7th International Conference on
Combinatorics, Cryptography, Computer Science and Computing
November. 16-17, 2022

Dengue Disease Prediction with Seasonality of Environmental Factors

Piyumi Chathurangika[1],Kushani De Silva, Sanjeewa Perera
Research & Development Center for Mathematical Modeling, Department of Mathematics, University of Colombo, Sri Lanka

Abstract

Dengue is a public health problem to many of the tropical countries. The dengue threat in Sri Lanka has a significant social and economic impact on the country and hence predicting dengue cases in the immediate future is of utmost importance. It is believed that the dengue mosquito breeding is highly affected by environmental factors. In this study, dengue cases are predicted based on environmental factors and the study is narrowed down to Colombo municipal council area, at which the most Dengue cases are reported island-wide.

Keywords: Wavelet Analysis, Principal Component Analysis, Principal Component Regression Analysis, per capita vector density

1 Introduction

Dengue is a mosquito born virus transmitted to humans by the mosquitoes of Aedes species. Dengue fever and Dengue Haemorrhagic fever account for approximately 390 million infections annually, including over 22,000 deaths worldwide [1]. Brazil, Vietnam, Peru and Philippines are the countries with most dengue cases. Sri Lanka also reports a significant number of annual dengue cases bearing over 60,000 infections reported in 2022 from January to September [2]. Dengue is a serologically confirmed disease in Sri Lanka since 1962. There had been several outbreaks in over a period of 50 years with frequent outbreaks after 1990. Understanding this situation, Sri Lankan government has prepared a dengue action plan to properly analyse and mitigate the disease spread. It shows the allocation of money and resources together with the steps taken to handle the situation.

Several past studies have analyzed disease dynamics to propose control strategies such as insecticide usage and awareness programs and educational strategies at primary level school education [3, 4]. The main focus of these studies are to identify the disease transmission from the perspective of both host and vector. Due to the unavailability of vaccines or drugs for the dengue virus, controlling strategies have become the only solution that has yet helped in curtailing the reported dengue cases. Thus, predictive models help a nation in forecasting upcoming outbreaks and help prepare for such outbreaks in advance. Dengue prediction models are extensively studied over the past decades, with statistical and weather data using various methods [5]. Mathematical modeling is one of many such methods of forecasting a disease emergence. Among many

[1]speaker

mathematical and epidemiological models, compartmental models are prominent as they give a descriptive visualization of virus transmission between vectors and hosts [6, 7]. Most of the existing compartmental models have used fixed parameters values, where as a few has considered the parameter variation approach. When dengue transmission is concerned, the parameters used in models are commonly considered as fixed parameters [8, 9, 11]. Several studies have used parameter estimation methods utilizing the actual dengue data [9, 10]. This study uses an approach of using external environmental factors to obtain parameter values in the transmission model incorporating vector genetics as well.

In the next section of this paper, a dengue transmission model is introduced to identify and illustrate the virus spread between human hosts and mosquito vectors. The environmental factors affecting the dengue disease transmission is identified and explained in the Section 3 together with the methods used to analyse such existing effects. The dengue prediction model built using a simulation method is explained in the last two sections.

2 Disease Transmission Dynamics

Dengue disease transmission is highly correlated with environmental factors. Therefore understanding the relationship between environmental factors and dengue emergence and building a method to predict the dengue incidences that may occur in the future is the main focus of this study. The human population can be divided into 3 compartments (susceptible, infected and recovered) and the vector population can be divided into 2 compartments (susceptible and infected). The transmission of the disease between the 5 compartments can be visualized by Figure 1.

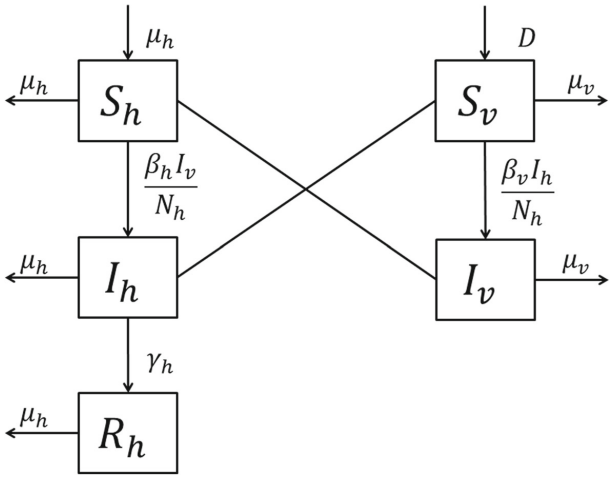

Figure 1: Dengue transmission compartmental schematic diagram

The disease transmission dynamics can mathematically be identified by means of ordinary differential equa-

tions as follows [12].

$$\frac{dS_h}{dt} = \mu_h(N_h - S_h) - \frac{\beta_h}{N_h} S_h I_v, \tag{1a}$$

$$\frac{dI_h}{dt} = \frac{\beta_h}{N_h} I_v S_h - (\mu_h + \gamma_h) I_h, \tag{1b}$$

$$\frac{dR_h}{dt} = \gamma_h I_h - \mu_h R_h, \tag{1c}$$

$$\frac{dS_v}{dt} = DN_v - \mu_v S_v - \frac{\beta_v}{N_h} I_h S_v, \tag{1d}$$

$$\frac{dI_v}{dt} = \frac{\beta_v}{N_h} I_h S_v - \mu_v I_v. \tag{1e}$$

The parameters and the variables used in this 5D model are as given in table 1.

Table 1: Parameters of the 5D model [12]

Parameter	Description
N_h	Total human population
N_v	Total vector population
S_h	Susceptible human population
I_h	Infected human population
R_h	Recovered human population
S_v	Susceptible vector population
I_v	Infected vector population
μ_h	host reproduction & mortality rate
μ_v	vector reproduction & mortality rate
γ_h	thost recovery rate
β_h	transmission rate from vector to host
β_v	transmission rate of from host to vector

For simplicity the above 5D model can be reduced to a 2D model considering vector and human population densities as well as the quasi equilibrium of the vector population [12]. The 2D model consisting of infected human population density dynamics and recovered human population dynamics is given by Equation 2.

$$\frac{dI}{dt} = \beta_h z \frac{\beta_v I}{\beta_v I + \mu_v}(1 - I - R) - (\mu_h + \gamma_h) I, \tag{2a}$$

$$\frac{dR}{dt} = \gamma_h I - \mu_h R, \tag{2b}$$

Table 2 gives the variables and the parameters of the 2D model.

Table 2: Parameters of the 2D model

Parameter	Description
I	Infected human population density
R	Recovered human population density
z	Per-capita vector density

3 Analysis of Environmental Factors

Since dengue is a vector borne disease, the behaviour of the disease highly depends on vector dynamics. The virus carrying aedes mosquito dynamics varies with many environmental factors owing to the breeding,

survival and biting behaviours [13, 14]. Thus, in this study we have considered, four environmental factors and the effect of each of them on dengue occurrences was analysed using Wavelet Analysis and Principal Component Analysis.

3.1 Wavelet Analysis (WA)

When vector biology is taken into account, it can be understood that the environmental factors affect dengue occurrences with certain time lags. The environmental factors considered in this study are, maximum temperature, minimum temperature, rainfall and humidity. In order to identify the time delays of the effect of each of these variables, cross wavelet analysis can be utilized.

The cross wavelet analysis uses time series data to identify existence of common power between two time series used [17]. Thus, in order to find the time lags the time series data of the dengue occurrences and the time series of each of the four environmental factors were used. In this study, a direct and an indirect cross wavelet analysis method was used to calculate the time lags. The results obtained were compared afterwards. The cross wavelet power spectrum of each of the environmental variables with reported dengue cases during the period from 2009-2016 are given in Figure 2.

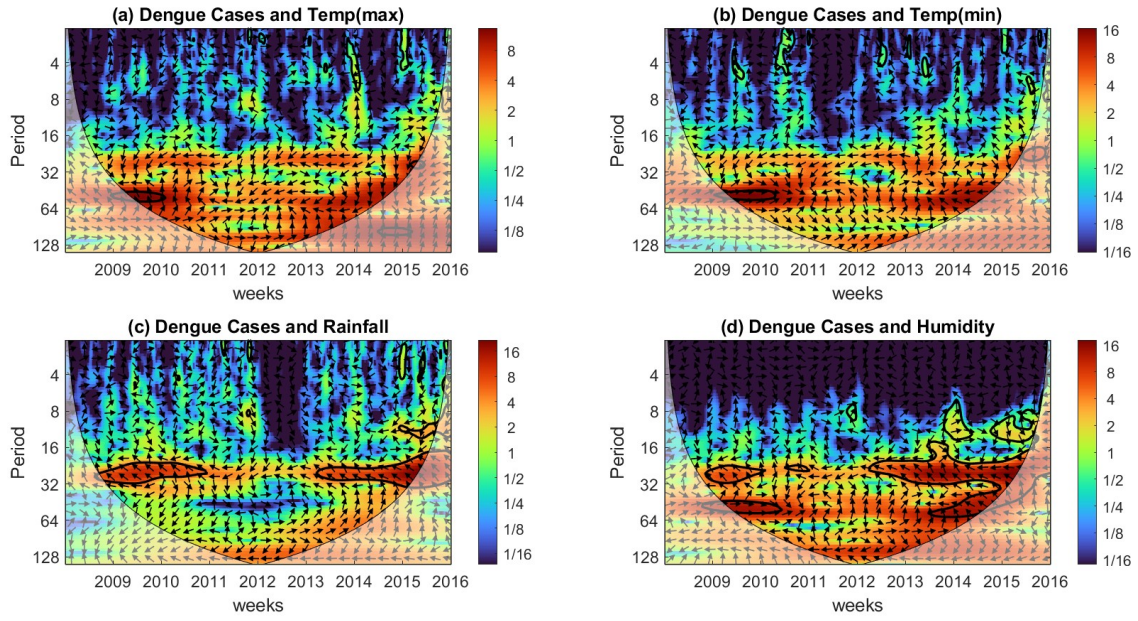

Figure 2: Cross wavelet power spectrum (a) between dengue cases and maximum temperature, (b) between dengue cases and minimum temperature, (c) between dengue cases and rainfall, (d) between dengue cases and humidity (2009-2016)

In the cross wavelet power spectrum, the significant high power between dengue time series and each environmental factor time series is depicted by the black thick line. Using the direct method, the angle of the arrows in the most significant area is measured and the time lag was calculated using the following formula.

$$\Delta t = \frac{\Delta\theta \times P}{2\pi} \tag{3}$$

Here Δt is the time delay, $\Delta\theta$ is the angle of the arrows and P is the periodicity.

The same analysis was conducted using an indirect method. In the indirect method the weekly time delays were utilized to change the direction of the arrows. The time at which the arrows in the high power regions faces right ($\theta = 0^0$) was taken to be the time lag.

Considering the power spectrum of each of the variables (Figure 2), a time lag between dengue occurrences and temperature values (maximum and minimum) could not be identified since the cross wavelet power spectrum does not identify any significant common power between the two time series (see figure 2a and 2b). For both rainfall and humidity 10 weeks time delays were calculated. The time delays obtained from both direct and indirect cross wavelet methods are the same (Table 3).

Table 3: Time lags (τ) of the environmental factors calculated using cross wavelet analysis

	Max Temp	Min Temp	Rainfall	Humidity
Direct wavelet Analysis	-	-	10	10
Indirect wavelet analysis	-	-	10	10

The common high power between dengue cases and temperature values could not be identified due to a low variability of temperature values in the CMC area. The maximum temperature though out the time period considered ranges from 27^0C to 34^0C (see Figure 3). This fact shows that the temperature in the CMC area is relatively warm and does not fluctuate beyond the range $27^0C - 34^0C$. The fluctuation of minimum temperature is within the range 22^0C to 29^0C (see Figure 4). These fluctuations are low when compared with the fluctuation of dengue cases during the considered period of time. These temperature ranges in the CMC area stays in the favourable range for dengue transmission though out an year [16]. Thus, the temperature fluctuation does not highly effect the disease transmission although it is a crutial condition. Due to this consistent favorability, the time lag of the temperature values are taken to be zero for this study.

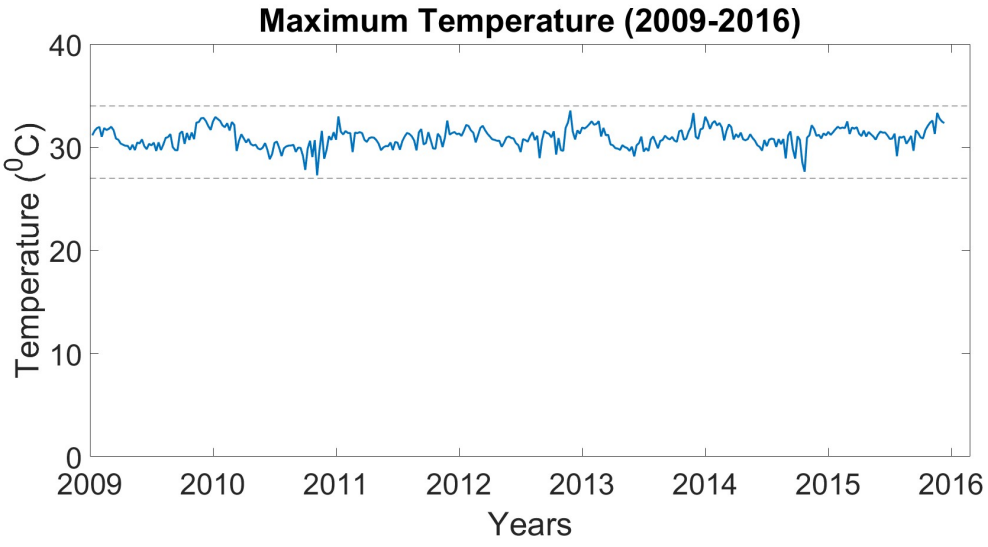

Figure 3: Variation of Maximum Temperature (2009-2016). The dotted line shows the maximum and minimum values of recorded maximum temperature

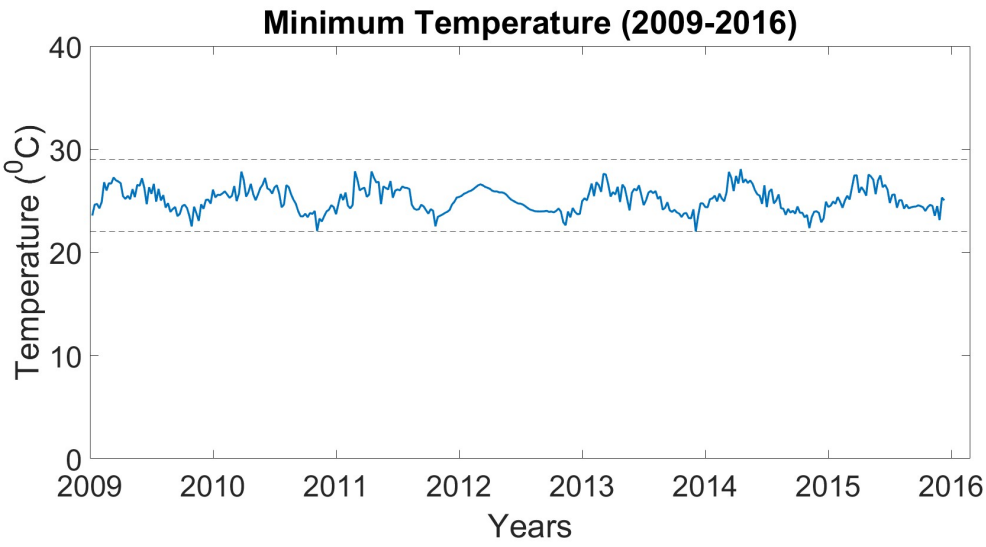

Figure 4: Variation of Minimum Temperature (2009-2016). The dotted line shows the maximum and minimum values of recorded minimum temperature

3.2 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a method of data transformation in such a way that a large set of variables are represented as a linear combination of smaller sets retaining most of the information and variance of the original data. By utilizing the method of PCA, one can determine the factors underlying the data by using the total variance of the data explained with fewer components [15].

The PCA for this study shows that the 4 environmental factors considered can be represented using only one principal component. By looking at the eigen vectors it was also seen that the highest correlation was seen between PC_1 and rainfall. The environmental factors can be represented using the PC_1 as given by Equation 3.

$$PC_1 = (-0.0236)MT_{(t)} + (-0.0199)mt_{(t)} + 0.9904R_{(t-\tau)} + 0.1348H_{(t-\tau)} \tag{4}$$

Here MT_t and mt_t are maximum temperature and minimum temperature at time t respectively and $R_{(t-\tau)}$ and $H_{(t-\tau)}$ are rainfall and humidity at time $t-10$ respectively. These time points are taken by the cross wavelet analysis. Therefore, PC_1 is a linear combination of the 4 environmental factors with their respective time lags.

3.3 Principal Component Regression Analysis (PCR)

Principal Component Regression (PCR) is used together with the PCA to minimize the multicollinearity between explanatory variables and dependant variable [18]. In this study, the explanatory variables are the environmental factors and the dependent variable is the reported dengue cases. The relationship with the environmental factors and the reported dengue cases, is found by a regression analysis by using the principal components selected. Since the selected principal component (PC_1) is a representation of all the 4 environmental factors, the PCR results give the relationship between reported dengue cases and the environmental factors with their respective time lags. The PCR equation obtained is,

$$I_{h(t)} = a_0 + a_1 PC_1 \quad \text{where } a_0 = 0.0003149 \text{ and } a_1 = 8.6726 \times 10^{-6} \tag{5}$$

4 Per Capita Vector Density

In the dengue transmission model given in section 2, there are several parameters. Out of these parameters, the most dynamic parameter can be identified as per-capita vector density. Per-capita vector density (z) is defined as the number of mosquitoes per person. The number of mosquitoes in any given environment changes frequently depending on many external factors. In this study the external factors affecting the dynamics of per-capita vector density are considered to be the environmental factors. Moreover, the dynamic behaviour of per-capita vector density highly affect the dengue occurrence since the virus is transmitted to humans by vectors. The relationship between per-capita vector density and human population density can be identified using incubation periods of the dengue virus (see Figure 5).

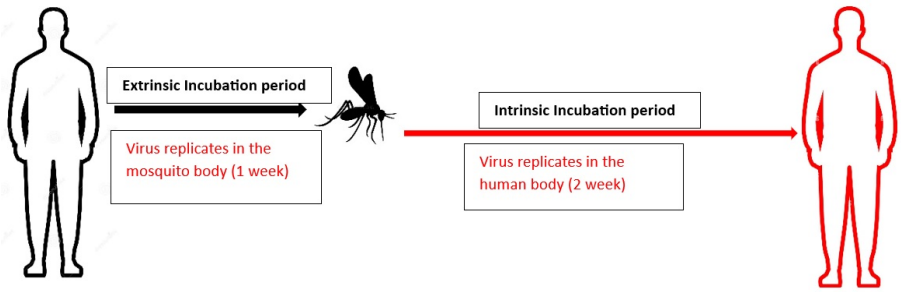

Figure 5: Transmission timeline of dengue virus from host to vector (extrinsic incubation period) and vector to host (intrinsic incubation period)

By taking into account the timeline of the incubation periods of both mosquito and human (see Figure 3), it can be concluded that $z_{(t-2)}$ affects $I_{h(t)}$ where t is measured in weeks. i.e.,

$$I_{h(t)} \propto z_{(t-2)} \tag{6}$$

Since infected vectors transmit the disease to humans, the per-capita vector density z is in fact the infected vector population per person. The infected vector population at time $(t-2)$ depends on infected human population at time $(t-3)$. Therefore we also have the relation,

$$z_{(t-2)} \propto I_{h(t-3)} \tag{7}$$

From these two relations we get the two equations,

$$I_{h(t)} = k_1 z_{(t-2)} \tag{8}$$

$$z_{(t-2)} = k_2 I_{h(t-3)} \tag{9}$$

From the Principal Component Analysis (PCA) the equation that gives the connection between dengue cases and environmental factors was obtained (Equation 5). Including this relationship in equation 8 and equation 9, we get

$$(a_0 + a_1 PC_1) = k_1 k_2 I_{h(t-3)} \tag{10}$$

Therefore, the function for per-capita vector density can be defined only using the environmental factors as given by Equation 10.

$$z_{(t-2)} = k(a_0 + a_1 PC_1) \quad \text{where} \quad k = \frac{1}{k_1} \tag{11}$$

Since the environmental factors are seasonal the value k is also seasonal. Thus, the parameter k can be defined as the seasonal k for this study.

5 Prediction of Dengue Incidences in CMC area

5.1 Simulation of Dengue Incidences

The results obtained from the PCR was used to simulate the dengue cases in the period from 2009 to 2016. For this simulation, the seasonal k value in equation 11 is used by selecting the increasing and decreasing patterns in the reported dengue cases. These fluctuation in the data were identified as the seasons in a year. For each selected season different seasonal k values were chosen using trial and error method. Then the obtained simulated lines were matched with the actual data to get an approximate seasonal k value. The simulated results are shown in Figure 6

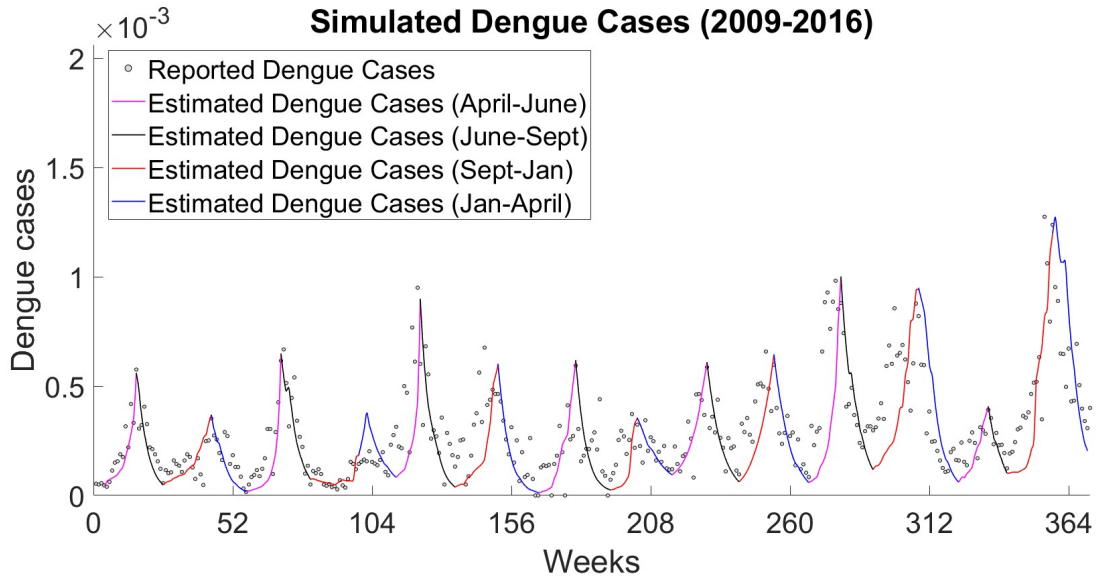

Figure 6: Simulated Dengue cases from 2009-2016: Each season selected are given in 4 different colors. The simulated lines are obtained with different seasonal k values.

5.2 Model Validation

In order to validate the method 70% of the data were used as the trial data and 30% were used as test data. The simulation results were obtained for the trial data and the seasonal k values were recorded. The recorded seasonal k values shows significant clusters (see Figure 7). Thus, the k value can be considered as significant for each season, which also confirms the seasonal variation of per-capita vector density resulting in seasonality of dengue occurrences.

Considering this clustering of k for each season, for the test data the mean of the k values were calculated separately for each season. Since this study used a trial and error method, the k values may contain errors. Therefore, the 90% confidence intervals (CI) were also calculated for each season using the Equation 11.

$$CI = \overline{X} \pm \alpha \frac{s}{\sqrt{n}} \tag{12}$$

Where,$\overline{x}$ =sample mean, α =confidence level value, s =sample standard deviation, n =sample size.

The mean seasonal k values and the 90% Confidence Intervals of each season are shown in Table 4.

The trial data results and test data results are depicted in Figure 8, in which the trial data simulations are given in colored lines and test data simulations are given in stared colored lines.

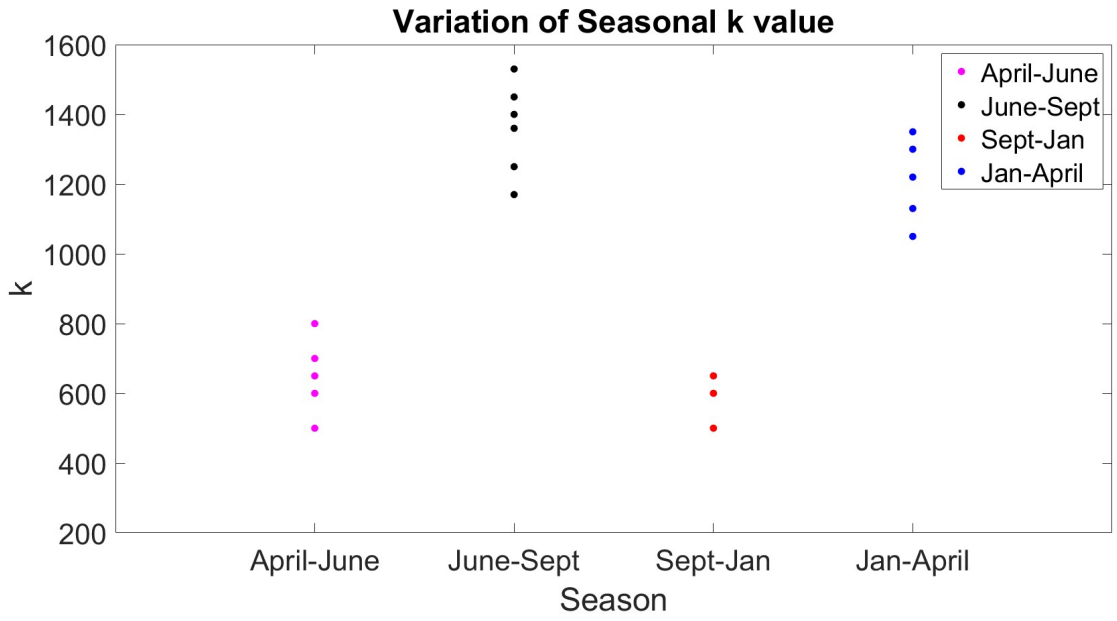

Figure 7: Seasonal k chosen are clustered according to the seasons in each year

Table 4: Mean and 90% confidence intervals for seasonal k

Season	$\overline{k}$	90% CI
January-April	650	[571.77161, 728.22839]
April-June	1365.714	[1290.34955, 1441.07845]
June-September	550	[509.86968, 590.13032]
September-January	1197.143	[1107.071, 1287.215]

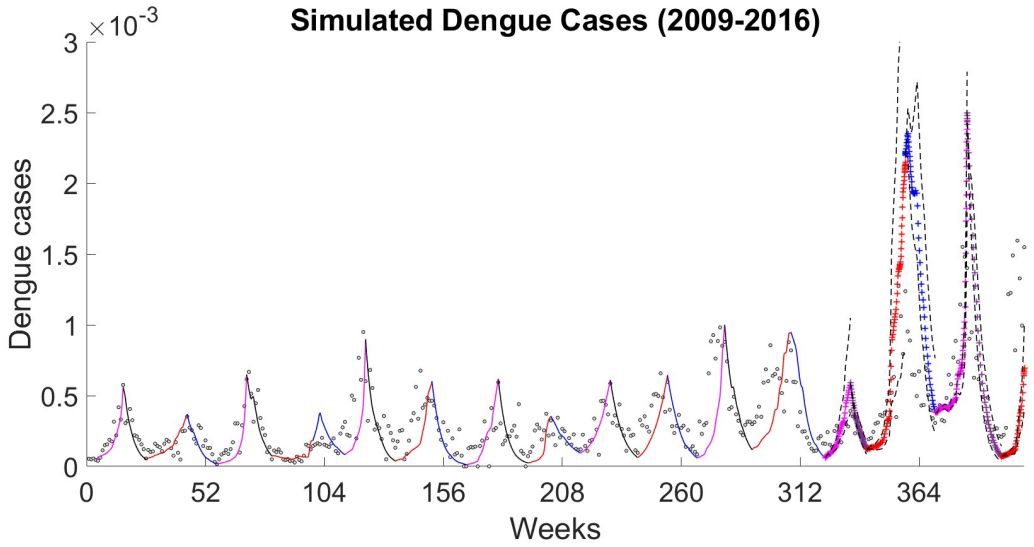

Figure 8: Predicted Dengue cases: 30% of the data set (test data) was predicted using the simulated results of 70% of data (trial data). The stared line shows the predicted dengue cases in each season obtained using the mean seasonal k value calculated using the trial data. The black dotted line shows the 90% confidence bounds for each season.

By looking at the test data it can be seen that the predicted dengue cases are not accurate after the second season in the test data set. For these predictions, in addition to k values the end point of each predicted season was considered to be the starting point of the next season. Therefore, the errors in each season gets

accumulated when this method is used for a longer period of time. Therefore, this method can only be used to predict dengue cases of not more than 2 upcoming seasons.

5.3 Prediction of Dengue Incidences of an Upcoming Season

Using the mean seasonal k value and the 90% confidence intervals the dengue cases in the 2016 April to June season was predicted as shown in Figure 9. For this prediction it is assumed that the environmental data in this season are not known. Since it was found that the environmental factors affect dengue cases with certain time lags, this prediction was done using those past environmental values only. From wavelet analysis it was found that the temperature values does not have any time lag. Therefore the mean of the temperature values calculated from past data were used for the prediction. Using the known rainfall and humidity data of the last 10 weeks with mean temperature values, the dengue incidences of the upcoming 10 weeks is predicted using the mean seasonal k values. The results are shown in Figure 9. In the figure the magenta stars represent the predicted dengue cases of the upcoming 10 weeks from April to June. The black dotted lines show the 90% confidence intervals.

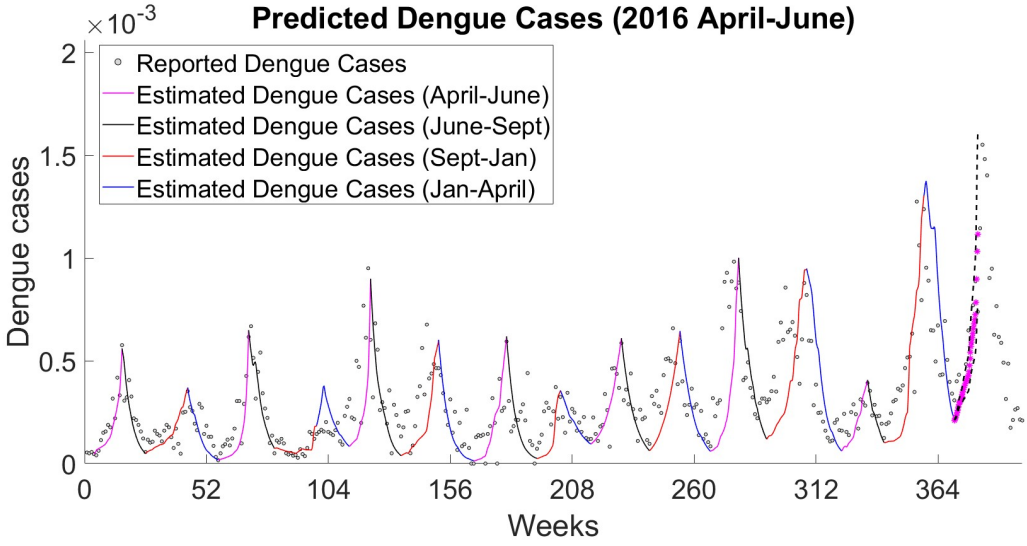

Figure 9: Prediction of Dengue cases for the April-June season in 2016 given in stared magenta line. The 90% confidence bound for this season is given in black dotted line

6 Discussion

Many studies conducted to analyze and predict dengue dynamics have used a variety of approaches. Sri Lanka is one of the countries that have been affected by Dengue for a longer time. In this work toward the the attempt to predict dengue cases, we first modelled the per-capita vector density using environmental factors. The time lags of the environmental factors identified in this model was found using cross-wavelet analysis and principal component analysis. Out of the four environmental factors, rainfall and humidity show a time lag of 10 weeks. Since the temperatures in the CMC area provide consistent and favorable conditions to breeding and spreading, the minimum and maximum temperatures do not have any time lags. The seasonal factors in the per-capita density was identified as k which also aligns with the seasonality of environmental factors.

Hence, the findings in this study show that the seasonal trend in reported dengue cases also occurs due to the variation of environmental factors with their respective time lags. Thus, using the proposed method, it is possible to predict both the expected trend and the number of dengue cases for the upcoming season using only the data of environmental factors.

References

[1] World Health Organization. (n.d.). Dengue and severe dengue. World Health Organization. Retrieved October 29, 2022, from https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

[2] Google. (n.d.). National Dengue Control Unit. Retrieved November 2, 2022, from https://datastudio.google.com/reporting/95b978f1-5c1a-44fb-a436-e19819e939c0/page/XRtTB

[3] Khatua, A., & Kar, T. K. (2020). Dynamical behavior and control strategy of a dengue epidemic model. The European Physical Journal Plus, 135(8), 1-22.

[4] Rather, I. A., Parray, H. A., Lone, J. B., Paek, W. K., Lim, J., Bajpai, V. K., & Park, Y. H. (2017). Prevention and control strategies to counter dengue virus infection. Frontiers in cellular and infection microbiology, 7, 336.

[5] Chen, S. C., & Hsieh, M. H. (2012). Modeling the transmission dynamics of dengue fever: implications of temperature effects. Science of the total environment, 431, 385-391.

[6] Van den Driessche, P., & Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical biosciences, 180(1-2), 29-48.

[7] Syafruddin, S., & Noorani, M. S. M. (2012). SEIR model for transmission of dengue fever in Selangor Malaysia. In International Journal of Modern Physics Conference Series (Vol. 9, pp. 380-389).

[8] Syafruddin, S., & Noorani, M. S. M. (2012). SEIR model for transmission of dengue fever in Selangor Malaysia. In International Journal of Modern Physics Conference Series (Vol. 9, pp. 380-389).

[9] Pandey, A., Mubayi, A., & Medlock, J. (2013). Comparing vector–host and SIR models for dengue transmission. Mathematical biosciences, 246(2), 252-259.

[10] Coudeville, L., Baurin, N., & Vergu, E. (2016). Estimation of parameters related to vaccine efficacy and dengue transmission from two large phase III studies. Vaccine, 34(50), 6417-6425.

[11] de los Reyes, A. A., & Escaner IV, J. M. L. (2018). Dengue in the Philippines: model and analysis of parameters affecting transmission. Journal of biological dynamics, 12(1), 894-912.

[12] Erandi, K. K. W. H., Perera, S. S. N., & Mahasinghe, A. C. (2021). Analysis and forecast of dengue incidence in urban Colombo, Sri Lanka. Theoretical Biology and Medical Modelling, 18(1), 1-19.

[13] Chaiphongpachara, T., Yusuk, P., Laojun, S., & Kunphichayadecha, C. (2018). Environmental factors associated with mosquito vector larvae in a malaria-endemic area in ratchaburi province, Thailand. The Scientific World Journal, 2018.

[14] Dom, N. C., Ahmad, A. H., Ishak, A. R., & Ismail, R. (2013). Assessing the risk of dengue fever based on the epidemiological, environmental and entomological variables. Procedia-Social and Behavioral Sciences, 105, 183-194.

[15] Pinto, E., Coelho, M., Oliver, L., & Massad, E. (2011). The influence of climate variables on dengue in Singapore. International journal of environmental health research, 21(6), 415-426.

[16] Tran, B. L., Tseng, W. C., Chen, C. C., & Liao, S. Y. (2020). Estimating the threshold effects of climate on dengue: a case study of Taiwan. International journal of environmental research and public health, 17(4), 1392.

[17] Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. Bulletin of the American Meteorological society, 79(1), 61-78.

[18] Liu, X., Liu, K., Yue, Y., Wu, H., Yang, S., Guo, Y., ... & Liu, Q. (2021). Determination of factors affecting dengue occurrence in representative areas of China: a principal component regression analysis. Frontiers in Public Health, 8, 603872.