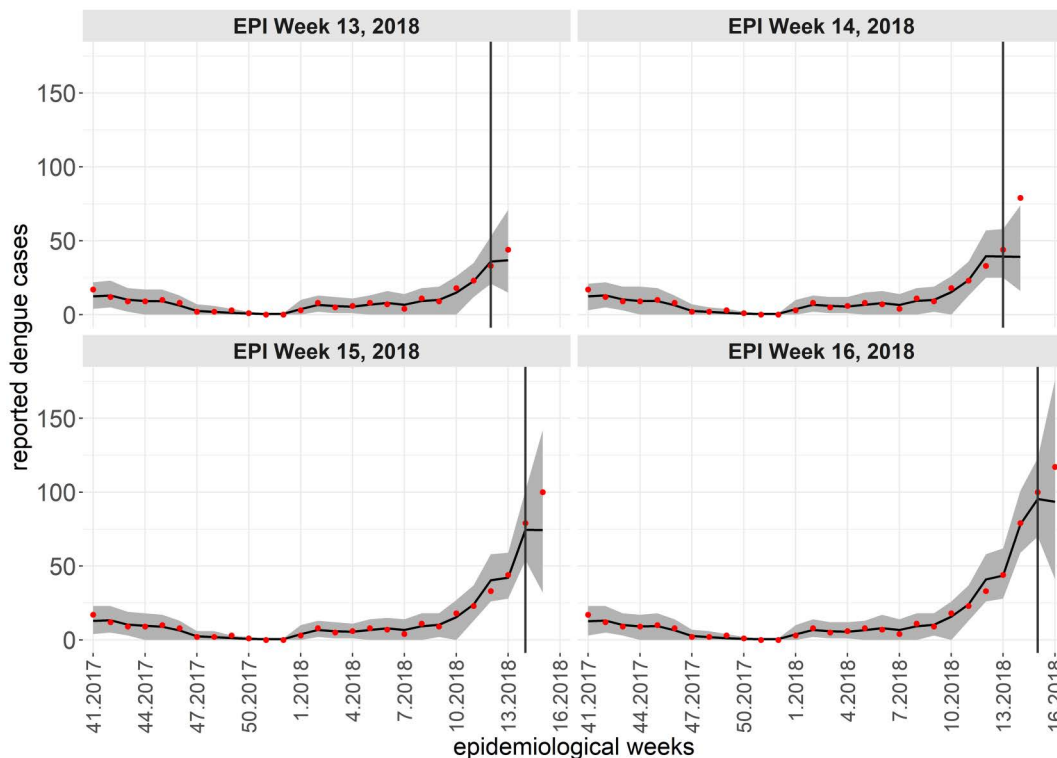


Fig 11. Fit and prediction of the average number of dengue cases for the Pajuçara neighborhood in period t_4 , which starts at the 45th epidemiological week of 2015. Forecasts one week ahead, in the period between the 13th and 16th epidemiological weeks of 2018, city of Natal, RN. Solid line: Point adjustment/forecast. Gray region: Credibility interval at the 95% credibility level for the adjustment /forecast. Red dots: Reported dengue case counts.

<https://doi.org/10.1371/journal.pntd.0012984.g011>

Models with a space-time dependency structure for epidemiological counts commonly assume responses with a Poisson distribution. However, such models are inadequate in contexts where the observation of zero counts is frequent (a problem referred to in the literature as “excess of zeros”). To address this limitation, several studies have made use of hurdle and zero-inflated count models, such as the Hurdle Poisson and Zero-inflated Poisson models adopted in this work [53–56]. Better fits were obtained for periods and neighborhoods in Natal with null counts of dengue cases by incorporating a predictor structure based on entomological and sociosanitary indicators, as well as space-time effects. The structural flexibility of the model enabled it to capture not only the periods and neighborhoods without any occurrences, but also those with high dengue case counts. A limitation of the models adopted in this study is their inability to distinguish between true zeros (actual absence of cases) and zeros due to underreporting. Freitas et al. [57] addressed this issue by employing a Markov-switching model with latent states representing the presence or absence of cases. Furthermore, estimating the transition matrix between states of absence and presence accommodates long sequences of zeros, as well as prolonged sequences of positive counts interspersed with zeros, which may not be effectively handled by conventional zero-inflated structures.

We assumed the simplifying hypothesis of proportionality between spatiotemporal effects for the frequency and occurrence of dengue cases in Natal. More general relationships between these effects can be explored. For instance, in a multivariate spatial context, Schmidt et al. [58] employed distinct specifications for the spatial structures associated with dengue, Zika, and chikungunya in the city of Rio de Janeiro from 2015 to 2016.

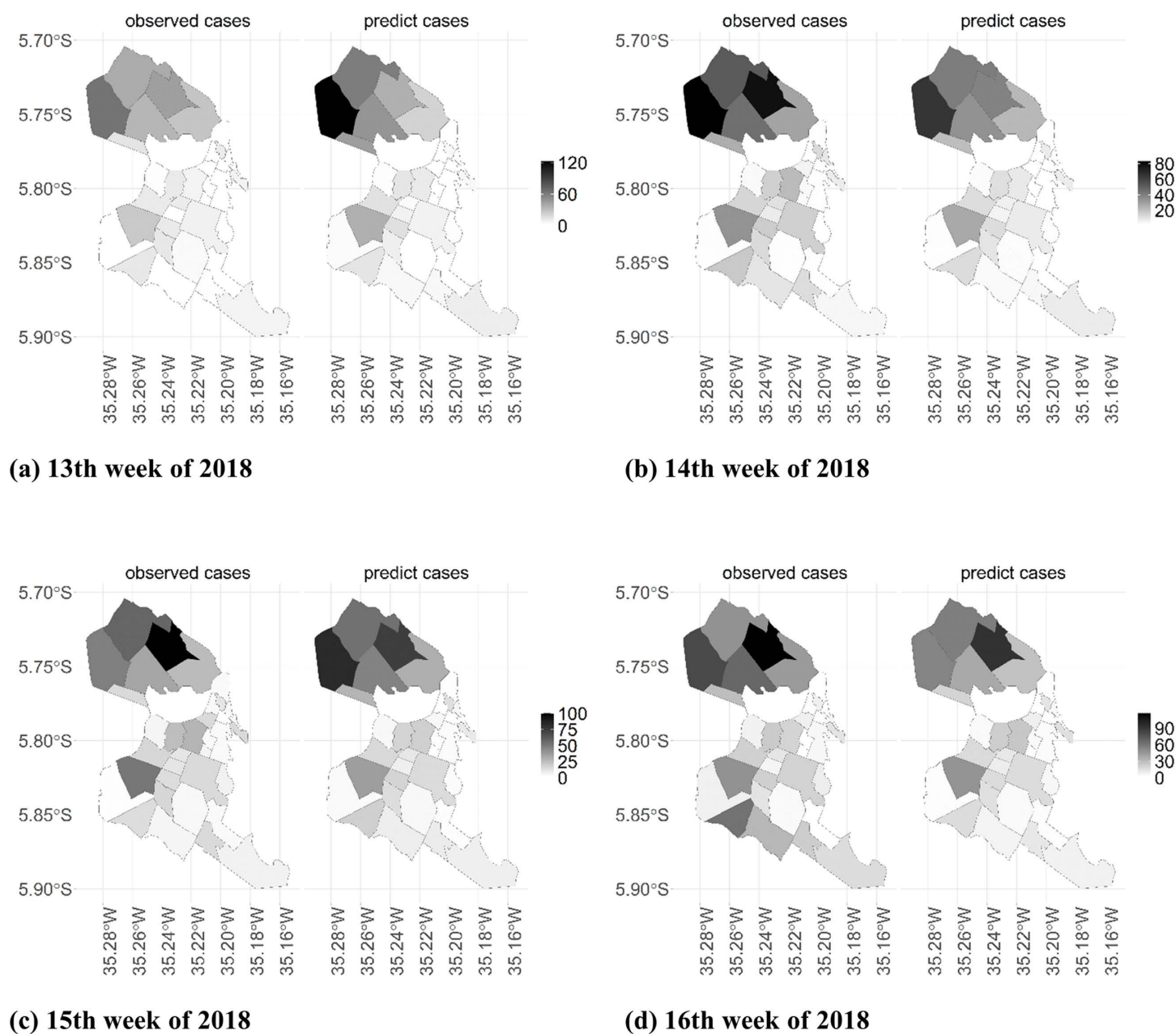


Fig 12. Stratification maps of observed and predicted dengue case counts for four epidemiological weeks ahead (between the 13th and 16th epidemiological weeks of 2018), city of Natal, RN.

<https://doi.org/10.1371/journal.pntd.0012984.g012>

From a predictive standpoint, it is noteworthy that our forecasts four steps ahead are conditioned on the lagged values of the entomological and climatic indicators. Therefore, the model relies on the previously observed values of the explanatory variables without relying on the establishment of hypothetical scenarios, except for the number of cases lagged by one week.

Finally, we emphasize that obtaining predictive maps of dengue cases serves as a tool for public health management services and epidemiological surveillance, enabling them to anticipate risk stratification within a city or metropolitan region, considering its heterogeneity. In this study, we used predictive maps to identify regions that subsequently exhibited high,

moderate, low or zero dengue rates. Additionally, quantifying the uncertainty associated with all estimates and forecasts through probability intervals with intuitive and practical interpretations allows for a comprehensive evaluation of the precision of the estimates obtained. The ability to temporally describe forecasts that account for the individual conditions of each neighborhood stands out as a valuable tool for public health management, facilitating more closely targeted and effective interventions.

Supporting information

S1 Table. Population distribution by neighborhood and region, 2015 – 2018. * The population by neighborhood is a projection based on the average annual growth rate calculated from the 2000 and 2010 census data.
(DOCX)

S2 Table. Epidemiological, entomological, climatic and sociosanitary indicators constructed at the neighborhood level in Natal. * Variables selected using the LASSO procedure.
(DOCX)

Acknowledgements

We thank the staff of the Natal Center for Zoonosis Control - in memory of Alexandre de Medeiros Tavares; the Dengue-target Network, and the ARBOALVO project team for their significant contributions to the discussion regarding territorial stratification of arboviruses.

Author contributions

Conceptualization: Mariane Branco Alves, Rafael Santos Erbisti, Aline Araújo Nobre, Taynãna César Simões, Nildimar Alves Honório.

Data curation: Mariane Branco Alves, Rafael Santos Erbisti, Aline Araújo Nobre, Taynãna César Simões, Alexandre de Medeiros Tavares (in memoriam), Márcia Cristina Melo, Rodrigo Moreira Pedreira, Jan Pierre Martins de Araújo.

Formal analysis: Mariane Branco Alves, Rafael Santos Erbisti, Aline Araújo Nobre, Taynãna César Simões, Nildimar Alves Honório.

Funding acquisition: Nildimar Alves Honório.

Investigation: Mariane Branco Alves, Rafael Santos Erbisti, Aline Araújo Nobre, Taynãna César Simões, Nildimar Alves Honório.

Methodology: Mariane Branco Alves, Rafael Santos Erbisti, Aline Araújo Nobre, Taynãna César Simões, Nildimar Alves Honório.

Software: Mariane Branco Alves, Rafael Santos Erbisti, Aline Araújo Nobre, Taynãna César Simões.

Writing – original draft: Mariane Branco Alves, Rafael Santos Erbisti, Aline Araújo Nobre, Taynãna César Simões, Nildimar Alves Honório.

Writing – review & editing: Mariane Branco Alves, Rafael Santos Erbisti, Aline Araújo Nobre, Taynãna César Simões, Marília Sá Carvalho, Nildimar Alves Honório.

References

1. Weaver SC. Arrival of chikungunya virus in the new world: prospects for spread and impact on public health. *PLoS Negl Trop Dis*. 2014;8(6):e2921. <https://doi.org/10.1371/journal.pntd.0002921> PMID: [24967777](https://pubmed.ncbi.nlm.nih.gov/24967777/)
2. Honório NA, Câmara DCP, Calvet GA, Brasil P. Chikungunya: an arbovirus infection in the process of establishment and expansion in Brazil. *Cad Saude Publica*. 2015;31(5):906–8. <https://doi.org/10.1590/0102-311XPE020515> PMID: [26083166](https://pubmed.ncbi.nlm.nih.gov/26083166/)

3. Lima-Camara TN. Emerging arboviruses and public health challenges in Brazil. *Rev Saude Publica*. 2016;50:36. <https://doi.org/10.1590/S1518-8787.2016050006791> PMID: [27355468](#)
4. Lowe R, Barcellos C, Brasil P, Cruz OG, Honório NA, Kuper H, et al. The Zika Virus Epidemic in Brazil: From Discovery to Future Implications. *Int J Environ Res Public Health*. 2018;15(1):96. <https://doi.org/10.3390/ijerph15010096> PMID: [29315224](#)
5. Consoli R, Lourenço-de-Oliveira R. Principais mosquitos de importância sanitária no Brasil Rio de Janeiro. Fiocruz; 1994. p. 225.
6. Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. *Elife*. 2015;4:e08347. <https://doi.org/10.7554/eLife.08347> PMID: [26126267](#)
7. Schatzmayr HG, Nogueira RM, Travassos da Rosa AP. An outbreak of dengue virus at Rio de Janeiro--1986. *Mem Inst Oswaldo Cruz*. 1986;81(2):245–6. <https://doi.org/10.1590/s0074-02761986000200019> PMID: [3587006](#)
8. Nogueira RMR, de Araújo JMG, Schatzmayr HG. Dengue viruses in Brazil, 1986–2006. *Rev Panam Salud Publica*. 2007;22(5):358–63. <https://doi.org/10.1590/s1020-49892007001000009> PMID: [18198045](#)
9. Nogueira RMR, Eppinghaus ALF. Dengue virus type 4 arrives in the state of Rio de Janeiro: a challenge for epidemiological surveillance and control. *Mem Inst Oswaldo Cruz*. 2011;106(3):255–6. <https://doi.org/10.1590/s0074-02762011000300001> PMID: [21655810](#)
10. Codeco CT, Oliveira SS, Ferreira DAC, Riback TIS, Bastos LS, Lana RM, et al. Fast expansion of dengue in Brazil. *Lancet Reg Health Am*. 2022;12:100274. <https://doi.org/10.1016/j.lana.2022.100274> PMID: [36776428](#)
11. Zanluca C, Melo VCA de, Mosimann ALP, Santos GIVD, Santos CNDD, Luz K. First report of autochthonous transmission of Zika virus in Brazil. *Mem Inst Oswaldo Cruz*. 2015;110(4):569–72. <https://doi.org/10.1590/0074-02760150192> PMID: [26061233](#)
12. Barbosa IR, Tavares A de M, Torres ÚP da S, Nascimento CA do, Moura MCB de M, Vieira VB, et al. Identification of surveillance and control priority areas for dengue and other arboviruses transmitted by *Aedes aegypti* in Natal-RN, Brazil: experience report. *Epidemiol Serv Saude*. 2017;26(3):629–38. <https://doi.org/10.5123/S1679-49742017000300020> PMID: [28977187](#)
13. Barbosa IR, Araújo L de F, Carlota FC, Araújo RS, Maciel IJ. Epidemiologia do dengue no Estado do Rio Grande do Norte, Brasil, 2000 a 2009. *Epidemiol Serv Saude*. 2012;21(1):149–57. <https://doi.org/10.5123/s1679-49742012000100015>
14. Silva J, Machado CJ. Effects of climate variables on dengue risk in Natal-RN by generalized models. *International Journal of Infectious Diseases*. 2018;73:183. <https://doi.org/10.1016/j.ijid.2018.04.3827>
15. de Aguiar DF, de Barros ENC, Ribeiro GS, Brasil P, Mourao MPG, Luz K, et al. A prospective, multicentre, cohort study to assess the incidence of dengue illness in households from selected communities in Brazil (2014–2018). *Int J Infect Dis*. 2021;108:443–53. <https://doi.org/10.1016/j.ijid.2021.04.062> PMID: [33894353](#)
16. Dos Santos LA, Dos Santos AFA, de Assis AG, da Costa Júnior JF, de Souza RP. Model to support intervention prioritization for the control of *Aedes aegypti* in Brazil: a multi-criteria approach. *BMC Public Health*. 2022;22(1):932. <https://doi.org/10.1186/s12889-022-13006-1> PMID: [35538565](#)
17. Sanchez-Gendríz I, de Souza GF, de Andrade IGM, Neto ADD, de Medeiros Tavares A, Barros DMS, et al. Data-driven computational intelligence applied to dengue outbreak forecasting: a case study at the scale of the city of Natal, RN-Brazil. *Sci Rep*. 2022;12(1):6550. <https://doi.org/10.1038/s41598-022-10512-5> PMID: [35449179](#)
18. Dalvi APR, Gibson G, Ramos AN Jr, Bloch KV, Sousa GDS de, Silva TLN da, et al. Sociodemographic and environmental factors associated with dengue, Zika, and chikungunya among adolescents from two Brazilian capitals. *PLoS Negl Trop Dis*. 2023;17(3):e0011197. <https://doi.org/10.1371/journal.pntd.0011197> PMID: [36928657](#)
19. Gardini Sanches Palasio R, Marques Moralejo Bermudi P, Luiz de Lima Macedo F, Reis Santana LM, Chiaravalloti-Neto F. Zika, chikungunya and co-occurrence in Brazil: space-time clusters and associated environmental-socioeconomic factors. *Sci Rep*. 2023;13(1):18026. <https://doi.org/10.1038/s41598-023-42930-4> PMID: [37865641](#)
20. Vanlerberghe V, Gómez-Dantés H, Vazquez-Prokopec G, Alexander N, Manrique-Saide P, Coelho G, et al. Changing paradigms in *Aedes* control: considering the spatial heterogeneity of dengue transmission. *Rev Panam Salud Publica*. 2017;41:e16. <https://doi.org/10.26633/RPSP.2017.16> PMID: [31391815](#)
21. PAHO – Pan American Health Organization. Technical document for the implementation of interventions based on generic operational scenarios for *Aedes aegypti* control. Washington, District of Columbia; 2019.
22. Carvalho MS, Honório NA, Garcia LMT, Carvalho LC de S. *Aedes aegypti* control in urban areas: A systemic approach to a complex dynamic. *PLoS Negl Trop Dis*. 2017;11(7):e0005632. <https://doi.org/10.1371/journal.pntd.0005632> PMID: [28749942](#)
23. Santos JPCD, Albuquerque HG, Siqueira ASP, Praça HLF, Pereira LV, Tavares A de M, et al. ARBOALVO: territorial stratification for definition of areas for prompt response by surveillance and timely control of urban arbovirus infections. *Cad Saude Publica*. 2022;38(3):e00110121. <https://doi.org/10.1590/0102-311X00110121> PMID: [35384994](#)
24. Freitas LP, Schmidt AM, Cossich W, Cruz OG, Carvalho MS. Spatio-temporal modelling of the first Chikungunya epidemic in an intra-urban setting: The role of socioeconomic status, environment and temperature. *PLoS Negl Trop Dis*. 2021;15(6):e0009537. <https://doi.org/10.1371/journal.pntd.0009537> PMID: [34143771](#)
25. Siqueira ASP, Praça HLF, Santos JPCD, Albuquerque HG, Pereira LV, Simões TC, et al. ArboAlvo: stratification method for territorial receptivity to urban arboviruses. *Rev Saude Publica*. 2022;56:39. <https://doi.org/10.11606/s1518-8787.2022056003546> PMID: [35649086](#)

26. Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math*. 1991;43(1):1–20. <https://doi.org/10.1007/bf00116466>
27. IBGE – Instituto Brasileiro de Geografia e Estatística. IBGE-Cidades; 2021. Available from: <https://cidades.ibge.gov.br/brasil/rn/natal/panorama>
28. Honório NA, Silva W da C, Leite PJ, Gonçalves JM, Lounibos LP, Lourenço-de-Oliveira R. Dispersal of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) in an urban endemic dengue area in the State of Rio de Janeiro, Brazil. *Mem Inst Oswaldo Cruz*. 2003;98(2):191–8. <https://doi.org/10.1590/s0074-02762003000200005> PMID: 12764433
29. IBGE – Instituto Brasileiro de Geografia e Estatística. Censo Brasileiro de 2010. Rio de Janeiro: IBGE; 2010.
30. Gao C, Wei P, Xie Z, Zhang P, Tahir MM, Toktonazarovich TK, et al. Genomic identification of the NF-Y gene family in apple and functional analysis of MdNF-YB18 involved in flowering transition. *Mol Breed*. 2024;45(1):4. <https://doi.org/10.1007/s11032-024-01524-2> PMID: 39726978
31. Almon S. The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*. 1965;33(1):178. <https://doi.org/10.2307/1911894>
32. Schwartz J. The distributed lag between air pollution and daily deaths. *Epidemiology*. 2000;11(3):320–6. <https://doi.org/10.1097/00001648-200005000-00016> PMID: 10784251
33. Zanobetti A, Wand MP, Schwartz J, Ryan LM. Generalized additive distributed lag models: quantifying mortality displacement. *Biostatistics*. 2000;1(3):279–92. <https://doi.org/10.1093/biostatistics/1.3.279> PMID: 12933509
34. Xiang J, Hansen A, Liu Q, Liu X, Tong MX, Sun Y, et al. Association between dengue fever incidence and meteorological factors in Guangzhou, China, 2005–2014. *Environ Res*. 2017;153:17–26. <https://doi.org/10.1016/j.envres.2016.11.009> PMID: 27883970
35. Gasparrini A. Distributed Lag Linear and Non-Linear Models in R: The Package dlnm. *J Stat Softw*. 2011;43(8):1–20. PMID: 22003319
36. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022.
37. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 1996;58(1):267–88. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
38. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Soft*. 2010;33(1). <https://doi.org/10.18637/jss.v033.i01>
39. da Costa ACC, Codeço CT, Krainski ET, Gomes MF da C, Nobre AA. Spatiotemporal diffusion of influenza A (H1N1): Starting point and risk factors. *PLoS One*. 2018;13(9):e0202832. <https://doi.org/10.1371/journal.pone.0202832> PMID: 30180215
40. Rue H, Martino S, Chopin N. Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2009;71(2):319–92. <https://doi.org/10.1111/j.1467-9868.2008.00700.x>
41. Congdon P. Bayesian models for categorical data. 1st ed. John Wiley and Sons; 2005.
42. WHO – World Health Organization. Disease Outbreak News; Dengue – Global situation; 2023. Available from: <https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498>
43. Kyle JL, Harris E. Global spread and persistence of dengue. *Annu Rev Microbiol*. 2008;62:71–92. <https://doi.org/10.1146/annurev.micro.62.081307.163005> PMID: 18429680
44. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013;496(7446):504–7. <https://doi.org/10.1038/nature12060> PMID: 23563266
45. Honório NA, Castro MG, Barros FSM de, Magalhães M de AFM, Sabroza PC. The spatial distribution of *Aedes aegypti* and *Aedes albopictus* in a transition zone, Rio de Janeiro, Brazil. *Cad Saude Publica*. 2009;25(6):1203–14. <https://doi.org/10.1590/s0102-311x2009000600003> PMID: 19503951
46. Bavia L, Melanda FN, de Arruda TB, Mosimann ALP, Silveira GF, Aoki MN, et al. Epidemiological study on dengue in southern Brazil under the perspective of climate and poverty. *Sci Rep*. 2020;10(1):2127. <https://doi.org/10.1038/s41598-020-58542-1> PMID: 32034173
47. Kakarla SG, Caminade C, Mutheneni SR, Morse AP, Upadhyayula SM, Kadir MR, et al. Lag effect of climatic variables on dengue burden in India. *Epidemiol Infect*. 2019;147:e170. <https://doi.org/10.1017/S0950268819000608> PMID: 31063099
48. Betanzos-Reyes ÁF, Rodríguez MH, Romero-Martínez M, Sesma-Medrano E, Rangel-Flores H, Santos-Luna R. Association of dengue fever with *Aedes spp.* abundance and climatological effects. *Salud Publica Mex*. 2018;60(1):12–20. <https://doi.org/10.21149/8141> PMID: 29689652
49. Cabrera M, Leake J, Naranjo-Torres J, Valero N, Cabrera JC, Rodríguez-Morales AJ. Dengue Prediction in Latin America Using Machine Learning and the One Health Perspective: A Literature Review. *Trop Med Infect Dis*. 2022;7(10):322. <https://doi.org/10.3390/tropicalmed7100322> PMID: 36288063
50. Lowe R, Lee SA, O'Reilly KM, Brady OJ, Bastos L, Carrasco-Escobar G, et al. Combined effects of hydrometeorological hazards and urbanisation on dengue risk in Brazil: a spatiotemporal modelling study. *Lancet Planet Health*. 2021;5(4):e209–19. [https://doi.org/10.1016/S2542-5196\(20\)30292-8](https://doi.org/10.1016/S2542-5196(20)30292-8) PMID: 33838736
51. Ferreira GS, Schmidt AM. Spatial modelling of the relative risk of dengue fever in Rio de Janeiro for the epidemic period between 2001 and 2002. *Braz J Probab Stat*. 2006;20:29–47. Available from: <http://www.jstor.org/stable/43601072>
52. Alves MB, Gamerman D, Ferreira MA. Transfer functions in dynamic generalized linear models. *Statistical Modelling*. 2010;10(1):03–40. <https://doi.org/10.1177/1471082x0801000102>

53. Lambert D. Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*. 1992;34(1):1. <https://doi.org/10.2307/1269547>
54. Hall DB. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics*. 2000;56(4):1030–9. <https://doi.org/10.1111/j.0006-341x.2000.01030.x> PMID: [11129458](https://pubmed.ncbi.nlm.nih.gov/11129458/)
55. Agarwal DK, Gelfand AE, Citron-Pousty S. *Environmental and Ecological Statistics*. 2002;9(4):341–55. <https://doi.org/10.1023/a:1020910605990>
56. Fernandes MV, Schmidt AM, Migon HS. Modelling zero-inflated spatio-temporal processes. *Statistical Modelling*. 2009;9(1):3–25. <https://doi.org/10.1177/1471082x0800900102>
57. Picinini Freitas L, Douwes-Schultz D, Schmidt AM, Ávila Monsalve B, Salazar Flórez JE, García-Balaguera C, et al. Zika emergence, persistence, and transmission rate in Colombia: a nationwide application of a space-time Markov switching model. *Sci Rep*. 2024;14(1):10003. <https://doi.org/10.1038/s41598-024-59976-7> PMID: [38693192](https://pubmed.ncbi.nlm.nih.gov/38693192/)
58. Schmidt AM, Freitas LP, Cruz OG, Carvalho MS. A Poisson-multinomial spatial model for simultaneous outbreaks with application to arboviral diseases. *Stat Methods Med Res*. 2022;31(8):1590–602. <https://doi.org/10.1177/09622802221102628> PMID: [35658776](https://pubmed.ncbi.nlm.nih.gov/35658776/)