RESEARCH

Prediction of Dengue Outbreaks Based on Disease Surveillance, Meteorological and Socio-Economic Data

Raghvendra Jain[1*], Sra Sontisirikit[2], Sopon Iamsirithaworn[3] and Helmut Prendinger[1]

*Correspondence:
raghavendra.jain@gmail.com
[1]National Institute of Informatics, Tokyo, Japan
Full list of author information is available at the end of the article

Abstract

Background: The goal of this research is to create a system that can use the available relevant information about the factors responsible for the spread of dengue and; use it to predict the occurrence of dengue within a geographical region, so that public health experts can prepare for, manage and control the epidemic. Our study presents new geospatial insights into our understanding and management of health, disease and health-care systems.

Methods: We present a machine learning-based methodology capable of providing forecast estimates of dengue prediction in each of the fifty districts of Thailand by leveraging data from multiple data sources. Using a set of prediction variables, we show an increase in prediction accuracy of the model with an optimal combination of predictors which include: meteorological data, clinical data, lag variables of disease surveillance, socioeconomic data and the data encoding spatial dependence on dengue transmission. We use Generalized Additive Models (GAMs) to fit the relationships between the predictors (with a lag of one month) and the clinical data of Dengue hemorrhagic fever (DHF) using the data from 2008 to 2012. Using the data from 2013 to 2015 and a comparative set of prediction models, we evaluate the predictive ability of the fitted models according to RMSE and SRMSE as well as using adjusted R-squared value, deviance explained and change in AIC.

Results: In this paper, we present a model which allows for combining different predictors to make forecasts with a lead time of one month and also describe the statistical significance of the variables used to characterize the forecast. The discriminating ability of the final model was evaluated against Bangkok specific *constant* threshold and WHO *moving* threshold of the epidemic in terms of specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV).

Conclusions: Along with the high accuracy of prediction, we also determine that for the prediction of dengue outbreaks within a district, the influence of dengue incidences and socioeconomic data from the surrounding districts is statistically significant. This suggests the influence of movement patterns of people and spatial heterogeneity of human activities on the spread of the epidemic.

Keywords: Dengue forecasting; Data-driven Epidemiology; Disease Surveillance; Generalized Additive Models (GAMs)

1 Background

Dengue, a mosquito-borne viral disease, is caused by four distinct, but closely related, serotypes of the virus [1, 2]. Recovery from infection by one of these four (DEN-1, DEN-2, DEN-3, and DEN-4) provides the infected person lifelong immunity against that particu-

lar serotype and cross-immunity to the other serotypes. The time duration for this cross-immunity is 6-12 months [1]. If the person is infected by other serotypes subsequently then the risk of severe dengue increases. The uninfected mosquitoes get the virus from infected humans and thus the later becomes the primary carrier, multiplier, and transmitter of the DENV (dengue virus).

Thailand began to experience Dengue fever in 1949 and it became pandemic in the country for the first time in 1958 in Bangkok [3]. The information about the clinically diagnosed cases of dengue fever(DF), dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS) are sent to the Epidemiology Department in Bangkok [4] via the Provincial Health Offices. The transmission of DENV which occurs through the bite of infected Aedes mosquitoes, principally Aedes aegypti, has dramatically increased in the last two decades [5] and occurrence of dengue fever is likely to rise [6] due to the combination of several direct and indirect factors. The causal dependency of some of the prominent factors on DENV transmission is shown shown in Figure 1.

These factors include density of infected mosquitoes, (level of) immunity of people to dengue serotypes, meteorology, human related factors e.g. housing type, population density and demographics, cleanliness etc. Several potential predictive indicators that contribute to the above mentioned factors have been described[23, 30–32]. For example, the density of infected mosquitoes depends on total mosquito density [13, 25] which depends on inside house breeding [24] and outside house breeding [23]. The inside house breeding depends on cleaning activity [27, 28] and housing type [18] among other factors whereas outside house breeding is primarily influenced by neglected garbage [22], population density [19–21] and rain [11]. The seasonal influence on rain [7, 8] and temperature [8–10] contributes to the mosquito density [9, 11, 23, 32, 33]. The density of infected mosquitoes in the region is influenced by dengue rate in the region [16, 17] and so is the number of people with immunity [12, 13] along with the population demographics [6, 14, 15].

Despite the numerous urban outbreaks of dengue with significant health and economic impact [34–37], the detailed surveillance for diagnosing dengue has been limited which makes it difficult to generate detailed information on its epidemiology [38, 39]. Moreover, there are currently no licensed vaccines or specific therapeutics for the treatment of the infected people. Thus, effective vector control interventions are the only way to control the transmission of dengue and other Aedes-borne arboviral diseases. A variety of dengue vector control strategies [6] have been adopted in different regions[40–43] but this has not stopped its rapid emergence and global spread [44].

The complexity of early warning systems (EWS) arise due to the involvement of various factors such as environmental, climatic or geographic ones along with the well-studied transmission patterns between the different animal, human or vector components. Prediction forms an important part of surveillance systems and more specifically in EWS. To predict the future outbreaks using information on the risk factors of the disease, epidemiological models have been proposed ([45] for a review). These prediction models help in decision making processes concerning control purposes and surveillance methods. A large number of them [46–48] focused on modeling climate impact (temperature and precipitation data) on dengue transmission. In a few studies, along with the climate data, other covariates were also used on a longitude-latitude grid with time lags relevant to dengue transmission. These covariates incorporated relevant socio-economic and environmental variables [49, 50], socio-geographical factors[51, 52], imported cases[53, 54] as well human movement patterns[55, 56]. We found that for the modeling process, many previous

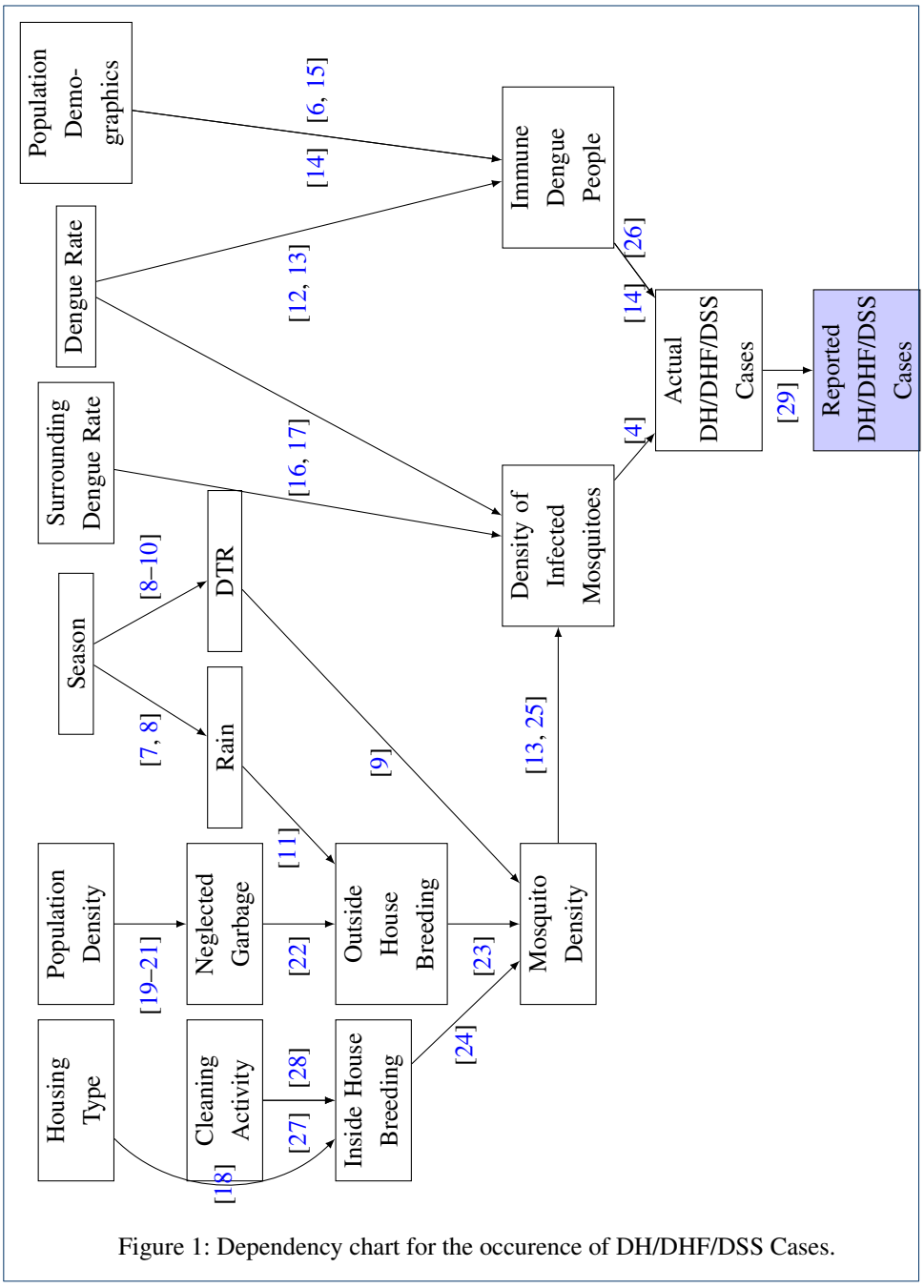


Figure 1: Dependency chart for the occurence of DH/DHF/DSS Cases.

studies have not sufficiently accounted for the *integration* of spatio-temporal features of the disease, its socio-environmental aspects and factors due to increased movements of people in a single prediction model. However, incorporating such information and understanding the relative importance of one risk factor over the other is important for their use in an early warning system.

Therefore, the goal of this research is to create a system that can use the available relevant information about the factors responsible for the spread of dengue and; use it to predict the occurrence of dengue within a geographical region, so that public health experts can prepare for, manage and control the epidemic. Our study presents new geospatial insights into our understanding and management of health, disease and health care systems. It yields practical results (e.g., results of value to a national public health, control, screening or prevention program, or local resource planning program along with serving to re-demonstrate a previously well-documented phenomenon.

2 Methods

2.1 Study Area

The dengue outbreak in Bangkok can influence the dengue situation for the whole country because Bangkok is a very crowded city located at the center of Thailand. In the 2010 census, the population of Bangkok was about 8.28 million, although only about 5.7 million were the registered residents. The population within the city during the day swells to about to 15 million [7] due to the commuters from the surrounding areas. During a winter season, the temperature in Bangkok is still high around 28-35 degree Celsius and there is rain in every season [57]. Figures 2 and 3 show the cumulative monthly rainfall and monthly diurnal temperature range (DTR) of Bangkok throughout the years ($2008-2015$). The diurnal temperature range (DTR) is the difference between the daily maximum and minimum temperature. The experimental evaluation has shown that DTR influences two important parameters underlying dengue virus (DENV) transmission by Aedes mosquito [9].

Moreover, many people who live in Bangkok travel to other provinces frequently, so these people may be the main carriers and multipliers of the virus that cause dengue outbreak in other parts of Thailand. The Bangkok city covers an area of 1,568.737 square kilometers and it is subdivided into 50 districts, which are further sub-divided into 169 sub-districts. The total population registered in Bangkok is 5,693,884 and more than three million unregistered people live in Bangkok. The average registered population of Bangkok districts in 2014 is 113845.

2.2 Data

Climate is one of the main factor related to dengue outbreak both locally and globally [58, 59]. Most of Thailand has a tropical wet and dry climate type, making the city amenable for Aedes mosquito to breed and spread in any season. The role of the variation in climatic factors on transmission dynamics and the geographic distribution of dengue has been well-studied [60]. The rainy season allows Aedes mosquitoes egg to grow into adult mosquitoes easier than in dry season. The increase of Aedes mosquitoes is directly affecting dengue cases in Thailand as shown in Figure 5.

Figure 4 shows the increase of DHF cases in Bangkok up to 2015. The average number of DHF cases in Bangkok from 2008 until 2015 is around 4,750 cases per year. The highest

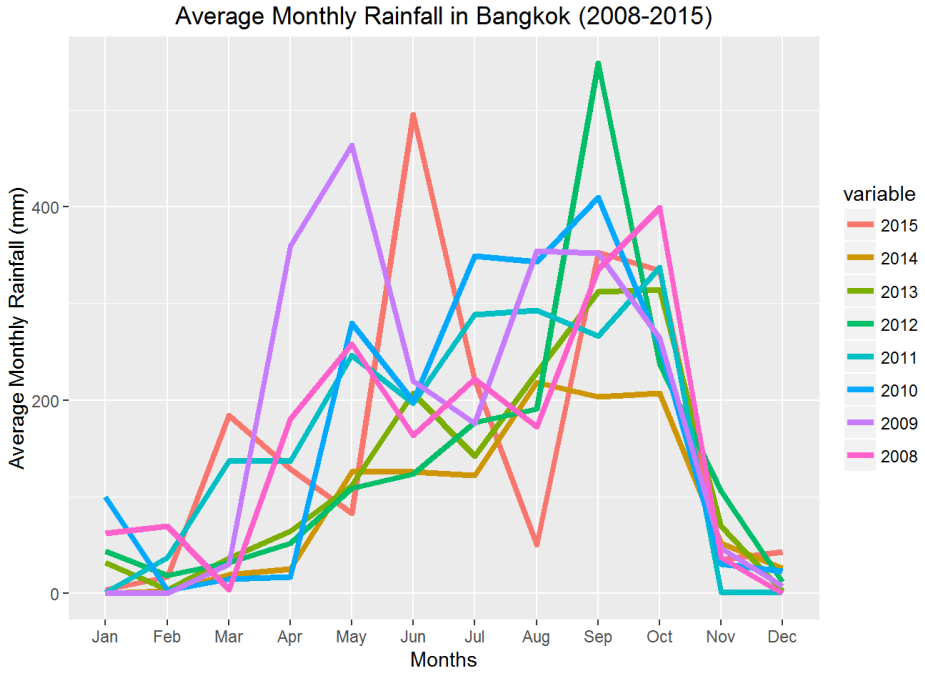


Figure 2: Rainfall within Bangkok throughout the years.

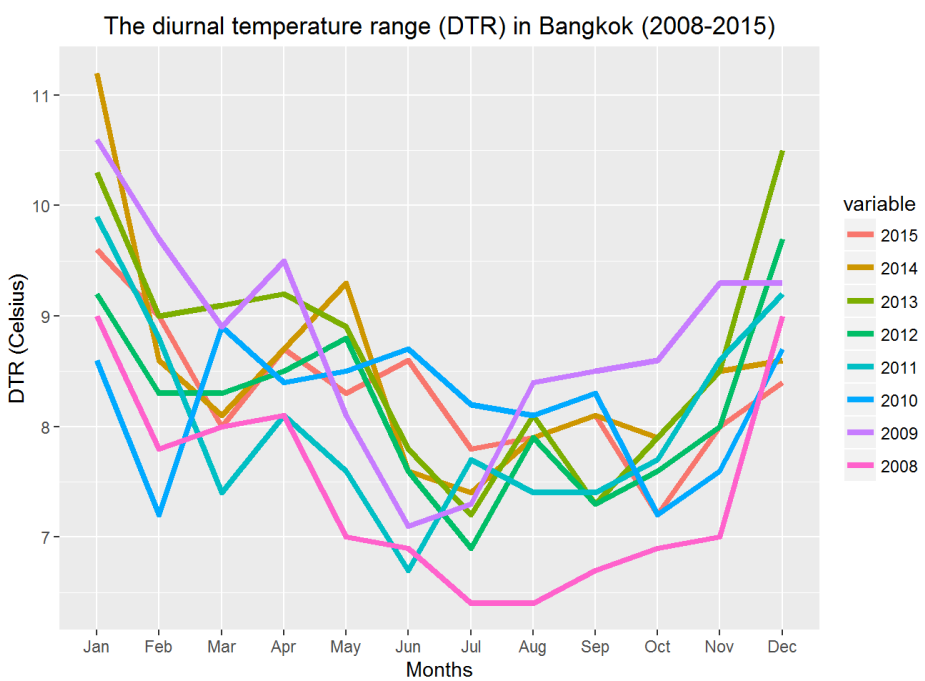


Figure 3: DTR within Bangkok throughout the years.

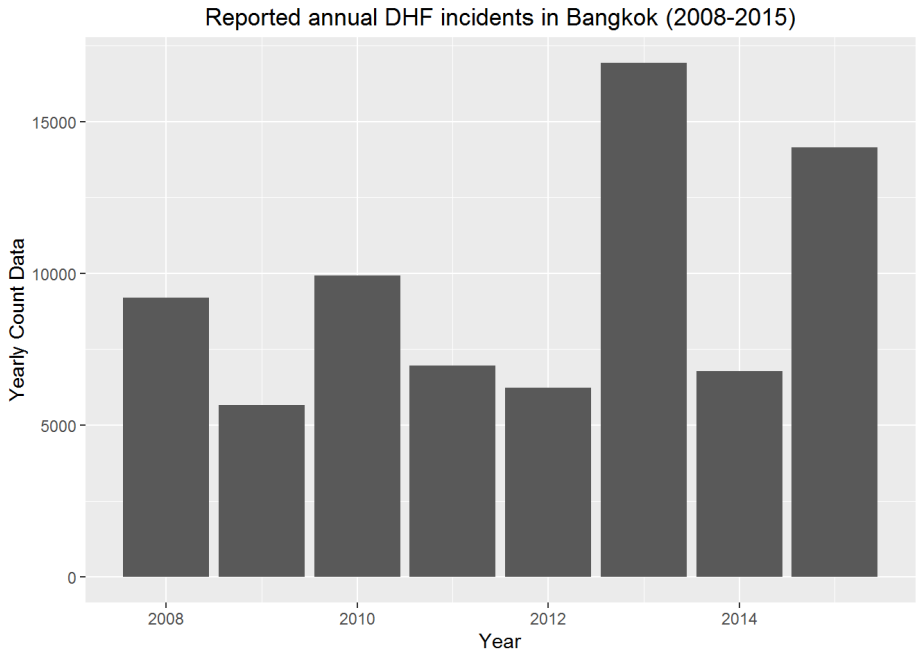


Figure 4: The DHF incidence peaked in 2013 and 2015. It shows that dengue outbreak increases every alternative year. Data from [61]

.

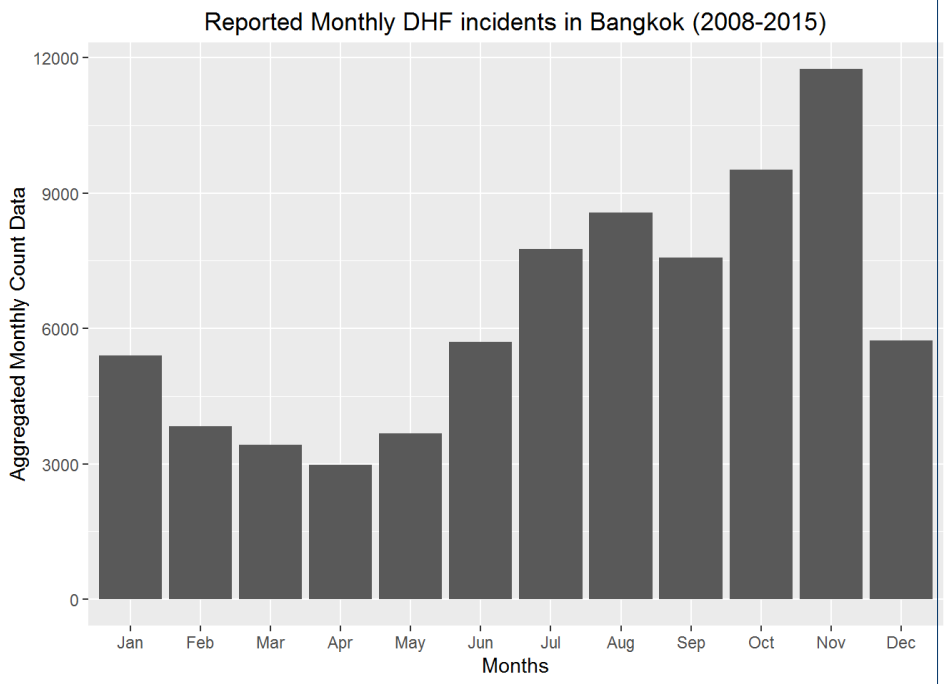


Figure 5: The DHF incidence peaked in October and November. Graph shows the amount of dengue patients for each month for the years 2008–2015. Rainy season continues from mid-May to mid-October. Data from [61]

.

number of DHF cases in Bangkok is 16942 which happened in 2013 followed by 14154 cases in 2015. As shown in Figure 5, the highest number of monthly DHF cases in Bangkok is 11752 in the month of November followed by 9511 cases in October.

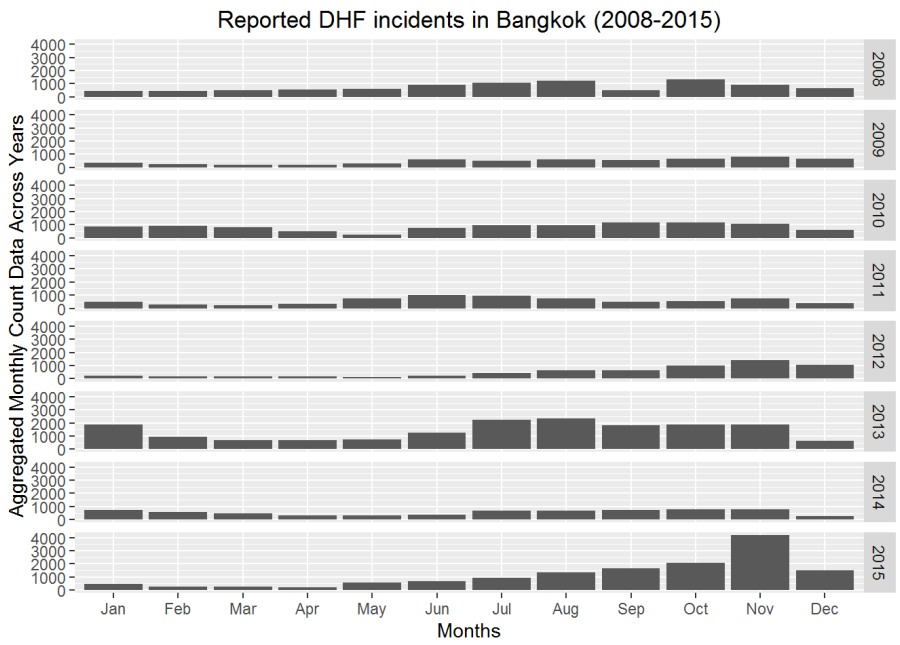


Figure 6: The DHF incidence peaks in the months of October and November. However in the year 2013, DHF outbreak happened for several months continuously.

Figure 6 shows the variation of dengue cases with months. In the rainy season (mid-May–October), dengue cases rise dramatically, and then they decrease suddenly in summer season (February—mid-May). The reason that November has the highest number of cases is that dengue virus in patients need 4–10 days for the incubation period. So most dengue patients in November are infected in the rainy season (October-November).

In this study, we have used aggregated monthly dengue cases (dengue hemorrhagic fever) at district level from the Department of Disease Control, Bangkok for the period 2008 – 2015. Rainfall data from 2008 – 2015 is provided by the Department of Drainage and Sewerage, Bangkok and the temperature data for the same period is provided by the Meteorological Department, Thailand. Records of cumulative monthly rainfall (mm) and monthly DTR readings (°C) were obtained. The meteorological data were merged to the aggregated total number of confirmed dengue cases per month in each district of Bangkok. Using the geographical information of different districts of Bangkok, this data of each target district was merged to the aggregated total number of confirmed dengue cases per month in each of its surrounding district. Thus, our data has (a) Dengue virus prediction variables (Monthly DHF incidents in a district + aggregated monthly DHF incident in the nearby districts) and (b) Aedes mosquito prediction variables (cumulative monthly rainfall in the Bangkok + mean monthly average diurnal temperature range (DTR) in Bangkok).

A model was developed and validated by dividing the data file into two data sets: one for training the model and another for testing and validation of the fitted model. Here we varied the period in both the data sets. For example, in one case of prediction the data from

January 2008 to December 2012 was used to train a model, and data from January 2013 to December 2015 was used for testing and validation of the fitted model. All the analyses were performed in R [62] using the mgcv package [63]. The details of the different such cases of predictions are listed in Table 1.

2.3 Statistical Analysis

These analyses focus on studying the following relationships:

- relationship of meteorological (DTR and rainfall) and socioeconomic data (monthly garbage collection in each district) to the time series of dengue incidences in that particular district.
- relationship of dengue transmission in a specific month in a district with the data of its past occurrences.
- relationship of dengue transmission in a specific month in a district with the data of past occurrences of dengue in its surrounding districts.

We created a set of prediction models encoding the above-mentioned relationships. Our geospatial/statistical method used in our work and the approach using Generalized Additive Models (GAM) to derive the insights are similar to the research study [64] conducted in Indonesia. However, along with evaluating different predictions models, we evaluate a different/unique hypothesis using a novel data-set; the feature-set used in the study is richer and the study area/country is different from the above-mentioned work.

The target of prediction was the cumulative dengue count in Bangkok (i.e. sum total of dengue incidences in all the 50 districts of the city) in a particular month of the year. According to the previous studies to determine the optimal lead time for dengue forecasts [65], there is evidence of increasing dengue cases in lag time of up to 4–20 weeks. Thus, similar to [64] we decided a priori, for meteorological variables, to use lag times with up to 4 months delay (i.e. 0–3 months) in the analysis. Since the dengue counts vary within and between the years, the count data is likely to be over-dispersed. Thus, rather than using the "standard" Poisson regression in which it is assumed that variance of count data is constant regardless of the expected value, we adopt a Quasi-Poisson regression in which the variance of count data (dengue counts) is assumed to be a linear function of the mean. To allow for over-dispersion a log-link function of dengue count data is used. To allow for non-linear response and exposure association between the predictors and the dengue incidences, cubic splines of 3 degrees of freedom was applied on the meteorological variables. The generalized additive model (GAM) can be expressed as:

$$log(C_{0,t}) \sim \alpha + \sum_{l=0}^{3} ns(\mathfrak{T}_{lt}, d = 3) + \sum_{l=0}^{3} ns(R_{lt}, d = 3) \tag{1}$$

where log() denotes the natural logarithm, C represents the total dengue count data, t denotes to time in months, $\mathfrak{T}$ represents DTR (°C), R denotes the mean monthly rainfall (mm) and l, denotes the lag variables, ns() denotes a natural cubic spline.

High correlation among predictor variables may give rise to singularity problems when fitting a statistical model. However, for GAMs, checking for collinearity is not sufficient. Since, we are now fitting smooth functions; it should be determined whether the smooth

function of one variable can be produced using a combination of the smooths of the other terms in the model. This is called checking for *concurvity*. We performed the concurvity check for all the predictor variables.

Disease surveillance data of each districts as predictor

Since the current number of dengue incidences are influenced by the number of cases in the past, to determine this period of influence we have considered two approaches. The first approach focuses on determining the optimal lag term for short-term lagged dengue incidence data. The auto-regressive patterns in dengue time series data were studied by fitting a GAM in which data up to a delay of 4 months was used (similar to what we did with the meteorological data). This model was to fit to assess the influence of past dengue incidence on current count independent of meteorological factors. The regression model can be expressed as:

$$log(C_{0,t}) \sim \alpha + \sum_{l=1}^{4} ns(C_{lt}, d = 3) \tag{2}$$

Second approach focused on assessing the risk of retrospective transmission (1–48 months) on dengue incidences. Prior studies conducted in the region have shown that cross-immunity for dengue virus serotypes may significantly alter the dengue transmission over a period [12, 66]. We hypothesize that this might partly explain bi-annual cyclic epidemic pattern of dengue occurrence in Bangkok as shown in Figure 4. Thus, we assess and estimate the risk of retrospective dengue transmission up to 1–30 months on current dengue transmission. To incorporate the effects delayed in time, the statistical model of DLNM[1] was used to describe the additional time dimension of this exposure-relationship[67].

The regression model can be expressed as:

$$log(C_{0,t}) \sim \alpha + \sum_{l=1}^{30} DLNM(C_{lt}, d = 4) \tag{3}$$

The output from both these approaches was then combined to determine the 'optimal lag' of disease surveillance.

Disease surveillance data from surrounding districts as predictor There is much research which concludes that many dengue cases that occur in urban areas are due to the factors such as high population density, inadequate housing, and inappropriate human behavioral practices [19–21]. Surveillance of Aedes mosquito density is important for construction models of dengue transmission, in order to prioritize areas and seasons for vector control. The 80% of larvae or pupa in house are from Aedes mosquito. A recent study [23] has explored the dengue occurrence in a region in relation to its surrounding regions. The study is conducted in near real-time using object-based and spatial metric approaches. The geospatial analysis conducted on the data acquired using Google search and advanced land observation satellite images suggests that the occurrence and spread of dengue cases are positively correlated with densely populated areas which are *surrounded by dense vegetation*. This further suggests that the spatial heterogeneity of human activities influence the

[1]The GitHub repository of R implementation of DLNM used for our analyses is available at https://github.com/gasparrini/dlnm

dengue epidemic. Thus, to determine the influence of spatial heterogeneity of human activities ongoing in nearby areas, we consider the data (both 'short-term' and 'long-term') from the dengue incidences of surrounding districts. The regression model can be expressed as:

$$log(C_{0,t}) \sim \alpha + \sum_{l=1}^{4} ns(S_{lt}, d = 3) + \sum_{l=0}^{30} DLNM(S_{lt}, d = 4) \quad (4)$$

where log() denotes the natural logarithm, C represents the total dengue count data, S represents the count data of dengue incidence occurred in the surrounding districts, t denotes to time in months, l denotes the lag variables and ns() denotes a natural cubic spline.

Waste disposal data from each district as predictor Previous studies have shown the spatial correlation of socioeconomic data and urbanization with dengue incidences [68, 69]. Since the waste disposal and landfill dumps are the spatial infrastructures of any modern city, we used the data about monthly garbage collection from each district as an indicator for social capital.

Thus using the above-mentioned predictors, we evaluated the following models:

(A) Seasonal Naïve model (each forecast is set to be equal to the last observed value from the same season of the year (e.g., the same month of the previous year).

(B) Meteorology Optimal model (includes lag of $0-4$ months for DTR and mean monthly rainfall)

(C) Optimal Lag Surveillance Model (includes lagged dengue count data for 1, 2 and 23^{rd} months)

(D) Optimal Met and Lag Surveillance Model (includes lagged meteorology data for $1-3$ months and lagged dengue count data for 1, 2 and 23^{rd} months)

(E) Optimal Representation Model Combination of (D) with lagged dengue count data of surrounding districts for 1^{st} and 2^{nd} months

(F) Social-economic data Included Combination of (E) with garbage collection data of each district as the social capital

For complex models (B)-(F), a necessary time latency of at least 1 month lag was evaluated to allow up to 1 month lead time in controlling the disease at forecast of epidemics.

3 Results and Discussion

In this section, we explain about the models (A-F) listed in last section. The performance of these model is shown in Table 1 and the visual analyses of the prediction are shown in Figure 11. The vertical axis in Figure 11 represents the number of DHF cases and the horizontal axis depicts the time in months from 2008-2012. The errors RMSE/SRMSE are calculated by testing on in-sample data. The prediction performance is compared over the same time periods (months $23-60$) to reduce the potential bias. This period is chosen because the final model uses lagged dengue data of past 23^{rd} month (as explained later in this section). The increase in adjusted R-squared value happens only if the new term improves the model more than what would be expected by chance (the higher the better). Deviance is a goodness-of-fit statistic for a statistical model (the higher the better) and the larger difference in AIC indicates stronger evidence for one model over the other (the lower the better). For comparing the difference in AIC, all the fitted models (B-F) are compared to

base model A (Seasonal Naïve model) . The discriminating ability of the the finally selected model was evaluated against Bangkok specific and WHO threshold of the epidemic in terms of specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV).

As a first step for model fitting, the data for lags of month 0–3 of Diurnal Temperature Range (DTR) and monthly average rainfall was used to analyze the (1) statistical significance of the model terms (2) association of DTR and rainfall with dengue cases. Table 5 shown the approximate statistical significance of the smooth terms for DTR and rainfall. As shown in Figure 7, the association of temperatures with dengue cases in lag 0,1 decreases when the DTR is less than 8 °C and then increases. However, for the lag 3, this association shows a slight increase and then decrease. For lag 4, it is observed that the association of DTR with dengue cases increases linearly when the temperature is less than 9 °C and then a strong drop is observed. Due to the statistical significance of these terms i.e. lags of month 0–3 for the meteorological data, we call the model comprising of them as 'Optimal Met' model (model B).

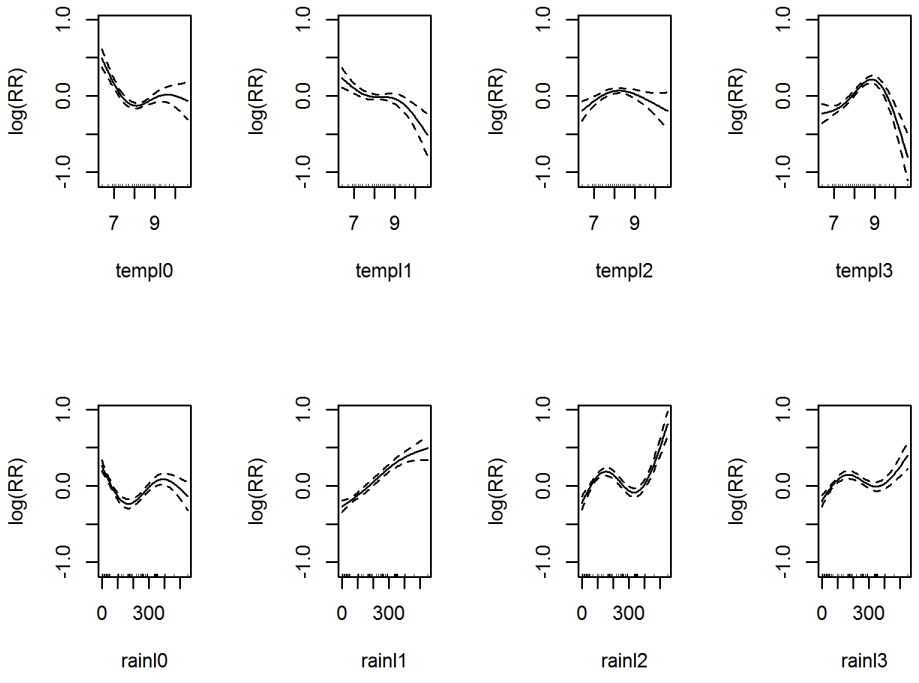


Figure 7: **Association between the meteorological variables and dengue over lags of 0–3 months**. Solid lines represent relative risks (RR) of dengue cases whereas the dotted lines depict the upper and lower limits of 95% confidence intervals. P-values and the initial analyses are listed in Table 5

.

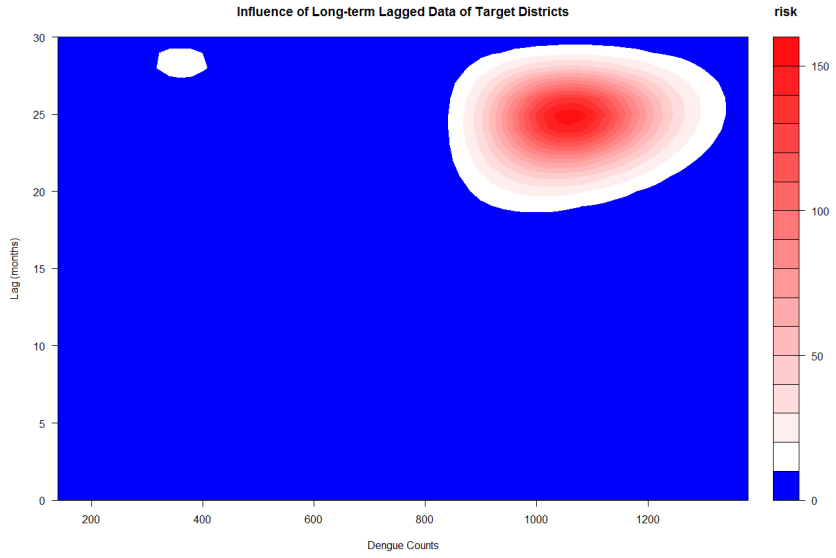
We calculated the cross-correlation of dengue cases with (a) dengue cases of lag of 0–3 months (b) DTR of lag of 0–3 months (c) rainfall of lag of 0–3 months and (d) the dengue

cases of lag of 0–3 months from the surrounding districts. Figure 13 shows the cross-correlation indicate that the highest positive association between dengue incidence and lagged dengue incidences were found at lag 0 (r, 0.667), with rainfall at lag 2 (r, 0.428) and with dengue incidences from surrounding districts at lag 1 (r, 0.514). There was a negative correlation of DTR with dengue incidence at all the lags. For lag 0 (r, -0.261), lag 1 (r, -0.373), lag 2 (r, -0.309) and lag 3 (r, -0.154) was observed.

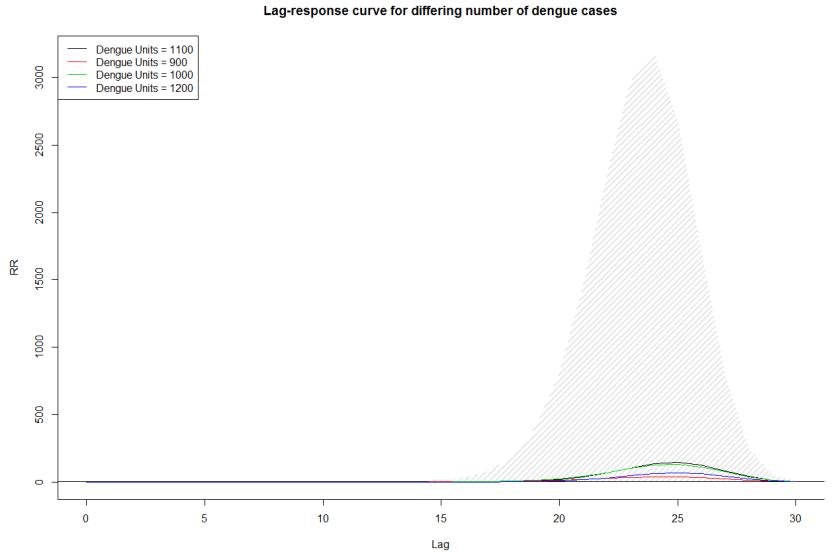
But in order to create an early warning system (EWS), it is required to give a certain lead time. Thus, in subsequent analysis all the data that we use shall have a lag of atleast one month. The second model that we fit consists solely the data of lagged dengue incidences. For short term lags, the data from past $1-4$ months was included. When the long-term lagged dengue incidence data was taken into account, the non-linear distributed lag models were used using dlnm package[70]. The influence of long-term lagged data of dengue incidences is shown using the contour plot in Figure 8a. It is observed that lagged long-term data has lower relative risks of transmission up to almost 2 years following a large outbreak in around lag 23. This suggests a negative feedback cyclic pattern. Figure 8b suggests that when an outbreak happens in a particular month then dengue risk in each of the following months will increase with a peak in next 23 months (Figure 8b). Figure 8b contains the lag-response curve for the differing number of dengue cases after an outbreak happens in a specified month. Thus based on these analyses, the optimal variables for the prediction models included dengue count at lag 1,2 for short-term lags and lag 23 for long term lag. We then combined the meteorological variables for lag greater than 1 month and dengue lag terms used in 'Optimal Lag Surveillance' model (model C) to determine their association with dengue incidences. This model is termed as 'Optimal Met and Lag Surveillance Model' (model D).

Since one of our hypotheses is that the significance of movement patterns of people and spatial heterogeneity of human activities on the spread of the epidemic in statistically significant. In other words, the dengue cases in a particular district are influenced by the dengue cases in there surrounding districts. To test the hypothesis, we determine how the occurrence of dengue in a target district is influenced by the occurrence of dengue in its surrounding districts. Both short-term and long-term lagged dengue cases data of the dengue incidences in surrounding districts was taken into account. For short-term lags, we considered the lagged data of past 1-4 months in which the data of lag 1 and 2 months was found to be statistically significant ($p < 0.05$). For long-term lags, the data up to the past 30 months was used. To determine the relative risks of transmission following an outbreak in a specified month, the non-linear distributed lag models were used using dlnm package [70]. For the sake of brevity, the lag-response curve for the differing number of dengue cases is not shown here. Based on the aforementioned analyses, the optimal lag terms of dengue incidences in surrounding districts were found to be for lag 1, lag 2 and lag 12 ($p \gg 0.05$). These smooth terms were combined with that of Model D and termed as 'Optimal Representation Model' (model E) according to Equation 5.

$$\begin{aligned} log(C_{0,t}) \sim \alpha + \sum_{l=0}^{3} ns(T_{lt}, d=3) + \sum_{l=0}^{3} ns(R_{lt}, d=3) + \sum_{l=1}^{2} ns(C_{lt}, d=3) + \\ \sum_{l=23}^{23} ns(C_{lt}, d=3) + \sum_{l=1}^{2} ns(SC_{lt}, d=3) + \sum_{l=12}^{12} ns(SC_{lt}, d=3) \end{aligned} \tag{5}$$



(a)



(b)

Figure 8: Retrospective transmission period is calculated to account for the influence of dengue incidences in each of the target districts (Fig 8a) and lag-response curve for an increase in various units of dengue incidences (Fig 8b).

where C represents the dengue counts in the target district, $\mathfrak{T}$ represents DTR (°C), R denotes the mean monthly rainfall (mm) and SC denotes the dengue counts in the surrounding districts.

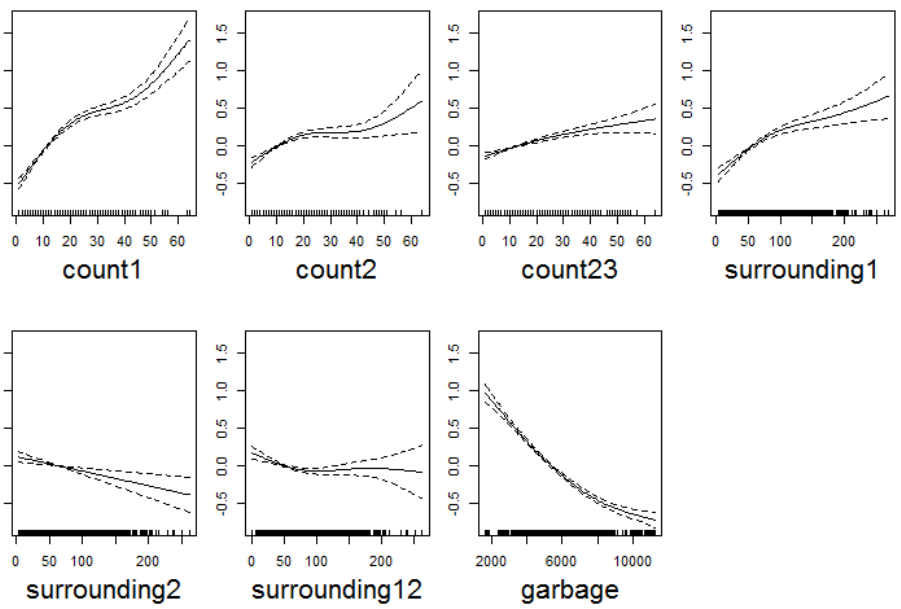


Figure 9: Association between past dengue count over optimal lags (lag 1,2 and 23) for each target district and lag 1,12 for surrounding district count data and garbage data with dengue outbreak. Solid lines represent relative risks (RR) of dengue cases whereas the dotted lines depict the upper and lower limits of 95% confidence intervals.

Our last model (F) included waste disposal data (as a socioeconomic indicator) from each district as a predictor. The association of the smooth terms with dengue transmission is shown in Figure 9. After the final model (F) is fit, the residuals are inferred; and their normality and residual autocorrelation are checked. Figure 10 shows that residual histograms (top-left) are symmetric and follow a unimodal distribution. The Q-Q plot of deviance residuals which are conditional on the fitted model coefficients and scale parameter (bottom-left) is close to a straight line. This suggests that the distributional assumptions of the model are satisfied. The plot of reported and predicted cases indicate linear relation (bottom-right). The partial autocorrelation function plot (top-right) does show significant autocorrelation for some long-term lags, but it was found that keeping those lagged terms increases the complexity of the model without any significant increase in prediction performance. Thus, they were chosen to be ignored in the model [2].

The simplest Seasonal Naïve model (A) that used month of the year data for making the forecasts showed the poorest performance when compared with the other models. The quality of fit and predictive ability increases with the subsequent models as shown (Table 1).

To evaluated the predictive performance of our final model (F), few external validation data sets were created for which the predictive performance is shown in Table 2. The model

[2]These include those additional terms that do not as well provide insights in favor of our hypothesis.

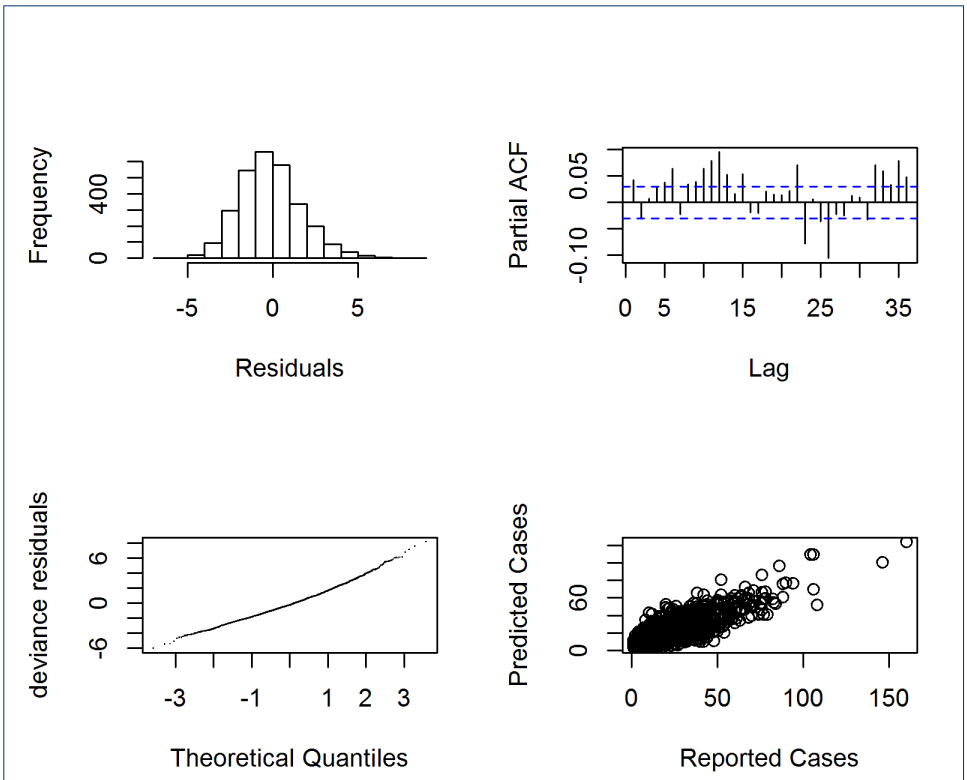


Figure 10: **Residual Diagnosis.** Top-left is the histogram of residuals; top-right is the partial ACF plot. On bottom-left is the Q-Q plot for the deviance residuals whereas the relationship between the reported and predicted cases is shown on bottom-right

.

Model Name	RMSE	SRMSE	R-sq.(adj)	Deviance Explained	Δ AIC
A: Seasonal Naïve	10.22	0.62	0.16	0.16	0
B: Meteorology Optimal	8.83	0.54	0.28	0.32	-492.64
C: Optimal Lag Surveillance Model	7.32	0.45	0.49	0.49	-2420.39
D: Optimal Met and Lag Surveillance Model	6.30	0.39	0.62	0.64	-2725.62
E: Optimal Representation Model	6.12	0.37	0.64	0.66	-2718.90
F: Social-economic data Included	6.10	0.37	0.64	0.73	-2713.86

Table 1: Predictive performance statistics of different models evaluated on the training data for the same time period (months $23-60$) to reduce the potential bias. The performance is measured on different metrics. The best model should have the lowest errors (RMSE, SRMSE) and have the best fit (measured in R-sq.(adj).), high deviance and low Δ AIC.

trained on the larger set of training data shows the better predictive performance as the influence of both, the direct and retrospective transmission could be learned in the model. For one the case in which model was trained on data from $2008-2014$ and evaluation performed for year 2015, the visual analysis is shown is provided in Figure 12.

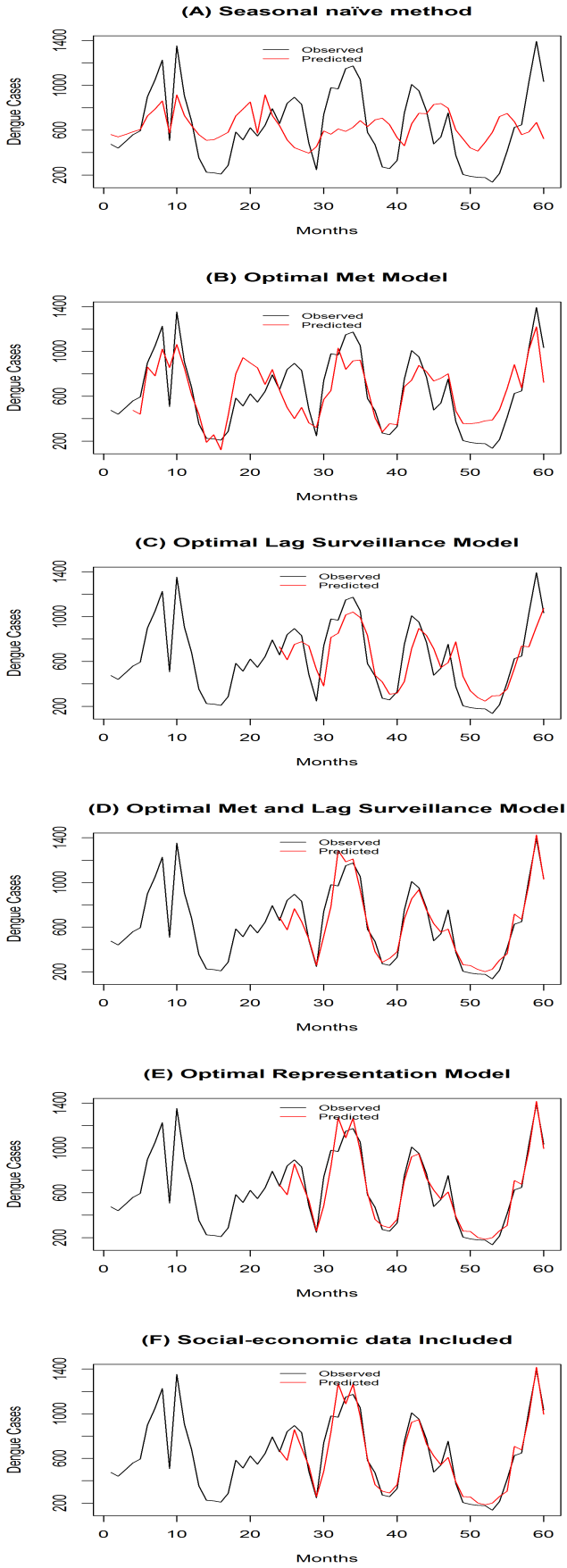


Figure 11: Monthly Observed and Predicted Dengue Cases from 2008-2012.

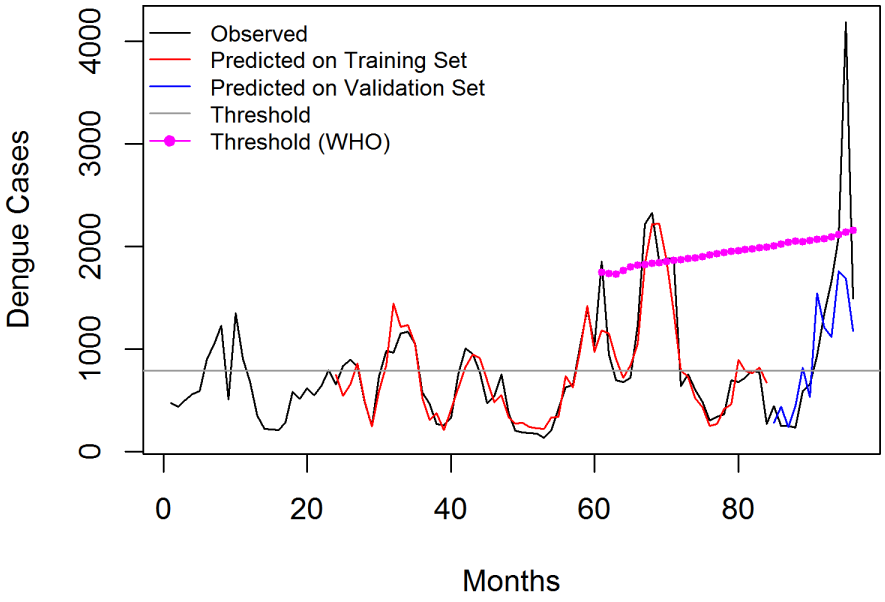


Figure 12: The final model (F) was trained on the data from 2008 – 2014 and the validation is performed on external data of year 2015.

.

Prediction Accuracy	**Epidemic constant threshold** monthly average cases (2008 – 2015) = 790 cases	**Epidemic moving threshold** (mean of a moving window over preeding 5 years + 2 SD)
Sensitivity	87.0	100.0
Specificity	92.60	57.14
PPV	95.72	90.62
NPV	78.80	100.0

Table 3: The discriminating ability of the final model (F).

Training Dataset	**In-sample**	**Out-Sample** (2014-2015)	**Out-Sample** (2014)	**Out-Sample** (2015)
2008-2013	0.41	0.62	0.35	0.60
2008-2014	0.41			0.54

Table 2: Predictive performance statistics of most optimal model evaluated on validation data sets is measured in SRMSE.

The results of the discriminating ability of model (F) evaluated against the constant threshold of the epidemic (average monthly dengue cases from 2008 – 2015) as well WHO moving threshold is shown in Table 3. The comparison is done in terms of Sensitivity (true positive rate), Specificity (true negative rate), positive and negative predictive values (PPV and NPV respectively). In other words, Sensitivity means the proportion of observed positives that were predicted to be positive. The Specificity is the proportion of observed negatives that were predicted to be negatives. The PPV is the proportion that answers the

question: 'How likely is it that this dengue outbreak will happen given that the prediction result is positive?' The NPV answers the question: 'How likely is it that outbreak does not happen given that the prediction result is negative?'. Because of the higher values of the WHO threshold, the NPV is 100% as compared to that of moderate values of constant threshold. However, it should be noted that Thailand suffered a big unexpected outbreak in year 2015 for which the quality of prediction results deteriorated.

4 Conclusion

In this study, the dengue incidences were predicted using a variety of data. The best model for the dengue prediction is the one that includes lagged meteorological data (rainfall, DTR), lagged dengue data of the target districts as well as their surroundings and the socioeconomic data. We proved that for the prediction of dengue outbreaks within a district, the influence of dengue incidences and socioeconomic data from the surrounding districts is statistically significant. Thus for forecasting dengue outbreaks and taking preventing measures, the epidemiologists and health authorities should consider the influence of movement patterns of people and spatial heterogeneity of human activities. The results also support previous studies that suggest temperature, precipitation, short and long term lagged incidences are related with dengue occurrence and its transmission.

A number of limitations are apparent for this study. First, the predictive model that finally selected, could explain only 73 per cent of the variation in the occurrence of dengue cases. The remaining 27 per cent unexplained variation could be due to the influence of other factors. The out-of-sample predictive performance was considerably worse than that of in-sample performance. Rather than dismissing it as a case of over-fitting, the case demands that we look into the facts and 'plausible causes' behind it. Thailand had an unexpected dengue outbreak in year 2015, the worst in last 20 years. It was the year for dengue outbreaks across Asia. Along with Thailand, other countries like Malaysia, the Philippines, Thailand, Taiwan, Vietnam and India were among the worst hit countries. According to the World Health Organization, Malaysia reported nearly 18% more cases from 2014, The Philippines reported an almost 50% rise in cases compared with 2014 and India reported double the cases as compared to the previous year. Thus, the characters of the validation dataset found less representation in the training dataset. Second, the study simply used monthly dengue aggregate data rather than using direct analysis of laboratory surveillance reports. The monthly dengue aggregate data constitutes only the laboratory confirmed dengue cases. But almost three-quarters of people who catch dengue have few or no symptoms. Despite being asymptomatic, these people may play a key role in the dengue transmission cycle. But our data does not represent such cases nor does it put weighted emphasis on people that suffer from multiple infections from different serotypes which puts one at greater risk for deadly severe dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS). Third, dengue severity is a key determinant of underreporting. There are several impediments of reporting dengue cases in Thailand and we need to incorporate the measure to estimate the underreporting of dengue inpatients on the district level. The results show that binary system of classifying months into outbreaks and non-outbreaks as proposed in the paper worked quite well in our evaluation. We need to add more data and collect a more diverse data. The diverse data may include House index (HI), Breteau index (BI), Container index (CI) and integrating the information from social media platforms to track the dengue incidences in real time will likely better the prediction.

Social media also has become a key component in understanding dengue. For example in 2015, popular actor Thrisadee 'Por' Sahawong died due to dengue infection. This case created a sense of panic which motivated people to understand the facts, symptoms, mechanism of virus transmission and the preventive measures including vaccination. We also aim to develop customized models for each individual district that includes demographic data, data from government surveys and above-mentioned additional features at a more granular level.

List of abbreviations

DENV: Dengue Virus; DF: Dengue Fever; DHF: Dengue Haemorrhagic Fever; DSS: Dengue Shock Syndrome; DTR: Diurnal Temperature Range; GAM: Generalized Additive Model; DLNM: Distributed Lag Non-Linear Models; RMSE: Root Mean Squared Error; SRMSE: Standard Root Mean Squared Error; PPV: Positive Predictive Value; NPV: Negative Predictive Value.

Declarations

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and materials
The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests
The authors declare that they have no competing interests.

Funding
No funds were obtained to carry out this project.

Author's contributions
RJ analyzed and interpreted the data. He also wrote the manuscript. SS collected and cleaned the data. SI helped make data acquisition task easier by coordinating between different sources. HP supervised the project. All authors read and approved the final manuscript.

Acknowledgements
Mrs. Sarinthorn Sontisirikit, Senior Public Health Officer of the Department of Disease Control) enabled surveying in the local communities of Kannayao, Lat Krabang , Thung khru, Saphansung and Jatujak districts of Bangkok.

Author details

[1]National Institute of Informatics, Tokyo, Japan. [2]Asian Institute of Technology, School of Engineering and Technology, Bangkok, Thailand. [3]Department of Disease Control Thirteenth Division, Bangkok, Thailand.

References

1. Gubler, D.J., Clark, G.G.: Dengue/dengue hemorrhagic fever: the emergence of a global health problem. Emerging infectious diseases **1**(2), 55 (1995)
2. WHO: Dengue Type. http://www.who.int/mediacentre/factsheets/fs117/en/. Accessed 2015-11-20 (2015)
3. Hammon, W.M., Sather, G.: Virological findings in the 1960 hemorrhagic fever epidemic (dengue) in thailand. The American journal of tropical medicine and hygiene **13**(4), 629–641 (1964)
4. Chareonsook, O., Foy, H., Teeraratkul, A., Silarug, N.: Changing epidemiology of dengue hemorrhagic fever in thailand. Epidemiology and infection **122**(01), 161–166 (1999)
5. Hesse, R.R.: Dengue virus evolution and virulence models. Clinical Infectious Diseases **44**(11), 1462–1466 (2007)
6. Wilder-Smith, A., Renhorn, K.-E., Tissera, H., Abu Bakar, S., Alphey, L., Kittayapong, P., Lindsay, S., Logan, J., Hatz, C., Reiter, P., *et al.*: Denguetools: innovative tools and strategies for the surveillance and control of dengue. Global health action **5**(1), 17273 (2012)
7. World Population Review: Bangkok Population 2015. http://worldpopulationreview.com/world-cities/bangkok-population/. Accessed 2016-01-18 (2015)
8. Stoddard, S.T., Wearing, H.J., Reiner Jr, R.C., Morrison, A.C., Astete, H., Vilcarromero, S., Alvarez, C., Ramal-Asayag, C., Sihuincha, M., Rocha, C., *et al.*: Long-term and seasonal dynamics of dengue in iquitos, peru. PLoS Negl Trop Dis **8**(7), 3003 (2014)
9. Lambrechts, L., Paaijmans, K.P., Fansiri, T., Carrington, L.B., Kramer, L.D., Thomas, M.B., Scott, T.W.: Impact of daily temperature fluctuations on dengue virus transmission by aedes aegypti. Proceedings of the National Academy of Sciences **108**(18), 7460–7465 (2011)
10. Carrington, L.B., Seifert, S.N., Armijos, M.V., Lambrechts, L., Scott, T.W.: Reduction of aedes aegypti vector competence for dengue virus under large temperature fluctuations. The American journal of tropical medicine and hygiene **88**(4), 689–697 (2013)
11. Nakhapakorn, K., Tripathi, N.K.: An information value based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence. International Journal of Health Geographics **4**(1), 13 (2005)
12. Adams, B., Holmes, E., Zhang, C., Mammen, M., Nimmannitya, S., Kalayanarooj, S., Boots, M.: Cross-protective immunity can account for the alternating epidemic pattern of dengue virus serotypes circulating in bangkok. Proceedings of the National Academy of Sciences **103**(38), 14234–14239 (2006)
13. Alto, B.W., Lounibos, L.P., Mores, C.N., Reiskind, M.H.: Larval competition alters susceptibility of adult aedes mosquitoes to dengue infection. Proceedings of the Royal Society of London B: Biological Sciences **275**(1633), 463–471 (2008)
14. WHO: Human Dengue Symptoms. http://www.who.int/denguecontrol/human/en/. Accessed 2015-11-19 (2015)
15. Cummings, D.A., Iamsirithaworn, S., Lessler, J.T., McDermott, A., Prasanthong, R., Nisalak, A., Jarman, R.G., Burke, D.S., Gibbons, R.V., *et al.*: The impact of the demographic transition on dengue in thailand: insights from a statistical analysis and mathematical modeling. PLoS medicine **6**(9), 999 (2009)
16. Halstead, S.B.: Dengue virus-mosquito interactions. Annu. Rev. Entomol. **53**, 273–291 (2008)
17. Esteva, L., Vargas, C.: Influence of vertical and mechanical transmission on the dynamics of dengue disease. Mathematical biosciences **167**(1), 51–64 (2000)
18. Favier, C., Schmit, D., Müller-Graf, C.D., Cazelles, B., Degallier, N., Mondet, B., Dubois, M.A.: Influence of spatial heterogeneity on an emerging infectious disease: the case of dengue epidemics. Proceedings of the Royal Society of London B: Biological Sciences **272**(1568), 1171–1177 (2005)
19. Chang, A.Y., Parrales, M.E., Jimenez, J., Sobieszczyk, M.E., Hammer, S.M., Copenhaver, D.J., Kulkarni, R.P.: Combining google earth and gis mapping technologies in a dengue surveillance system for developing countries. International journal of health geographics **8**(1), 49 (2009)
20. Knudsen, A.B., Slooff, R.: Vector-borne disease problems in rapid urbanization: new approaches to vector control. Bulletin of the World Health Organization **70**(1), 1 (1992)
21. Troyo, A., Fuller, D.O., Calderón-Arguedas, O., Solano, M.E., Beier, J.C.: Urban structure and dengue incidence in puntarenas, costa rica. Singapore journal of tropical geography **30**(2), 265–282 (2009)
22. Arunachalam, N., Tana, S., Espino, F., Kittayapong, P., Abeyewickrem, W., Wai, K.T., Tyagi, B.K., Kroeger, A., Sommerfeld, J., Petzold, M.: Eco-bio-social determinants of dengue vector breeding: a multicountry study in urban and periurban asia. Bulletin of the World Health Organization **88**(3), 173–184 (2010)
23. Sarfraz, M.S., Tripathi, N.K., Kitamoto, A.: Near real-time characterisation of urban environments: a holistic approach for monitoring dengue fever risk areas. International Journal of Digital Earth **7**(11), 916–934 (2014)
24. Syarifah, N., Rusmatini, T., Djatie, T., Huda, F.: Ovitrap ratio of aedes aegypti larvae collected inside and outside houses in a community survey to prevent dengue outbreak, bandung, indonesia, 2007. Proc Assoc Southeast Asian Nations Congr Trop Med Parasitolol **3**, 116–120 (2008)
25. Scott, T.W., Morrison, A.C.: Aedes aegypti density and the risk of dengue-virus. Ecol. Aspects Appl. Genet. Modi. Mosq **2**, 187 (2003)
26. Reiter, P., *et al.*: Yellow fever and dengue: a threat to europe. Euro Surveill **15**(10), 19509 (2010)
27. Chareonviriyaphap, T., Akratanakul, P., Nettanomsak, S., Huntamai, S.: Larval habitats and distribution patterns of aedes aegypti (linnaeus) and aedes albopictus (skuse), in thailand. (2003)
28. Raju, K., Sokhi, B.: Application of gis modeling for dengue fever prone area based on socio-cultural and environmental factors–a case study of delhi city zone. Int Arch Photogramm Remote Sens Spat Inf Sci **37**, 165–170 (2008)
29. 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 **496**(7446), 504–507 (2013)
30. Leitmeyer, K.C., Vaughn, D.W., Watts, D.M., Salas, R., Villalobos, I., Ramos, C., Rico-Hesse, R., *et al.*: Dengue virus structural differences that correlate with pathogenesis. Journal of virology **73**(6), 4738–4747 (1999)
31. Runge-Ranzinger, S., Horstick, O., Marx, M., Kroeger, A.: What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends? Tropical Medicine & International Health **13**(8), 1022–1041 (2008)

32. Thammapalo, S., Chongsuvivatwong, V., Geater, A., Dueravee, M.: Environmental factors and incidence of dengue fever and dengue haemorrhagic fever in an urban area, southern thailand. Epidemiology and Infection **136**(01), 135–143 (2008)
33. Sulaiman, S., Pawanchee, Z.A., Arifin, Z., Wahab, A.: Relationship between breteau and house indices and cases of dengue/dengue hemorrhagic fever in kuala lumpur, malaysia. American Mosquito Control Association **12**, 494–496 (1996)
34. Gubler, D.J.: Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. Trends in microbiology **10**(2), 100–103 (2002)
35. Gubler, D.J.: Cities spawn epidemic dengue viruses. Nature medicine **10**(2), 129–130 (2004)
36. Anuradha, S., Singh, N., Rizvi, S., Agarwal, S., Gur, R., Mathur, M.: The 1996 outbreak of dengue hemorrhagic fever in delhi, india. (1998)
37. Vaughn, D.W.: Invited commentary: Dengue lessons from cuba. American Journal of Epidemiology **152**(9), 800–803 (2000)
38. Ooi, E.-E., Gubler, D.J.: Dengue in southeast asia: epidemiological characteristics and strategic challenges in disease prevention. Cadernos de saude publica **25**, 115–124 (2009)
39. Guzmán, M.G., Kourí, G.: Dengue diagnosis, advances and challenges. International journal of infectious diseases **8**(2), 69–80 (2004)
40. Erlanger, T., Keiser, J., Utzinger, J.: Effect of dengue vector control interventions on entomological parameters in developing countries: a systematic review and meta-analysis. Medical and veterinary entomology **22**(3), 203–221 (2008)
41. Horstick, O., Runge-Ranzinger, S., Nathan, M.B., Kroeger, A.: Dengue vector-control services: how do they work? a systematic literature review and country case studies. Transactions of the Royal Society of Tropical Medicine and Hygiene **104**(6), 379–386 (2010)
42. Vanlerberghe, V., Toledo, M., Rodriguez, M., Gomez, D., Baly, A., Benitez, J., Van Der Stuyft, P.: Community involvement in dengue vector control: cluster randomised trial. Bmj **338**, 1959 (2009)
43. Luz, P.M., Vanni, T., Medlock, J., Paltiel, A.D., Galvani, A.P.: Dengue vector control strategies in an urban setting: an economic modelling assessment. The Lancet **377**(9778), 1673–1680 (2011)
44. Tatem, A.J., Hay, S.I., Rogers, D.J.: Global traffic and disease vector dispersal. Proceedings of the National Academy of Sciences **103**(16), 6242–6247 (2006)
45. Racloz, V., Ramsey, R., Tong, S., Hu, W.: Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. PLoS neglected tropical diseases **6**(5), 1648 (2012)
46. Degallier, N., Favier, C., Menkes, C., Lengaigne, M., Ramalho, W.M., Souza, R., Servain, J., Boulanger, J.-P.: Toward an early warning system for dengue prevention: modeling climate impact on dengue transmission. Climatic Change **98**(3-4), 581–592 (2010)
47. Naish, S., Dale, P., Mackenzie, J.S., McBride, J., Mengersen, K., Tong, S.: Climate change and dengue: a critical and systematic review of quantitative modelling approaches. BMC infectious diseases **14**(1), 167 (2014)
48. Hii, Y.L., Zhu, H., Ng, N., Ng, L.C., Rocklöv, J.: Forecast of dengue incidence using temperature and rainfall. PLoS neglected tropical diseases **6**(11), 1908 (2012)
49. Lowe, R., Bailey, T.C., Stephenson, D.B., Graham, R.J., Coelho, C.A., Carvalho, M.S., Barcellos, C.: Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in brazil. Computers & Geosciences **37**(3), 371–381 (2011)
50. Banu, S., Hu, W., Hurst, C., Tong, S.: Dengue transmission in the asia-pacific region: impact of climate change and socio-environmental factors. Tropical Medicine & International Health **16**(5), 598–607 (2011)
51. Toledo, M.E., Rodriguez, A., Valdés, L., Carrión, R., Cabrera, G., Banderas, D., Ceballos, E., Domeqc, M., Peña, C., Baly, A., *et al.*: Evidence on impact of community-based environmental management on dengue transmission in santiago de cuba. Tropical Medicine & International Health **16**(6), 744–747 (2011)
52. Tipayamongkholgul, M., Lisakulruk, S.: Socio-geographical factors in vulnerability to dengue in thai villages: a spatial regression analysis. Geospatial health **5**(2), 191–198 (2011)
53. Shang, C.-S., Fang, C.-T., Liu, C.-M., Wen, T.-H., Tsai, K.-H., King, C.-C.: The role of imported cases and favorable meteorological conditions in the onset of dengue epidemics. PLoS neglected tropical diseases **4**(8), 775 (2010)
54. Sang, S., Gu, S., Bi, P., Yang, W., Yang, Z., Xu, L., Yang, J., Liu, X., Jiang, T., Wu, H., *et al.*: Predicting unprecedented dengue outbreak using imported cases and climatic factors in guangzhou, 2014. PLoS neglected tropical diseases **9**(5), 0003808 (2015)
55. Stoddard, S.T., Forshey, B.M., Morrison, A.C., Paz-Soldan, V.A., Vazquez-Prokopec, G.M., Astete, H., Reiner, R.C., Vilcarromero, S., Elder, J.P., Halsey, E.S., *et al.*: House-to-house human movement drives dengue virus transmission. Proceedings of the National Academy of Sciences **110**(3), 994–999 (2013)
56. Reiner Jr, R.C., Stoddard, S.T., Scott, T.W.: Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal. Epidemics **6**, 30–36 (2014)
57. World Weather Online: Bangkok average temperature and rain from year 2000 to 2012. http://worldpopulationreview.com/world-cities/bangkok-population/. Accessed 2016-01-18 (2012)
58. Johansson, M.A., Cummings, D.A., Glass, G.E.: Multiyear climate variability and dengue—el nino southern oscillation, weather, and dengue incidence in puerto rico, mexico, and thailand: a longitudinal data analysis. PLoS Med **6**(11), 1000168 (2009)
59. Morin, C.W., Comrie, A.C., Ernst, K.: Climate and dengue transmission: evidence and implications. Environmental Health Perspectives (Online) **121**(11-12), 1264 (2013)
60. Thai, K.T., Anders, K.L.: The role of climate variability and change in the transmission dynamics and geographic distribution of dengue. Experimental Biology and Medicine **236**(8), 944–954 (2011)
61. of Vector-Borne Diseases Thailand, B.: Dengue fever situation in Thailand. http://www.thaivbd.org. Accessed 2015-12-1 (2015)
62. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2016). R Foundation for Statistical Computing. https://www.R-project.org/
63. Wood, S.N.: Thin-plate regression splines. Journal of the Royal Statistical Society (B) **65**(1), 95–114 (2003)
64. Ramadona, A.L., Lazuardi, L., Hii, Y.L., Holmner, Å., Kusnanto, H., Rocklöv, J.: Prediction of dengue outbreaks based on disease surveillance and meteorological data. PloS one **11**(3), 0152688 (2016)
65. Hii, Y.L., Rocklöv, J., Wall, S., Ng, L.C., Tang, C.S., Ng, N.: Optimal lead time for dengue forecast. PLoS Negl Trop Dis **6**(10), 1848 (2012)

66. Reich, N.G., Shrestha, S., King, A.A., Rohani, P., Lessler, J., Kalayanarooj, S., Yoon, I.-K., Gibbons, R.V., Burke, D.S., Cummings, D.A.: Interactions between serotypes of dengue highlight epidemiological impact of cross-immunity. Journal of The Royal Society Interface **10**(86), 20130414 (2013)
67. Gasparrini, A., Armstrong, B., Kenward, M.G.: Distributed lag non-linear models. Statistics in medicine **29**(21), 2224–2234 (2010)
68. Mondini, A., Chiaravalloti-Neto, F.: Spatial correlation of incidence of dengue with socioeconomic, demographic and environmental variables in a brazilian city. Science of the Total Environment **393**(2), 241–248 (2008)
69. Wu, P.-C., Lay, J.-G., Guo, H.-R., Lin, C.-Y., Lung, S.-C., Su, H.-J.: Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical taiwan. Science of the total Environment **407**(7), 2224–2233 (2009)
70. Gasparrini, A.: Distributed lag linear and non-linear models in r: the package dlnm. Journal of statistical software **43**(8), 1 (2011)

Appendix

Table 4: Approximate significance of smooth terms depicting lagged meteorological data.

Results				
P-Value	**Lag 0**	**Lag 1**	**Lag 2**	**Lag 3**
Mean Monthly DTR	5.26e-16	0.000261	0.000921	2e-16
Cumulative Monthly Rainfall	8.96e-15	2e-16	2e-16	8.77e-10
R-sq(adj.)	0.283			
Deviance Explained	31.9%			
RMSE	8.462			
SRMSE	0.52			

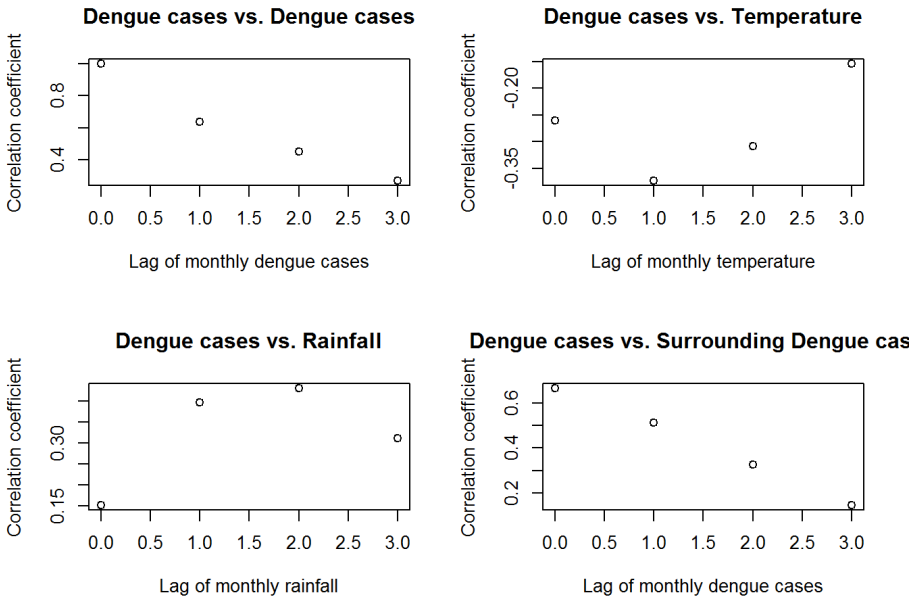


Figure 13: Cross-correlation between the Outcome and the Predictors.

Table 5: Approximate significance of smooth terms depicting lagged meteorological data.

Results				
P-Value	**Lag 0**	**Lag 1**	**Lag 2**	**Lag 3**
Mean Monthly DTR	5.26e-16	0.000261	0.000921	2e-16
Cumulative Monthly Rainfall	8.96e-15	2e-16	2e-16	8.77e-10
R-sq(adj.)	0.283			
Deviance Explained	31.9%			
RMSE	8.462			
SRMSE	0.52			