

Knowledge discovery and dengue forecasting applied in a four-year dataset collected at Natal/RN – Brazil

Ignacio Sanchez-Gendriz (✉ ignacio.gendriz@lais.huol.ufrn.br)

Laboratory of Technological Innovation in Health (LAIS), Hospital Universitário Onofre Lopes, Federal University of Rio Grande do Norte (UFRN)

Gustavo Fountuora

Advanced Nucleus of Technological Innovation (NAVI), Federal Institute of Rio Grande do Norte (IFRN)

Ion de Andrade

Laboratory of Technological Innovation in Health (LAIS), Hospital Universitário Onofre Lopes, Federal University of Rio Grande do Norte (UFRN)

Adrião Duarte Doria Neto

Department of Computer and Automation, Federal University of Rio Grande do Norte (UFRN)

Alessandre de Medeiros Tavares

Municipal Health Department, Zoonoses Control Center, Natal/RN, Brazil

Daniele Barros

Laboratory of Technological Innovation in Health (LAIS), Hospital Universitário Onofre Lopes, Federal University of Rio Grande do Norte (UFRN)

Antonio Higor Freire de Moraes

Advanced Nucleus of Technological Innovation (NAVI), Federal Institute of Rio Grande do Norte (IFRN)

Leonardo J. Galvão-Lima

Laboratory of Technological Innovation in Health (LAIS), Hospital Universitário Onofre Lopes, Federal University of Rio Grande do Norte (UFRN)

Ricardo Alessandro de Medeiros Valentim

Laboratory of Technological Innovation in Health (LAIS), Hospital Universitário Onofre Lopes, Federal University of Rio Grande do Norte (UFRN)

Research Article

Keywords: Ovitrap, Dengue, Deep Learning, LSTM

Posted Date: October 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-955820/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Dengue is recognized as a health problem, it causes significant impacts on health worldwide, affecting millions of people each year. A useful method of dengue vector surveillance is to count *Aedes aegypti* eggs deposited in spatially distributed ovitraps. The present work uses a database collected in 397 ovitraps distributed in the municipality of Natal/RN – Brazil. The number of eggs in each ovitrap was counted weekly, for four years (2016 - 2019) and was analyzed jointly with the dengue incidence in the same period. Our results confirms that dengue incidence seems to be related to socioeconomic status on Natal's municipality. Using a deep learning model, we predict the incidence of new dengue cases based on data obtained from the previous week of dengue or in the number of eggs present in the ovitraps. The analysis shows that ovitrap data allows earlier detection (4-6 weeks) than dengue cases itself (1 week). The results confirm that quantifying *Aedes aegypti* eggs can be valuable for planning health actions.

Background

Dengue is recognized as the most important human disease caused by arbovirus, studies estimate 10 thousands deaths and 100 million symptomatic infections each year in approximately 125 countries [1], [2]. Particularly, Brazil contributed to 55 % of dengue cases reported in the Americas in the last 3 decades [3]. *Aedes aegypti* is the vector for dengue as well as other diseases (yellow fever, chikungunya, and Zika virus) [4]. Monitoring and control of Aedes infestation is a valuable health action to prevent dengue outbreaks. An efficient way for monitoring levels of Aedes is through ovitraps [5][6]. The ovitraps are special holders built to collect mosquito eggs [6]. Counting the eggs deposited in spatially distributed ovitraps can serve as a proxy for levels of Aedes infestation, and allows determining geographic distribution, density and seasonality of the vector [7]. Although ovitrap are not a direct measurement for adult mosquito density, this method represent a good estimator [6][8].

However, studies using ovitrap data for direct prediction of dengue incidence are scarce [8], in this sense a gap can be highlighted for studies that address weekly dengue time series forecasts. The reasons may be based on the multifaceted dynamics of the disease itself, and by the complex relations between mosquito incidence and the risk of human infections. Particularly, have been reported that Dengue Incidence seems to be influenced by the visit to places where contact with infected mosquitos is probable, regardless the distance to subjects' residence to that places [9]. Although the relationships of Dengue Incidence and socioeconomic status have been addressed elsewhere [10][11][12][13], studies for specific cities may shed light into this complex relations.

The main objectives of the present work are twofold. First is to extract understandings of dengue disease from complete four-year data, sampled weekly at Natal/RN – Brazil (Natal) city. Second is to train models that allow dengue forecasting for the Natal city, both by using past samples of dengue cases or previous values of *Aedes aegypti* eggs count (ovitrap data). In the study we opted to use Long Short Term Memory (LSTM), a neural network model that has been used elsewhere [14]–[16] for dengue time series forecasting, outperforming traditional methods such as Random Forest and Lasso Regression [14]. The

novelty of our study related to dengue time series forecasting is to use ovitrap data as predictor in conjunction with LSTM model.

The data analyzed in the study includes spatially distributed ovitrap data and Dengue Incidence reported by neighborhoods. A compilation of the relevant results obtained in the study is highlighted below:

- Was estimated a 1-year seasonality for dengue incidence and vector incidence (quantified by egg depositions) in the analyzed data.
- The time lag estimated between vector and dengue was 4 weeks.
- In Natal city Dengue Incidence shows strong relation with neighborhood socioeconomic status.
- Using dengue cases reported from previous weeks to forecasting dengue incidence for the next week allowed train LSTM models that show encouraging performance (goodness of fit estimated by a correlation coefficient of 0.92).
- Forecasting dengue incidence using ovitrap data as predictor shows performance (goodness of fit estimated by a correlation coefficient of 0.87) similar to using dengue as predictor.
- The advantage of using ovitrap data is the possibility of earlier detection (six to four weeks in advance) of dengue outbreaks than when the number of dengue cases itself is used as a predictor (one week in advance).
- Accumulated values for 1-year duration temporal windows show strong relation between ovitrap data and dengue incidence.

The presented findings highlight the relevance of ovitrap for vector monitoring as well as for planning health actions at city level. The use of Deep Learning models and data mining approaches could contribute substantially with epidemiologist and public health specialist to overcome and manage dengue related problems.

Results

Heatmap representation for Egg Density Index (EDI) and Dengue incidence, grouped by districts. Figure 1 shows in left panels heatmaps for EDI (bottom) and Dengue incidence (top) for neighborhoods grouped by districts. Some observations can be highlighted from the visual inspection of these heatmaps. First, an annual seasonal variation seems to be apparent both for dengue incidence and EDI data. Second, trends of EDI increase appear in antecedence of dengue incidence increases. Third, grouping neighborhoods by districts shows difference between them, for example East and South, are the districts with less dengue incidence, and on the other hand, North and West have higher cases through the studied period.

In the right panels, of Figure 1, two different scatter plots can be observed, right top panel represent Principal Component Analysis (PCA) obtained from socioeconomic variables (total income, population, income by persons, see more details on Methods section). The 1st and 2nd components (PC1 and PC2) explain 100 % of the variance of the data. The left bottom panel is obtained by scatter plot between

accumulated dengue incidence and the 1st component (PC1, 61.82 % of the explained variance) obtained for the PCA projection mentioned above. A Pearson's correlation coefficient computed between PC1 and Dengue Incidence (transformed by log10 operation), shows a significant negative relation between the variables ($r = -0.69$, $p < 0.001$). A plot of the values of income per capita for each neighborhood shows that income almost separate perfectly neighborhoods into two different groups, Group 1: East-South, Group 2: North-West (see Supplementary Figure 1). The exceptions are the low-income neighborhoods 'Mãe Luiza' and 'Alecrim' that regionally belong to the East-South group but have socioeconomic profiles compatible with the North-West group. Supplementary Figure 1 also illustrates a negative relation between Dengue Incidence and Income when analyzed by districts, that is, districts with the lowest income by person have the highest Dengue Incidence.

Seasonality and lag between dengue and ovitrap data, quantified by Discrete Fourier transform and cross correlation. The Discrete Fourier transform (DFT) of the mean values for Dengue Incidence and EDI were computed to estimate the periodicity of both time series respectively, see Supplementary Figure 2. The peak for DFT for both analyzed time series estimated a seasonality of 52 weeks, which coincide with the number of epidemiological weeks for one year, thus results indicate a 1-year periodicity for Dengue Incidence and vector (egg density) at a city level. Also, cross correlation was used to estimate the time lag between mean EDI and mean Dengue Incidence. The time lag estimated by cross correlation resulted in 4 weeks. These results suggest that an increase in EDI precedes dengue increase in around 1 month at city level.

Predicting aggregated dengue incidence for Natal City. The time series for dengue incidence and EDI for all neighborhoods were aggregated, thus resulting in two time series respectively, which could be used as a global indicator for dengue occurrence and Aedes incidence for the Natal city. In this case, were trained LSTM models for forecasting aggregated dengue values for the whole city. As predictor, were used either aggregated dengue values or aggregated EDI values. The models were trained with the following samples of the time series used as a predictor (referencing i the target sample of the dengue time series):

- $i - 1$ previous sample (1 past sample),
- $i - 3, i - 2, i - 1$ previous samples (3:1 past samples),
- $i - 4, i - 3, i - 2$ previous samples (4:2 past samples),
- $i - 5, i - 4, i - 3$ previous samples (5:3 past samples),
- $i - 6, i - 5, i - 4$ previous samples (6:4 past sample).

These models were named by the nomenclature listed below:

- $D_{i-j} \rightarrow D_i$: dengue ($i - j$ past samples) for predicting dengue at the target week i .
- $O_{i-j} \rightarrow D_i$: ovitrap index ($i - j$ past samples) for predicting dengue at the target week i .

Figure 2 illustrates the performance for the trained models for dengue forecasting for the aggregated values.

The error of the models was quantified by RMSE (Root Mean Square Error) and the goodness of fit by correlation coefficient (r) between known values and predicted values. The plot RMSE versus r indicates that the 2 best ranked models were $D \rightarrow D$, when the input consider the $i - 1$ previous sample and $O \rightarrow D$, when the input consider $i - 6$, $i - 5$, $i - 4$ previous samples. Also, can be observed from Figure 2 that the performance of the models $D \rightarrow D$ decrease when older samples are used for prediction. In contrast, this was not the case for $O \rightarrow D$ models, where the best performance was achieved with the older samples of the predictor. It is worth noting that the RMSE for the $O \rightarrow D$ models are lower if compared with the values of the $D \rightarrow D$ models, principally when older samples are used for prediction.

A detailed look for the response of the two models with the best predictions is represented in Figure 3. The predicted response for these models follows closely the true values, which is quantified by a correlation coefficient close to 0.9 (0.92 and 0.89 respectively, $p < 0.001$) and $RMSE < 5$ in both cases.

Analyzing aggregated values broken down by annual periods. Figure 4 illustrates that time lag and cross correlations could change over the years; suggesting that a complex dynamic could trigger underlying links between the vector increase estimated by ovitraps and the incidence of dengue. Not every year in which there was an increase in egg density, there was an increase in dengue cases (2017). However, the other years in which there was an increase in dengue cases were preceded by an increase in the egg count in the ovitraps.

These results encouraged new analyzes, aiming to explore other factors that could be related with dengue incidence on Natal city. Thus, other data such as precipitations and dengue hospitalizations were included in the analysis. Since dengue hospitalizations is recorded with a monthly sample rate, all analyzed time series were accumulated by month periods, the normalized values for Precipitations (V1), EDI (V2), Dengue Incidence (V3) and Dengue Hospitalizations (V4) are illustrated in Figure 5. It is worth noting from visual inspection of this figure that the time series pairs (Dengue Incidence and Hospitalizations) and (Precipitation and EDI) seems to present similar patterns of evolution through the studied period. The similarity between the mentioned time series was estimated by computing the correlation coefficient between samples selected randomly (100 times), for all possible pairwise combinations of the four time series. For more details about similarity estimation mentioned above see Methods section. This similarity estimation between the time series pairs was plotted in bar graph, also included in Figure 5. As expected, the pairs and shows the higher score for similarity, followed by the similarity of the pair (EDI vs Dengue Incidence).

However, if we observe time series plotted in Figure 5 some interesting points can be highlighted. For instance, one might ask if the years with highest accumulate for a given time series correspond with the years with highest accumulate for other time series. It can be noticed that for V3 and V4 this statement it is not fulfilled, the year with the highest accumulates for Dengue Incidence is 2019, since for Dengue Hospitalizations the highest accumulates was reached for 2018 (see Supplementary Figure 3). For expand this analysis further, we compute the accumulated values for 1 year sliding temporal windows, by using a one-month step for all time series represented in Figure 5. As results were obtained another four

time series, which were plotted in Figure 6. By using the accumulated values for 1 year sliding windows, were also estimated the similarity between all pairwise combinations, and included as a bar graph that also appears in Figure 6. The analysis based on the yearly accumulated values presented in Figure 6 indicates that the pair (EDI vs Dengue Incidence) have the highest similarity, followed by pairs (Dengue Incidence vs Dengue Hospitalizations) and (EDI vs Dengue Hospitalizations). Another interesting behavior of the time series in Figure 6 is that increase in Precipitation precedes increase in EDI, which precedes Dengue Incidence, which finally seems to be aligned with Dengue Hospitalizations.

Discussion

Seasonality and time lag. The 1-year seasonality detected from a visual inspection of heatmaps represented in Figure 1 was confirmed by computing DFT of mean values of Dengue Incidence and EDI. This type of periodicity in dengue occurrence and *Aedes aegypti* population have been reported elsewhere [17]–[20]. However, it is worth noting that yearly periodicities for dengue cases do not necessarily implies that all years have the same level of incidence. For instance, for 2017 the levels of dengue incidence for all neighborhoods of Natal city were considerably lower if compared with the other recorded years. This fact could be related to several aspects such as the presence of susceptible people for the main circulating serotype [21], since Dengue is caused by 4 different virus serotypes [22]. Other aspects such as the complex interaction between environmental drivers and the 4 dengue serotypes could occur [23]. Also, the low incidence of dengue cases for 2017 was reported lower than expected for Brazil and Colombia [24], this could be related to previous human population infection with zika virus in that regions.

The cross correlation estimates 1 month of time lag between EDI and Dengue Incidence, which is consistent with the expected elapsed time from mosquito eggs depositions to adult phase and finally virus transmission to humans, similar results have been reported elsewhere [8]. The high correlation and the possibility of anticipation of the severity of the epidemic with a time lapse of four weeks makes ovitrap monitoring extraordinarily important for the timely adoption of contingency measures against dengue. This contributes to the early detection of the epidemic, thus facilitating its controllability.

At the same time, the significant correlation found at the city level, it is not necessarily expected at the local neighborhood level [8]. This suggests that dengue is a disease of eminently municipal scope, a fact that points to the need for the confrontation of dengue to have a systemic character in the territorial space in which a given community lives, and not only at the local level where housing is located.

Dengue and socioeconomic status in the city of Natal. The heatmaps represented in Figure 1 illustrates that neighborhoods from North and West districts presents the highest dengue occurrence, and neighborhoods from East and South account the lowest dengue incidence. The city of Natal is stratified into neighborhoods with a markedly differentiated socioeconomic status, which can be observed for Supplementary Figure 1. Also, it can be pointed from that figure that poor districts have higher incidence of dengue. This results agree with previous studies suggesting that Dengue Incidence are associated with

lower socioeconomic status [11]. Also studies report that poverty conditions could be related to factors that increase the risk of human exposure to *Aedes aegypti* [20].

Also, the visualization based on heatmaps such as those presented in Figure 1 shows utility for gaining insights for the dynamics and evolution of vector and dengue incidence by localities (in our case neighborhoods) through a long-term recorded period.

Performance of models trained for predicting at aggregated values. Deep learning (DL) has been applied in several areas of research in the last decade, with extraordinary results. Here we applied LSTM models. Models trained using dengue as predictor for predicting values for the next week ($D \rightarrow D$) obtain best results among all tested models. However, it is relevant to note that using ovitrap data for predicting aggregated Dengue Incidence ($O \rightarrow D$) shows similar performance to $D \rightarrow D$. This suggests the relevance of egg monitoring at global scale of the city. The performance obtained for aggregated values remarks the usefulness of ovitraps for planning health actions at city level. One can note that although $D \rightarrow D$ shows better performance, $O \rightarrow D$ shows comparable results, but with the advantage of anticipation. By using EDI time series as predictor, the best performance was obtained for six to four weeks before the target week for dengue incidence. These facts highlight the importance of ovitrap monitoring for early epidemic risk detection and, therefore, point to the possibility of delineating health actions to prevent dengue outbreaks.

The dynamics of the spread of dengue fever, contrary to the traditional way of being perceived as underlying local and peri-domiciliary conditions of infection, has shown to be a result of municipal dynamics, probably produced by urban mobility of people, but also of the infected vector itself. This urban mobility is crucial for the set of existing requirements for the outbreak of a dengue epidemic to be set in motion [25][26].

The local variables of vector infestation and increased presence of eggs in the ovitraps, should be understood, therefore, not as local triggers, but as municipal triggers, which can put into dialogue regions with large numbers of vectors and little viral circulation with others where viral circulation is already established even though there is low infestation of mosquitoes.

The connection between this set of conditions for the outbreak of the epidemic that are generated locally, will be completed by the exercise of urban life, characterized by the existence of a community whose exercise of life and work takes place without territorial limits defined within that municipality and expressed by urban mobility. The local infestation interests therefore the city as a whole, in a complex dynamic whose exercise brings together links in the epidemiological chain converting the local risk into a municipal risk.

Accumulated values for 1 year sliding windows. Comparing the time series for accumulated values obtained for 1 year sliding windows provide two interesting results, that will be discussed next. First, when analyzed the time series Precipitation (V1), EDI (V2), Dengue Incidence (V3) and Dengue Hospitalizations (V4) the higher correlations were obtained from pair V_{12} and pair V_{34} (see Figure 5). These relations are consistent with expectations since Precipitation creates favorable conditions for

Aedes aegypti reproduction and higher Dengue Incidence favors probabilities of Dengue Hospitalizations. Nevertheless, when analyzed the respective time series for the accumulated values for 1-year windows, a subtle relation appears. That is, the strongest relation appear between EDI and Dengue Incidence for yearly accumulates. This suggest that higher accumulates of *Aedes aegypti* eggs over 1-year periods present strong relation with Dengue Incidence, which once again point egg monitoring as relevant variable, but this time from a long-term perspective.

Second, Figure 6 also suggest that increase in yearly accumulated Precipitation precedes increase in accumulated egg depositions (measured by EDI), which precedes Dengue Incidence and Hospitalizations.

Overall considerations

Concluding remarks. Here we explore four year dataset composed by *Aedes aegypti* eggs counted for 397 spatially distributed ovitraps in Natal city, sampled weekly. Also was analyzed the dengue incidence reported for Natal's neighborhoods. Dengue incidence of the neighborhoods of Natal city show positive relation with socioeconomic indicators of poverty. Annual trends were quantified for vector and dengue incidence, and a time lag of 4 weeks was estimated between these variables. Early detection of dengue outbreaks may be possible based on ovitrap data, four to six weeks in advance. Accumulated values for annual temporal windows evidence strong correlation between *Aedes aegypti* egg depositions and dengue incidence for the Natal city. Our work shows the importance of continuous recording of dengue incidence for long periods and signifies the relevance of ovitrap monitoring.

Future work. Further studies directed to dengue prediction should include human mobility data as a predictor and circulating serotypes. In addition, upcoming studies could apply methods for exploring causal relations between vector proliferation and dengue incidence. Finally, subsequent studies may focus on vector incidence forecasting, and use the prediction for supporting and planning actions for controlling *Aedes aegypti* vector proliferation.

Methods

Database description. Dengue cases incidence registered in each neighborhood of Natal city, sampled weekly (52 epidemiological weeks a year) between 2016 – 2019 was used as target for forecasting. The source for dengue data was the Notifiable Diseases Information System (SINAN, according to the acronym in Portuguese). Also, ovitrap egg counts for *Aedes aegypti*, collected every week for 397 ovitraps and reported by the Zoonoses' Center of Natal municipality were used in the study. See Supplementary Figure 5 for geographical distribution of ovitraps at Natal municipality.

Ovitrap indexes. Ovitrap Positivity Index (OPI) and Egg Density Index (EDI) are entomological indices commonly used for *Aedes aegypti* monitoring. The OPI is defined as the ratio between the number of traps with at least one egg and the total units installed and successfully retrieved, the EDI is calculated as the ratio between the number of eggs totalized for a given area divided by the quantity of ovitraps

respective to that area [6]. In the present study we use EDI computed by neighborhoods, since this index have higher discretization than OPI, which could be helpful for dengue incidence forecasting purposes.

Heatmap visualizations and scatter plots for socioeconomic variables. Heat maps constructed based on EDI and dengue occurrence were used for visualizing the variation and dynamics of vector incidence and dengue cases (Figure 1) through the monitored period. For the heatmaps, rows indicate neighborhoods, columns indicate weeks, and the colors indicate relative values for respective variables visualized. Ovitrap indexes and dengue occurrence visualized as colored heat maps allowed to obtain an easily interpretable image for gaining insights on possible relations between variations in the number of Aedes eggs and dengue outbreaks. For the PCA presented in the right top panel of Figure 1 were used the variables population, total income by neighborhood and income by person, all variables were obtained from the last census performed in Brazil (2010).

Models for Dengue incidence forecasting. Several traditional models have been used in the literature for dengue time series forecasting, for example Artificial Neural Networks, Random Forest, Lasso Regression, Generalized Additive Models and Autoregressive Models [14], [27]–[29]. Recent advances in Deep Learning (DL) methods have shown remarkable performance of these algorithms in different areas of applications [30], been Convolutional Neural Networks (CNNs) very popular for image analysis and computer vision [31] and, Long-Short Term Memory (LSTM) for sequence and time series analysis [32]. Particularly LSTM is an excellent candidate model for dengue forecasting, which have been used for this task outperforming traditional machine learning methods [14]. In this section we briefly explain LSTM fundamentals and describe how the models were configured for dengue forecasting.

Long Short-Term Memory Networks (LSTMs), are a special type of Recurrent Neural Networks (RNN), which can learn long-term dependencies while dealing with the problem of vanishing/exploding gradient [32], [33]. LSTM are used in several applications related to sequential data, such as Natural Language Processing, Time series prediction, computer vision among others [32]. The architecture for all the LSTM networks trained in the study were the same, and was composed by 1) an input layer, 2) an LSTM layer with 100 hidden units, 3) a fully connected layer and 4) a regression layer. The networks were trained for 250 epochs, using a mean squared error (MSE) loss function and a Nesterov Adam optimizer, similar to LSTM model trained in [14].

LSTM for forecasting dengue aggregated values for the Natal city.

Aggregated values for dengue incidence and ovitrap data were used for training LSTM models, the target for forecasting will be either dengue incidence or EDI for the next week. The input to the models will be represented as D, when the predictor is dengue incidence or O when the input to the model is ovitrap index, that is EDI. As the target for prediction were used dengue incidence (D) or ovitrap index – EDI (represented by O). The models were trained with either the last sample for the predictor, or with the last 3 samples for the predictor. The nomenclature $D \rightarrow D$, means that the models was trained with dengue incidence as predictor (past samples) and dengue incidence (next week) as target. Other possible

combinations are $D \rightarrow O$, $O \rightarrow D$ and $O \rightarrow O$. Models trained for ovitrap data forecasting were also evaluated for complementing the discussion described in following sections.

Computing Accumulated values for 1 year sliding windows. Based on the times series of Precipitation, EDI, Dengue Incidence and Dengue Hospitalizations for monthly samples (Figure 5) were computed the Accumulated values for 1 year sliding windows, which are illustrated in Figure 6. The procedure used to compute accumulated values for 1 year sliding windows is detailed in Supplementary Figure 4. Basically, for a given time series , all samples for a 1-year length widows are summed up, obtaining accumulated values respective to that 1-year window. By sliding the 1-year window, and then computing the respective accumulated values was obtained a new time series , as presented in the next equation:

$$y(j) = \sum_{i=j}^{j+11} x(i), j = 0, 1, \dots, 36$$

To estimate the relationship between the time series analyzed, either for series of monthly samples or for accumulated 1-year window series, Spearman correlation coefficient [34] was computed between all pairwise time series combinations. The relation between the mentioned time series was estimated by computing the correlation coefficient between 25 samples selected randomly (100 times), for all possible pairwise combinations of the four time series.

Declarations

Acknowledgements

The authors thank the professionals from the Zoonoses Center of Natal for all relevant discussions and contributions to the research.

Author contributions

R.A.M.V and A.M.T. idealized the research. I.S.G., and G.F. analyzed the data; I.S.G. and A.D.D.N. fit the models; I.S.G. and I.A. wrote the original draft; all authors interpreted and discuss the results; all authors contributed written feedback. All authors wrote the manuscript. R.A.M.V, A.H.F.M and D.B. supervised the project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper

References

- [1] J. P. Messina *et al.*, “The current and future global distribution and population at risk of dengue,” *Nat. Microbiol.*, vol. 4, no. 9, pp. 1508–1515, 2019, doi: 10.1038/s41564-019-0476-8.
- [2] J. P. Messina, O. J. Brady, D. M. Pigott, J. S. Brownstein, A. G. Hoen, and S. I. Hay, “A global compendium of human dengue virus occurrence,” *Sci. Data*, vol. 1, pp. 1–6, 2014, doi: 10.1038/sdata.2014.4.
- [3] P. C. G. Nunes, R. P. Daumas, J. C. Sánchez-Arcila, R. M. R. Nogueira, M. A. P. Horta, and F. B. dos Santos, “30 years of fatal dengue cases in Brazil: a review,” *BMC Public Health*, vol. 19, no. 1, p. 329, Dec. 2019, doi: 10.1186/s12889-019-6641-4.
- [4] P. M. de A. Zanotto and L. C. de C. Leite, “The Challenges Imposed by Dengue, Zika, and Chikungunya to Brazil,” *Front. Immunol.*, vol. 9, no. August, p. 1964, Aug. 2018, doi: 10.3389/fimmu.2018.01964.
- [5] J. V. Oliveira Noletto *et al.*, “Use of ovitraps for the seasonal and spatial monitoring of *Aedes* spp. in an area endemic for arboviruses in Northeast Brazil,” *J. Infect. Dev. Ctries.*, vol. 14, no. 04, pp. 387–393, Apr. 2020, doi: 10.3855/jidc.12245.
- [6] C. T. Codeço *et al.*, “Surveillance of *Aedes aegypti*: Comparison of House Index with Four Alternative Traps,” *PLoS Negl. Trop. Dis.*, vol. 9, no. 2, pp. 1–23, 2015, doi: 10.1371/journal.pntd.0003475.
- [7] E. Barreto, M. C. Resende, A. E. Eiras, and P. C. Demarco Júnior, “Evaluation of the baited ovitrap with natural attractant for monitoring *Aedes* spp. in Dili, capital of East Timor,” *Cien. Saude Colet.*, vol. 25, no. 2, pp. 665–672, Feb. 2020, doi: 10.1590/1413-81232020252.12512018.
- [8] D. Albuquerque, R. C. Pinto, M. Sadahiro, and B. P. Lima, “Relationship between local presence and density of *Aedes aegypti* eggs with dengue cases: a spatial analysis approach,” vol. 23, no. 11, pp. 1269–1279, 2018, doi: 10.1111/tmi.13150.
- [9] S. T. Stoddard *et al.*, “House-to-house human movement drives dengue virus transmission,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 3, pp. 994–999, Jan. 2013, doi: 10.1073/pnas.1213349110.
- [10] S. Ryan *et al.*, “Socio-Ecological Factors Associated with Dengue Risk and *Aedes aegypti* Presence in the Galápagos Islands, Ecuador,” *Int. J. Environ. Res. Public Health*, vol. 16, no. 5, p. 682, Feb. 2019, doi: 10.3390/ijerph16050682.

- [11] R. M. Zellweger *et al.*, "Socioeconomic and environmental determinants of dengue transmission in an urban setting: An ecological study in Nouméa, New Caledonia," *PLoS Negl. Trop. Dis.*, vol. 11, no. 4, p. e0005471, Apr. 2017, doi: 10.1371/journal.pntd.0005471.
- [12] E. C. Farinelli, O. S. Baquero, C. Stephan, and F. Chiaravalloti-Neto, "Low socioeconomic condition and the risk of dengue fever: A direct relationship," *Acta Trop.*, vol. 180, pp. 47–57, Apr. 2018, doi: 10.1016/j.actatropica.2018.01.005.
- [13] M. Kikuti *et al.*, "Spatial Distribution of Dengue in a Brazilian Urban Slum Setting: Role of Socioeconomic Gradient in Disease Risk," *PLoS Negl. Trop. Dis.*, vol. 9, no. 7, p. e0003937, Jul. 2015, doi: 10.1371/journal.pntd.0003937.
- [14] E. Mussumeci and F. Codeço Coelho, "Large-scale multivariate forecasting models for Dengue - LSTM versus random forest regression," *Spat. Spatiotemporal. Epidemiol.*, vol. 35, p. 100372, Nov. 2020, doi: 10.1016/j.sste.2020.100372.
- [15] J. Xu *et al.*, "Forecast of Dengue Cases in 20 Chinese Cities Based on the Deep Learning Method," *Int. J. Environ. Res. Public Health*, vol. 17, no. 2, p. 453, Jan. 2020, doi: 10.3390/ijerph17020453.
- [16] A. Doni and T. Sasipraba, "LSTM-RNN Based Approach for Prediction of Dengue Cases in India," *Ingénierie des systèmes d'Inf.*, vol. 25, no. 3, pp. 327–335, Jun. 2020, doi: 10.18280/isi.250306.
- [17] M. Churakov, C. J. Villabona-Arenas, M. U. G. Kraemer, H. Salje, and S. Cauchemez, "Spatio-temporal dynamics of dengue in Brazil: Seasonal travelling waves and determinants of regional synchrony," *PLoS Negl. Trop. Dis.*, vol. 13, no. 4, p. e0007012, Apr. 2019, doi: 10.1371/journal.pntd.0007012.
- [18] E. Schwartz *et al.*, "Seasonality, Annual Trends, and Characteristics of Dengue among Ill Returned Travelers, 1997–2006," *Emerg. Infect. Dis.*, vol. 14, no. 7, pp. 1081–1088, Jul. 2008, doi: 10.3201/eid1407.071412.
- [19] M. G. Guzmán and G. Kouri, "Dengue: an update," *Lancet Infect. Dis.*, vol. 2, no. 1, pp. 33–42, Jan. 2002, doi: 10.1016/S1473-3099(01)00171-2.
- [20] A. J. Monaghan *et al.*, "On the Seasonal Occurrence and Abundance of the Zika Virus Vector Mosquito *Aedes Aegypti* in the Contiguous United States," *PLoS Curr.*, 2016, doi: 10.1371/currents.outbreaks.50dfc7f46798675fc63e7d7da563da76.
- [21] W. P. Vargas, H. Kawa, P. C. Sabroza, V. B. Soares, N. A. Honório, and A. S. de Almeida, "Association among house infestation index, dengue incidence, and sociodemographic indicators: surveillance using geographic information system," *BMC Public Health*, vol. 15, no. 1, p. 746, Dec. 2015, doi: 10.1186/s12889-015-2097-3.

- [22] J. P. Hanley *et al.*, “Immunotranscriptomic profiling the acute and clearance phases of a human challenge dengue virus serotype 2 infection model,” *Nat. Commun.*, vol. 12, no. 1, pp. 1–14, 2021, doi: 10.1038/s41467-021-22930-6.
- [23] Q. A. ten Bosch, B. K. Singh, M. R. A. Hassan, D. D. Chadee, and E. Michael, “The Role of Serotype Interactions and Seasonality in Dengue Model Selection and Control: Insights from a Pattern Matching Approach,” *PLoS Negl. Trop. Dis.*, vol. 10, no. 5, p. e0004680, May 2016, doi: 10.1371/journal.pntd.0004680.
- [24] R. K. Borchering *et al.*, “Impacts of Zika emergence in Latin America on endemic dengue transmission,” *Nat. Commun.*, vol. 10, no. 1, p. 5730, Dec. 2019, doi: 10.1038/s41467-019-13628-x.
- [25] A. Wesolowski *et al.*, “Impact of human mobility on the emergence of dengue epidemics in Pakistan,” *Proc. Natl. Acad. Sci.*, vol. 112, no. 38, pp. 11887–11892, Sep. 2015, doi: 10.1073/pnas.1504964112.
- [26] R. Bomfim *et al.*, “Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas,” *J. R. Soc. Interface*, vol. 17, no. 171, p. 20200691, Oct. 2020, doi: 10.1098/rsif.2020.0691.
- [27] O. S. Baquero, L. M. R. Santana, and F. Chiaravalloti-Neto, “Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models,” *PLoS One*, vol. 13, no. 4, pp. 1–12, 2018, doi: 10.1371/journal.pone.0195065.
- [28] H. M. Aburas, B. G. Cetiner, and M. Sari, “Dengue confirmed-cases prediction: A neural network model,” *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4256–4260, Jun. 2010, doi: 10.1016/j.eswa.2009.11.077.
- [29] N. Zhao *et al.*, “Machine learning and dengue forecasting: Comparing random forests and artificial neural networks for predicting dengue burden at national and sub-national scales in Colombia,” *PLoS Negl. Trop. Dis.*, vol. 14, no. 9, pp. 1–16, 2020, doi: 10.1371/journal.pntd.0008056.
- [30] Ian Goodfellow, Yoshua Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: The MIT Press, 2016.
- [31] A. Dhillon and G. K. Verma, “Convolutional neural network: a review of models, methodologies and applications to object detection,” *Prog. Artif. Intell.*, vol. 9, no. 2, pp. 85–112, Jun. 2020, doi: 10.1007/s13748-019-00203-0.
- [32] G. Van Houdt, C. Mosquera, and G. Nápoles, “A review on the long short-term memory model,” *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, doi: 10.1007/s10462-020-09838-1.
- [33] K. Smagulova and A. P. James, “Overview of Long Short-Term Memory Neural Networks,” in *Modeling and Optimization in Science and Technologies*, vol. 14, Springer International Publishing, 2020, pp. 139–153.

Figures

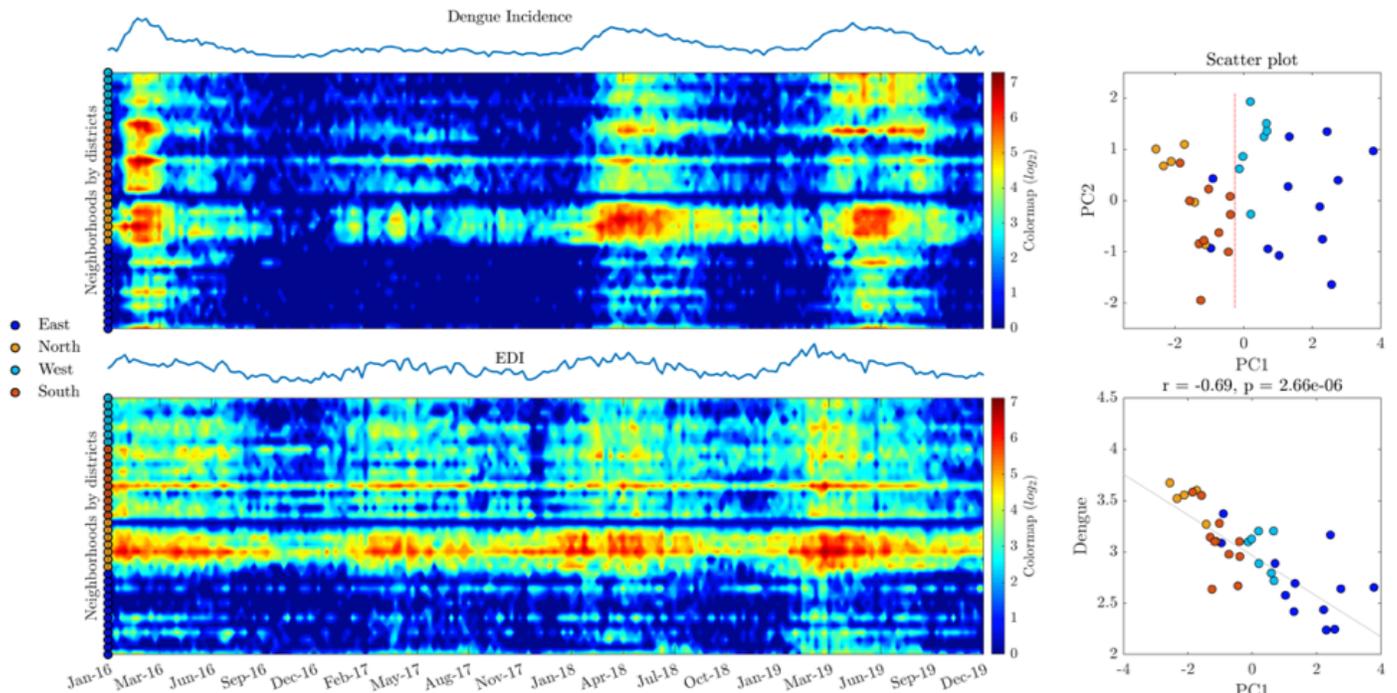


Figure 1

Evolution of dengue incidence and EDI and the relation of accumulated dengue incidence by neighborhoods and socioeconomic variables. a) Representation of heatmaps for EDI and dengue incidence, neighborhoods grouped by districts. b) Top panel: scatter plot of PC1 vs PC2 for PCA of socioeconomic variables; bottom panel: PC1 vs accumulated Dengue Incidence (log2 transformed) by neighborhoods.

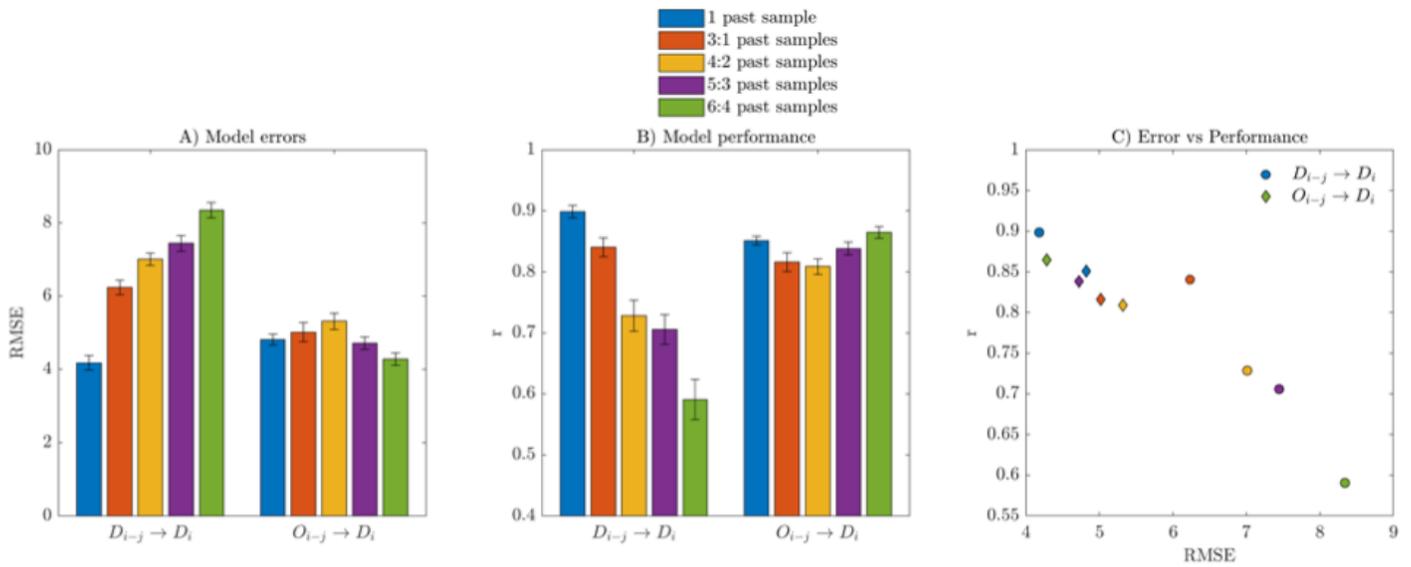


Figure 2

Evaluation for LSTM models performance for dengue forecasting based on aggregated time series, the models were trained and tested 30 repeated times. Bars indicate mean values and whisker indicate standard error, both for RMSE and r metrics.

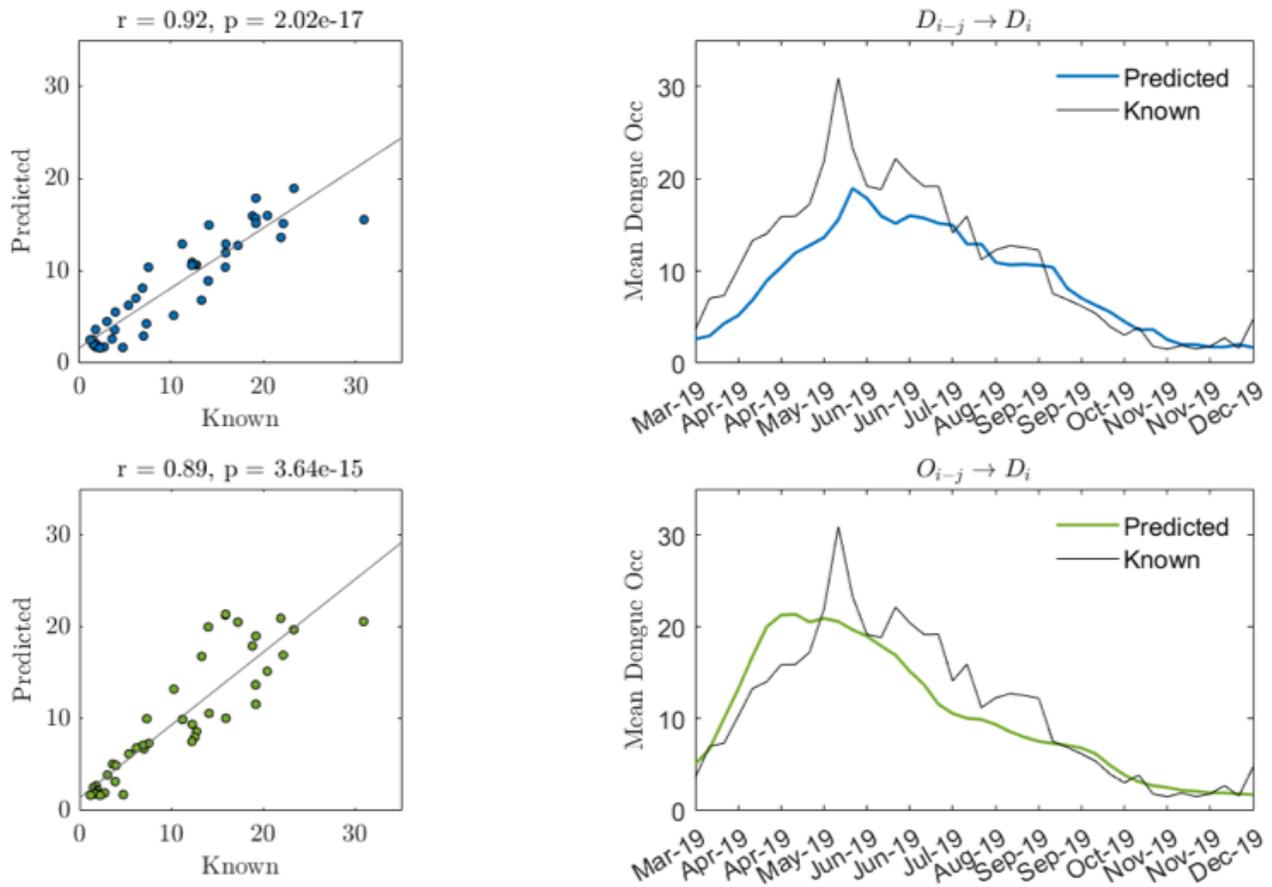


Figure 3

Predictions of the LSTM models trained and tested 30 repeated times. Left panels illustrate scatter plots Known vs Predicted values. Right panels illustrate the time series of the mean of the predicted values for the 30 repetitions and the known values for a test set samples.

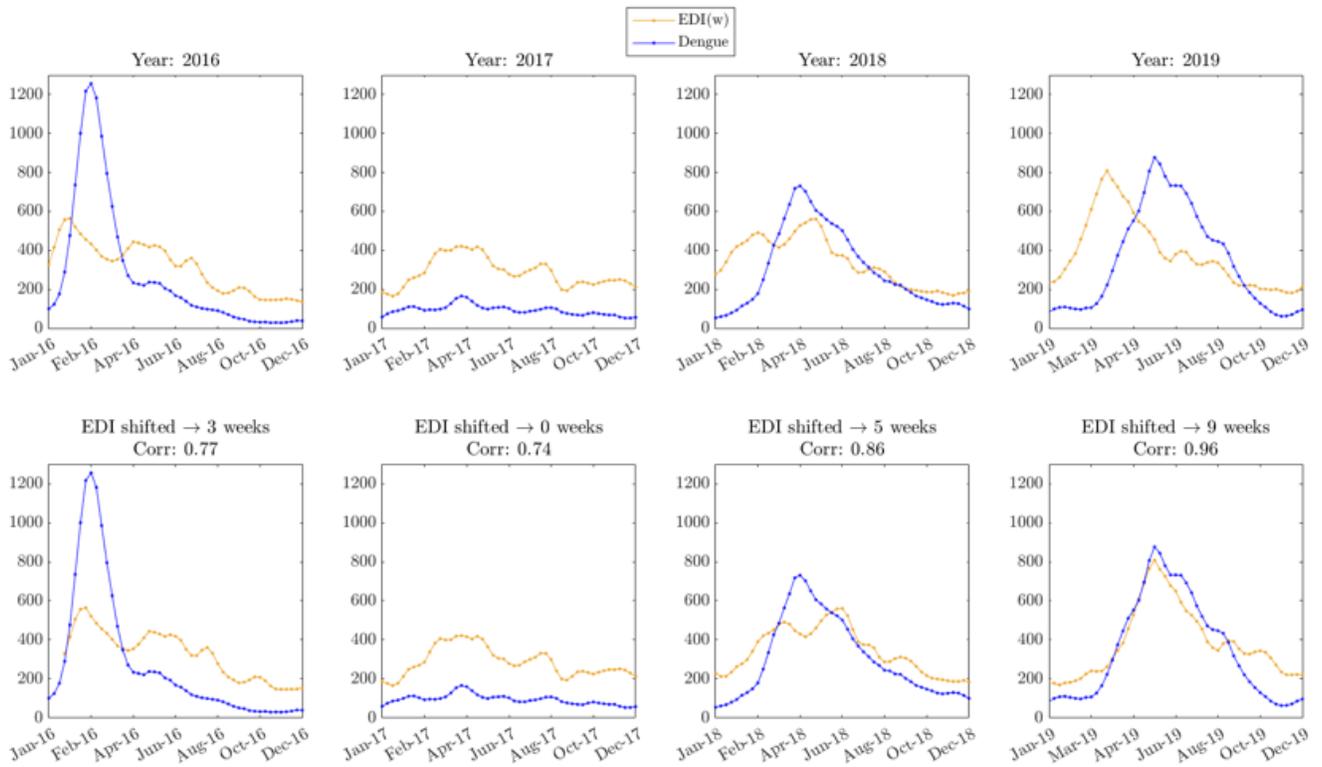


Figure 4

Dengue and EDI relations by studied years.

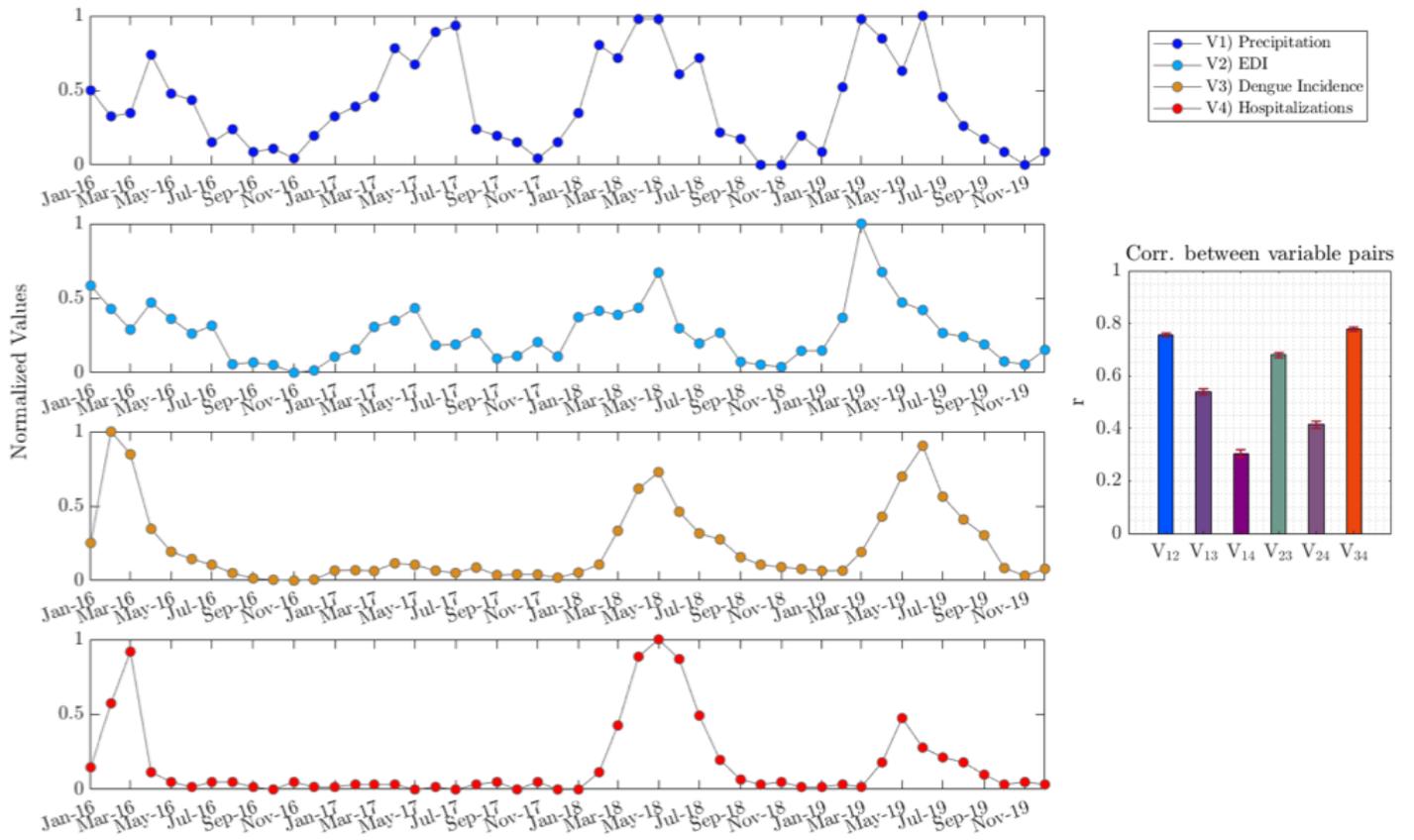


Figure 5

Time series by monthly samples.

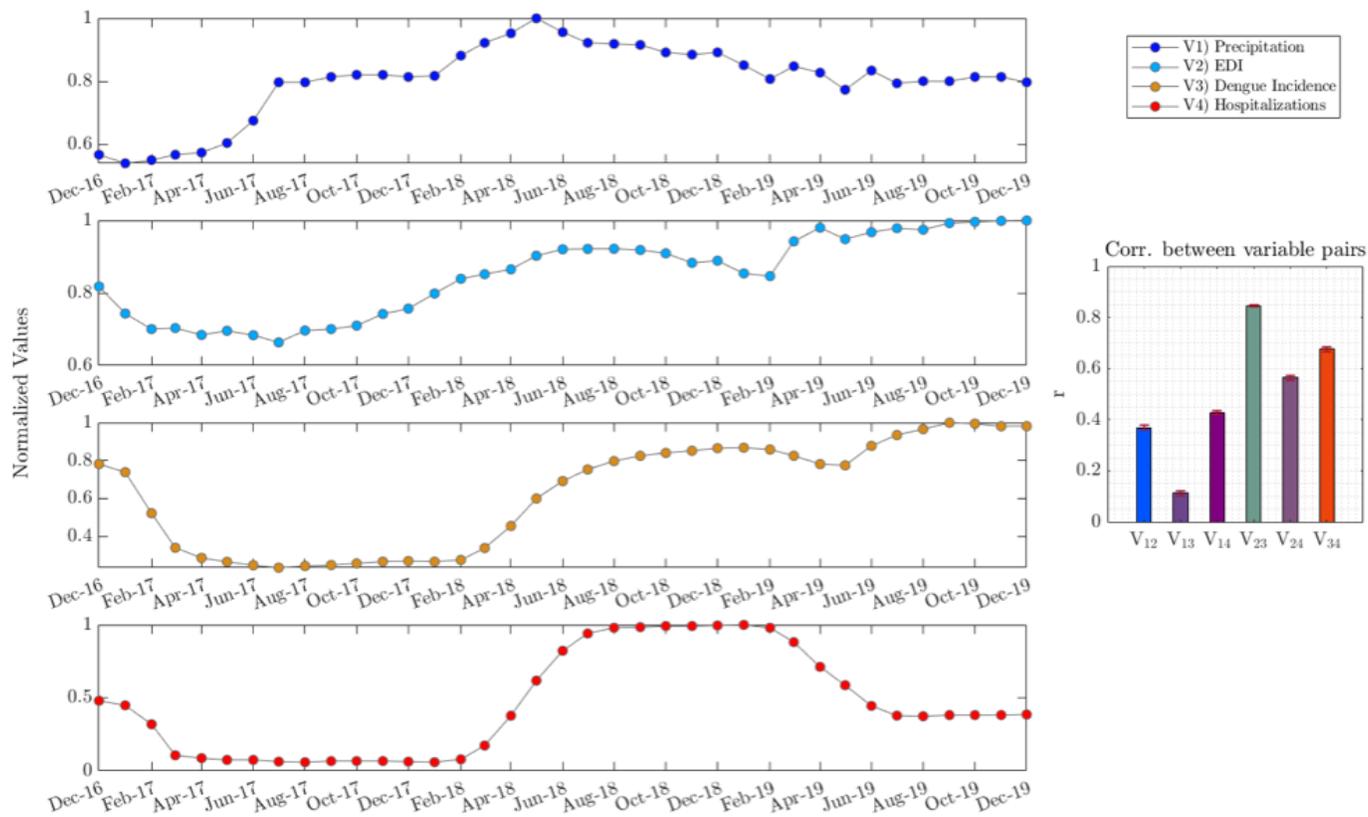


Figure 6

Correlations of accumulated values for 1 Yr sliding windows.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SuppResearchSquare202101.docx](#)