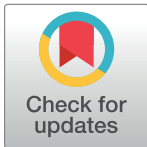


## RESEARCH ARTICLE

## Forecasting dengue fever in Brazil: An assessment of climate conditions

Lucas M. Stolerman<sup>1,2\*</sup>, Pedro D. Maia<sup>3</sup>, J. Nathan Kutz<sup>3</sup>

**1** Programa de Computação Científica (PROCC)/Fiocruz, Rio de Janeiro, RJ, Brazil, **2** IMPA, Rio de Janeiro, RJ, Brazil, **3** Department of Applied Mathematics, University of Washington, Seattle, Washington, United States of America

\* [lstolerman@eng.ucsd.edu](mailto:lstolerman@eng.ucsd.edu)

## Abstract

Local climate conditions play a major role in the biology of the *Aedes aegypti* mosquito, the main vector responsible for transmitting dengue, zika, chikungunya and yellow fever in urban centers. For this reason, a detailed assessment of periods in which changes in climate conditions affect the number of human cases may improve the timing of vector-control efforts. In this work, we develop new machine-learning algorithms to analyze climate time series and their connection to the occurrence of dengue epidemic years for seven Brazilian state capitals. Our method explores the impact of two key variables—frequency of precipitation and average temperature—during a wide range of time windows in the annual cycle. Our results indicate that each Brazilian state capital considered has its own climate signatures that correlate with the overall number of human dengue-cases. However, for most of the studied cities, the winter preceding an epidemic year shows a strong predictive power. Understanding such climate contributions to the vector's biology could lead to more accurate prediction models and early warning systems.

## OPEN ACCESS

**Citation:** Stolerman LM, Maia PD, Kutz JN (2019) Forecasting dengue fever in Brazil: An assessment of climate conditions. PLoS ONE 14(8): e0220106. <https://doi.org/10.1371/journal.pone.0220106>

**Editor:** Abdallah M. Samy, Faculty of Science, Ain Shams University (ASU), EGYPT

**Received:** July 26, 2018

**Accepted:** July 9, 2019

**Published:** August 8, 2019

**Copyright:** © 2019 Stolerman 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:** All relevant data underlying the results of this study are available within the paper and its Supporting Information files.

**Funding:** The authors received no specific funding for this work.

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

## Introduction

Dengue Fever is a tropical mosquito-borne viral disease present in more than 110 countries and a current threat to half of the world population [1, 2]. The dengue virus is primarily transmitted to humans through infected *Aedes aegypti* mosquitoes. This main disease vector is well adapted to urban environments, which allow viruses to spread easily through cities. In addition, local climate conditions play a critical role in the development of vector populations in major urban centers.

The first cases of dengue in Brazil date from the end of the 19<sup>th</sup> century, and despite the elimination of the *Aedes aegypti* in 1955, the mosquito was reintroduced in the country in the 70s. A historically important outbreak occurred in 1981 in Boa Vista, in the state of Roraima, following several outbreaks in Central America involving the DENV-1 and DENV-4 serotypes [3, 4]. Since then, dengue has become one of the major public health problems in Brazil, with several epidemics reported yearly across the country. While dengue symptoms are usually limited to fever and muscle/joint pain, some develop more severe forms of the disease such as hemorrhagic fever or shock syndrome.

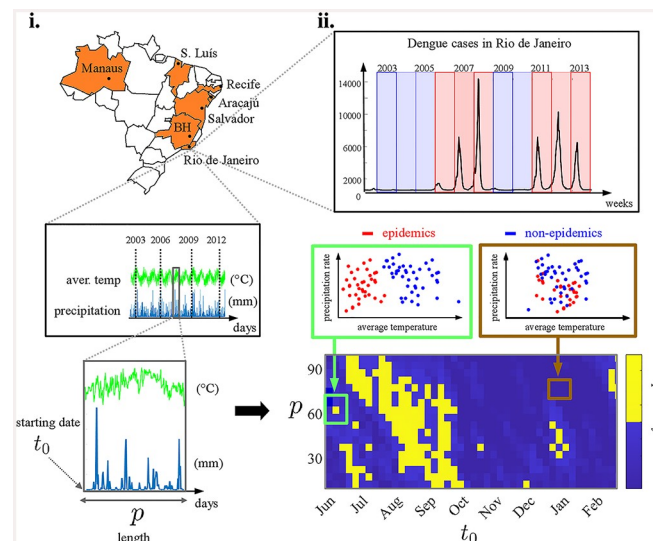
The proliferation of *Aedes aegypti* and the sustained transmission of dengue are influenced by a complex interplay of multi-scale factors such as the circulation of different serotypes [5, 6], the movement of infected and susceptible humans within a city [7, 8], and mosquito population size. There is also a growing body of evidence showing that local climate conditions such as temperature and precipitation may highly influence the biology of the mosquito [9–12]. Complicating our understanding is the fact that several cities exhibit an intricate alternation between epidemic and non-epidemic years. This suggests that climate conditions that favor dengue transmission are more complex than generally appreciated [13, 14].

In this work, we analyze climate and epidemiological data from seven major Brazilian cities (Aracajú, Belo Horizonte, Manaus, Recife, Rio de Janeiro, Salvador and São Luís), which had epidemic and non-epidemic years in the recent past. Fig 1 is a schematic overview of our method. We estimate the correlation of climate conditions in different epochs preceding epidemic periods using a data-driven methodology based on machine learning algorithms for clustering and classification [15, 16] known as Support Vector Machines (SVM) [17, 18], which were applied to climate variables that are key to the life cycle of the mosquito. We also explore the predictability of our method combining periods of high association between climate conditions and dengue epidemics with different prediction approaches. The insights of this work may help tailor public health policies for each different city by increasing vector control measures during neglected critical epochs and ultimately improving the forecasting of dengue epidemic years—which would allow the public health system to make earlier logistic preparations or mosquito eradication programs.

## Materials and methods

### Description of epidemiological and climate datasets

All epidemiological data utilized in this work were taken from the publicly available datasets of the Brazilian Notifiable Diseases Information System (SINAN, [19]). This includes the total



**Fig 1. Schematic overview.** We analyze time series data for climate variables from seven Brazilian state capitals (Aracajú, Belo Horizonte, Manaus, Recife, Rio de Janeiro, Salvador, and São Luís) and their connection to dengue epidemic years. (i) Illustrative example showing data from Rio de Janeiro. Two parameters define the epochs in which climate conditions are considered: the starting date  $t_0$  (month/day) and period length of  $p$  (days). (ii) We locate periods along the year where the separability between epidemic and non-epidemic climate is higher. Keeping track of signature differences at key epochs may significantly improve dengue forecasting in the upcoming years.

<https://doi.org/10.1371/journal.pone.0220106.g001>

number of dengue cases per year (from 2002 to 2017) for all Brazilian state capitals. While we cannot be sure that all dengue cases occurred within the area measured by the climate variables, we are confident that the numbers reported are sufficient for disambiguating between a dengue and non-dengue year. A year is conventionally classified as an *epidemic* year for a given city if the incidence of dengue is above 100 cases per 100,000 inhabitants in the period January–December and classified as a *non-epidemic* year otherwise, based on the Brazilian Ministry of Health classifications of dengue incidences [20]. In order to find critical climate signatures that may have contributed to the epidemic outcomes, we restrict ourselves to seven state capitals that displayed at least 3 epidemic years and 3 non-epidemic years in the period 2002–2012. This allowed us to investigate the correlation between distinct climate conditions and the complicated alternations between epidemic and non-epidemic years over time. The climate data utilized in this work was obtained from the National Institute of Meteorology (INMET) [21] and included time series for the average temperature (in Celsius) and precipitation (in millimeters) for the state capitals Aracajú, Belo Horizonte, Manaus, Recife, Salvador, and São Luís (from 1/1/2001 to 12/31/2012) and for Rio de Janeiro (from 1/1/2002 to 12/31/2013).

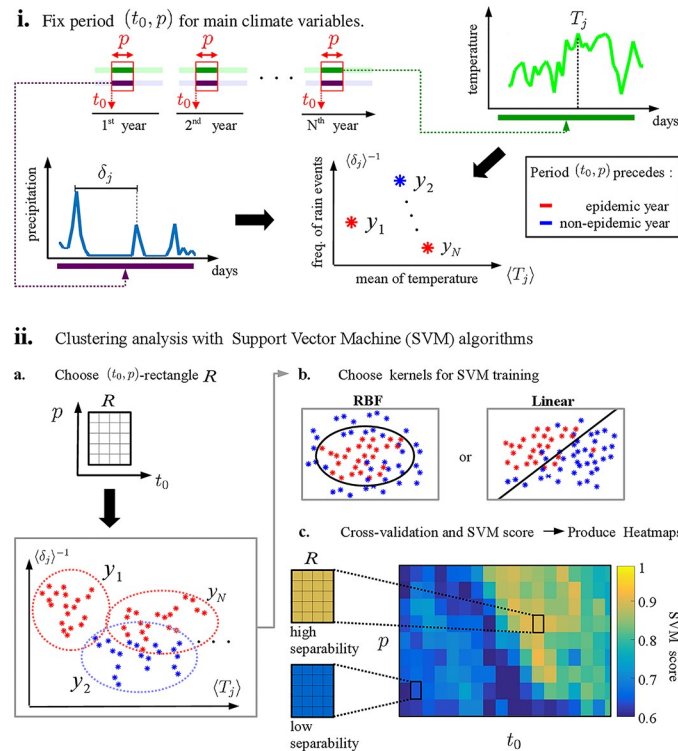
### Defining periods of critical climate conditions for dengue

In this work, we investigate the correlation of climate conditions on dengue epidemics at different periods along the yearly cycle. We let  $(t_0, p)$  denote a sampling period of  $p$  days starting at the date  $t_0$ . Then, for a fixed period, we evaluate a score quantifying the discrepancy between climate conditions in epidemic years and non-epidemic years. See Fig 1 for an illustrative example using data from the city of Rio de Janeiro: periods with high climate *separability* between epidemic years (red dots) and non-epidemic years (blue dots) might be of critical importance to the cycle of the urban mosquito population and consequently, to the occurrence of dengue in the following year.

In what follows, we define the *SVM scores* as a proxy for the cluster separability. Our method highlights potentially critical periods for the occurrence of dengue. Finally, since dengue outbreaks in Brazil typically take place between March–May in a given year, we limit the range of  $(t_0, p)$  from June (of the previous year) to May.

### SVM scores for cluster separability

Our SVM score for measuring discrepancies between climate conditions in epidemic/non-epidemic years is based on a supervised learning technique for classification. Fig 2 outlines the main steps of our SVM algorithm: (i) For a fixed  $(t_0, p)$  interval, we evaluate two climate indicators—the arithmetic mean of the average temperature  $\langle T_j \rangle$  and average frequency of rain events  $\langle \delta_j \rangle^{-1}$ , where  $\delta_j$  represents time intervals between consecutive peaks on precipitation data (see Fig 2i). We find the precipitation local maxima in the time series using Matlab's *find-peaks* function, calculate the time intervals  $\delta_j$  between them, and define the precipitation rate as the average peak interval. No specific thresholds were used in this step. (ii) We label the climate indicators in a 2D plot as an epidemic year (red) or as a non-epidemic year (blue) according to our dengue outbreak criteria. (iii) We repeat the process for  $t_0$  and  $p$  within a rectangular range  $R$  in the parameter space. Then we have a collection of red/blue points (dashed ellipses in Fig 2iia). In our simulations, the rectangular range  $R$  was  $5 \times 6$ , i.e, spanning 5 consecutive starting dates ( $t_0$ ) and 6 consecutive duration lengths ( $p$ ). In this work we use both Linear and Radial Basis Function (RBF) kernels for the SVM training step on the  $(t_0, p)$ -rectangles  $R$ . We cross-validated the climate indicators (red/blue dots) in the  $t_0 \times p$  period by subsampling 80% of the dataset and testing the classification accuracy in the remaining 20%. We

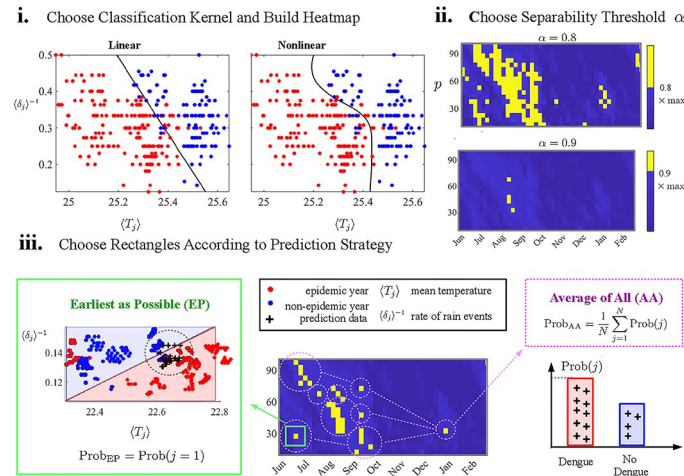


**Fig 2. Outline of SVM methodology.** A supervised learning technique for classification: (i) We calculate and plot mean of average temperature  $\langle T_j \rangle$  and frequency of rain events  $(\delta_j)^{-1}$  for a fixed  $(t_0, p)$  interval of all years, using red and blue colors or periods preceding epidemic and non-epidemic years respectively. (ii)(a) For each  $(t_0, p)$  interval of the rectangle  $R$  (called  $(t_0, p)$ -rectangle), we apply (i) to obtain a *cloud* (dashed circles) of points in the plane, for each year. (b) Linear and RBF kernels are used to execute the SVM train/test and cross-validation routines. (c) The SVM score for  $R$  is obtained. We plot  $t_0 \times p$  heatmaps with Regions of High and Low SVM scores, which indicates where temperature and precipitation are better correlated with the occurrence of dengue.

<https://doi.org/10.1371/journal.pone.0220106.g002>

evaluated the percentage of correctly classified test points and define the *SVM score* as the average accuracy after re-sampling the training/test data for 100 trials. No normalizations steps were used within the SVM steps, i.e., the climate indicators were simply the average temperature values and precipitation rates. Finally, we plot heatmaps (Fig 2iic) of the SVM scores for different  $(t_0, p)$ -rectangles within a range of  $t_0$  and  $p$  values. We remark that high/low SVM scores are consistently associated with separable/overlapping clusters of epidemic vs non-epidemic points (red vs blue dots). Thus, the SVM score is a good proxy for the geometrical separability of the clusters. We postulate that periods with high SVM scores might be of critical importance to the cycle of the urban mosquito population and consequently, to predict the occurrence of dengue in a given out-of-sample test year. The  $t_0$  values range from June 1<sup>st</sup> to February 21<sup>st</sup> and  $p$  ranges from 10–100 days (except for Rio de Janeiro, ranging from 5–100 days), which completely covers (from June 1<sup>st</sup> to May 31<sup>st</sup>) the periods that may influence dengue outbreaks.

The down-selection to the two parsimonious variables is consistent with well established and commonly used techniques such as LASSO and model/variable selection through information criteria such as AIC (Akaike Information Criteria) and BIC (Bayesian Information Criteria). These methods specifically penalize the number of predictive terms so that a parsimonious model is selected. In the application here, the two variables selected generalize their predictive power across all the different cities despite the different specific patterns of



**Fig 3. Outline of the prediction method.** For each state capital, we calculate the dengue probability for an out-of-sample year using the remaining 10 years as a training set: the user (i) chooses between a linear/nonlinear (RBF) classification kernel to build a heatmap of SVM<sub>score</sub> for a wide range of  $t_0$  and  $p$  values, (ii) selects  $(t_0, p)$  rectangles with  $SVM_{score} \geq \alpha \times \max(SVM_{score})$  for a threshold parameter  $\alpha$ , and (iii) computes the probability of dengue occurrence in the testing year using the *Earliest as Possible (EP)* strategy or the *Average of All (AA)* strategy. EP uses only the first selected rectangle (boxed in green) while AA takes an average of the probabilities of all selected rectangles (circled in magenta). See text for details.

<https://doi.org/10.1371/journal.pone.0220106.g003>

clustering (See Supplementary Information). More broadly, the down selection is consistent with the philosophy of the Pareto optimal solution, or Occam’s razor: explain the majority of observed data with an interpretable, parsimonious model. See [S1 Appendix](#) for details.

### Out-of-sample prediction

Our training dataset for each state capital consists of 11 years (2002–2011) of temperature and precipitation time series. Due to the small number of years available and due to methodological constraints (that require a certain number of both epidemic and non-epidemic years in the training set), we can select 10 years for training and test/predict the remaining out-of-sample year with a few different strategies. This is effectively a *leave-one-out* cross-validation procedure enforced by the limited number of years in the dataset. Ideally, one would like to use a more sophisticated cross-validation procedure, but most other methods require substantially more data, i.e. number of dengue versus non-dengue years. [Fig 3](#) illustrates the steps below:

1. **Choose SVM kernel and compute heatmap:** The user should choose between a linear/nonlinear (RBF) kernel to classify the climate data in the  $\langle T_j \rangle \times \langle \delta_j \rangle^{-1}$  plane. See [Fig 3i](#) for an example. This classifier will provide an SVM score (color-coded in the heatmap) for each  $(t_0, p)$ -rectangle.
2. **Choose the SVM threshold  $\alpha$ :** Once the SVM heatmap is ready, we must select the  $(t_0, p)$  rectangles that will be used to predict the testing year. We introduce a threshold parameter  $\alpha \in [0, 1]$  and pick rectangles with  $SVM_{score} \geq \alpha \times \max(SVM_{score})$ . [Fig 3ii](#) shows that higher values of  $\alpha$  diminishes the number of selected rectangles in the  $t_0 \times p$  plane.
3. **Choose a prediction strategy:** [Fig 3iii](#) illustrates the last choice needed to compute the probability of dengue occurrence in the testing year.
  - *Earliest as Possible (EP)*: this strategy uses the rectangle in the  $t_0 \times p$  plane with earliest  $t_0$ , and in case of a tie, it chooses the one with the lowest  $p$ . We denote the index of this



rectangle as  $j = 1$  (see Fig 3). It then computes the dengue probability, denoted by  $\text{Prob}(j = 1)$ , as the fraction of those test climate data points that fall into the dengue  $\langle T_j \rangle \times \langle \delta_j \rangle^{-1}$  semi-space. We address the evaluated quantity as the EP probability

$$\text{Prob}_{\text{EP}} = \text{Prob}(j = 1). \quad (1)$$

- *Average of All (AA)*: this strategy computes the probability of dengue occurrence in the testing year using all  $N$  selected  $(t_0, p)$  rectangles in step (ii) and taking an average of their probabilities. We address the AA probability as

$$\text{Prob}_{\text{AA}} = \frac{1}{N} \sum_{j=1}^N \text{Prob}(j). \quad (2)$$

Predicting the dengue outcome of an out-of-sample year requires choosing (i) a classification kernel (linear vs nonlinear), (ii) a threshold  $\alpha$  value (0.9, 0.95 or 1), and (iii) a strategy for calculating the probability of dengue occurrence (EP vs AA). Probability values above/below 0.5 led to epidemic/ non-epidemic predictions, respectively.

The results were then summarized in *confusion matrices* containing all four types of correct/wrong predictions: True Positives (TP), True Negatives (TN)/False Positives (FP), and False Negatives (FN). Our prediction accuracy

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (3)$$

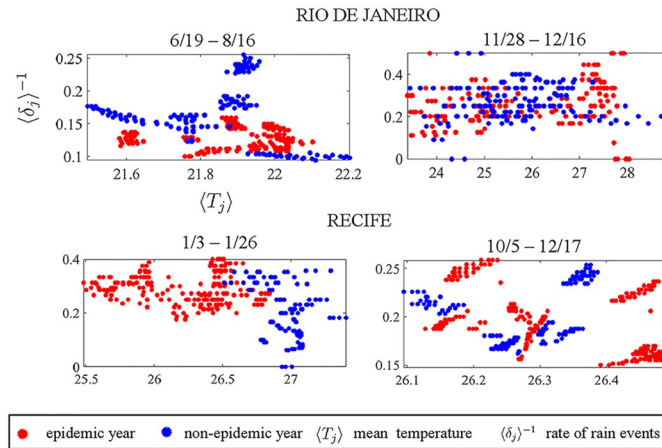
was the outcome was the outcome measure by which we compared the different prediction methods.

## Results

In this section, we highlight significant differences between climate conditions during epidemic/non-epidemic years for a period starting at day  $t_0$  and duration of  $p$  days along the yearly cycle. Before delving into our dengue prediction results, it is highly informative to interpret high/low SVM scores for distinguishing epidemic and non-epidemic correlations. Fig 4 demonstrates the clustering of data, or lack thereof, for Rio de Janeiro and Recife, considering all 11 years of training data (Figs C and D in the S1 Appendix show similar results for all state Capitals). The left side of the panel shows representative data for time windows achieving a high correlation score. Remarkably, the red (epidemic) and blue (non-epidemic) dots are well separated and distinguishable by visual inspection. On the other hand, the right side shows data structures with low correlation scores. Note the significant overlap between the red and blue dots, suggesting that using this region for prediction of an epidemic is highly suspect. This result illustrates that each city has a unique pattern of clustering that can be capitalized on in order to provide predictive metrics for dengue epidemic years.

### Validation of the training dataset (2002–2012)

The different choices in (i) SVM kernel (linear or RBF), (ii) SVM threshold  $\alpha$  (0.9, 0.95 or 1), and (iii) prediction strategy (EP or AA) lead to 12 possible confusion matrices for each state capital. We report the best choices for each city in Table 1 and leave the full report of our results for the SI. For the 5 state capitals where the EP strategy had best results, we found their respective *EP-windows*, i.e., a median date-range that comprises all EP-chosen rectangles used in the prediction. For state capitals where the AA strategy performed better, we highlighted the *AA-months* that were common to all out-of-sample predictions. Finally, we showed specific



**Fig 4. Examples of high and low cluster separability plots with the full training dataset.** For each state Capital, we selected special time windows in which there was a clear separation between climate signatures preceding epidemic and non-epidemic years. This picture illustrates the cases of Rio de Janeiro and Recife. The Left side of the panel shows distinct data separation, while in the right side the climate variables seem to be poorly distinguishable, therefore not suitable for dengue prediction. This separability notion is made quantitatively precise by the SVM scores (see text for details). Examples for the other capitals can be found in the [S1 Appendix](#).

<https://doi.org/10.1371/journal.pone.0220106.g004>

climate signatures for state capitals with good EP predictions. In what follows, we compute the heatmaps as described in the methods section (see also [Fig 2](#)). Here we present the best prediction results for Rio de Janeiro and Salvador and leave the details for the other capitals in the [S1 Appendix](#).

**Rio de Janeiro.** Fig H in the [S1 Appendix](#) shows the best prediction result for Rio de Janeiro using (i) an RBF kernel, (ii) an SVM threshold of  $\alpha = 1$ , and (iii) the EP-strategy to calculate the outbreak probability. Most EP-chosen rectangles occurred in the winter and in the spring. The corresponding EP-window ranged between June 19<sup>th</sup> and September 25<sup>th</sup>, when most Epidemic years (all except 2012) had average temperatures above 23 Celsius and precipitation rates below 0.15 (see prediction tables at the [S1 Appendix](#) for details). All years except 2010 (FP) and 2012 (FN) were correctly predicted (82% accuracy).

**Salvador.** Fig I (top) in the [S1 Appendix](#) shows the best prediction result for the city of Salvador using (i) an RBF kernel, (ii) an SVM threshold of  $\alpha = 0.95$ , and (iii) the AA-strategy to calculate the outbreak probability. The  $(t_0, p)$  rectangles used in the prediction covered most of the year but were especially clustered around December-February (boxed in magenta). All years except 2002 (FN) and 2010 (FN) were correctly predicted (82% of accuracy).

**Table 1. Best prediction results: Training set.** We report the choices of SVM kernel, threshold, and strategy that resulted in highest prediction accuracy for each state capital, along with their respective EP-Windows or AA-months. \*Similar results were found with the AA-strategy for Belo Horizonte. \*\*Both strategies gave good results for Salvador. See text and [S1 Appendix](#) for details.

Capital	Kernel	$\alpha$	Strategy	Accuracy	EP-Window/AA-months
Aracajú	RBF	0.9	EP	91%	Jun 1–19
Belo Horizonte	RBF	1	EP	73%*	Jun 13–Aug 25
Manaus	Linear	0.95	AA	64%	Aug–Oct
Recife	Linear	1	AA	82%	Dec–Jan
Rio de Janeiro	RBF	1	EP	82%	Jun 19–Sept 25
Salvador	Linear/RBF	0.9/0.95	EP/AA	73%/82%	Aug 30–Dec 11**
São Luís	RBF	1	AA	82%	Dec–Mar

<https://doi.org/10.1371/journal.pone.0220106.t001>

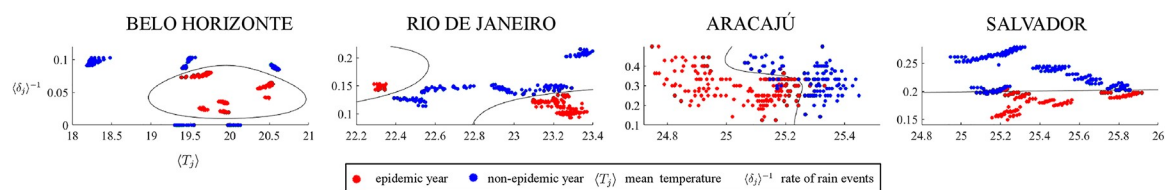
Predictions using (i) a linear kernel, (ii)  $\alpha = 0.9$ , and (iii) the EP-strategy also gave good results (highlighted in Fig I (bottom) in the SI appendix). Eight years were correctly predicted (73% accuracy) but the years of 2008 (FP), 2010 (FN) and 2012 (FN) were not. The EP strategy was just slightly less accurate than the AA strategy, yielding EP-windows within August 30<sup>th</sup> and December 11<sup>th</sup> (spring and summer). The epidemic years typically showed lower precipitation rates in the selected EP-rectangles.

**Other Capitals.** Each state Capital has an optimal choice of SVM kernel,  $\alpha$  value and prediction strategy, as Table 1 shows. In the S1 Appendix, we present the different prediction results for Aracajú, Belo Horizonte, Manaus, Recife and, São Luís. See Figs E, F, G, and J in the S1 appendix for a comprehensive description of their prediction outcomes.

**SVM classification and climate signatures.** Fig 5 shows the corresponding favorable climate conditions for all capitals with predictive EP-periods. The EP prediction strategy uses only one rectangle from the  $t_0 \times p$  heatmap, i.e., the one with the lowest  $t_0$ . This allows us to show the specific temperature and rain values that distinguished epidemics and non-epidemic years in that EP window. Contrastingly, the AA strategy averages over several rectangles throughout the entire year, making the analysis of specific climate conditions for each window impractical. The EP rectangles occur in June (winter) for the first three capitals and in the spring/summer for Salvador. The classifiers (curves in black) take very distinct shapes for the different cities. For Belo Horizonte, the different clusters were separated by an ovoid-shape kernel and most epidemic years had a precipitation rate between [0.02,0.08]. For Rio de Janeiro, most epidemic years have average temperatures above 23° C and the clusters are separated by a hyperboloid-shape kernel. In Aracajú, an S-shape kernel separates the clusters around a temperature threshold of 25.2° C. Finally, favorable climate conditions for dengue epidemics occur in Salvador during the spring for a frequency of rain events below 0.2. It is hard to infer specific relationships between optimal temperature and rain events however, because the cities have significantly different sizes, geography, vegetation, topography and other factors that might impact the mosquito development. Rio de Janeiro, for instance, exhibits a vast array of sub-regions ranging from highly-populated urban centers to forests [22].

### Predictions for the holdout dataset (2013–2017)

We used the model trained with data from earlier years 2002–2012 to predict dengue outcomes in a holdout dataset (usually from 2013–2017, but may vary depending on data availability). See S1 Appendix for details. It should be noted that approximately seven months after



**Fig 5. Favorable climate conditions for epidemics in predictive EP-periods.** The EP strategy uses the rectangle in the  $t_0 \times p$  plane with earliest  $t_0$ . Four capitals exhibited highly predictive EP rectangles, and we show the corresponding epidemic vs non-epidemic climate conditions. Belo Horizonte: EP-window from June 13th to August 25th. Most epidemic years had a precipitation rate in the interval [0.02,0.08] and different clusters were separated by an ovoid-shape kernel. Rio de Janeiro: EP-window ranged between June 19th and September 25th. Most epidemic years had average temperatures above 23° Celsius and precipitation rates below 0.15. Clusters were separated by a hyperboloid-shape kernel. Aracajú: EP-window from June 1st–19th. There is a clear separability between dengue and no-dengue regarding a temperature threshold around 25.2° Celsius. Clusters were separated by an S-shape kernel. Salvador: EP-windows from August 30th–December 11th. Clusters were separated by a single linear threshold of  $\langle \delta_i \rangle^{-1}$  below 0.2. The picture shows climate signatures considering training years 2003–2012 for Rio de Janeiro and 2002–2011 for the other capitals.

<https://doi.org/10.1371/journal.pone.0220106.g005>



**Table 2. Best prediction results: Holdout dataset.** \*We obtained 80% accuracy (4/5) with  $\alpha = 0.9$  for both EP and AA strategies and both RBF and linear kernels. \*\* We obtained 100% accuracy (5/5) with  $\alpha = 0.95$ , EP strategy and RBF kernel.

Capital	Accuracy
Aracajú	60% (3/5)
Belo Horizonte	60% (3/5)
Manaus	80% (4/5)
Recife	60% * (3/5)
Rio de Janeiro	67% (2/3)
Salvador	80%** (5/5)
São Luís	100% (3/3)
<b>TOTAL</b>	<b>74% (23/31)</b>

<https://doi.org/10.1371/journal.pone.0220106.t002>

submission of the manuscript, SINAN released new dengue data for the years 2013-2017 [19]. Table 2 shows the accuracy for each state capital using the corresponding kernel, parameters, and strategy defined in the training step. The state capital of São Luís exhibited the best accuracy (100% corresponding to 3 correct predictions from a total of 3 test years), followed by Manaus and Salvador (80% accuracy corresponding to 4 correct predictions from a total of 5 test years). For Rio de Janeiro, Aracajú, Belo Horizonte and Recife, we obtained accuracies below 70%. Overall, we obtained a 74% accuracy considering the predictions from the 7 state capitals, correctly predicting the outcome of 23 out of 31 experiments.

## Discussion

Understanding how *Ae. aegypti* mosquitoes respond to climate conditions is crucial for developing climate-based early warning systems for dengue prediction. While several works report and quantify how climate may influence the mosquito development on a weekly scale [23–25], we suggest that long-term effects occurring even months before the outbreaks may also play an important role. We developed a new data-driven method, based on Support Vector Machine (SVM) algorithms to identify, in a systematic manner, a set of critical periods and climate signatures in the annual cycle that may be decisive for the development of dengue epidemic years. We applied our methods to temperature and precipitation time series data for seven state capitals in Brazil where there was a significant alternation between epidemic and non-epidemic years in the recent past. We explored a few strategies to estimate the predictive power of our method, and the most accurate results for each state capital led to interesting time periods and climate patterns associated with the occurrence of dengue epidemics.

### Critical seasons for each state capital

In accordance to other reported studies [13, 26, 27], we obtained strong evidence that the correlation between climate and epidemics varies significantly across different state capitals, thus rejecting simplistic or universal explanations involving temperature and rain precipitation in urban centers. Remarkably, the average temperature and the frequency of precipitation showed a strong predictive power throughout the winter season for the cities of Aracajú, Belo Horizonte, Manaus and Rio de Janeiro (see Table 3). As a consequence, intensifying mosquito control campaigns during the winter season may prove an interesting epidemic control strategy, especially due to the smaller size of the vector populations during that period. In Brazil, the national and local campaigns are usually restricted to spring and summer periods [28, 29]. In fact, the Brazilian government announced that a special task force for fighting mosquitoes

**Table 3. Season highlights.** We found that each state capital has its own preferred seasons in which climate signatures may impact Dengue occurrence. Peaks of dengue happen typically during the fall (March–May).

Capital	Winter	Spring	Summer
Aracajú	×		
Belo Horizonte	×		
Manaus	×	×	
Recife			×
Rio de Janeiro	×	×	
Salvador		×	×
São Luís			×

<https://doi.org/10.1371/journal.pone.0220106.t003>

was to be formed November 3<sup>rd</sup>, 2016 [30]. We believe this starting date to be too late since critical climate conditions were detected in some cities even 9 months prior to epochs with higher dengue incidence.

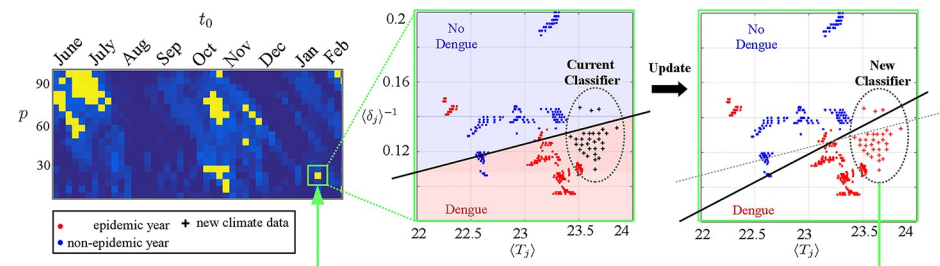
### Appending new data and updating our method

Our training sets and our classifiers used in intermediate methodological steps should be updated as new climate/epidemic data are made available. See Fig 6 for a schematic representation of how new climate data (black crosses) should be assimilated by the training dataset to improve the separability within the SVM heatmap and increase the statistical robustness of our prediction method. Thus, our method and its accuracy should be continuously updated in time to provide more reliable separability regions and accurate climate-based forecasts.

### Impacts of climate variables on the *Aedes aegypti* life cycle

Temperature and precipitation are important environmental factors affecting all biological processes of the *Ae. aegypti*. In fact, there are even precise mathematical expressions relating developmental rates with temperature [12, 31]. The rates at which mosquitoes acquire and transmit viruses are also temperature-dependent [32–35]. Precipitation events in their turn are extremely important for dengue transmission [36, 37]. The abundance of *Ae. aegypti* is regulated by rainfall during the water-dependent stages (egg, larva and, pupa), which provides breeding sites and stimulates egg hatching [38, 39].

The relations between lower temperatures, rain, and size of the mosquito population are usually studied in countries with temperate climates, where excessive rain propitiate egg



**Fig 6. Appending data for further analysis.** For a high scored  $(t_0, p)$ -rectangle (green box), we plot the respective climate indicators with their epidemic/non-epidemic (red/blue) labels. A classifier is used to predict the outcome of newly available climate data (black crosses). Depending on the outcome, the new data is appended to the SVM-training set. This procedure will also update the SVM score and the importance of the chosen  $(t_0, p)$ -rectangle for dengue prediction.

<https://doi.org/10.1371/journal.pone.0220106.g006>

hatching but the lower temperature might prove fatal for the larvae [40, 41]. The Brazilian tropical climate, however, may present adequate temperatures for vector proliferation even in the winter. Thus, we conjecture that winter rain-events may play an important role in the first mosquito generation in that year. A larger initial population, when compounded over several reproductive cycles, could lead to an epidemic outbreak in the summer. As shown in Fig 5, favorable climate conditions in Belo Horizonte, Rio de Janeiro, and Salvador are mostly a function of the rain events. The development of the mosquito population in the winter season is not a central subject of epidemic studies in tropical countries, and our work suggests that this can be a promising avenue for future studies.

Kesorn et al. (2015) [42] recently addressed a decade-long limitation of dengue surveillance systems, namely, that environmental factors can be unreliable and degrade the predictions when applied to areas with similar climate. The prediction accuracy of their model increased dramatically when, instead of using climate parameters in a classical framework, they utilized the *Ae. aegypti* female and larvae mosquito infection rates. Our work, on the other hand, was able to successfully predict dengue years using solely climate variables. This raises an important question: how reliable are climate parameters for dengue prediction? One possible explanation is that these parameters are reliable only at coarser spatial scales, and the large distances between cities in a continental country such as Brazil lead to meaningful climate differences. Another explanation is that our methodological innovations did improve the reliability of local climate factors; Kesorn et al. (2015) dismissed temperature as a good predictor by visual inspection of its time series, while we allow a wide range of time-lags linking temperature and future outcomes. It would be interesting to see if our approach could improve the reliability of climate signatures in other contexts.

Daily changes in temperature are known to affect the efficiency of the *Ae. aegypti* [43, 44]. Kesorn et al. (2015) also showed that the infection rates for the female *Ae. aegypti* and larvae correlate strongly with the number of human-reported dengue cases. On this regard, our method is agnostic as to which specific mechanisms led to an increase in the number of human cases. The factors above may be the missing link between climate variables and observed human cases. However, as also pointed out by the authors, it is not always possible to obtain data on mosquito infection rates. To the best of our knowledge, there are no available data on female and larvae mosquito infection rates for the Brazilian cities that we studied. Moreover, it would be extremely challenging to obtain a single infection rate on our spatial scale, especially for large capitals such as Rio de Janeiro and Salvador.

## Machine learning for dengue prediction

There is a broad array of methods to examine the influence of climate variables on dengue outbreaks: wavelet-analysis for time series [26], autoregressive integrated moving average (ARIMA) models [45], fuzzy association rule mining techniques [14], rule-based classifiers [46], Bayesian methods [47, 48] and others. See Raclouz et al. [49] for a systematic literature review. More recently, a number of machine learning methods emerged to address the prediction of dengue outbreaks. Baqueiro et al. [50] produced a comprehensive comparison of generalized additive models (GAMs), artificial neural networks (ANNs) and seasonal autoregressive integrated moving average models (SARIMA) for the city of São Paulo. They obtained accurate predictions for dengue within a one-month time window. Our method provides larger time windows, and thus more time for implementing disease surveillance or outbreak prevention measures. In a similar study, Guo et al. [51] analyzed climate data from Guangdong, China, to forecast dengue outbreaks using support vector regression (SVR) algorithm, step-down linear regression model, gradient boosted regression tree algorithm (GBM), negative binomial

regression model (NBM), least absolute shrinkage and selection operator (LASSO) linear regression models. The authors explored a four-year time series of weekly dengue cases, which can be a major issue for their prediction routine if the data is not reported in a timely fashion. Their SVR algorithm exhibited the best prediction performance with a 12-week time-window, which was also effective in other regions of China. Their results reported SVM-based models as highly predictive tools for dengue epidemics, but we provide additional plots for key climate signatures (see Fig 5). Our major methodological innovation is to frame the dengue-forecasting problem within an SVM setting that localizes the important periods for dengue prediction and their associated climate patterns. Other machine learning methods used recently include C-Support Vector Classification (C-SVC) [52], Random Forests [53], Decision Tree-Based Approaches [54] and even the curious combination of ARIMA models with Google Trends data [55]. Due to the nuanced and complex differences between the specific settings, we will leave a more detailed comparison of the methods for future works.

### Limitations of our methods

There are several limitations to our work and all of our results must be interpreted with caution and parsimony. We also acknowledge that using a binary threshold for classifying a year as epidemic/non-epidemic is somewhat arbitrary, but we decided to abide by the convention established by the Brazilian Ministry of Health. Moreover, we did not consider several other factors believed to be important for explaining dengue dynamics in details, such as circulation of different strains of the dengue virus [5, 6], human mobility within and among the cities [7, 8, 56, 57], human demographic dynamics [58, 59] and global warming and climate changes [60, 61]. Therefore it is important to acknowledge that there might be potential confounding between epidemic years and the coincidence of favorable climate conditions, given that other processes are not represented in the model.

Finally, we acknowledge that our machine learning method is agnostic as to which sequence of events was responsible for increasing/decreasing the number of human dengue cases from year to year. To be an effective vector, mosquitoes must have a high vector competence and vectorial capacity. The first refers to their ability to receive a disease agent microorganism from the reservoir host and then later transmit the infectious agent to another susceptible host. The vectorial capacity includes a number of factors like vector competence, mosquito population density, host preferences, biting rate, immunity of the mosquitoes, and others. All these factors may have been affected by the climatic differences from year to year. While we cannot disambiguate *which* changes occurred, our predictive windows along the yearly cycle may provide insight as to *when* they occurred.

### Future work

Since our methods led to promising forecasting results for various capitals of Brazil, in the future we would like to apply the same approach to other cities and climate-datasets worldwide. We also hope to better compare our methods with other machine learning techniques in future works. With respect to the definition of epidemic and non-epidemic years, we acknowledge that labeling years as dengue vs non-dengue might be too coarse and further insight might be gained by a richer categorization of epidemic years. Exploring the use of more categories or classes (such as high/medium/low years) may be an interesting approach for future studies.

### Conclusion

Epidemic control of dengue is one of the most urgent public health challenges in tropical countries such as Brazil. A better understanding of the multi-scale and long-term effects of

climate conditions on the development of *Aedes aegypti* populations is crucial for improving the timing of vector-control efforts and other policies. In this work, we show that two specific climate variables—mean of temperature and frequency of precipitation—may be crucial for dengue prediction in Brazil. Remarkably, for Aracajú, Belo Horizonte, Manaus, and, Rio de Janeiro, a prediction can be made approximately six to nine months before the epidemic outbreak, which usually takes place in the months of March-May. Interestingly, the summer season in Rio de Janeiro offers little insight into this matter, since the data of years with and without dengue are qualitatively similar from a climate perspective. Yet public strategies have typically been enacted and decided during this time period, which is both too late and does not leverage the predictive capabilities of the climate data. This work also highlights that climate patterns with predictive success are quite distinct from city to city. This is large to be expected as climatic effects, such as proximity to the ocean, to the jungle/forest, dense populations, etc will likely play a significant role in how precipitation and temperature affect the growth of the disease vector *Aedes aegypti*.

## Supporting information

**S1 Appendix. Contains supplemental data and additional information regarding the article.**

(PDF)

**S2 Appendix. Contains supplemental data regarding Dengue incidence and missing climate data between 2013 and 2017.**

(PDF)

**S1 File. Contains data and Matlab codes.**

(ZIP)

## Acknowledgments

LMS, PDM and JNK would like to acknowledge Prof. Stefanella Boatto from Federal University of Rio de Janeiro (UFRJ), Joshua L. Proctor from the Institute of Disease Modelling, Prof. Cláudia T. Codeço (PROCC/FIOCRUZ), Prof. Roberto.I. Oliveira (IMPA), and Micheal Getz (UCSD) for their insightful comments and enthusiastic support of this work. A significant part of this project was done during Lucas' Ph.D. at the Instituto Nacional de Matemática Pura e Aplicada (IMPA/Brazil), and his internship at the University of Washington funded by the Brazilian National Council of Research (CNPq).

## Author Contributions

**Conceptualization:** Lucas M. Stolerman, Pedro D. Maia, J. Nathan Kutz.

**Data curation:** Lucas M. Stolerman.

**Formal analysis:** Lucas M. Stolerman.

**Methodology:** Lucas M. Stolerman.

**Supervision:** J. Nathan Kutz.

**Writing – original draft:** Lucas M. Stolerman, Pedro D. Maia.

**Writing – review & editing:** Lucas M. Stolerman, Pedro D. Maia, J. Nathan Kutz.



## References

1. World Health Organization. dengue and severe dengue. WHO Factsheet No 117. Geneva. 2015; Available from: <http://www.who.int/mediacentre/factsheets/fs117/en/>.
2. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. *Nature*. 2013; 496, 504–507. <https://doi.org/10.1038/nature12060> PMID: 23563266
3. Figueredo LTM. Dengue in Brazil: Past, Present and Future Perspectives. *dengue Bulletin*. 2003; 27, 25–33.
4. Fares RC, Souza KP, Añez G, Rios M. Epidemiological Scenario of dengue in Brazil. *Biomed research international*. 2015. <https://doi.org/10.1155/2015/321873> PMID: 26413514
5. Rabaa MA, Simmons CP, Fox A, Le MQ, Nguyen TTT, et al. dengue virus in sub-tropical northern and central Viet Nam: Population immunity and climate shape patterns of viral invasion and maintenance. *PLoS Negl Trop Dis*. 2013; 7 (12) e2581. <https://doi.org/10.1371/journal.pntd.0002581> PMID: 24340118
6. Raghwani J, Rambaut A, Holmes EC, Hang VT, Hien TT, et al. Endemic dengue associated with the co-circulation of multiple viral lineages and localized density-dependent transmission. *PLoS Pathog*. 2011 7: e1002064. <https://doi.org/10.1371/journal.ppat.1002064> PMID: 21655108
7. Adams B, Kapan DD. Man bites mosquito: understanding the contribution of human movement to vector-borne disease dynamics. *PLoS one*. 2009 4(8), e6763. <https://doi.org/10.1371/journal.pone.0006763> PMID: 19707544
8. Stoleran LM, Coombs D, Boatto S. SIR-Network Model and Its Application to dengue Fever. *SIAM Journal on Applied Mathematics* *SIAM Journal on Applied Mathematics*. 2015. 75(6), 2581–2609. <https://doi.org/10.1137/140996148>
9. Watts DM, Burke DS, Harrison BA, Whitmire RE, Nisalak A. Effect of temperature on the vector efficiency of *Aedes aegypti* for dengue 2 virus. *Am J Trop Med Hyg*. 1987. 36: 143–152. <https://doi.org/10.4269/ajtmh.1987.36.143> PMID: 3812879
10. Foo LC, Lim TW, Lee HL, Fang R. Rainfall, abundance of *Aedes* and dengue infection in Selangor, Malaysia. *Southeast Asian J Trop Med Pub Health*. 1985. 16: 560–568.
11. Hopp MJ, Foley JA. Global-scale relationships between climate and the dengue fever vector, *Aedes aegypti*. *Clim Change*. 2001. 48: 441–463. <https://doi.org/10.1023/A:1010717502442>
12. Lana RM, Morais MM, Lima TFMd, Carneiro TGdS, Stoleran LM, dos Santos JPC, et al. (2018) Assessment of a trap based *Aedes aegypti* surveillance program using mathematical modeling. *PLoS ONE* 13(1): e0190673.
13. Adde A, Roucou P, Mangeas M, Ardillon V, Desenclos J-C, Rousset D, et al. Predicting dengue Fever Outbreaks in French Guiana Using Climate Indicators. *PLOS Negl Trop Dis*. 2016; 10: e0004681. <https://doi.org/10.1371/journal.pntd.0004681> PMID: 27128312
14. Buczak AL, Baugher B, Babin SM, Ramac-Thomas LC, Guven E, Elbert Y, et al. Prediction of high incidence of dengue in the Philippines. *PLOS Neglected Tropical Diseases*, 2014; 8:24:e2771. <https://doi.org/10.1371/journal.pntd.0002771>
15. Murphy KP. *Machine learning: a probabilistic perspective*. MIT press; 2012 Sep 7.
16. Bishop CM. *Pattern recognition. Machine Learning*. 2006; 128.
17. Cortes C, Vapnik V. Support-vector networks. *Machine learning*. 1995. 20(3), 273–297. <https://doi.org/10.1023/A:1022627411411>
18. Burges CJ. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery* 2.2. 1998. 121–167. <https://doi.org/10.1023/A:1009715923555>
19. Brazilian National Surveillance System (SINAN) Total number of dengue cases in state capitals Available from (website in Portuguese): <http://portalsinan.saude.gov.br/dados-epidemiologicos-sinan>
20. Ministry of Health, Epidemiological Report—dengue Fever (January to June, 2008). Available from (website in Portuguese): [http://bvsmms.saude.gov.br/bvs/publicacoes/informe\\_epidemiologico\\_dengue\\_janeiro\\_junho\\_2008.pdf](http://bvsmms.saude.gov.br/bvs/publicacoes/informe_epidemiologico_dengue_janeiro_junho_2008.pdf)
21. Brazilian National Institute of Meteorology (INMET) Temperature and precipitation time series. Available from (website in Portuguese): <http://www.inmet.gov.br/projetos/rede/pesquisa/>
22. Lana R. M., Carneiro T. G., Honório N. A., & Codeco C. T. (2014). Seasonal and nonseasonal dynamics of *Aedes aegypti* in Rio de Janeiro, Brazil: Fitting mathematical models to trap data. *Acta tropica*, 129, 25–32. <https://doi.org/10.1016/j.actatropica.2013.07.025> PMID: 23933186
23. Pessanha JEMP, Caiiffa WT, Almeida MCDM, Brandao ST, Proietti FA. Diffusion pattern and hotspot detection of dengue in Belo Horizonte, Minas Gerais, Brazil. *Journal of tropical medicine*. 2012. <https://doi.org/10.1155/2012/760951>

24. Honório NA, Codeço CT, Alves FC, Magalhães MDA, Lourenço-de-Oliveira R. Temporal distribution of *Aedes aegypti* in different districts of Rio de Janeiro, Brazil, measured by two types of traps. *Journal of Medical Entomology*. 2009; 46(5), 1001–1014. <https://doi.org/10.1603/033.046.0505> PMID: 19769029
25. Dibo MR, Chierotti AP, Ferrari MS, Mendonça AL, Chiaravalloti Neto F. Study of the relationship between *Aedes (Stegomyia) aegypti* egg and adult densities, dengue fever and climate in Mirassol, state of São Paulo, Brazil. *Memorias do Instituto Oswaldo Cruz*. 2008; 103(6), 554–560. <https://doi.org/10.1590/s0074-02762008000600008>
26. Liao CM, Huang TL, Lin YJ, You SH, Cheng YH, Hsieh NH, Chen WY. Regional response of dengue fever epidemics to interannual variation and related climate variability. *Stochastic Environmental Research and Risk Assessment*. 2015; 29(3), 947–958. <https://doi.org/10.1007/s00477-014-0948-6>
27. Johansson MA, Dominici F, Glass GE. Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Negl Trop Dis*. 2009; 3(2), e382. <https://doi.org/10.1371/journal.pntd.0000382>
28. Brazilian Ministry of Health. Promotion of national mobilization effort against *Aedes Aegypti* in 2013 (in Portuguese). Available from: <http://www.brasil.gov.br/saude/2013/11/governo-lanca-nova-campanha-de-mobilizacao-contra-dengue>
29. Brazilian Ministry of Health. Promotion of national mobilization effort against *Aedes Aegypti* in 2016 (in Portuguese). Available from: <http://www.brasil.gov.br/governo/2016/02/dilma-visita-rio-de-janeiro-no-dia-nacional-de-mobilizacao-zika-zero>
30. Brazilian Ministry of Health. Promotion of national mobilization effort against *Aedes Aegypti* for 2017. Available from: <http://www.brazilgovnews.gov.br/news/2016/11/government-promotes-national-mobilisation-effort-against-aedes-aegypti>
31. Focks D, Haile D. Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): simulation results and validation. *J Med Entomol*. 1993; 30(6):1018–1028. <https://doi.org/10.1093/jmedent/30.6.1018> PMID: 8271243
32. Alto B. W., & Bettinardi D. (2013). Temperature and dengue virus infection in mosquitoes: independent effects on the immature and adult stages. *The American journal of tropical medicine and hygiene*, 88 (3), 497–505. <https://doi.org/10.4269/ajtmh.12-0421> PMID: 23382163
33. Alto BW, Bettinardi D. (2013). Temperature and dengue virus infection in mosquitoes: independent effects on the immature and adult stages. *The American journal of tropical medicine and hygiene*, 88 (3), 497–505. <https://doi.org/10.4269/ajtmh.12-0421> PMID: 23382163
34. Peña-García VH, Triana-Chávez O, Arboleda-Sánchez S. (2017). Estimating Effects of Temperature on dengue Transmission in Colombian Cities. *Annals of global health*, 83(3-4), 509–518. <https://doi.org/10.1016/j.aogh.2017.10.011> PMID: 29221523
35. Mordecai EA, Cohen JM, Evans MV, Gudapati P, Johnson LR, Lippi CA et al. (2017). Detecting the impact of temperature on transmission of zika, dengue, and chikungunya using mechanistic models. *PLoS neglected tropical diseases*, 11(4), e0005568. <https://doi.org/10.1371/journal.pntd.0005568> PMID: 28448507
36. Choi Y., Tang C. S., McIver L., Hashizume M., Chan V., Abeyasinghe R. R., . . . & Huy R. (2016). Effects of weather factors on dengue fever incidence and implications for interventions in Cambodia. *BMC public health*, 16(1), 241. <https://doi.org/10.1186/s12889-016-2923-2> PMID: 26955944
37. Xu L., Stige L. C., Chan K. S., Zhou J., Yang J., Sang S., . . . & Lu L. (2017). Climate variation drives dengue dynamics. *Proceedings of the National Academy of Sciences*, 114(1), 113–118. <https://doi.org/10.1073/pnas.1618558114>
38. Silva FD, Santos AMD, Corrêa RDGCF, Caldas ADJM. (2016). Temporal relationship between rainfall, temperature and occurrence of dengue cases in São Luís, Maranhão, Brazil. *Ciencia & saude coletiva*, 21(2), 641–646. <https://doi.org/10.1590/1413-81232015212.09592015>
39. Xu L, Stige LC, Chan KS, Zhou J, Yang J, Sang S et al (2017). Climate variation drives dengue dynamics. *Proceedings of the National Academy of Sciences*, 114(1), 113–118. <https://doi.org/10.1073/pnas.1618558114>
40. Tsuda Y., & Takagi M. (2001). Survival and development of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) larvae under a seasonally changing environment in Nagasaki, Japan. *Environmental Entomology*, 30(5), 855–860. <https://doi.org/10.1603/0046-225X-30.5.855>
41. Valdez, L. D., Sibona, G. J., & Condat, C. A. (2017). Impact of rainfall on *Aedes aegypti* populations. arXiv preprint arXiv:1711.07164.
42. Kesorn K, Ongruk P, Chomposri J, Phumee A, Thavara U, Tawatsin A, et al. (2015) Morbidity Rate Prediction of Dengue Hemorrhagic Fever (DHF) Using the Support Vector Machine and the *Aedes aegypti* Infection Rate in Similar Climates and Geographical Areas. *PLoS ONE* 10(5): e0125049.

43. Watts D. M., Burke D. S., Harrison B. A., Whitmore R. E., & Nisalak A. (1987). Effect of temperature on the vector efficiency of *Aedes aegypti* for dengue 2 virus. *The American journal of tropical medicine and hygiene*, 36(1), 143–152. <https://doi.org/10.4269/ajtmh.1987.36.143> PMID: 3812879
44. Lambrechts L., Paaijmans K. P., Fansiri T., Carrington L. B., Kramer L. D., Thomas M. B., & Scott T. W. (2011). Impact of daily temperature fluctuations on dengue virus transmission by *Aedes aegypti*. *Proceedings of the National Academy of Sciences*, 108(18), 7460–7465. <https://doi.org/10.1073/pnas.1101377108>
45. Luz PM, Mendes BV, Codeço CT, Struchiner CJ, Galvani AP. (2008). Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *The American journal of tropical medicine and hygiene*, 79(6), 933–939. <https://doi.org/10.4269/ajtmh.2008.79.933> PMID: 19052308
46. Bakar AA, Kefli Z, Abdullah S, Sahani M (2011). Predictive models for dengue outbreak using multiple rulebase classifiers. In *Electrical Engineering and Informatics (ICEEI), 2011 International Conference on* (pp. 1-6). IEEE.
47. Lowe R, Bailey T, Stephenson D, Jupp T, Graham R, Coelho CA et al. Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Computers & Geosciences*. 2011; 37(3), 371–381. <https://doi.org/10.1016/j.cageo.2010.01.008>
48. Lowe R, Bailey T, Stephenson D, Jupp T, Graham R, et al. The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statist Med*. 2012; 32: 864–883 <https://doi.org/10.1002/sim.4375>
49. Raclouz V, Ramsey R, Tong S, Hu W. Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. *PLoS Negl Trop Dis*. 2012; 6(5), e1648. <https://doi.org/10.1371/journal.pntd.0001648> PMID: 22629476
50. Baquero O. S., Santana L. M. R., & Chiaravalloti-Neto F. (2018). Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PLoS one*, 13(4), e0195065. <https://doi.org/10.1371/journal.pone.0195065>
51. Guo P., Liu T., Zhang Q., Wang L., Xiao J., Zhang Q., . . . & Ma W. (2017). Developing a dengue forecast model using machine learning: a case study in China. *PLoS neglected tropical diseases*, 11(10), e0005973. <https://doi.org/10.1371/journal.pntd.0005973> PMID: 29036169
52. Rahmawati, D., & Huang, Y. P. (2016, July). Using C-support vector classification to forecast dengue fever epidemics in Taiwan. In *2016 International Conference on System Science and Engineering (ICSSE)* (pp. 1-4). IEEE.
53. Ong J., Liu X., Rajarethinam J., Kok S. Y., Liang S., Tang C. S., . . . & Yap G. (2018). Mapping dengue risk in Singapore using Random Forest. *PLoS neglected tropical diseases*, 12(6), e0006587. <https://doi.org/10.1371/journal.pntd.0006587> PMID: 29912940
54. Gambhir S., Kumar Y., Malik S., Yadav G., & Malik A. (2019). Early Diagnostics Model for Dengue Disease Using Decision Tree-Based Approaches. In *Pre-Screening Systems for Early Disease Prediction, Detection, and Prevention* (pp. 69–87). IGI Global.
55. Teng Y., Bi D., Xie G., Jin Y., Huang Y., Lin B., . . . & Tong Y. (2017). Dynamic forecasting of Zika epidemics using Google Trends. *PLoS One*, 12(1), e0165085. <https://doi.org/10.1371/journal.pone.0165085> PMID: 28060809
56. Wesolowski A, Qureshi T, Boni MF, Sundsøy PR, Johansson MA, Rasheed SB, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. *Proceedings of the National Academy of Sciences*. 2015; 112(38), 11887–11892. <https://doi.org/10.1073/pnas.1504964112>
57. Barmak DH, Dorso CO, Otero M, Solari HG. dengue epidemics and human mobility. *Physical Review E*. 2011; 84(1), 011901. <https://doi.org/10.1103/PhysRevE.84.011901>
58. Cummings DA, Iamsirithawor S, Lessler JT, McDermott A, Prasanthong R, Nisalak A, et al. The impact of the demographic transition on dengue in Thailand: insights from a statistical analysis and mathematical modeling. *PLoS Med*. 2009; 6(9), e1000139. <https://doi.org/10.1371/journal.pmed.1000139> PMID: 19721696
59. Mondini A, Chiaravalloti-Neto F. Spatial correlation of incidence of dengue with socioeconomic, demographic and environmental variables in a Brazilian city. *Science of the Total Environment*. 2008; 393(2), 241–248. <https://doi.org/10.1016/j.scitotenv.2008.01.010> PMID: 18262225
60. Cazelles B, Chavez M, McMichael AJ, Hales S. Nonstationary influence of El Niño on the synchronous dengue epidemics in Thailand. *PLoS Med*. 2005; 2(4), e106. <https://doi.org/10.1371/journal.pmed.0020106>
61. Banu S, Guo Y, Hu W, Dale P, Mackenzie JS, Mengersen K, et al. Impacts of El Niño Southern Oscillation and Indian Ocean Dipole on dengue incidence in Bangladesh. *Scientific report*. 2015; 5.