



OPEN

Correlations between COVID-19 and dengue obtained via the study of South America, Africa and Southeast Asia during the 2020s

Paula Bergero¹, Laura P. Schaposnik^{2,3}✉ & Grace Wang⁴

A dramatic increase in the number of outbreaks of dengue has recently been reported, and climate change is likely to extend the geographical spread of the disease. In this context, this paper shows how a neural network approach can incorporate dengue and COVID-19 data as well as external factors (such as social behaviour or climate variables), to develop predictive models that could improve our knowledge and provide useful tools for health policy makers. Through the use of neural networks with different social and natural parameters, in this paper we define a *Correlation Model* through which we show that the number of cases of COVID-19 and dengue have very similar trends. We then illustrate the relevance of our model by extending it to a Long short-term memory model (LSTM) that incorporates both diseases, and using this to estimate dengue infections via COVID-19 data in countries that lack sufficient dengue data.

Since its emergence, COVID-19 has been studied in correlation with many other infectious diseases, most notably influenza, meningitis, and measles¹. Caused by a coronavirus known as SARS-CoV-2, COVID-19 has become a worldwide pandemic that has, as of April of 2022, over 500 million confirmed cases and over six million deaths have been reported globally by WHO. Although a large amount of research has been done both on dengue as well as on COVID-19, not much work has been done on their correlation, and thus this shall be the focus of the present paper. In particular, building on², we study such correlation worldwide using machine learning techniques, and explore different directions in which the results we deduce could be used by policy makers in the future, especially since the syndemic and possible simultaneous transmission presents a significant challenge to public health efforts in those areas³.

Prominent in Latin American and Asian countries, dengue is a viral disease that is transmitted by mosquitoes, especially those of species *Aedes aegypti* (Aa)⁴. Dengue is a highly seasonal, multi-annual disease, most prevalent before and after rainy seasons. Individuals who recover from dengue have long-term immunity against that specific serotype (homotypic), and they also have short-term immunity against dengue of a different serotype (heterotypic)⁵. Moreover, dengue is widely considered to be the most important mosquito-borne viral disease and is very wide spread, covering mostly tropical and sub-tropical areas, between the January isotherm and the July isotherm of 10 degrees Celsius. Temperature, rainfall, and population density were all factors shown to be associated with the number of dengue infections, which explains the regions where it is most prevalent⁶.

Recently, infections of dengue and COVID-19 have begun to be considered together, particularly in South America (e.g.²). Despite the year 2020 being an epidemiological complex year with the pandemic of COVID-19, the fatality rate from dengue was recorded as 0.04%, the lowest in the past decade⁴. Interestingly, however, there seems to be a persistence of higher than expected number of dengue cases in endemic areas, occurring simultaneously with intense COVID transmission. Although it has been brought up in September of 2021 that there may be cross-immunity between the two diseases⁷, a population-based cohort study in December of 2021 found that individuals with prior dengue actually had an increased risk of clinically apparent COVID-19⁸. Thus, there is likely not any cross-immunity between the two diseases.

¹Universidad Nacional de La Plata, CCT, La Plata, Argentina. ²University of Illinois, Chicago, USA. ³All Souls College, Oxford, UK. ⁴PRIMES-MIT, Boston, USA. ✉email: schapos@uic.edu

We shall organize our paper in the following way: after introducing some background in section “[Background](#)”, we define different external variables that we then use in a neural network model for Brazil in section “[A correlation model](#)”. In particular, we consider holidays in section “[Impact of holidays](#)” and climate factors such as temperature and humidity in section “[Impact of humidity and temperature](#)”. By adding in these factors, we

- are able to **create a model that predicts and correlates COVID-19 to**
 - **holidays; climate, and dengue.**

After introducing and describing our novel model, we show its utility by applying it to other nations or regions in section “[The efficacy of the correlation model](#)”. In particular, we use our model to study the data from Peru and Colombia and

- show the **model’s predictive qualities** by using COVID-19 data to predict dengue in countries that lack sufficient dengue data.

Other factors such as Latent Heat Flux and peaks correlations are also considered in section “[Conclusion and applications](#)”. Finally, since our Machine Learning model focused solely on external variables, we take a more time-series based approach and

- introduce a **Long short-term memory model (LSTM)** in section “[Recurrent neural network and time series analysis](#)”, obtaining a more comprehensive look at predicting future values for both diseases than does the standard neural network.

The results and observations drawn from our work are summarized in section “[Final remarks](#)”, where future directions of research are also outlined. The datasets used during the current study are available in the repository⁹.

Background

Machine Learning has been incredibly useful in countless topics, and COVID-19 is no exception. Ever since the beginning of the pandemic, researchers have been using machine learning to predict future outbreaks or possible correlations with other factors. In the present paper we are specifically looking at the correlation of COVID-19 with dengue. Hence, we shall dedicate this section to introducing the background needed in Machine Learning, neural networks and LSTM models.

Machine learning and COVID-19. Because Machine Learning is such a powerful tool in making predictions and finding correlations, researchers have been using it in studies about COVID-19 since the beginning of the pandemic. As early as April of 2020, in fact, researchers were finding that specific models, such as MLP (Multi-layered perceptron) and ANFIS (adaptive network-based fuzzy interference system)—both artificial neural networks—had promising results for estimating the severity of the pandemic. In the beginning of 2020, researchers also tested several models to estimate time-series data, finding that whilst the logistic model outperformed the others, these fitted models were hard to generalize after a 30-day period¹⁰. More recently, Machine Learning begun to be used to classify different types of COVID-19 since it can also test different aspects, such as the severity of someone’s COVID-19 symptoms. In particular, researchers have suggested a hybrid machine learning/deep learning model that predicts COVID-19 severity from CT images¹¹.

Our work on the correlation between dengue and COVID-19 will be done firstly by using a neural network to study data that was collected of COVID-19 and dengue. By incorporating parameters that are possibly related to these two infectious diseases such as holiday seasons, temperature, or rainfall, we are able to draw strong correlations that then serve to build a predictive model. These models then shall generate loss curves and prediction graphs to find out what factors are the most prominent in these disease cases. A standard neural network has input layers, output layers, and one or more hidden layers as shown in Fig. 1.

The learning structure used here has one input layer, four output layers using the relu activation function, and an output layer using the linear activation function. The predicted infection number is given by the output node with the greatest activation. The neural network was trained using a training set of size 10 epochs and testing using a cross-validations set of size approximately 150 epochs. To understand the model we consider its loss function using the Mean Absolute Error function in section “[A correlation model](#)”, which was calculated by averaging the absolute distances between the predicted and actual values. Data for dengue fever has been collected from the [Pan-American Health Organization Dataset on Dengue](#), and data for COVID-19 has been collected from the World Health Organization [dashboard on COVID-19](#).

LSTM models. The correlation between dengue and COVID-19 will also be studied through a Recurrent Neural Network (RNN), in an effort to predict future values of the number of infections based on past observed data. Specifically, we shall consider a LSTM model (Long short-term memory), which is particularly useful especially in cases where there may be time lags, and which is often used in analyzing stock prices and the like. In the setting of our paper, we use the LSTM model to generate a model for each disease based solely on its past observed values and then we incorporate the time series values for the other disease as well as other external factors.

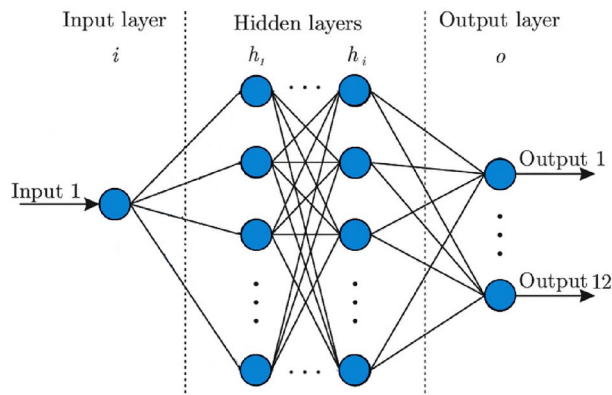


Figure 1. An example of a neural network with i hidden layers, 1 input layer and 12 output layers.

The LSTM model is more complicated than a standard neural network in that it can process sequences of data and not solely datapoints. Because these recurrent types of neural networks have “loops” (hence, recurrent) around them, the information is able to persist, something that standard neural networks cannot necessarily do. Essentially, a LSTM model contains a cell, input gate, and output gate as shown in Fig. 2. But it also consists of a forget gate, something that traditional neural networks do not have. The input gate inserts new information into the cell, and the output gate passes along the new, updated information. The forget gate, on the other hand, forgets useless information, which is especially important because LSTM is used for long-term processing. In our LSTM model, we have an input layer, a single hidden layer, and an output layer.

Correlations via datasets

In order to understand the data sets from the [Pan-American Health Organization Dataset on Dengue](#) and the [World Health Organization dashboard on COVID-19](#) which we will consider in this paper, we shall build some correlation plots to gauge the data at hand.

We shall begin our work with the largest country in South America, and the most prominent with respect to COVID-19 and dengue, which is Brazil. Since the number of COVID-19 cases is relatively large, one may take the base 10 logarithms of all the data to understand the sets better: a graph of the base 10 logarithm for new COVID-19 cases per week vs the base 10 logarithm for new dengue cases per week can be seen in Fig. 3 (left).

Along this paper we shall only use data from epidemiological weeks 30 to 100, which corresponds to the count beginning in January 2020. This is done since outside of that range the datasets are too irregular, and thus including those values would not be representative of the actual trend. Through Fig. 3 one can deduce some initial properties of the correlation between dengue and COVID-19 in South America. In the correlation graph for Brazil in Fig. 3 (left), one can see that the two diseases have a roughly positive correlation. A similar study of the datasets for Peru and Colombia are presented in Fig. 3 (middle) and (right) respectively.

It is interesting to note that the graph for Peru is very random, whereas the graph for Colombia shows slightly negative correlation, suggesting that the notion that the correlation between the diseases could differ by country. Returning the attention to Brazil, we shall consider a graph from the years 2020–2022 for dengue and COVID-19 infections in Brazil, where the horizontal axis is the epidemiological week, counted from the beginning of 2020, and the vertical axis represents the number of cases in millions as shown in Fig. 4. One can then compare the

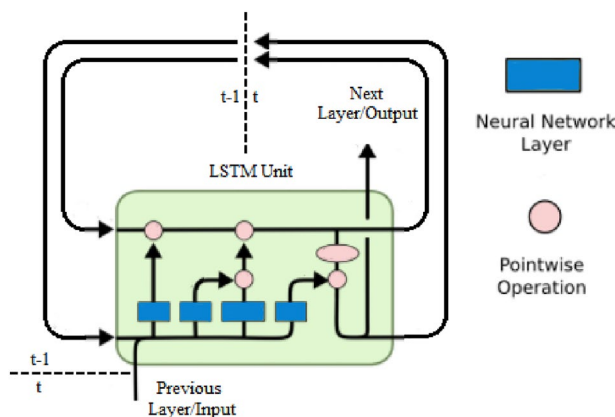


Figure 2. An example of an LSTM model, illustrated in the style of¹².

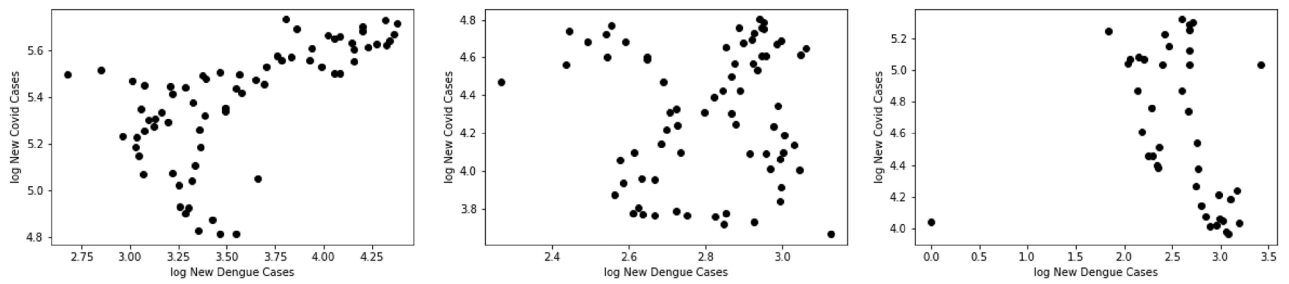


Figure 3. Dengue and COVID-19 cases plotted for Brazil (left), Peru (middle), Colombia (right). The data is normalized via taking the base 10 logarithm, and the graph is representative of what happens with the entire population.

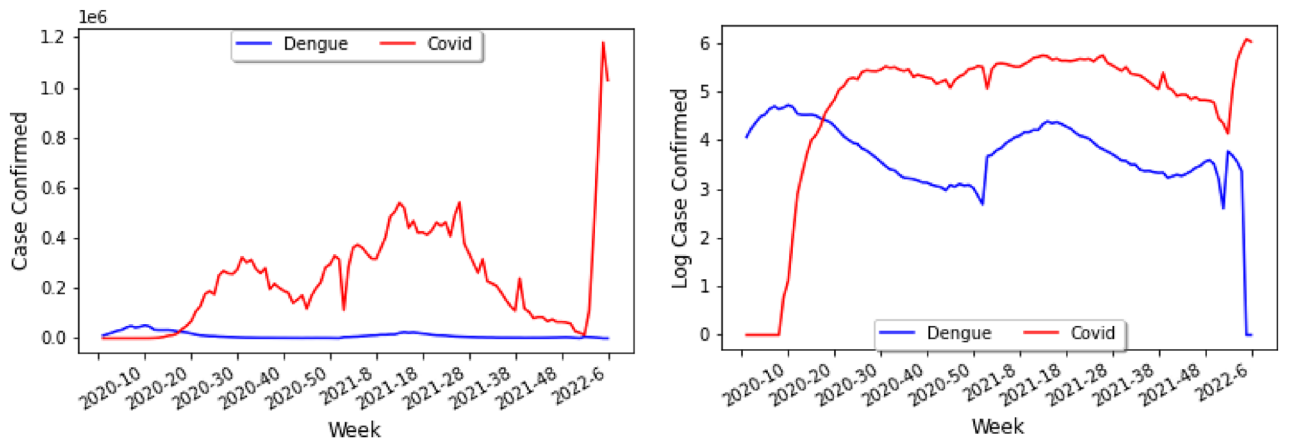


Figure 4. Plot of data of reported cases from the Pan-American Health Organization Dataset on Dengue and the World Health Organization dashboard on COVID-19 (left) and its log plot (right).

number of COVID-19 cases (red) and the number of dengue cases (blue), as shown in Fig. 4 (left), noting that between week 60 and week 80, the peak of COVID-19 generally coincides with the peak of dengue, a period that corresponds to the year 2021. After week 100 the number of COVID-19 cases spikes up, likely due to the recent increases of Omicron. In Fig. 4 (left) the data has not been normalized, and thus one can see the number of COVID-19 cases being much greater than the number of dengue cases.

Since the number of COVID-19 cases is of a greater magnitude than the number of dengue cases, it becomes difficult to see a definite trend. It is thus useful to consider the base 10 log of the number of COVID-19 and dengue cases as shown in Fig. 4 (right) to note that the peaks of dengue generally coincide with the peaks of COVID-19.

Through Fig. 4 (right) in the previous page, one can see that from approximately from week 30 to week 100 the increase or decrease of cases for COVID-19 and dengue are correlated. In particular, one can see that in the beginning of year 2020, there was a much greater number of dengue infections than COVID-19 cases, since the pandemic had not emerged yet. After week 100, or roughly the beginning of year 2022, the Omicron variant resulted in the number of COVID-19 cases increasing dramatically. Interestingly, both the number of COVID-19 and dengue experience a drop around weeks 50 and 100 (during 2021). Whilst this may be due to the fact that the period coincides with the wintertime and the end of the year, this may be counterintuitive because one would think that infectious diseases would spike during holiday times. Finally, it should be noted that in Fig. 4 (right) the similarity found in the graph could have been caused by the discovery of the two diseases, at the time when the individual was dented and thought it was just covid, but they could also have an arbovirus, and so the notifications were recorded at the same time.

A similar study to the one done above can be performed for the dataset of Perú, leading to Fig. 5 which shows the data against the epidemiological weeks for Perú, where the number of cases for each disease seem to follow the same general trend.

In order to understand other parameters which might be influencing the cases of both COVID-19 and dengue, we shall consider the particular influence of humidity and temperature (for further work on temperature and deep learning the reader should refer, e.g., to¹³ and¹⁴).

Following the style from the previous analysis, we can draw the explicit correlations between the various parameters and the number of cases of either COVID-19 or dengue. In particular Fig. 6 shows the correlation between COVID-19/Dengue and Temperature/Humidity, where the black dots represent the points for COVID-19, and the red dots represent the points for dengue. By analyzing the graphs one can deduce that temperature would be a better predictor than humidity for the number of cases.

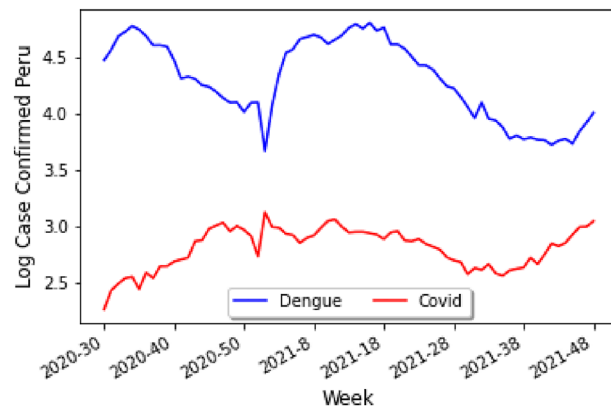


Figure 5. Weekly correlations between COVID-19 and dengue for Perú.

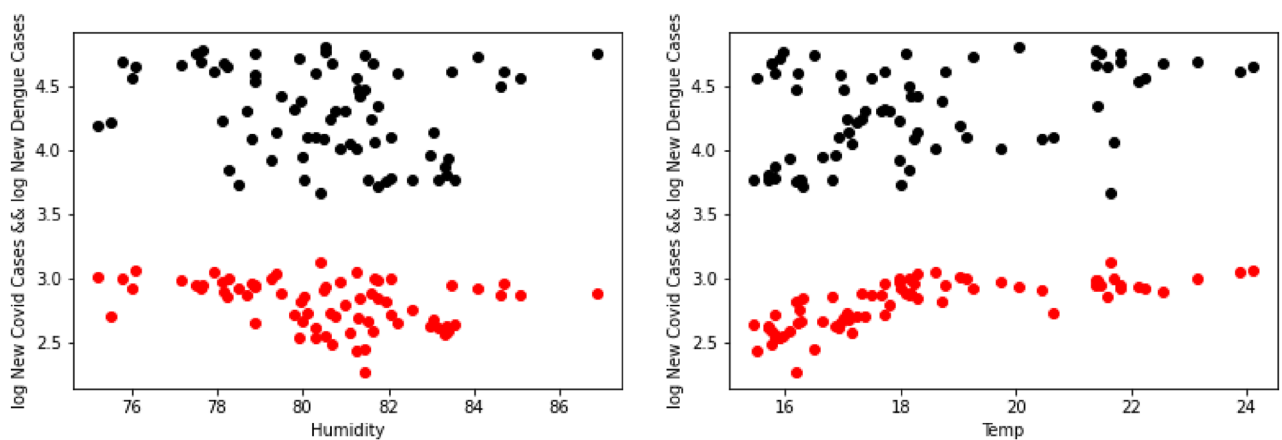


Figure 6. Humidity correlations (left) and temperature correlations (right) for Perú's dataset.

A correlation model

In order to study the correlation between COVID-19 and dengue in further detail, we shall build a neural network, the *Correlation Model*, for our model (see section “[Background](#)” for background). Our model has an input layer consisting of the various parameters we add, four hidden layers within the network, and an output layer.

As with the previous section, we shall begin our study with the dataset from Brazil. The *Correlation Model* will generate a prediction for the number of COVID-19 cases, once additional variables are taken into consideration. We have built the first version of the model to include:

- the number of dengue infections,
- the boolean variable of whether the week contains a “holiday”, and
- quantitative climate factors (temperature, humidity, and rainfall).

In what follows we shall describe the role each variable plays within the correlation study, and how we are able to incorporate such variables into our *Correlation Model*. The overall model is design following the workflow diagram given below in Fig. 7 which illustrates the process through which we are able to incorporate dengue and COVID-19 data into our *Correlation Model*.

Impact of holidays. The impact of public holidays on epidemic spreads has long been studied, from generic epidemic standpoints¹⁵ to directly related to the latest COVID-19 epidemic¹⁶. It is thus natural to consider how holidays influence the correlation between COVID-19 and dengue, and to study this question we shall consider the Brazilian dataset and a parameter which tracks whether or not a week contains an important Brazilian holiday. We firstly study the loss curve for this model, plotted in Fig. 8, where the blue line represents the training loss and the red line represents the validation loss. The list of holidays used can be found in Fig. 9, and our parameter has a value of 1 during the weeks that contain the holidays in 2020, 2021, and 2022, and 0 otherwise. As in usual neural networks, the data has been split randomly into a training set and a cross-validation set, with our epoch size of 150 and the loss function based on the Mean Absolute Error.

Since one of the most useful things that the Keras model can do is predict new values, we used `model.predict()` to do so within our setting. Specifically, our *Correlation Model* then predicts the number of cases for COVID-19

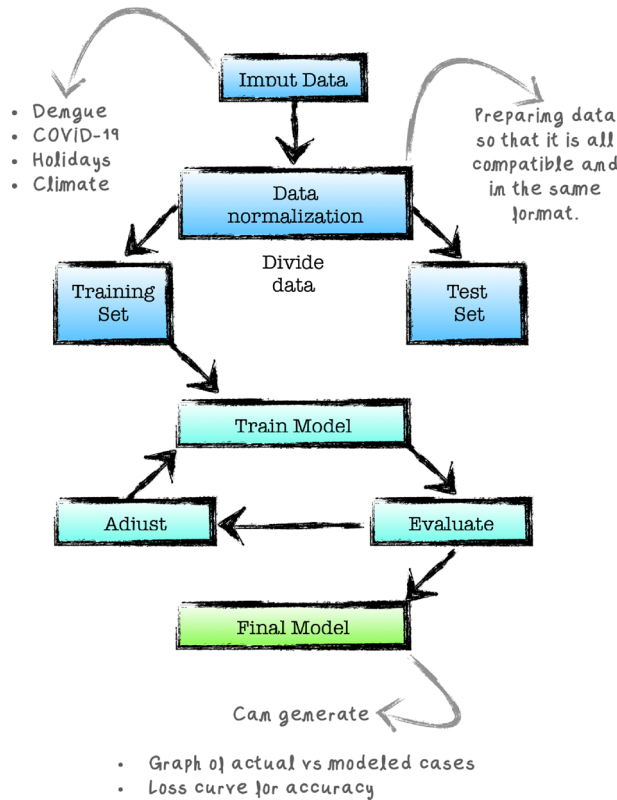


Figure 7. The workflow diagram of our correlation model.

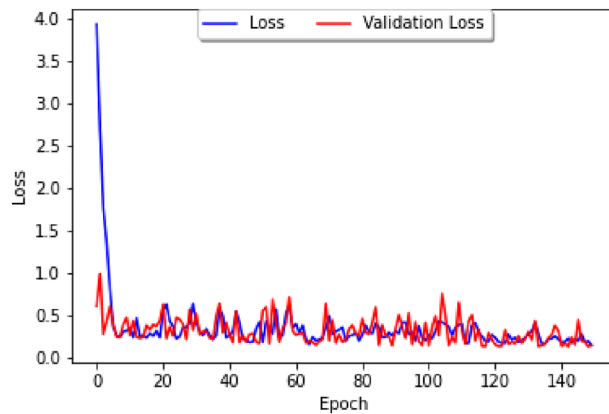


Figure 8. A graph of the loss curve when the holiday data is first added.

leading to the visualization in Fig. 10 (left), where the blue line is the predicted value for any given week using the prediction algorithm, and the red line is the actual value in the dataset.

Notice that the vertical axis runs from 4.8 to 5.6. Recall that we have taken the base-10 logarithm of the weekly cases for COVID-19 and dengue as a means of normalization and for the number of cases to be of the same approximate magnitude. In Fig. 10 (left), the prediction has the same general trend as the actual data and can therefore indicate where the approximate peaks and valleys will be. Yet, the prediction has a smaller amplitude than the actual data, and thus the prediction would seem to diminish the variance of the initial data. To account for this, we can define a new variable C to be the *Contraction ratio*: the ratio of the average distance from the mean for the actual data to the average distance from the mean for the predictions. We shall see the importance of this variable in the upcoming sections as well as in the conclusion.

In order to test our model, one repeats the process for the prediction done before and generates a new test graph shown in Fig. 10 (right), which presents the study on the cross-validation set, not the training set. In this case, the blue line represents the prediction model on the validation set, and the red line is the actual data in the validation set. Since not as many weeks were included in the test set in Fig. 10 (right), the relationship between

	Brazil	Peru	Colombia	Cambodia	Kenya
New Year's Day	January 1	January 1	January 1	January 1	January 1
Victory over Genocide Day	--	--	--	January 7	--
Epiphany	--	--	January 10	--	--
Shrove Tuesday	March 1	--	--	--	--
International Womens Day	--	--	--	March 8	--
St. Joseph's Day	--	--	March 21	--	--
Brazilian Carnival	February 21 - 26	--	--	--	--
Good Friday	April 2	April 2	April 15	--	April 15
Easter	April 4	April 4	--	--	April 18
Maunder Thursday	--	--	April 14	--	--
Khmer New Year*	--	--	--	April 14-15	--
President Kibaki State Funeral	--	--	--	--	April 19
Visak Bochea Day	--	--	--	April 26	--
Tiradentes Day	April 21	--	--	--	--
Labour Day	May 1	May 1	May 1	May 1	May 2
Idd-ul-Fitr	--	--	--	--	May 3
Kings birthday	--	--	--	May 14	--
Ascension Day Holiday	--	--	May 30	--	--
Madaraka Day	--	--	--	--	June 1
King's mother birthday	--	--	--	June 18	--
Feast of Corpus Christi	June 16	--	June 20	--	--
Sacred Heart	--	--	June 27	--	--
Idd-ul-Azha	--	--	--	--	July 10
Saint Peter and Saint Paul	--	June 29	July 4	--	--
Independence Day	September 7	July 28-29	July 20	November 9	--
Battle of Boyoca	--	--	August 7	--	--
Assumption Day	--	--	August 15	--	--
Santa Rosa de Lima	--	August 30	--	--	--
Constitutional Day	--	--	--	September 24	--
Ancestors Day	--	--	--	October 5-7	--
Utamaduni Day	--	--	--	--	October 10
Our Lady of Aparecida's day	October 12	--	--	--	--
Commemoration Day of King's F	--	--	--	October 15	--
Day of Races	--	--	October 17	--	--
Mashujaa Day	--	--	--	--	October 20
King's Coronation Day	--	--	--	October 29	--
Battle of Angamos	--	November 1	--	--	--
All Souls/Saints Day	November 2	--	November 7	--	--
Republic Day	November 15	--	--	--	--
Water Festival Ceremony	--	--	--	November 18-20	--
Independence of Cartagena	--	--	November 14	--	--
Immaculate Conception Day	--	December 8	December 8	--	--
Jamhuri Day	--	--	--	--	December 12
Christmas Day	December 25	December 25	December 25	--	December 25
Boxing Day	--	--	--	--	December 26

Figure 9. Table of holidays considered during the first pandemic year (* August 17–21 in 2020, as a replacement holiday).

the two curves is less apparent than that of Fig. 10 (left). However, one can still see that the prediction and the actual data for the validation set have the same approximate trend, and that like in the training set, the prediction values do not vary as much as the actual data values.

Impact of humidity and temperature. Research has shown that climate¹⁷ and in particular increasing temperatures influence COVID-19 transmission¹⁸, as well as for other viruses¹⁹. Therefore, in what follows we shall add climate factors such as temperature or humidity to our model. A characteristic of dengue is that it is

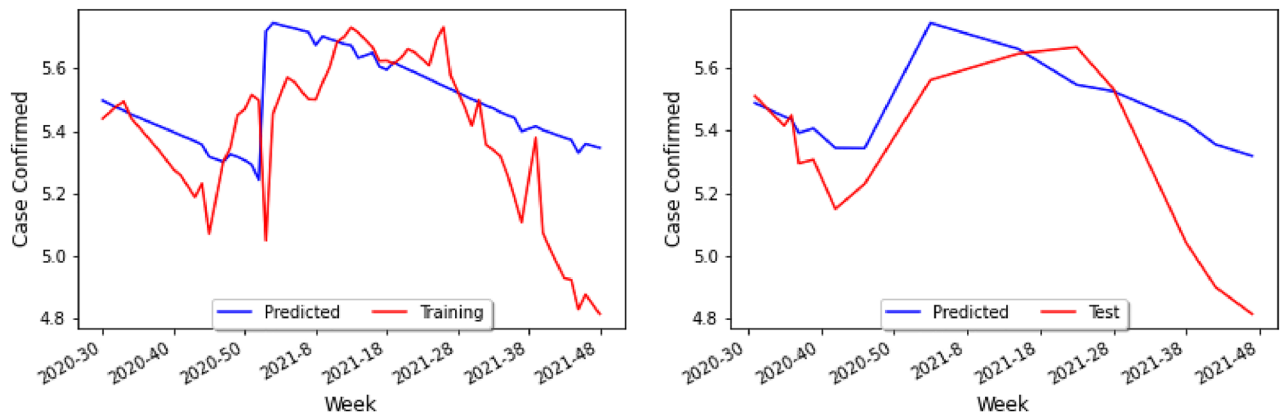


Figure 10. A graph of the actual data and the predicted data for holidays on the training set (left) and the test set (right).

most prevalent around rainy seasons, and thus one should expect higher value of humidity would correspond to a greater number of dengue cases and we shall look into this correlation below.

The dataset obtained on Brazilian climate contains daily maximum and minimum temperature and humidity for major Brazilian cities, so in order to fit it to our model, we can take the average of the temperature and humidity over each epidemiological week. In addition to the factors of temperature and humidity, we shall weight the data based on population, so that the greater the population, the more heavily it is weighted. To perform our study we consider the two biggest cities in Northern and Southern Brazil, which are Salvador and Sao Paulo, respectively. Unlike the study done in Fig. 8, one can observe that the loss is generally greater, and thus the model generates less precise results, leading to the conclusion that that climate factors are not as correlated with COVID-19 cases as holiday factors are.

A graph of the predicted data based on our neural network model on the training set can be obtained, as shown in Fig. 11 (left). In particular, one can see that the model predicts the actual numbers for COVID-19 rather well, as the two graphs have approximately the same amplitude and the same mean. Finally, one can repeat this for the test set, and find that, similar to before, the model in Fig. 11 (right) predicts the general trend of ups and downs fairly well for the test set. The weighted data according to the populations of Northern and Southern Brazil, is given by approximately 43% and 57%, respectively, leading to the loss curve model in Fig. 12.

Having studied our Correlation Model with climate parameters and holiday parameters separately, we shall now consider both sets of parameters together, leading to the comparison chart featured in Fig. 13 which describes the loss curve for both climate and holidays:

Once the parameters for temperature and humidity are added, the loss curve of Fig. 13 decreases more gradually. In contrast, the loss curve from Fig. 8 decreased very rapidly. The actual training data and the values predicted by the Keras model are then plotted in Fig. 14. In particular, in Fig. 14 (left) the prediction has the same general peaks and valleys as the actual data. However, whilst the prediction seems to have roughly the same amplitude as the data, the prediction is shifted down approximately 0.2 on the logarithmic scale. In Fig. 10 (left), the prediction had roughly the same mean value and a smaller amplitude. The model predictions with the actual test data are plotted in Fig. 14 (right): compared with Fig. 10 (right), the predictions in better seem to correlate with the trend of the actual data. This graph resembles Fig. 14 (left) in that the blue line has the same approximate trend as the red line, but the predictions has a smaller numerical values.

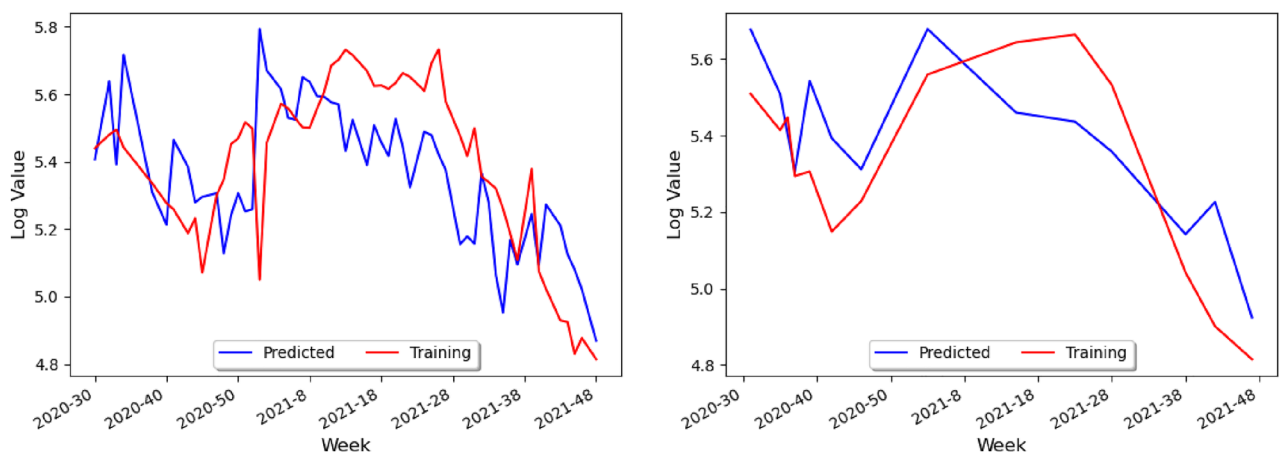


Figure 11. Actual data and predicted data for only climate on the training set (left) and the test set (right).

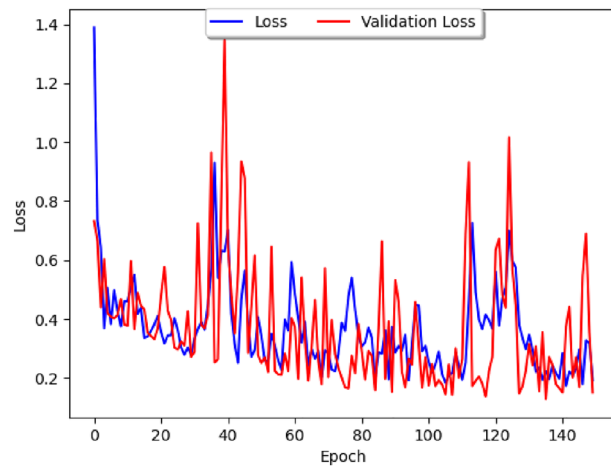


Figure 12. The loss curve for only climate factors.

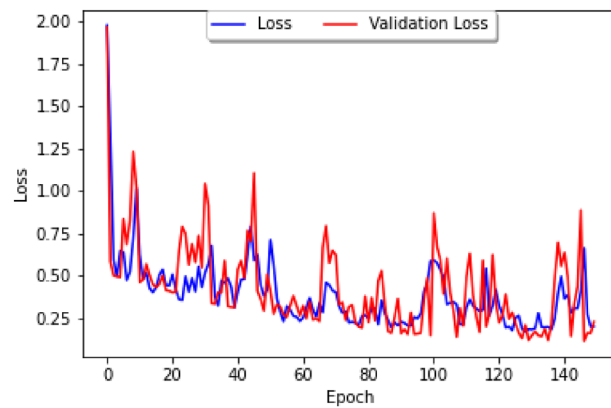


Figure 13. The loss curve for both climate and holidays. Data on the number of reported cases is from [Pan-American Health Organization Dataset on Dengue](#) and the [WHO dashboard on COVID-19](#). Data on Brazilian climate is from <https://www.visualcrossing.com/weather-history/brazil>.

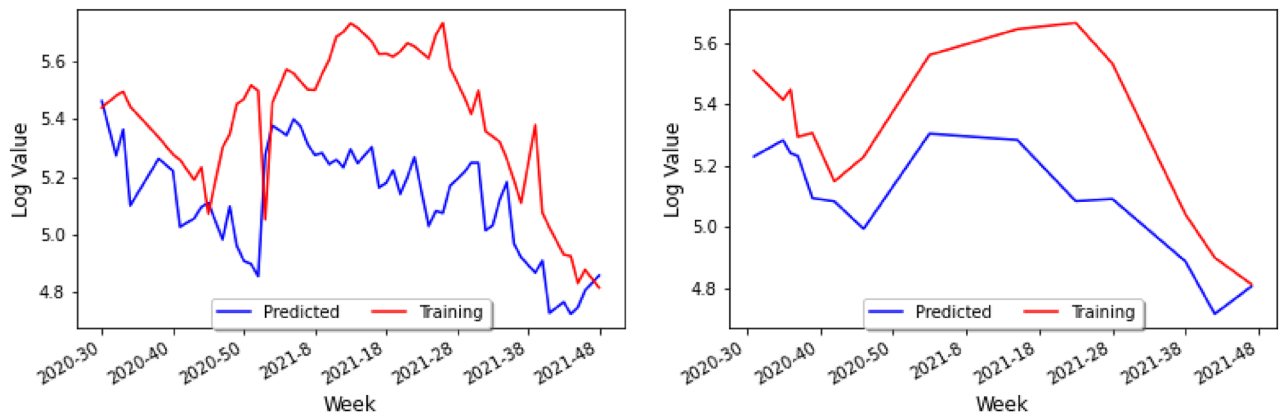


Figure 14. Actual data and the predicted data for Brazil when both holiday and climate factors are considered on the training set (left) and the test set (right).

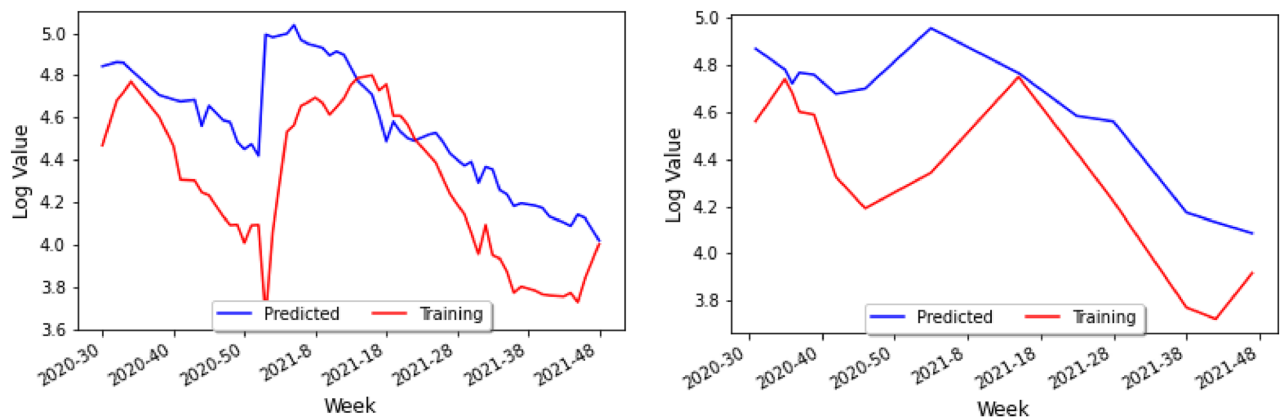


Figure 15. A graph of the actual data and the predicted data for Perú, for when both holiday and climate factors are considered on the training set (left) and the test set (right).

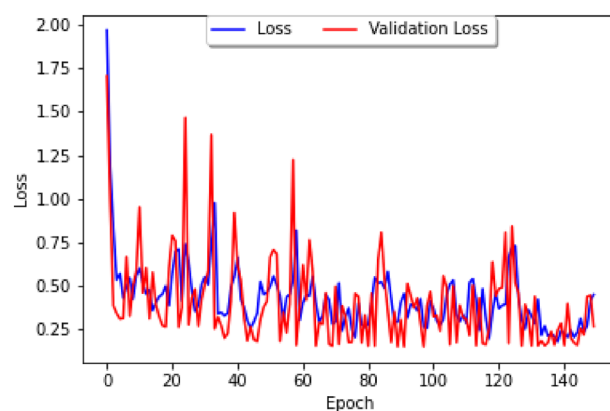


Figure 16. Loss curve of the neural network model for Perú.

The efficacy of the correlation model

We shall dedicate this section to the study of how well our Correlation Model can serve to understand data from other South American countries. For this, the model shall be re-trained for datasets from Peru and Colombia. Then, we shall perform a reverse study, flipping the parameters in order to predict the number of dengue cases on countries that do not have dengue data readily available such as Cambodia and Kenya, countries from Southeast Asia and Africa that have significant dengue.

South American nations. Now that we have the model tested on Brazil, we can use it on other nations as well to see if the model predictions match up with the actual data. The procedure shall be repeated for Peru and Colombia, which are both countries where both dengue and COVID-19 are very prevalent and frequent.

The correlation model on Perú. According to the [Pan-American Health Organization Dataset on Dengue](#), Peru had 49,274 cases of Dengue Fever in 2021 and has had 20,491 cases of dengue during January-June 2022, setting Peru as the second country with the highest dengue cases in South America, just after Brazil. Therefore, it would be meaningful to see what correlates with the number of dengue or COVID-19 cases in this country. In order to study the influence of holidays within the dataset from Perú, we replicated our Brazilian study but with the new set of national holidays, a list of which has been included in Fig. 9.

Because both Peru and Brazil are South American countries, they celebrate very similar national holidays allowing us to do further comparisons of the datasets. Furthermore, for our climate variables we consider the data from [Visual Crossing](#) to obtain data on humidity, temperature, and precipitation. Since Lima is the capital and largest city, with around one-third of the nation's population, we have chosen it for our climate data, allowing us to derive the loss curve and prediction curve for Peru shown in Figs. 15 and 16.

The correlation model on Colombia. We shall finally replicate our method for Colombia's dataset. In this case, however, because of incomplete data for dengue from approximately week 30 to week 60, we shall only focus on the period between week 60 to week 100, and consider the national holidays, a list of which has been included in Fig. 9.

Following a similar approach as in the previous sections, we derive the loss curve for Colombia's Correlation Model as shown in Fig. 17. In order to add the variables corresponding to climate factors, we consider the data from [Visual Crossing](#) to obtain data on humidity and temperature across the nation. Since Bogotá is the both Colombia's capital and largest city, we chose it to be a representative for our climate data in order to do a parallel study to the one done for Lima in the previous section. Then, by using our Correlation Model both on the training set and on the test set, we are able to evaluate its accuracy in predicting infection cases across Colombia as shown in Fig. 18.

It should be noted that for Brazil, Peru and Colombia, the abrupt drop in cases between weeks 52 y 104 could be due to the low reporting rate associated with the new year's celebrations, which has been seen when studying other diseases²⁰. Within our setting, we attempted to mitigate such effect by taking weekly averages when performing our studies. Finally, it is also important to note that dengue data has sometimes been thought to have been underreported, and in particular to have been undermined by the beginning of the COVID-19 pandemic²¹, and this is why it is sometimes useful to leave the first weeks of the data unused.

Using the model to predict infections in Africa or Southeast Asia

Having developed our Correlation Model and trained it with the datasets from several South American countries, we can use the model to predict the number of dengue cases for countries in Africa or Southeast Asia, especially since those regions also have a high number of dengue cases. In particular, whilst not much data can be found on recent dengue cases for those countries, it is known that the regions of Southeastern Asia have a significant amount of cases of dengue fever, which creates a heavy economic burden on those countries²². Thus, by *reversing* the Correlation Model to use COVID-19 as a parameter to predict dengue cases over time, one can obtain very useful information for future policy making.

Predicting dengue cases in Cambodia. We shall first consider dengue in Cambodia, a Southeastern Asia country. In a recent update as of April of 2022, there have been a total of 457 cases, including one death, which remains much lower than the normal range of cases from 2015 to 2021²³. In our model, we shall use South American countries for the training set, and then Cambodia for the test set.

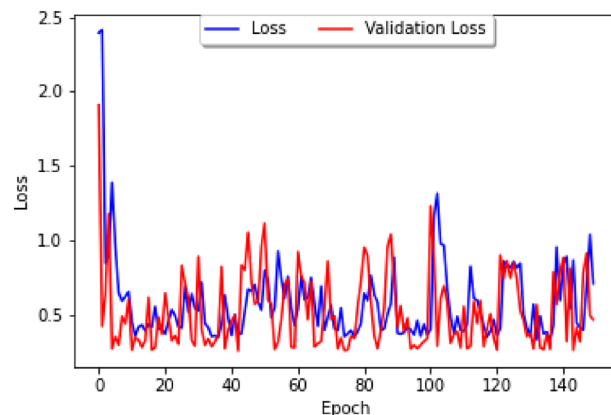


Figure 17. Loss curve for Colombia.



Figure 18. A graph of the actual data and the predicted data for Colombia, for when both holiday and climate factors are considered on the training set (left) and the test set (right).

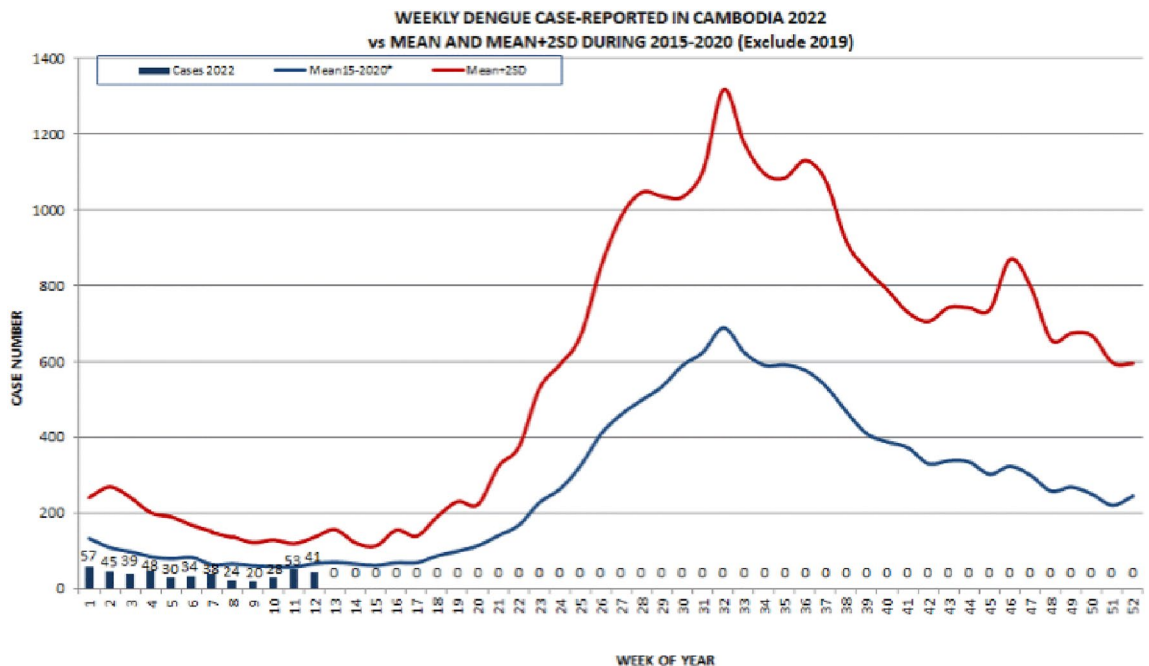


Figure 19. Dengue cases reported weekly in 2022 vs Mean and Mean+2SD during 2015–2020, *excluding 2019 in Cambodia. Source: National Dengue Surveillance System (NDSP/CNM/MOH)—World Health Organizations²⁴, Figure 1.

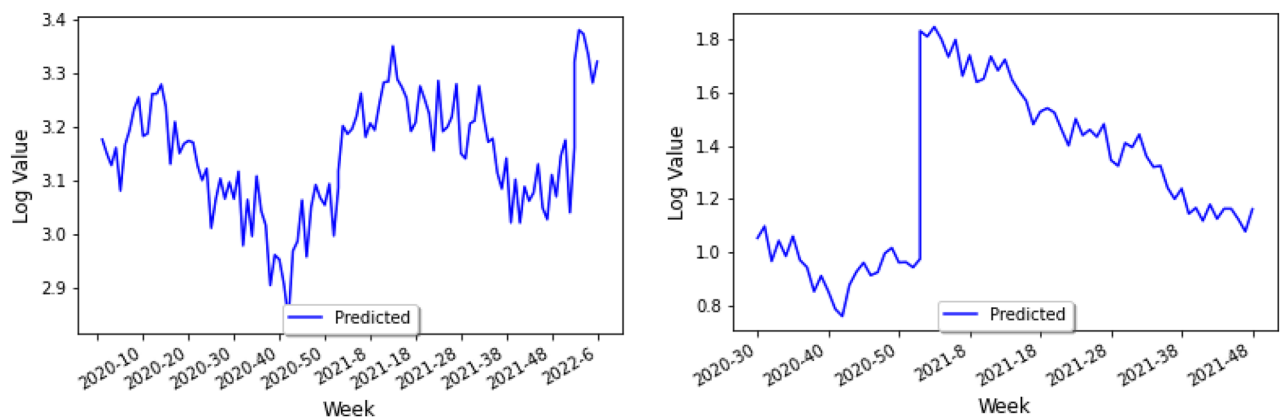


Figure 20. Predicting Cambodia's dengue cases using Peru as the training set (left), and using both Peru and Brazil's data as training set (right).

It should be noted that many of the holidays celebrated in Cambodia are quite different from those celebrated in South American countries, and thus we have included the new corresponding list in Fig. 9. In addition, we shall also include a climate variable built through the data for Cambodia found in Visual Crossing. Following the same style as before, we shall take Phnom Penh, the capital and largest city in Cambodia as our city of study. Finally, it should be noted that instead of using the log of the number of cases, we shall perform our study by considering the parameter

$$\log \left(\frac{\text{Number of COVID-19 or dengue cases}}{\text{Country population in millions}} \right)$$

This kind normalization, dividing by the nation's population, gives a better idea of the prevalence of that disease. By taking Peru as the country for the training set, one obtains the following prediction of the number of dengue cases in Cambodia (plotted in Fig. 19) with respect to time shown in Fig. 20 (left). Moreover, by adding Brazil into the training set, one can confirm similar trends with the previous figure, and around the same times of year, as shown in Fig. 20 (right).

It becomes interesting to compare the above analysis with the average trend of dengue fever in Cambodia from 2015–2020, from WHO's April 2022 Update of dengue in the Western Pacific²³. In particular, one should

note the similarities in trend, even though the peaks seen in Figs. 20 (right) and 19, which correspond to different times of the year.

In what follows we shall perform a similar study using Kenya's data set, and include an analysis of such study in the last section of the paper.

Predicting dengue cases in Kenya. In what follows, it is useful to have in mind the number of COVID-19 infections from the beginning of 2020 up until the beginning of 2022 as given in Fig. 21, with data used from 0–1. We shall continue showing the utility of our model by considering the data sets for Kenya, one of the countries in Eastern Africa with a significant amount of dengue cases. By applying our Correlation Model trained with South American countries on Kenya, one can predict infection peaks at roughly epidemiological weeks 60 and 110 and valleys at epidemiological weeks 50 and 100 as shown in Fig. 22.

Having seen how many external variables, such as celebration of holidays or temperature or humidity, affect the prevalence of COVID-19 or dengue, in what follows, we shall expand on our Correlation Model and build a Recurrent Neural Network and perform a time series analysis (see section “LSTM models” for background on LSTM models).

Recurrent neural network and time series analysis

We shall dedicate this section to expand our Correlation Model to build upon a Recurrent Neural Network. To train this new model, we shall use Peru's dataset, since it has the most complete data for dengue and thus provides an ideal starting point. Moreover, it should be noted that the number of epochs that one chooses for the model has a large impact on the model: the smaller the number of epochs, the more the model is likely to under-fit; Equivalently, the larger the number of epochs, the more the model tends to over-fit. In what follows we shall will assign the first 80% of the data, roughly until October 2021, for the training set and the remaining 20% for the test set, and use MinMaxScaler, so that the calculations are done with all the data from 0 to 1.

Through the data of Fig. 21, we can inverse of our original rescaled model first, and then consider the base 10 logarithm for all of the values, to obtain a prediction of COVID-19 infections using a time series analysis for Peru as shown in Fig. 23.

It should be noted that the horizontal scale is measured in days, and thus from our model one expects the number of COVID-19 cases to decrease gradually after the spike in January of 2022, which is reasonable considering that Omicron caused a large increase in the number of cases.

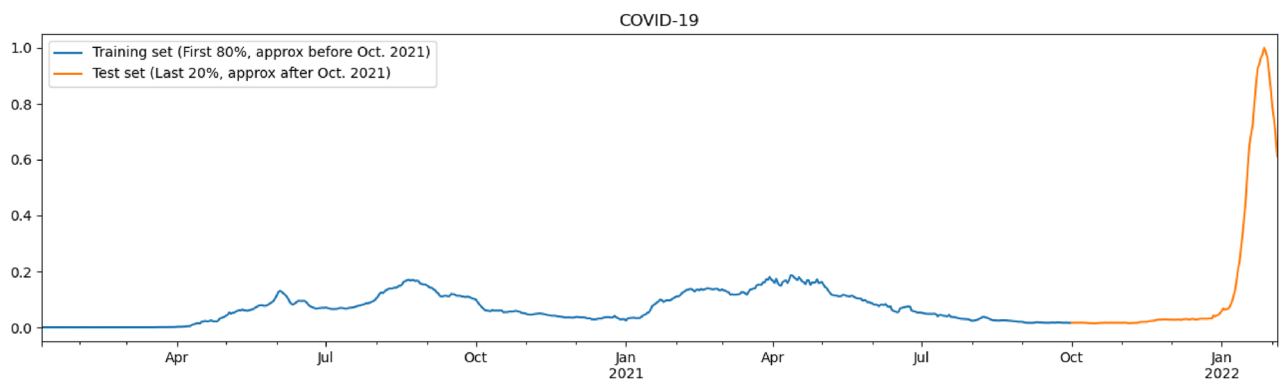


Figure 21. A scaled look at COVID-19 numbers over the past two years.

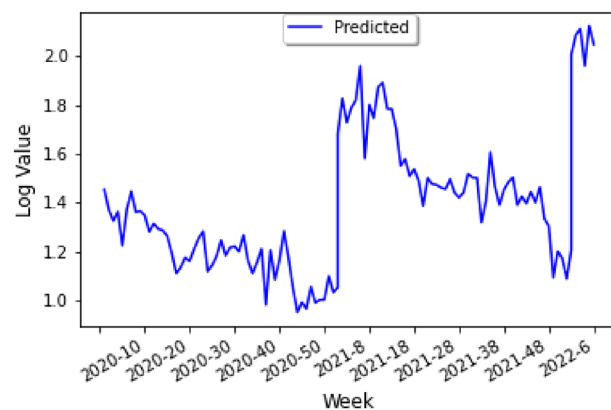


Figure 22. Predicting Kenya's dengue cases using Peru and Brazil as the training set.

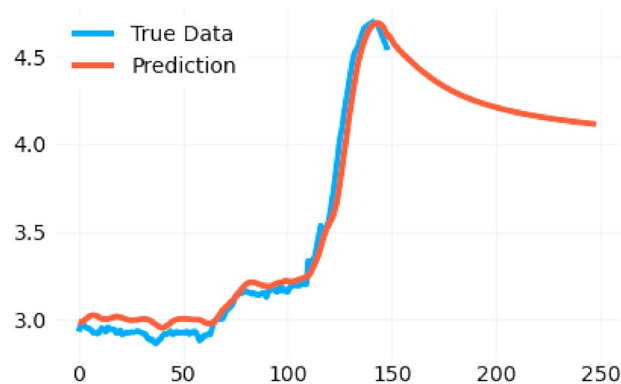


Figure 23. Predicting COVID-19 infections using time series.

A similar study can then be done for the prediction of dengue and COVID using LSTM, which we shall illustrate in Fig. 24. In the case of dengue, one can see in Fig. 24 (left) that with enough epochs the prediction fits the true data quite well.

The case of dengue infections has been considered in Fig. 24 (middle), where the blue spike downwards is actually probably due to a missing value of sorts and may not be an actual zero. Considering this, the model gets most of the general trends pretty accurately. After the beginning of 2022, this model predicts that the number of dengue infections will initially rise and then fall, which we can compare with Fig. 24 (left), in which we have incorporated both epidemics. Essentially, there are two models here; one using past values for both diseases to predict COVID-19, and the other using past observed values for both to predict dengue. These are updated simultaneously while the program is running. Again, we have used MinMaxScaler to scale the numbers from 0 to 1, then inverse scaled them back and took the logarithm of all value to make the details more clear.

Conclusion and applications

In this paper we have built a neural network Correlation Model as well as a recurrent Correlation model for a neural network to take advantage of the correlation between dengue and COVID-19 to deduce predictions on future infections of both diseases. Such correlations drawn should also present interesting lines of research for other viruses which present correlations. Moreover, there are many directions in which our models could further be expended. Since it is outside the scope of this paper to perform those extensions here, we shall instead give a brief summary of the directions we believe would be very fruitful to explore in upcoming work.

In particular, we shall consider Latent Heat Flux and whether it could be a better correlator than other climate factors, such as humidity, temperature, or precipitation. Finally, we shall also propose a notion of peak indices which could be the ideal setting when finding correlations between viral epidemics. We conclude this note by summarizing our methods and highlighting our findings.

Latent heat flux. From our model building analysis, it would seem that the time series of transmissions of COVID-19 and dengue could be better correlated to the latent heat flux (LHF), which is the amount of energy moving between the air and the land due to evaporation or condensation. In particular, the research of heat and mass transfers has been identified as a possible area that could lead to progress in reducing COVID-19²⁵. A sufficient dataset on latent heat flux could not be found because the datasets that are available online stopped at the early-2010s and were not daily or weekly, but there are many environmental factors correlated with LHF or just COVID-19 and dengue which could be taken into consideration. Instead. Some of the ones we believe could be best suited for finding correlations are the following:

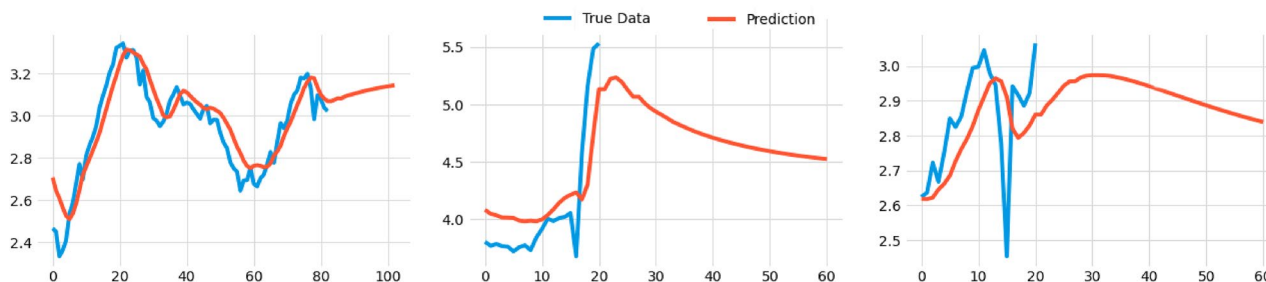


Figure 24. Predictions of infections using a time series analysis and LSTM model for dengue (left), COVID-19 (middle), and for predicting dengue infections using an LSTM model that incorporates both diseases (right).

- Sea Surface Temperature²⁶—unfortunately, no data on this either;
- Temperature change;
- Humidity—the higher the humidity, the higher the latent heat flux tends to be;
- Wind speed—the greater the wind speed, the lower the latent heat flux usually is;
- Atmospheric pollution impacts²⁵—for this, use the Visibility data from the Visual Crossing dataset.

From the above, one can add the new parameters of

- maximum temperature,
- minimum temperature,
- wind speed, and
- visibility

for the study of Perú's dataset (since that was the country that had the most complete data) leading to the graph of the loss curve and the prediction models shown in Fig. 25.

Peak index predictions. It should be very useful to understand correlations between viral epidemics through some sort of *peak index*, which can be defined in a general sense as the week index of where the largest number of COVID-19 or dengue cases occur. In particular, the following two perspectives could be incorporated to our Correlation Model to expand on our work:

- **A relative index.** We define an epidemiological week to be a peak if it has a greater number of cases than both the week before and after it. We define a function $\rho(t)$ to be the function that represents whether EW t is a peak. Thus,

$$\rho(t) = 1 \text{ if } Nc[t] \geq \alpha Nc[t - 1] \text{ and } Nc[t] \geq \alpha Nc[t + 1] \\ = 0 \text{ otherwise}$$

- **An absolute index.** We define a function $\rho(t)$ to be the function that represents whether EW t is a peak. Formally, this is given by

$$\rho(t) = 1 \text{ if } Nc \geq \alpha \cdot P \\ = 0 \text{ otherwise,}$$

where Nc represents the number of new infections for a particular disease and P represents the total population for a particular country, and α is a coefficient that differs among various diseases. For example, for COVID-19, this could be if the new cases is at least $\frac{1}{10,000}$ of the country's population. For dengue, this could be if the number of new cases is at least $\frac{1}{1,000,000}$ of the country's population (since there are currently far less cases for dengue than for COVID-19)

Final remarks. By considering datasets from different South American countries on COVID-19 and dengue cases, we have been able to build a Correlation Model which allows us to study COVID-19 and dengue cases in areas of the world where data is not fully available. In particular our study, we have been able to observe the following.

- When considering the importance of holidays as a parameter, the prediction curve has approximately the same mean as the actual data and yet has a smaller amplitude, and it seems as if the prediction essentially diminished the variance of the initial data as seen in Fig. 10 (left).
- When parameters for temperature and humidity are added, the loss curve seems to decrease more gradually as shown in Fig. 13. In contrast, the loss curve from Fig. 4A decreased very rapidly.
- By only adding in climate factors into the neural network results in a gradually descending loss curve, but with very big amplitude as shown in Fig. 4A. Combining this with holiday factors seems to take care of this issue.

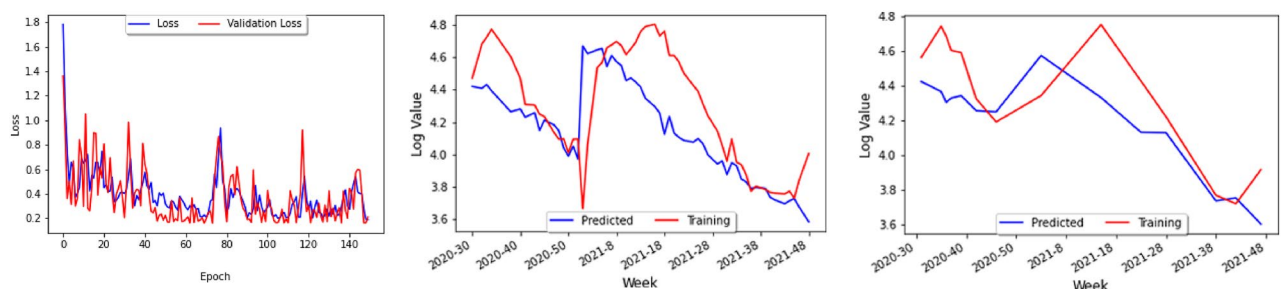


Figure 25. (Left) Loss curve, (Center) Training predictions, (right) Test predictions.

- When considering combined holiday and climate factors, the prediction seems to have roughly the same amplitude as the data, but the prediction is shifted down approximately 0.2 on the logarithmic scale as shown in Fig. 10 (left). However, interestingly the prediction had roughly the same mean value and a smaller amplitude.
- The sequence length is also a highly important factor that may change the prediction: indeed, when changing the 14-day sequence length to 35-days, the prediction for COVID-19 is to decrease (like before) and then rise slowly.

Finally, it would be interesting to see the effects of the COVID-19 lockdown on dengue²⁷. Many say that because the lockdown occurred so close to the peak of the mosquito season, the transmission of dengue increased²⁸, as the lockdown affected *Aedes* mosquito surveillance and vector control procedures²⁹. In this setting, then mosquito larvae populations would have increased, and thus the number of people infected with dengue would increase because they lived in close proximity with mosquitoes. It should be noted that the number of dengue cases could have been severely underestimated, simply due to the fact that health resources were being focused on COVID-19 instead or that people who were infected with dengue were unwilling to get a formal diagnosis due to the lockdown³⁰.

Data availability

The datasets used during the current study are available in the GitHub repository <https://github.com/lauraschaposnik/Covid19-Dengue>.

Received: 8 August 2022; Accepted: 11 January 2023

Published online: 27 January 2023

References

- Berberian, G. Dengue at the beginning of the COVID-19 pandemic in Argentina. *Arch. Argent. Pediatr.* **119**, 131–138 (2021).
- Bergero, P. & Guisoni, N. Mathematical modelling of dengue-COVID-19 coinfection: a first approximation. *Revista Argentina De Salud Publica* **13**, 05 (2021).
- Rahman, T., Sobur, A., Islam, S., Toniolo, A. & Nazmul Hussain Nazir, K. H. M. Is the COVID-19 pandemic masking dengue epidemic in Bangladesh?. *J. Adv. Vet. Anim. Res.* **7**(2), 218–219 (2020).
- Pan American Health Organization. Epidemiological Update: Dengue in the context of COVID-19 (2020).
- Murphy, B. R. & Whitehead, S. S. Immune response to dengue virus and prospects for a vaccine. *Annu. Rev. Immunol.* **29**, 587–619 (2011).
- Murray, N. E., Quam, M. B. & Wilder-Smith, A. Epidemiology of dengue: past, present and future prospects. *Clin. Epidemiol.* **5**, 299–309 (2013).
- Silvestre, O. M. *et al.* Previous dengue infection and mortality in coronavirus disease 2019 (COVID-19). *Clin. Infect. Dis.* **73**(5), e1219–e1221 (2020).
- Nicolete, V. C. *et al.* Interacting epidemics in Amazonian Brazil: prior dengue infection associated with increased coronavirus disease 2019 (COVID-19) risk in a population-based cohort study. *Clin. Infect. Dis.* **12**, 2021 (2019).
- Bergero, P., Schaposnik, L. P. & Wang, G. Github data repository. <https://github.com/lauraschaposnik/Covid19-Dengue> (2022).
- Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., Rabczuk, T. & Atkinson, P. M. Covid-19 outbreak prediction with machine learning. *medRxiv* (2020).
- Chierigato, M. *et al.* A hybrid machine learning/deep learning COVID-19 severity predictive model from CT images and clinical data. *Sci. Rep.* **12**, 4329 (2022).
- Olah, C. Understanding lstm networks. *github* (2015).
- Demongeot, J., Flet-Berliac, Y. & Seligmann, H. Temperature decreases spread parameters of the new COVID-19 case dynamics. *Biology* **9**(5), 94 (2020).
- Oshinubi, K., Amakor, A., Peter, O. J., Rachdi, M. & Demongeot, J. Approach to COVID-19 time series data using deep learning and spectral analysis methods. *Aims Bioeng.* **9**, 1–21 (2022).
- Eames, K. T. D. The influence of school holiday timing on epidemic impact. *Epidemiol. Infect.* **142**(9), 1963–1971 (2014).
- Chen, S. *et al.* Buying time for an effective epidemic response: The impact of a public holiday for outbreak control on COVID-19 epidemic spread. *Engineering* **6**(10), 1108–1114 (2020).
- Chen, S. *et al.* Climate and the spread of COVID-19. *Sci. Rep.* **11**(1), 1–6 (2021).
- Notari, A. Temperature dependence of COVID-19 transmission. *Sci. Total Environ.* **763**, 144390 (2021).
- Lowen, A. C., Mubareka, S., Steel, J. & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathog.* **3**(10), e151 (2007).
- Silva, M. M. O. *et al.* Accuracy of dengue reporting by national surveillance system, Brazil. *Emerg. Infect. Dis.* **22**(2), 336 (2016).
- Xinting, L. *et al.* Dengue outbreaks in the COVID-19 era: Alarm raised for Asia. *PLoS Negl. Trop. Dis.* **15**(10), e0009778 (2021).
- Shepard, D. S., Undurraga, E. A. & Halasa, Y. A. Economic and disease burden of dengue in Southeast Asia. *PLoS Negl. Trop. Dis.* **7**, e2055 (2013).
- World Health Organization. Update on the Dengue situation in the Western Pacific Region (2022).
- World Health Organization. Dengue situation update January 640 (WHO, 2021).
- Trancossi, M., Carli, C., Cannistraro, G., Pascoa, J. & Sharma, S. Could thermodynamics and heat and mass transfer research produce a fundamental step advance toward and significant reduction of SARS-COV-2 spread?. *Int. J. Heat Mass Transf.* **170**, 120983 (2021).
- Wu, R., Kirtman, B. & Pegion, K. Surface latent heat flux and relationships with SST in CFS. *Geophys. Res. Lett.* <https://doi.org/10.1029/2007GL030751> (2007).
- Chen, Y. *et al.* Measuring the effects of COVID-19-related disruption on dengue transmission in southeast Asia and Latin America: A statistical modelling study. *Lancet Infect. Dis.* **3**, 1. [https://doi.org/10.1016/S1473-3099\(22\)00025-1](https://doi.org/10.1016/S1473-3099(22)00025-1) (2022).
- Lu, X. *et al.* Dengue outbreaks in the COVID-19 era: Alarm raised for Asia. *PLoS Negl. Trop. Dis.* **15**, e0009778 (2021).
- Daniel Reegan, A., Rajiv Gandhi, M., Cruz Asharaja, A., Devi, C. & Perumal Shanthakumar, S. COVID-19 lockdown: Impact assessment on *Aedes* larval indices, breeding habitats, effects on vector control programme and prevention of dengue outbreaks. *Heliyon* **6**(10), e05181 (2020).
- Borre, F., Borri, J. I., Cohen, Y. Z., Gasparoto, M. & Gurung, T. B. Impact of the COVID-19 pandemic on infectious diseases in Brazil: A case study on dengue infections. *Epidemiologia* **3**(1), 97–115 (2022).

Acknowledgements

Grace Wang and Laura P. Schaposnik would like to express their deep gratitude to the MIT PRIMES Program for this wonderful opportunity to conduct research, and to Fidel A. Schaposnik for introducing them to the work of Paula Bergero which led to this collaboration.

Author contributions

All authors contributed equally to the planning of the paper. Much of the work was done by Grace Wang under the supervision of L.P.S. through PRIMES-MIT program 2022.

Funding

PB is member of the Scientific Career of Consejo Nacional de Investigaciones Científicas y Técnicas of Argentina (CONICET). The work of LPS is partially supported by a Simons Fellowship, NSF CAREER Award DMS 1749013 and NSF FRG Award 2152107, and by All Souls college, Oxford, UK. The manuscript was finalized with support by the National Science Foundation under grant No. DMS-1928930, while LPS was in residence at the Simons Laufer Mathematical Science Institute (previously known as MSRI) Berkeley, California, during the Fall 2022 semester.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.P.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023