International Journal of
Environmental Research and Public Health

MDPI

Article

Forecasting Weekly Dengue Cases by Integrating Google Earth Engine-Based Risk Predictor Generation and Google Colab-Based Deep Learning Modeling in Fortaleza and the Federal District, Brazil

Zhichao Li

Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; lizc@igsnrr.ac.cn

Abstract: Efficient and accurate dengue risk prediction is an important basis for dengue prevention and control, which faces challenges, such as downloading and processing multi-source data to generate risk predictors and consuming significant time and computational resources to train and validate models locally. In this context, this study proposed a framework for dengue risk prediction by integrating big geospatial data cloud computing based on Google Earth Engine (GEE) platform and artificial intelligence modeling on the Google Colab platform. It enables defining the epidemiological calendar, delineating the predominant area of dengue transmission in cities, generating the data of risk predictors, and defining multi-date ahead prediction scenarios. We implemented the experiments based on weekly dengue cases during 2013–2020 in the Federal District and Fortaleza, Brazil to evaluate the performance of the proposed framework. Four predictors were considered, including total rainfall (R_{sum}), mean temperature (T_{mean}), mean relative humidity (RH_{mean}), and mean normalized difference vegetation index ($NDVI_{mean}$). Three models (i.e., random forest (RF), long-short term memory (LSTM), and LSTM with attention mechanism (LSTM-ATT)), and two modeling scenarios (i.e., modeling with or without dengue cases) were set to implement 1- to 4-week ahead predictions. A total of 24 models were built, and the results showed in general that LSTM and LSTM-ATT models outperformed RF models; modeling could benefit from using historical dengue cases as one of the predictors, and it makes the predicted curve fluctuation more stable compared with that only using climate and environmental factors; attention mechanism could further improve the performance of LSTM models. This study provides implications for future dengue risk prediction in terms of the effectiveness of GEE-based big geospatial data processing for risk predictor generation and Google Colab-based risk modeling and presents the benefits of using historical dengue data as one of the input features and the attention mechanism for LSTM modeling.

Keywords: dengue risk prediction; big geospatial data; Google Earth Engine; cloud deep learning; Google Colab

Citation: Li, Z. Forecasting Weekly Dengue Cases by Integrating Google Earth Engine-Based Risk Predictor Generation and Google Colab-Based Deep Learning Modeling in Fortaleza and the Federal District, Brazil. *Int. J. Environ. Res. Public Health* **2022**, *19*, 13555. https://doi.org/10.3390/ijerph192013555

Academic Editor: Paul B. Tchounwou

Received: 2 October 2022
Accepted: 18 October 2022
Published: 19 October 2022

1. Introduction

Dengue fever, one of the mosquito-borne diseases, is mainly distributed in tropical and subtropical urban and semi-urban areas worldwide [1,2]. Mosquito prevention and control is still the main measure for dengue epidemic prevention due to the lack of dengue vaccines. Dengue risk prediction is an important basis that permits providing valuable information for mosquito control decision-making. Due to global climate change, urbanization, and urban population growth, the need for efficient, accurate, and timely dengue risk prediction is even more urgent [3,4].

Different climate factors derived from weather stations, such as precipitation, temperature, and relative humidity, have been used in urban dengue risk prediction as they affect the life cycles, survival rates, and biting rates of *Aedes* mosquitoes, as well as the virus

incubation period, thereby affecting the spatio-temporal patterns of dengue epidemics [5–8]. However, weak spatial representation of weather stations and massive data downloading and analysis confront dengue risk prediction with challenges [9]. In addition, vegetation indices, such as the normalized difference vegetation index (NDVI) and enhanced vegetation index (EVI) [10–12], have been used in dengue risk prediction as they could be regarded as a proxy of vector presence [13]; however, the time-consuming downloading and processing of satellite images and ready-to-use datasets to compute vegetation indices also make dengue risk forecasting a challenge. In this case, it is crucial to propose an efficient method for facilitating the generation of dengue risk predictors.

The advancement of geospatial data cloud computing platforms facilitates the identification of dengue risk predictors. For example, Google Earth Engine (GEE) (Mountain View, CA, USA) hosts multi-petabyte satellite images (e.g., MODIS, Landsat and Sentinel) and global scale ready-to-use datasets on different topics (e.g., climate, land cover, cropland, urbanization and population), provides different algorithms for image preprocessing, spatial and temporal analysis, and image classification, and supports parallel computation [14–17], which provides unprecedented opportunities for effectively generating data of dengue risk predictors. Moreover, it can accept the upload of external geographic information system (GIS) vector data for targeting specific study areas that permits generating the data of risk predictors according to the disease data, often collected from administrative unit-based statistics [16,17]. It is thus clear that depending on the GEE platform, selecting appropriate geospatial data and identifying the risk predictors according to the epidemic area and disease data are key issues for dengue risk prediction.

Diverse artificial intelligence approaches have been used in dengue risk prediction that include, but are not limited to, generalized additive model (GAM) [18–20], random forest (RF) [10,21,22], support vector machine (SVM) [23], artificial neural networks (ANNs) [6,10,19], and long-short-term memory (LSTM) [9,24–26]. Among them, LSTM is proposed to automatically identify the characteristics of long-term trends and short-term fluctuations of time series and become popular in dengue risk prediction. Moreover, it often couples with other mechanisms to simplify the dengue risk prediction or enhance the model accuracy in real-world applications. For example, LSTM with transfer learning enables the transfer of pre-trained LSTM features from one study area to another if the two areas have comparable climate and environmental conditions and dengue epidemics [24]. LSTM with attention mechanisms (namely LSTM-ATT) adds an attention layer after LSTM architecture and can assign a weight for each hidden state, which attends to the different sequence steps for improving the power of exploiting information [26]. In addition, the parameter time step of the LSTM model, which is the length of the input time series, makes it possible to explain the effect of climate and environmental conditions on dengue risk in the past [9,24]. Deep learning approaches also play important roles in the prediction of other infectious diseases, such as the use of the neural network for weekly Zika risk prediction [27] and the use of the gated recurrent unit (GRU) for predicting the weekly influenza cases at both the city-level and state-level [28]. Despite these successful applications of deep learning models in infectious disease risk prediction, model training is still a major issue that is time-consuming and very computationally intensive.

Cloud deep learning is a good choice to accelerate model training with distributed hardware, and dengue risk prediction can benefit greatly from computing resources. For example, Google Colab (Mountain View, CA, USA), a free web platform with free resources of the Google servers, provides a serverless Jupyter notebook for interactive development, supports deep learning frameworks, and enables model training and evaluation using Tensorflow [29,30]. To date, the applications of Google Colab in public health are still limited, which has not been used in dengue risk prediction yet.

In this context, this study aims to propose a framework by integrating GEE and Google Colab for efficient and accurate city-level dengue risk prediction. Specifically, it expects to effectively generate the dengue risk predictors using big geospatial data cloud computing based on the GEE platform, and accurately predict dengue risk based on different artificial

intelligence models on Google Colab. This study expects to show the potential of integrating GEE and Google Colab in public health.

2. The Framework for Dengue Risk Prediction Based on GEE and Google Colab

A framework of efficient and accurate city-level dengue risk prediction is proposed, which includes two parts (Figure 1). Part (a) represents the steps of generating risk predictors using big geospatial data cloud computing hosted on the GEE platform. Part (b) represents the steps of defining multi-date ahead forecast scenarios, model construction, training, and evaluation using Google Colab.

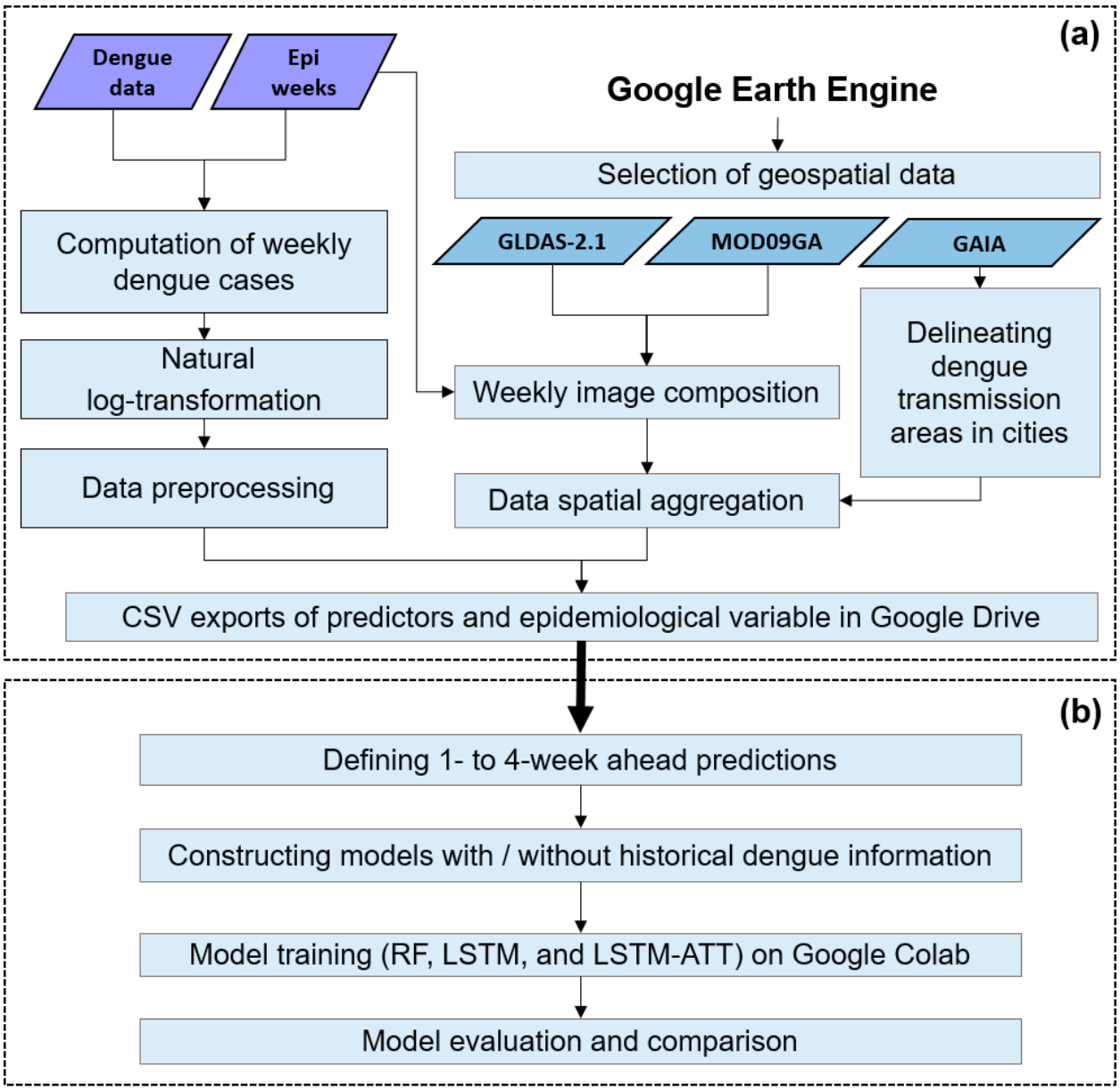

Figure 1. The proposed framework for city-level dengue risk prediction by integrating the GEE platform and Google Colab. Part (**a**) shows the steps of generating risk predictors using big geospatial data cloud computing based on the GEE platform. Part (**b**) represents the steps of defining multi-week ahead forecast scenarios, model construction, training, and evaluation using Google Colab.

Part (a) includes: (1) defining the epidemiological weeks according to an epidemiological calendar and generating the dengue-dependent factor by counting the number of dengue cases per week and naturally log-transformed weekly dengue case plus 1 to obtain

a more stationary dependent factor; (2) considering epidemiological week and generating a suite of weekly image composites by stacking sub-daily or daily images between the beginning date and the end date of epidemiological weeks and computing a value per pixel using an algorithm (e.g., minimum, maximum, mean or sum); (3) delineating the main area of dengue transmission by creating a buffer zone of 1 km around the impervious surface area and generating a time series for each predictor by spatially aggregating the values of pixels covering the buffer zone. Here, the maximum flight range of dengue vectors (i.e., 1 km for *Aedes aegypti* and *Aedes albopictus*) is considered as the distance of the buffer zone [31–33]. We assume that the majority of human activities are within the impervious surface area and *Aedes* mosquitoes prefer to live near people; (4) combining the time series of risk predictors and dengue epidemiological variable to generate a dataset in CSV format that is saved in Google Drive (Figure 1a).

Part (b) includes (1) defining multi-week ahead prediction scenarios by considering time lags in advance of dengue epidemics (i.e., 1- to 4-week ahead predictions); (2) constructing different artificial intelligence models by taking historical dengue time series as one of the input features or not; (3) training models and evaluating the performance of models by computing evaluation indices and model comparison (Figure 1b).

2.1. Models

Two deep learning models (i.e., LSTM and LSTN-ATT) were considered in this study to predict the dengue risk. In addition, we used a classical machine learning model (i.e., RF) as a baseline model to provide a comparison for understanding the performance of LSTM and LSTM-ATT.

2.1.1. LSTM

The conventional recurrent neural network (RNN) is susceptible to gradient vanishing and gradient explosion and is incapable of learning nonlinear relationships from long sequences [34]. LSTM is proposed to automatically identify the characteristics of long-term trends and short-term fluctuations of time series. A typical LSTM cell has three gates: an input gate, an output gate, and a forget gate. The forget gate (f_t) uses the sigmoid activation function to selectively forget the irrelevant information [34]:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (1)$$

where σ denotes the sigmoid function, W_f denotes the weight matrix of the unit, h_{t-1} denotes the hidden state at the previous time $t-1$, and b_f denotes the bias parameter of the unit.

The input gate (i_t) decides what to memorize using the sigmoid and tanh functions:

$$i_t = \sigma(W_i \cdot [h_{t-1},\ x_t] + b_i) \quad (2)$$

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (4)$$

where W_i and b_i represent the weight matrix and bias of the unit, respectively, C_t denotes the state of the memory unit.

The output gate (o_t) determines the exact output from the current cell and updates the historical state:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

where W_o and b_o are the weight matrix and bias of the unit. We also add a single neuron to the last layer of LSTM for obtaining the predicted label.

2.1.2. LSTM with Attention Mechanism

Although LSTM can protect and control the state of neurons by manipulating the input gate, output gate, and forget gate, solving the problems of traditional RNN neurons, it cannot provide a certain degree of interpretability for the importance of different input types of data. Here, we introduced the attention mechanism, a technology that allows the model to focus on important information and fully learn to absorb it [35]. Taking self-attention as an example, it contains three important matrices: Q, K, and V. Q is the query vector of the word, K is the "checked" vector, and V is the content vector. In this study, we introduced the attention mechanism after the last layer of LSTM model [36,37]. Suppose the embedding dimension of each hidden state is d, then $H \in R^{N \times d}$. We can obtain query, key, and value from the following projections:

$$Q = H * W^Q \tag{7}$$

$$K = H * W^K \tag{8}$$

$$V = H * W^V \tag{9}$$

where W^Q, W^K and $W^V \in R^{d \times d}$ are the weight matrices.

We then calculate attention weight and attend the value to obtain the final output by the following equation, which enables judging the importance of different hidden states:

$$H' = Softmax\left(\frac{Q * K^T}{\sqrt{d}}\right) * V \tag{10}$$

which can judge the importance of different hidden states.

2.1.3. RF

RF is a non-linear ensemble method based on decision trees. By introducing bootstrap aggregation (i.e., bagging), multiple decision trees can be integrated and are then combined to create a predictor based on the mean of each voting result from each decision tree [38]. RF is highly interpretable because the model is a tree-like diagram where each node has a partitioning rule.

2.2. Model Evaluation

The model accuracy is evaluated based on the predicted and observed weekly dengue cases by computing the root mean squared error (*RMSE*) and mean absolute error (*MAE*) as follows [39]:

$$RMSE = \sqrt{\frac{1}{n}\sum(o_i - y_i)^2}, \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|o_i - y_i|, \tag{12}$$

where o_i represents the observed value for epi week i, and y_i represents the predicted value for epi week i. The larger the indices' values, the worse the model effect.

2.3. Multi-Date Ahead Prediction Scenarios

In this study, four prediction scenarios (i.e., 1- to 4-week-ahead) with two groups of input features (i.e., modeling with climate and environmental factors, and modeling with historical dengue data and climate and environmental factors) were used.

3. Experiments

To evaluate the performance of the proposed framework, we selected the Federal District of Brazil and Fortaleza as study sites. The Federal District of Brazil is the smallest federative unit without municipalities, and intense urban land expansion and population growth make dengue a major public health issue in recent years [40–42]. Moreover, Fort-

aleza is the fifth most densely populated city in Brazil and exhibits great socioeconomic inequality where intense dengue epidemics occurred in recent decades due to population concentration and urban inequality [25,41,43,44]. The experiments of city-level forecasting of weekly dengue cases were implemented as follows: (1) generating the time series of weekly dengue cases; (2) delineating the dengue transmission areas and generating the time series of risk predictors based on the GEE platform; (3) constructing, training and evaluating multi-date ahead and multi-scenario models. The experimental steps are detailed hereafter.

3.1. Generating the Time Series of Weekly Dengue Cases

We obtained the dengue cases data for the Federal District and Fortaleza from a publicly available dataset, namely Arboviral disease record data-Dengue and Chikungunya, Brazil [41], which collects the dengue notification cases submitted to the Notifiable Diseases Information System in Brazil (SINAN) from 2013 to 2020. We counted the number of dengue cases per epidemiological week and generated a time series of 418 weekly dengue cases per city. We then calculated the natural log-transformation for weekly dengue cases plus 1 to obtain a more stable time series.

3.2. Delineating the Dengue Transmission Areas and Generating the Time Series of Risk Predictors Based on the GEE Platform

We used the impervious map of 2017 derived from the annual maps of global artificial impervious area (GAIA), a dataset containing the annual change in global impervious surface during 1985–2018 with 30 m spatial resolution [45]. We then defined a buffer zone of 1 km around the imperious surface to delimitate the predominant area of dengue transmission for each city (Figure 2). Due to no significant dynamic changes in the impervious surface during 2013–2020, we used the data of 2017 to characterize the predominant area of dengue transmission.

Four climate and environmental factors were used as risk predictors in this study, including total rainfall (R_{sum}), mean temperature (T_{mean}), mean relative humidity (RH_{mean}), and mean normalized difference vegetation index ($NDVI_{mean}$). To generate the time series with a weekly temporal resolution for each factor during 2013–2020, we first selected the data covering the predominant area of dengue transmission (Figure 2) with daily or sub-daily temporal resolution. R_{sum}, T_{mean} and RH_{mean} were derived from a global, ready-to-use dataset of land surface states and fluxes, namely Global Land Data Assimilation System Version 2.1 (GLDAS-2.1) [46], and $NDVI_{mean}$ was derived from MODIS MOD09GA dataset [47]. Then, all raster layers between the beginning date and the end date of each epidemiological week were selected, and a weekly composite was generated using a composition algorithm. Here, the sum was used to compute the total precipitation per pixel for each week, and the mean was used to compute the average status of temperature, relative humidity, and NDVI per pixel for each week. A total of 418 weekly composites were generated for each climate and environmental factor. Finally, for each weekly composite, we obtained a mean value by spatially aggregating the pixel values covering the buffer zone. A time series of 418 values was generated for each climate and environmental factor. Detailed information on big geospatial data used in this study is presented in Table 1.

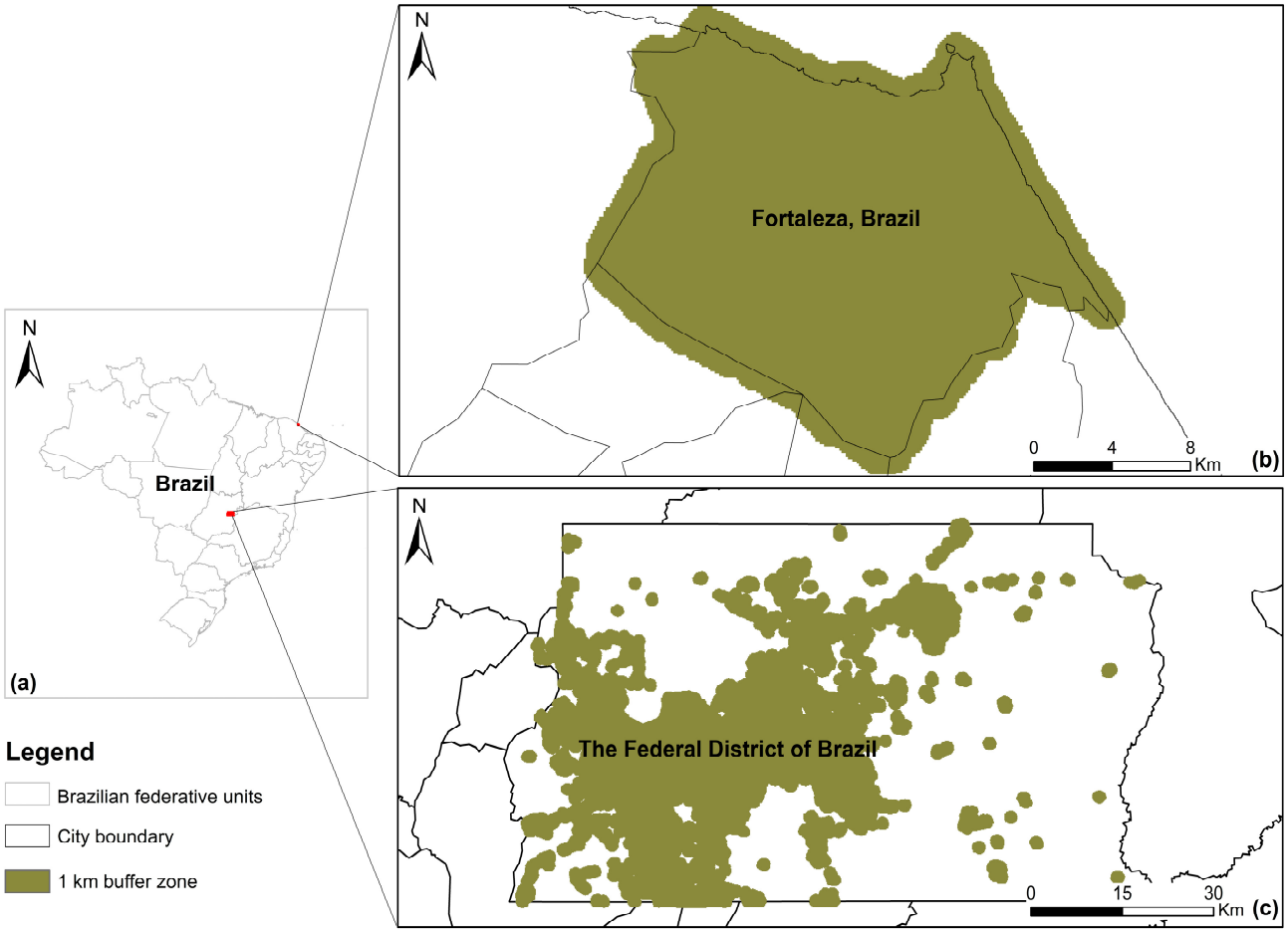

Figure 2. The geographical distribution of the study sites (**a**) and the main area of dengue transmission in the Federal District (**b**) and Fortaleza (**c**) in Brazil.

Table 1. The big geospatial data used to delineate the main area of dengue transmission in cities and generate the time series of dengue risk predictors in this study.

Dengue Risk Predictors and Epidemiological Variables		Data Sources	Spatial Resolution	Temporal Resolution	Period
Climate	Precipitation per week (R_{sum}) Mean temperature per week (T_{mean}) Mean relative humidity per week (RH_{mean})	GLDAS-2.1	27,000 m	3-hourly	2000 to present
Environment	Mean NDVI per week ($NDVI_{mean}$)	MOD09GA	500 m	Daily	2000 to present
Epidemiology	Number of dengue cases per week	Brazilian arboviral disease by [41]	City-level	Weekly	2013–2020
Dengue transmission areas	1 km buffer around the imperious surface in cities	GAIA	30 m	Annual	2017

3.3. Model Construction, Training, and Evaluation Using Google Colab

The time series of R_{sum}, T_{mean}, RH_{mean}, $NDVI_{mean}$, and naturally log-transformed weekly dengue cases were combined to generate the dataset for model training and evaluation. Datasets of the Federal District and Fortaleza were split respectively, taking data in the

first 326 weeks for model training and the rest for model evaluation. To fully compare the performance of RF, LSTM, and LSTM-ATT, we considered time lags in advance of dengue epidemics and defined 1- to 4-week ahead prediction scenarios with two groups of input features (i.e., climate and environmental factors with and without historical dengue data. A total of 24 models were built. RF models were implemented using an open-source package scikit-learn [48] for Python 3.7, and LSTM and LSTM-ATT models were implemented based on Tensorflow 2.8.2 and Python 3.7. All experiments were run in Google Colab. RF models were trained using default parameters, and the parameters of LSTM and LSTM-ATT models are listed in Table 2.

Table 2. Summary of the parameters in LSTM and LSTM-ATT models. The time step denotes the length of the input features to make the prediction. The loss function quantifies the differences between the predicted value and the ground truth. The number of units is the number of neurons in the LSTM layer. The epoch is the number of completed training, while all data in the training set are used. Batch size represents the size of data in each batch for training the model. The learning rate refers to the step rate for updating parameters in backpropagation. Optimizer is the updating algorithm. Attention size is the embedding dimension of the weight matrix in the attention layer.

Parameters	LSTM without Dengue Cases	LSTM with Dengue Cases	LSTM-ATT without Dengue Cases	LSTM-ATT with Dengue Cases
Time step	12	12	12	12
Loss function	MSE	MSE	MSE	MSE
Number of units	64	64	64	64
Epoch	1000	1500	1500	2000
Batch size	12	12	12	12
Learning rate	0.005	0.003	0.005	0.003
Optimizer	Adam	Adam	Adam	Adam
Attention Size	-	-	64	64

4. Results

4.1. Time Series of Climate and Environmental Factors and Weekly Dengue Cases

Figure 3 presents the temporal pattern of weekly dengue cases in the Federal District and Fortaleza in Brazil during 2013–2020. Dengue outbreaks could be observed each year for the two cities, and similar temporal patterns were observed, with the epidemic season mainly from February to May each year.

4.2. Outcomes of RF, LSTM and LSTM-ATT Modeling

Figure 4 and Table A1 show the prediction accuracies in the Federal District and Fortaleza in Brazil. Figures 5 and 6 present the predicted and observed curves of weekly dengue cases during 2019–2020 in the Federal District and Fortaleza, respectively. We can draw the following general conclusions:

1. RF models frequently have higher prediction errors than LSTM and LSTM-ATT models, and introducing historical dengue data as one of the input features can improve the performance of RF models for 1- to 4-week ahead predictions (see the yellow lines in Figure 4). Most of the predicted curves of RF models differed greatly from those of the observed cases, especially on the dataset of the Federal District of Brazil (Figure 5c).
2. All the blue and red lines are relatively stable in Figure 4, which suggests a slight difference in the accuracies of the 1- to 4-week ahead predictions for both LSTM and LSTM-ATT models.
3. The red curves with squares are frequently above the blue curves with squares in Figure 4, which suggests that LSTM modeling can benefit from an attention mechanism, which can further achieve performance lift in most cases. Similarly, the red curves with triangles are frequently above the blue curve with triangles in Figure 4,

which suggests that LSTM-ATT modeling can also benefit from an attention mechanism.

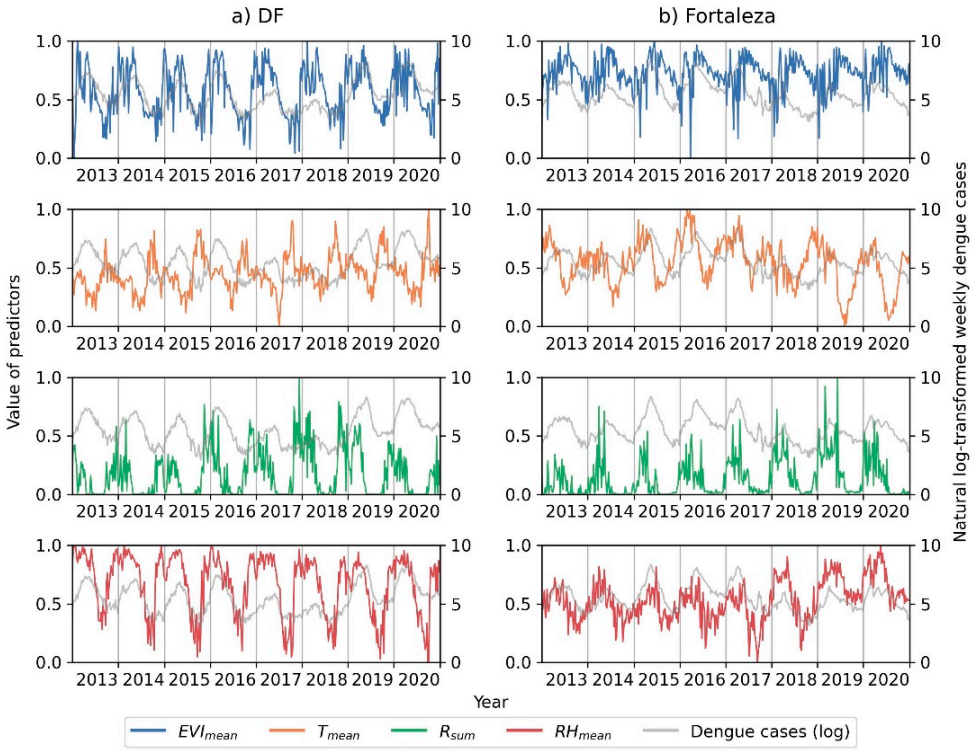

Figure 3. Illustration of the time series of climate and vegetation factors and natural log-transformed weekly dengue cases during 2013–2020 in the Federal District (**a**) and Fortaleza (**b**) in Brazil.

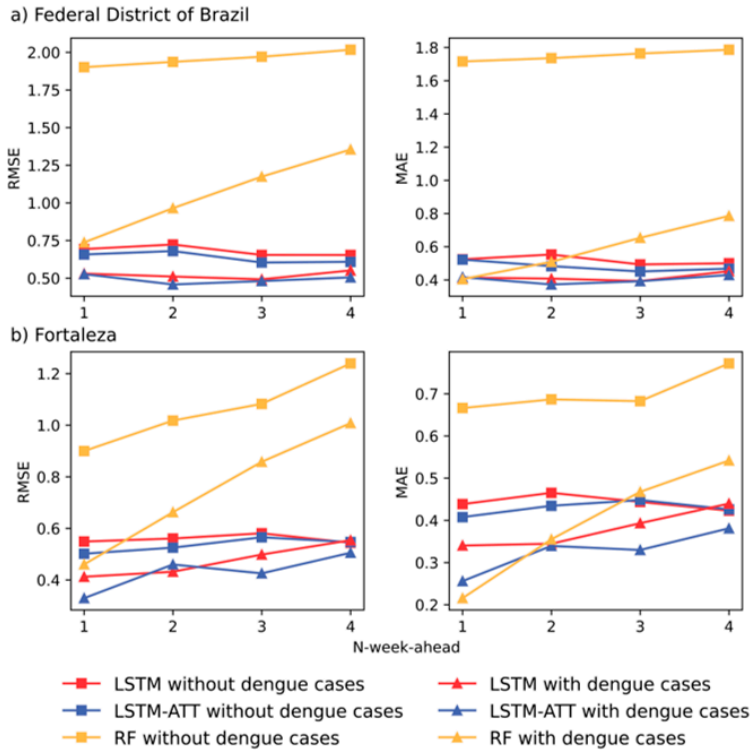

Figure 4. Accuracy comparison of multi-step ahead RF, LSTM and LSTM-ATT modeling with two groups of input features (i.e., with or without historical dengue data) using RMSE and MAE.

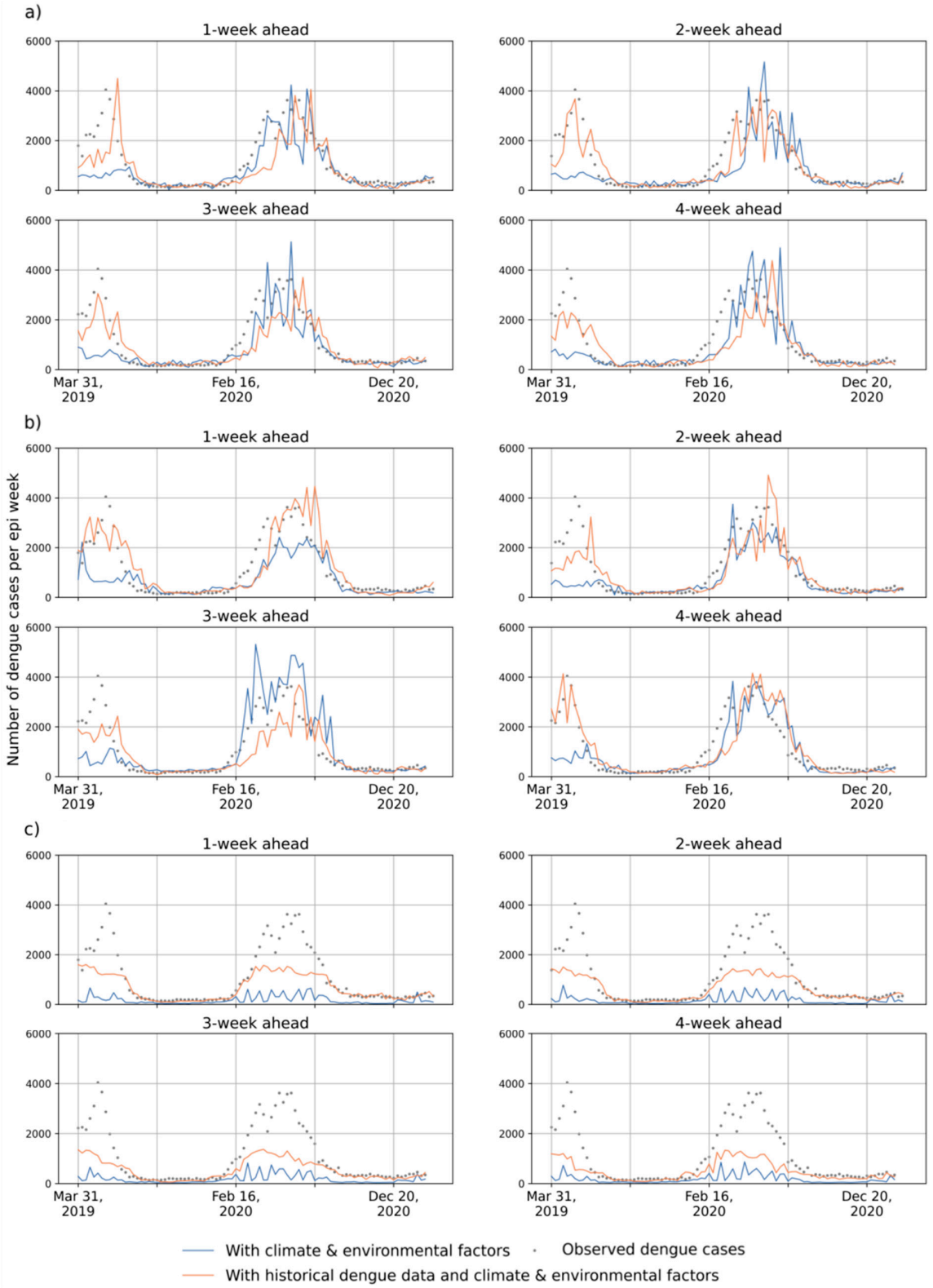

Figure 5. Illustration of the 1- to 4-week-ahead prediction with two groups of the input features for the dataset of the Federal District using (**a**) LSTM, (**b**) LSTM-ATT, and (**c**) RF, respectively. The grey points represent the number of observed cases per week. Orange curves represent the number of predicted cases per week with historical dengue data, climate factors, and vegetation factors. Blue curves represent the number of predicted cases per epi week with climate and vegetation factors.

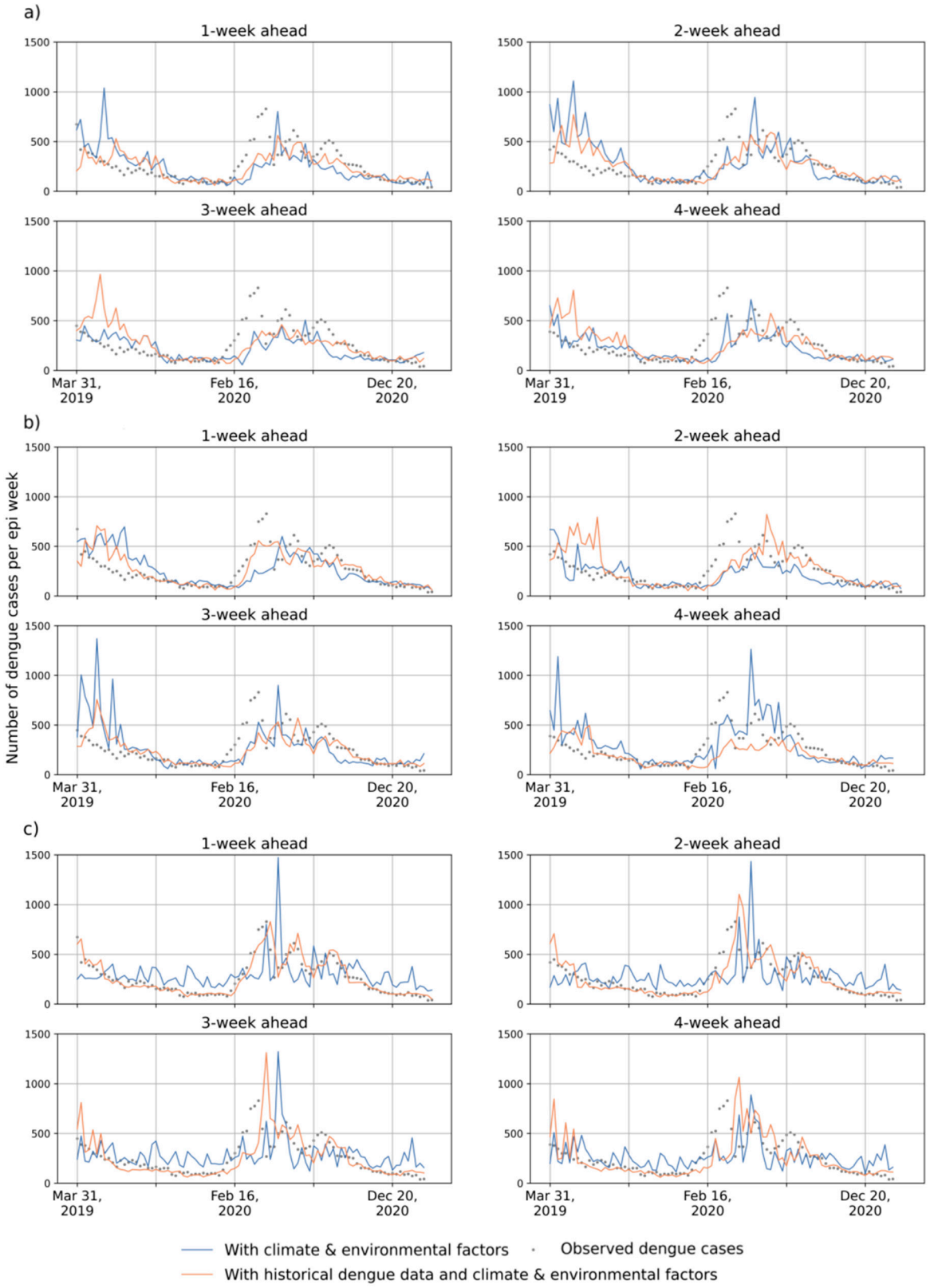

Figure 6. Illustration of the 1- to 4-week-ahead prediction with two groups of the input features for the Fortaleza using (**a**) LSTM, (**b**) LSTM-ATT, and (**c**) RF, respectively. The grey points represent the number of observed cases per epi week. Orange curves represent the number of predicted cases per epi week with historical dengue data, climate factors, and vegetation factors. Blue curves represent the number of predicted cases per epi week with climate and vegetation factors.

The red curves with squares are frequently above the red curves with triangles in Figure 4, which suggests that LSTM modeling can benefit from using historical dengue data as one of the input features. Similar results could be observed for LSTM-ATT modeling, where the blue curves with squares are frequently above the blue curves with triangles. Moreover, the predicted curves are more stable compared to those of the predictions only with climate and environmental factors (Figures 5 and 6).

5. Discussion

This study developed an efficient, accurate, and timely framework for city-level prediction of weekly dengue cases by integrating the GEE-based generation of risk predictors time series, historical dengue data, and Google Colab-based modeling. This study demonstrates the potential of multi-source geospatial data and cloud computing for the generation of dengue risk predictors and the power of Google Colab for developing various machine learning and deep learning models to predict dengue risk in advance.

The GEE platform allows assessing big geospatial data freely, avoids downloading and preprocessing multi-source data, and improves the efficiency and timeliness of generating the time series of dengue risk predictors using parallel data processing in the Google Cloud. A recent review of big geospatial data and data-driven models for dengue risk prediction shows that the GEE platform hosts global-scale satellite images and ready-to-use products on different topics that provide sufficient data sources to identify diverse driving factors at different spatial (e.g., health units, neighborhood, city, state/province, and country) and temporal scales (e.g., weekly and monthly) [49]. Additionally, a GEE-based web application has been developed to support malaria early warning by effectively generating environmental factors (e.g., daytime and nighttime land surface temperature, vegetation indices, and total precipitation) at the district scale in Ethiopia [16], which confirmed the effectiveness of the GEE platform in disease early warning.

We found that LSTM-ATT modeling with historical time series of weekly dengue cases and other driving factors frequently outperformed RF and LSTM models. Previous studies also confirmed that historical dengue data of a specific epidemic area provide the temporal characteristics of dengue transmission and help to improve the prediction accuracy of future dengue cases for the area and its neighboring areas [24,49,50]. In addition, a previous study forecasted the monthly dengue incidence rates for 20 provinces in Vietnam and confirmed that LSTM-ATT frequently outperformed other deep learning models [26]. Despite the good performance of LSTM and LSTM-ATT models in this study, RF has been used in several studies as it is easy to be implemented. Some facts limit the prediction accuracy of RF. It cannot quantify the relationships between the time series of dengue cases and risk predictors using a specific equation and may suffer from extrapolation problems, with predicted values being hard to be beyond the range of that in the training set [10]. That is to say, an underestimation of dengue cases can be observed while unprecedented outbreaks occur. By contrast, LSTM is capable of capturing the long-term dependency and non-linearity in the complex system of dengue transmission and permits adjusting the parameter time step to better quantify the impact of climate and environmental factors on dengue transmission; however, LSTM models lose information due to passing information across several sequence steps, and thus it will be worse in a long sequence. The problem can be mitigated by integrating LSTM with the attention mechanism, which enhances the power of information exploration by creating output for each sequence step. Besides, attention also provides a certain degree of interpretability for the importance of different hidden states [26].

The experiments in the Federal District and Fortaleza in Brazil confirmed the feasibility of the proposed framework to a certain extent; however, there are some limitations while using the proposed framework in real-world applications. For example, many climate and environmental factors could be considered in the dengue risk prediction based on the GEE platform; however, other important factors could not be quantified, such as the cycle of dengue genotype, population mobility, mosquito population, and immune status [51–54].

Especially, population mobility, an important factor influencing the spatial spread of dengue fever between geographical units (e.g., between cities or provinces), has been characterized using mobile phone data and public transportation data and used as one of the input features to improve the accuracy of dengue risk prediction in recent studies [25,52]. Future studies should consider how to explore big geospatial data to characterize the population mobility proxy and use it together with GEE-based climate and environmental factors as input features to model the future dengue risk. Moreover, other deep learning models dealing with time series risk prediction, such as GRU and Transformer, have been used in risk prediction of infectious diseases [26,55], which should be integrated into the proposed framework to understand the performance and time consumption. Lastly, this study solely adapted to the prediction of dengue cases 1 to 4 weeks in advance, which provides information on the temporal dynamics of dengue risk. Future studies should consider how to integrate more needs for dengue prevention and control into the proposed framework, such as predicting the peak intensity and peak timing of dengue epidemics and dengue outbreaks [25,56].

This study highlights the potential of the GEE platform and Google Colab in dengue risk prediction and also shows the benefits of using historical dengue data as one of the input features and attention mechanisms while LSTM modeling, which has important implications for future dengue risk prediction in terms of improving the effectiveness and accuracy of future dengue risk prediction. Early and accurate information on dengue transmission risk can guide the decision-making for dengue prevention and control and allow more time for the implementation of strategies.

6. Conclusions

Climate change, urbanization, and population growth highlight the importance of efficient and accurate dengue risk prediction. Multi-source data downloading and processing to identify dengue risk predictors and significant time and computational resources consumed to develop deep learning models locally make dengue risk prediction a challenge. This study used GEE-based geospatial data cloud computing to effectively generate the time series of climate and environmental factors and proposed accurate forecasting of weekly dengue cases based on LSTM modeling and an attention mechanism using Google Colab, considering total precipitation, mean temperature, mean relative humidity, mean NDVI, and historical dengue cases as input features. Our findings show the great potential of GEE-based geospatial data analysis and Google Colab-based deep learning modeling for facilitating dengue risk prediction for broader use in public health.

Funding: This research was funded by the Key Research Program of Frontier Sciences (QYZDB-SSW-DQC005) of the Chinese Academy of Sciences (CAS), the Strategic Priority Research Program (XDA19040301) of the CAS, and the Institute of Geographic Sciences and Natural Resources Research (IGNSRR), Chinese Academy of Sciences (CAS) (E0V00110YZ).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The data of weekly dengue cases and risk predictors and the codes for RF, LSTM and LSTM-ATT modeling are available online at https://github.com/geeigsnrr/Dengue-risk-forecasting-IJERPH (accessed on 9 October 2022).

Appendix B

Table A1. The RMSE and MAE values of multi-step ahead RF, LSTM, and LSTM-ATT models with two groups of input features for the Federal District and Fortaleza in Brazil.

Models			Federal District of Brazil		Fortaleza	
			RMSE	MAE	RMSE	MAE
LSTM	LSTM without dengue cases	1-week	0.6929	0.5238	0.5488	0.4385
		2-week	0.7231	0.5527	0.5606	0.4652
		3-week	0.6540	0.4930	0.5808	0.4437
		4-week	0.6535	0.4999	0.5439	0.4229
	LSTM with dengue cases	1-week	0.5299	0.4137	0.4123	0.3403
		2-week	0.5105	0.4074	0.4317	0.3446
		3-week	0.4920	0.3931	0.4983	0.3932
		4-week	0.5509	0.4519	0.5539	0.4397
	LSTM-ATT without dengue cases	1-week	0.6570	0.5222	0.5017	0.4077
		2-week	0.6798	0.4827	0.5251	0.4345
		3-week	0.6037	0.4505	0.5653	0.4481
		4-week	0.6083	0.4677	0.5467	0.4260
	LSTM-ATT with dengue cases	1-week	0.5265	0.4162	0.3292	0.2560
		2-week	0.4579	0.3723	0.4598	0.3394
		3-week	0.4805	0.3920	0.4254	0.3298
		4-week	0.5049	0.4295	0.5053	0.3811
RF	RF without dengue cases	1-week	1.9010	1.7157	0.8998	0.6661
		2-week	1.9362	1.7358	1.0179	0.6866
		3-week	1.9702	1.7637	1.0827	0.6824
		4-week	2.0166	1.7864	1.2393	0.7717
	RF with dengue cases	1-week	0.7371	0.4041	0.4601	0.2156
		2-week	0.9646	0.5087	0.6626	0.3550
		3-week	1.1738	0.6531	0.8581	0.4675
		4-week	1.3540	0.7855	1.0079	0.5417

References

1. Horstick, O.; Tozan, Y.; Wilder-Smith, A. Reviewing dengue: Still a neglected tropical disease? *PLoS Negl. Trop. Dis.* **2015**, *9*, e0003632. [CrossRef] [PubMed]
2. Bhatt, S.; Gething, P.W.; Brady, O.J.; Messina, J.P.; Farlow, A.W.; Moyes, C.L.; Drake, J.M.; Brownstein, J.S.; Hoen, A.G.; Sankoh, O.; et al. The global distribution and burden of dengue. *Nature* **2013**, *496*, 504–507. [CrossRef]
3. Ryan, S.J.; Carlson, C.J.; Mordecai, E.A.; Johnson, L.R. Global expansion and redistribution of Aedes-borne virus transmission risk with climate change. *PLoS Negl. Trop. Dis.* **2019**, *13*, e0007213. [CrossRef]
4. Yang, S.; Kou, S.C.; Lu, F.; Brownstein, J.S.; Brooke, N.; Santillana, M. Advances in using Internet searches to track dengue. *PLoS Comput. Biol.* **2017**, *13*, e1005607. [CrossRef]
5. Withanage, G.P.; Viswakula, S.D.; Nilmini Silva Gunawardena, Y.I.; Hapugoda, M.D. A forecasting model for dengue incidence in the District of Gampaha, Sri Lanka. *Parasites Vectors* **2018**, *11*, 262. [CrossRef] [PubMed]
6. Polwiang, S. The time series seasonal patterns of dengue fever and associated weather variables in Bangkok (2003–2017). *BMC Infect. Dis.* **2020**, *20*, 208. [CrossRef] [PubMed]
7. Estallo, E.L.; Benitez, E.M.; Lanfri, M.A.; Scavuzzo, C.M.; Almirón, W.R. MODIS Environmental Data to Assess Chikungunya, Dengue, and Zika Diseases Through Aedes (Stegomia) aegypti Oviposition Activity Estimation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 5461–5466. [CrossRef]
8. Jain, R.; Sontisirikit, S.; Iamsirithaworn, S.; Prendinger, H. Prediction of dengue outbreaks based on disease surveillance, meteorological and socio-economic data. *BMC Infect. Dis.* **2019**, *19*, 272. [CrossRef]
9. Li, Z.; Gurgel, H.; Xu, L.; Yang, L.; Dong, J. Improving Dengue Forecasts by Using Geospatial Big Data Analysis in Google Earth Engine and the Historical Dengue Information-Aided Long Short Term Memory Modeling. *Biology* **2022**, *11*, 169. [CrossRef]
10. Zhao, N.; Charland, K.; Carabali, M.; Nsoesie, E.O.; Maheu-Giroux, M.; Rees, E.; Yuan, M.; Garcia Balaguera, C.; Jaramillo Ramirez, G.; Zinszer, K. 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.* **2020**, *14*, e0008056. [CrossRef]

11. Buczak, A.L.; Baugher, B.; Moniz, L.J.; Bagley, T.; Babin, S.M.; Guven, E. Ensemble method for dengue prediction. *PLoS ONE* **2018**, *13*, e0189988. [CrossRef]
12. Chen, Y.; Ong, J.H.Y.; Rajarethinam, J.; Yap, G.; Ng, L.C.; Cook, A.R. Neighbourhood level real-time forecasting of dengue cases in tropical urban Singapore. *BMC Med.* **2018**, *16*, 129. [CrossRef]
13. Marti, R.; Li, Z.; Catry, T.; Roux, E.; Mangeas, M.; Handschumacher, P.; Gaudart, J.; Tran, A.; Demagistri, L.; Faure, J.-F.; et al. A Mapping Review on Urban Landscape Factors of Dengue Retrieved from Earth Observation Data, GIS Techniques, and Survey Questionnaires. *Remote Sens.* **2020**, *12*, 932. [CrossRef]
14. Tamiminia, H.; Salehi, B.; Mahdianpari, M.; Quackenbush, L.; Adeli, S.; Brisco, B. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS J. Photogramm. Remote Sens.* **2020**, *164*, 152–170. [CrossRef]
15. Amani, M.; Ghorbanian, A.; Ahmadi, S.A.; Kakooei, M.; Moghimi, A.; Mirmazloumi, S.M.; Moghaddam, S.H.A.; Mahdavi, S.; Ghahremanloo, M.; Parsian, S.; et al. Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5326–5350. [CrossRef]
16. Wimberly, M.C.; Nekorchuk, D.M.; Kankanala, R.R. Cloud-based applications for accessing satellite Earth observations to support malaria early warning. *Sci. Data* **2022**, *9*, 208. [CrossRef]
17. Frake, A.N.; Peter, B.G.; Walker, E.D.; Messina, J.P. Leveraging big data for public health: Mapping malaria vector suitability in Malawi with Google Earth Engine. *PLoS ONE* **2020**, *15*, e0235697. [CrossRef]
18. Carvajal, T.M.; Viacrusis, K.M.; Hernandez, L.F.T.; Ho, H.T.; Amalin, D.M.; Watanabe, K. Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan Manila, Philippines. *BMC Infect. Dis.* **2018**, *18*, 183. [CrossRef]
19. Baquero, O.S.; Santana, L.M.R.; Chiaravalloti-Neto, F. Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PLoS ONE* **2018**, *13*, e0195065. [CrossRef]
20. Liu, D.; Guo, S.; Zou, M.; Chen, C.; Deng, F.; Xie, Z.; Hu, S.; Wu, L. A dengue fever predicting model based on Baidu search index data and climate data in South China. *PLoS ONE* **2019**, *14*, e0226841. [CrossRef]
21. Mussumeci, E.; Codeco Coelho, F. Large-scale multivariate forecasting models for Dengue—LSTM versus random forest regression. *Spat Spatiotemporal Epidemiol.* **2020**, *35*, 100372. [CrossRef]
22. Benedum, C.M.; Shea, K.M.; Jenkins, H.E.; Kim, L.Y.; Markuzon, N. Weekly dengue forecasts in Iquitos, Peru; San Juan, Puerto Rico; and Singapore. *PLoS Negl. Trop. Dis.* **2020**, *14*, e0008710. [CrossRef]
23. Liu, K.; Yin, L.; Zhang, M.; Kang, M.; Deng, A.-P.; Li, Q.-L.; Song, T. Facilitating fine-grained intra-urban dengue forecasting by integrating urban environments measured from street-view images. *Infect. Dis. Poverty* **2021**, *10*, 40. [CrossRef]
24. Xu, J.; Xu, K.; Li, Z.; Meng, F.; Tu, T.; Xu, L.; Liu, Q. Forecast of Dengue Cases in 20 Chinese Cities Based on the Deep Learning Method. *Int. J. Env. Res. Public Health* **2020**, *17*, 453. [CrossRef]
25. Bomfim, R.; Pei, S.; Shaman, J.; Yamana, T.; Makse, H.A.; Andrade, J.S., Jr.; Lima Neto, A.S.; Furtado, V. Predicting dengue outbreaks at neighbourhood level using human mobility in urban areas. *J. R. Soc. Interface* **2020**, *17*, 20200691. [CrossRef]
26. Nguyen, V.-H.; Tuyet-Hanh, T.T.; Mulhall, J.; Minh, H.V.; Duong, T.Q.; Chien, N.V.; Nhung, N.T.T.; Lan, V.H.; Minh, H.B.; Cuong, D.; et al. Deep learning models for forecasting dengue fever based on climate data in Vietnam. *PLoS Negl. Trop. Dis.* **2022**, *16*, e0010509. [CrossRef]
27. Akhtar, M.; Kraemer, M.U.G.; Gardner, L.M. A dynamic neural network model for predicting risk of Zika in real time. *BMC Med.* **2019**, *17*, 171. [CrossRef]
28. Aiken, E.L.; Nguyen, A.T.; Viboud, C.; Santillana, M. Toward the use of neural networks for influenza prediction at multiple spatial resolutions. *Sci. Adv.* **2021**, *7*, eabb1237. [CrossRef]
29. Carneiro, T.; NóBrega, R.V.M.D.; Nepomuceno, T.; Bian, G.B.; Albuquerque, V.H.C.D.; Filho, P.P.R. Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access* **2018**, *6*, 61677–61685. [CrossRef]
30. Bisong, E. Google colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*; Bisong, E., Ed.; Apress: Berkeley, CA, USA, 2019; pp. 59–64.
31. Tsunoda, T.; Cuong, T.C.; Dong, T.D.; Yen, N.T.; Le, N.H.; Phong, T.V.; Minakawa, N. Winter refuge for Aedes aegypti and Ae. albopictus mosquitoes in Hanoi during Winter. *PLoS ONE* **2014**, *9*, e95606. [CrossRef]
32. Maciel-de-Freitas, R.; Neto, R.B.; Gonçalves, J.M.; Codeço, C.T.; Lourenço-de-Oliveira, R. Movement of dengue vectors between the human modified environment and an urban forest in Rio de Janeiro. *J. Med. Entomol.* **2006**, *43*, 1112–1120. [CrossRef]
33. Lacroix, R.; Delatte, H.; Hue, T.; Reiter, P. Dispersal and survival of male and female Aedes albopictus (Diptera: Culicidae) on Réunion Island. *J. Med. Entomol.* **2009**, *46*, 1117–1124. [CrossRef]
34. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
35. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–15. [CrossRef]
36. Li, Y.; Zhu, Z.; Kong, D.; Han, H.; Zhao, Y. EA-LSTM: Evolutionary attention-based LSTM for time series prediction. *Knowl. -Based Syst.* **2019**, *181*, 104785. [CrossRef]
37. Wang, Y.; Huang, M.; Zhu, X.; Zhao, L. Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 606–615.
38. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; Routledge: London, UK, 2017.

39. Hyndman, R.J.; Koehler, A.B. Another look at measures of forecast accuracy. *Int. J. Forecast.* **2006**, *22*, 679–688. [CrossRef]
40. Li, Z.; Gurgel, H.; Li, M.; Dessay, N.; Gong, P. Urban Land Expansion from Scratch to Urban Agglomeration in the Federal District of Brazil in the Past 60 Years. *Int. J. Environ. Res. Public Health* **2022**, *19*, 1032. [CrossRef]
41. da Silva Neto, S.R.; Tabosa de Oliveira, T.; Teixiera, I.V.; Medeiros Neto, L.; Souza Sampaio, V.; Lynn, T.; Endo, P.T. Arboviral disease record data—Dengue and Chikungunya, Brazil, 2013–2020. *Sci. Data* **2022**, *9*, 198. [CrossRef]
42. Drumond, B.; Angelo, J.; Xavier, D.R.; Catao, R.; Gurgel, H.; Barcellos, C. Dengue spatiotemporal dynamics in the Federal District, Brazil: Occurrence and permanence of epidemics. *Cien Saude Colet* **2020**, *25*, 1641–1652. [CrossRef]
43. MacCormack-Gelles, B.; Lima Neto, A.S.; Sousa, G.S.; Nascimento, O.J.; Machado, M.M.T.; Wilson, M.E.; Castro, M.C. Epidemiological characteristics and determinants of dengue transmission during epidemic and non-epidemic years in Fortaleza, Brazil: 2011–2015. *PLoS Negl. Trop. Dis.* **2018**, *12*, e0006990. [CrossRef]
44. Charlesworth, S.M.; Kligerman, D.C.; Blackett, M.; Warwick, F. The Potential to Address Disease Vectors in Favelas in Brazil Using Sustainable Drainage Systems: Zika, Drainage and Greywater Management. *Int. J. Environ. Res. Public Health* **2022**, *19*, 2860. [CrossRef] [PubMed]
45. Gong, P.; Li, X.; Wang, J.; Bai, Y.; Chen, B.; Hu, T.; Liu, X.; Xu, B.; Yang, J.; Zhang, W.; et al. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sens. Environ.* **2020**, *236*, 111510. [CrossRef]
46. Rodell, M.; Houser, P.R.; Jambor, U.; Gottschalck, J.; Mitchell, K.; Meng, C.-J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* **2004**, *85*, 381. [CrossRef]
47. Vermote, E.; Wolfe, R. MOD09GA MODIS/Terra Surface Reflectance Daily L2G Global 1kmand 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC. Available online: https://doi.org/10.5067/MODIS/MOD09GA.006 (accessed on 15 July 2022).
48. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
49. Li, Z.; Dong, J. Big Geospatial Data and Data-Driven Methods for Urban Dengue Risk Forecasting: A Review. *Remote Sens.* **2022**, *14*, 5052. [CrossRef]
50. Zhang, Y.; Wang, T.; Liu, K.; Xia, Y.; Lu, Y.; Jing, Q.; Yang, Z.; Hu, W.; Lu, J. Developing a Time Series Predictive Model for Dengue in Zhongshan, China Based on Weather and Guangzhou Dengue Surveillance Data. *PLoS Negl. Trop. Dis.* **2016**, *10*, e0004473. [CrossRef]
51. McGough, S.F.; Clemente, L.; Kutz, J.N.; Santillana, M. A dynamic, ensemble learning approach to forecast dengue fever epidemic years in Brazil using weather and population susceptibility cycles. *J. R. Soc. Interface* **2021**, *18*, 20201006. [CrossRef]
52. Kiang, M.V.; Santillana, M.; Chen, J.T.; Onnela, J.P.; Krieger, N.; Engo-Monsen, K.; Ekapirat, N.; Areechokchai, D.; Prempree, P.; Maude, R.J.; et al. Incorporating human mobility data improves forecasts of Dengue fever in Thailand. *Sci. Rep.* **2021**, *11*, 923. [CrossRef]
53. Sanchez, L.; Vanlerberghe, V.; Alfonso, L.; Marquetti, M.d.C.; Guzman, M.G.; Bisset, J.; van der Stuyft, P. Aedes aegypti larval indices and risk for dengue epidemics. *Emerg. Infect. Dis.* **2006**, *12*, 800–806. [CrossRef]
54. Ong, J.; Aik, J.; Ng, L.C. Short Report: Adult Aedes abundance and risk of dengue transmission. *PLoS Negl. Trop. Dis.* **2021**, *15*, e0009475. [CrossRef]
55. Shahid, F.; Zameer, A.; Muneeb, M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos Solitons Fractals* **2020**, *140*, 110212. [CrossRef] [PubMed]
56. Bracher, J.; Ray, E.L.; Gneiting, T.; Reich, N.G. Evaluating epidemic forecasts in an interval format. *PLoS Comput. Biol.* **2021**, *17*, e1008618. [CrossRef] [PubMed]