Trabalho apresentado no XLI CNMAC, Unicamp - Campinas - SP, 2022.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics
Preprint

Temperature-based Dengue Outbreaks Modelling with Exogenous Variables

Juan V. Bogado [1]
Universidad Nacional de Caaguazú, Coronel Oviedo, Paraguay
Universidad Nacional de Asunción, Asunción, Paraguay
Diego H. Stalder [2], Christian E. Schaerer [3]
Universidad Nacional de Asunción, Asunción, Paraguay
Max Ramírez Soto [4], Denisse Champin [5]
Facultad de Ciencias de la Salud, Universidad Tecnológica del Perú, Lima, Perú

Abstract. Dengue fever is an endemic disease, present in tropical and subtropical regions, transmitted by the *Aedes Aegypti* mosquito vector. It has recently appeared in non-tropical regions with dry weather. This represents a setback for advanced temperature-based reference models, since mosquitos reproductive cycle does not necessarily match with the outbreaks. This situation indicates that other variables are also involved in epidemic outbreaks. In this work we propose to include a component that capture this process, whether entomological, environmental or related to population mobility, and include it to the reference model by adding a Gaussian function to the formulation of humans (β_h) and vectors (β_v) transmission rate. The parameters to be adjusted for this function were evaluated by a probabilistic model selection experiment. The parameters for this function are u, σ and k. The results indicate that, our model outperforms the reference model, and that additional information about outbreaks can be obtained from the new parameters. .

Keywords. Dengue, Temperature-based Models, SIR-SI.

1 Introduction

Dengue fever is a viral disease transmitted mainly by the *Aedes Aegypti* mosquito acting as a vector. This disease causes hospitalizations and even death. Its endemic characteristic makes it a public health problem. Dengue fever is now endemic in Africa, America, Asia and the Western Pacific. In South America, there has been a dramatic increase in cases in countries such as Colombia, Ecuador, Paraguay, Peru[6], Venezuela and Brazil [4].

Statistical and mathematical models have been proposed in order to capture the characteristics of the outbreaks [2, 3]. Compartmental models e.g. SIR, SEIR, SIS, are the classic modeling strategy and allows characterize an outbreak by ranking the population into different compartments. Several correlations [5] have been found which are use to predict the characteristics of the outbreaks and thus explain this phenomenon. Climate-related variables (temperature, humidity, rainfall) have been studied with promising results in Asian countries [6]. However, these models will fail to fit the data (when the consider only temperature), because they do not have enough

[1]juan.vicente.bm@pol.una.py
[2]dstalder@ing.una.py
[3]cschaer@pol.una.py
[4]max.ramirez@upch.pe
[5]dchampin@utp.edu.pe
[6]Between 2004 and 2017, a constant increase in Dengue cases has been observed in Lima.

flexibility to consider other variables involved in the transmission process such as overwintering of mosquito eggs, population mobility and tourism. Furthermore, with the expansion of the epidemic to desert regions, temperature data do not necessarily coincide with the reproductive cycles of the vector. Therefore other factor that can trigger the outbreak such as the arrival of infected vectors and/or people [7]. This work propose the use of well characterized Gaussian function to capture the missing effect.

2 Methodology

In this work, the model known as SIR-SI[7] has been used. This model has a transmission rate that is temperature dependent, similarly to Lee et al. [6]. The region of study, Lima, Peru is part of the Sechura desert, therefore to apply the same model we introduce a Gaussian function to capture the outbreak dynamics. Then the free parameters are estimated using a log-normal likelihood function to optimize the model predictions using a differential evolution algorithm [9]. With all possible combinations of the free parameters, a model selection experiment was performed to define which combination gives better predictions. Finally, the best fitting curves of the reference model against the new model are presented.

2.1 Dataset

The weekly cases of Dengue organized in 43 districts of the province of Lima were obtained from the National Center for Epidemiology, Disease Prevention and Control (CDC Peru). For this work, we add all the cases of the districts weekly obtaining the total outbreaks of Lima per year for the years 2017, 2019, and 2020.

2.2 Mathematical Models

Let $S_h \in \mathbb{N}$, $I_h \in \mathbb{N}$ and $R_h \in \mathbb{N}$ be the fraction of Susceptible, Infected and Recovered humans (host), respectively. $S_v \in \mathbb{N}$ and $I_v \in \mathbb{N}$ are the fraction of Susceptible and Infected mosquitoes (vectors). Then the SIR-SI model is defined by the following set ordinary differential equations (ODE):

$$\frac{dS_h}{dt} = -\beta_v S_h I_v \tag{1}$$

$$\frac{dI_h}{dt} = \beta_v S_h I_v - \gamma I_h \tag{2}$$

$$\frac{dR_h}{dt} = \gamma I_h \tag{3}$$

$$\frac{dS_v}{dt} = -\beta_h S_v I_h - \mu S_v \tag{4}$$

$$\frac{dI_v}{dt} = \beta_h S_v I_h - \mu I_v, \tag{5}$$

where $\beta_v \in \mathbb{R}$ is the transmission rate from vector to host and $\beta_h \in \mathbb{R}$ from host to vector; $t \in \mathbb{R}$ is the time, $\gamma \in \mathbb{R}$ is the recovery rate for hosts, $\mu \in \mathbb{R}$ is the death rate for vectors. This set of equations are solved as initial value problem by setting:

$$S_{h_0} = 1 - \frac{I_{h_0}}{N_h}, \quad I_{h_0} = \frac{1}{N_h}, \quad R_{h_0} = 0, \quad S_{v_0} = 1 - \frac{I_{v_0}}{N_v}, \quad I_{v_0} = \frac{1}{N_v}, \tag{6}$$

[7]SIR-SI: Susceptible-Infective-Recovered for human populations; Susceptible-Infective for vector populations

where N_h and N_v are human and vector populations respectively and representing the values at time $t = 0$. Furthermore, according to the reference model $N_v = 2N_h$. In the literature the most common free parameters are N_h, β_h, β_v, γ and μ.

Reference Model: Lee et al. (2018) assume that μ, β_h and β_v are defined by a functions of the temperature $T(t)$, i.e. $b_{(T(t))} \in \mathbb{R}$, $b_{h(T(t))} \in \mathbb{R}$ and $b_{v(T(t))} \in \mathbb{R}$ and $\mu_{v(T(t))} \in \mathbb{R}$ (see Figure 1). Therefore the transmission rates are defined as follows:

$$\beta_{h(t)} = x_1\, b_{(T(t))}\, b_{h(T(t))} \tag{7}$$

$$\beta_{v(t)} = x_2\, b_{(T(t))}\, b_{v(T(t))}, \tag{8}$$

where the two main free constant parameters are $x_1 = \{x \in \mathbb{R}, 0 < x < 1\}$ and $x_2 = \{x \in \mathbb{R}, 0 < x < 1\}$.

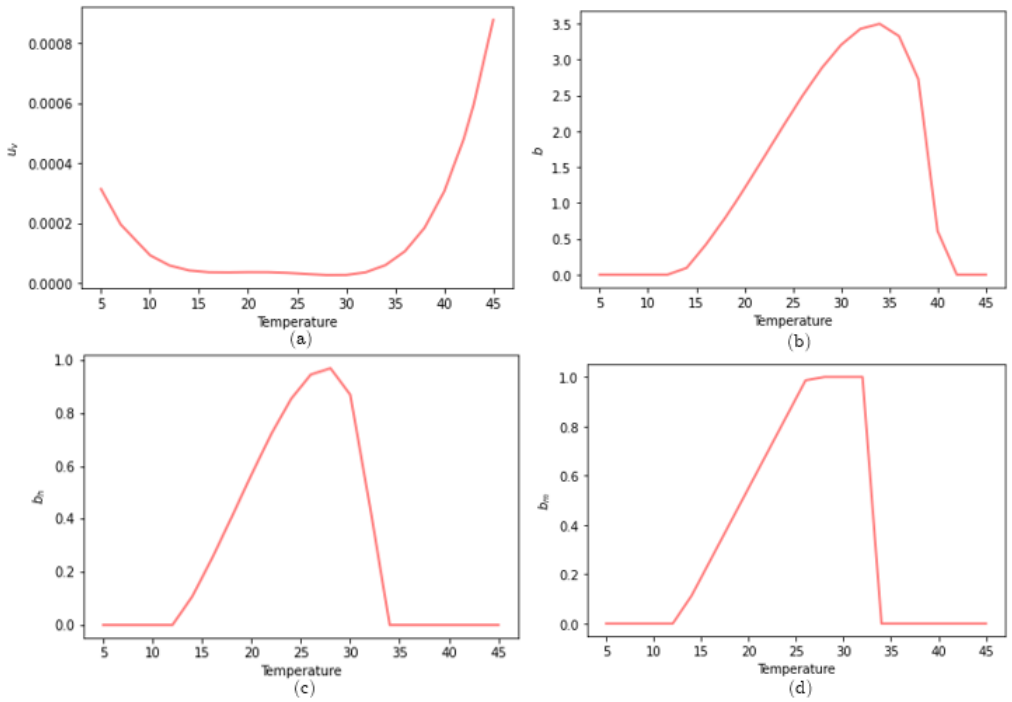

Figure 1: (a), (b), (c), and (d) show temperature-based functions for μ, b, b_h, and b_m, respectively. Adapted from [6].

Note that to fit this model N_h, x_1, and x_2 are let free.

Model with β_{ex} : Since the outbreaks peak does not necessary match with the week the highest transmission rate $b_{(T(t))}$, $b_{h(T(t))}$ and $b_{v(T(t))}$. This work replaces the constants x_1 and x_2 by a time dependent variable $\beta_{ex(t)} \in \mathbb{R}$ using a Gaussian function. Thus, the computation of $\beta_{h(t)}$ and $\beta_{v(t)}$ are redefined as:

$$\beta_{h(t)} = \beta_{ex(t)}\; b_{(T(t))}\; b_{h(T(t))} \tag{9}$$

$$\beta_{v(t)} = \beta_{ex(t)}\; b_{(T(t))}\; b_{v(T(t))}, \tag{10}$$

where β_{ex} is a time dependent Gaussian function:

$$\beta_{ex} := ke^{-\frac{(t-u)^2}{2\sigma^2}}, \tag{11}$$

where $k \in \mathbb{R}$ is a constant, $u \in \mathbb{R}$ is the mean (the week with which the transmission is triggered), and $\sigma^2 \in \mathbb{R}$ the variance of the function. For the fit we use likelihood as a cost function for optimization. Note that this model can have more free parameters which are N_h, u, σ and k.

One common assumption to fit the model to a given set of weekly reported cases (an outbreak) is that the observational errors follow a normal distribution (likelihood) i.e. least square error. Thus the normal distribution should return the maximum value when the model prediction is perfect i.e. optimal (the maximum likelihood estimation, MLE). The optimal parameters can be estimated by minimizing the sum of the negative log-likelihood (SNLL) as follows:

$$SNLL := -\sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(v_i-u_i)^2}{2\sigma^2}}\right), \tag{12}$$

where σ^2 is the mean of cases, u_i is the prediction for the time t, n is the number of observations, and v_i represent the observed cases at time t. Note that to obtain the model predictions the ODE should be approximated numerically for each step of the optimization algorithm.

2.3 Model Selection

This subsection presents results of the model selection experiment. This experiments seeks to determine which parameters of the Gaussian function should be left free to have an reasonable optimal solution. Four experiments were set up,

1. Model 1: let free u and $\sigma = 1$, $k = 1$,
2. Model 2: let free u, σ and $k = 1$,
3. Model 3: let free u, k and $\sigma = 1$,
4. Model 4: let free u , σ and k.

Then given a set of observations, each best fit solutions is compared to select which model explains better the data. The comparison was made considering probabilistic statistical measures that attempt to quantify the models performance:

1. The optimal value of the SNLL is called Maximum Likelihood Estimation (MLE). However, this function does not penalize solutions where the model complexity is increased i.e. with more free parameters in the model.
2. Akaike Information Criteria (AIC) includes a penalization for the model complexity as follow:
$$AIC := 2q - 2\ln(MLE) \tag{13}$$
where q is the number of free parameters [1]. Although, it does not take account of the uncertainty in the model parameters.
3. Bayesian Information Criteria (BIC), which is defined as:
$$BIC := q\ln(n) - 2\ln(MLE), \tag{14}$$
where q is the number of model parameters and n the number of observations.

The AIC, BIC and MLE were used to assess the quality of the best fit of each model [8]. The lower the values, the better the fit.

3 Numerical Results

This section presents the numerical experiments results. Table 1 presents the model selection evaluation. In the first column, we have different models to select, then the evaluation metrics and in the following columns, we have the metric values for different years. Lower values are the best ones. In all cases, Model 4 gives better results for AIC, BIC, and MLE. In some cases, it is up to twice as good as other methods despite being more complex in terms of the number of parameters.

Table 1: Evaluation of the fit of the models using AIC, BIC and MLE.

Model	Metric	2017	2019	2020
Model 1	AIC	1463.0973	2690.7484	1438.6672
	BIC	1463.3999	2691.0511	1438.9698
	MLE	730.5486	1344.3742	718.3336
Model 2	AIC	1438.3085	2248.1844	1439.1038
	BIC	1438.9136	2248.7896	1439.7089
	MLE	717.1542	1122.0922	717.5519
Model 3	AIC	1466.7228	1952.3625	1424.4774
	BIC	1467.3279	1952.9676	1425.0826
	MLE	731.3614	974.1812	710.2387
Model 4	AIC	286.9677	1201.5911	263.0954
	BIC	287.8754	1202.4988	264.0032
	MLE	140.4838	597.7955	128.5477

As shown in Figure 2, the adjust of the infected curve to the observed data is substantially improved by incorporating β_{ex}, fitting according to Model 4. This model allows to adjust the three parameters of the Gaussian function that determines the parameter β_{ex}. These parameters (u, σ and k) are the ones that define the shape of β_{ex} which are then multiplied by b and $b_{h,v}$ introducing the necessary process to correctly model the outbreak. The values of β_h in the benchmark model and β_h with β_{ex} included, in addition to β_{ex}, can be seen in Figure 3.

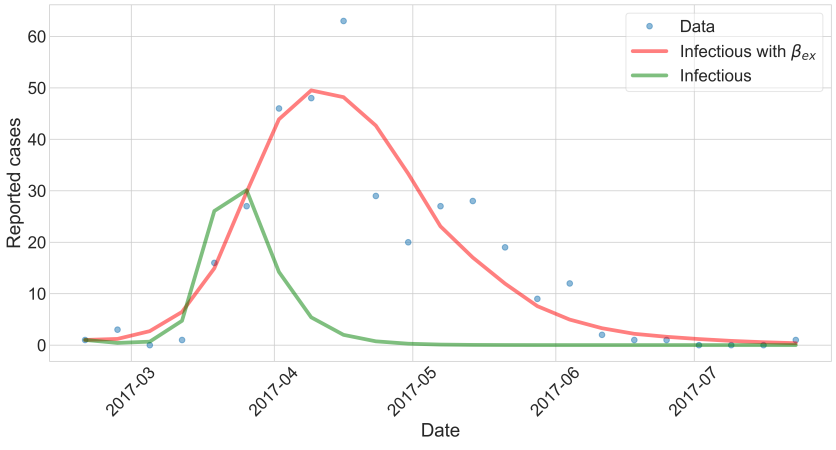

Figure 2: Data from the SIR-SI model with the exogenous variable and the climatic conditions adjusted for year 2017. Adjustment without β_{ex} (in green) and adjustment with β_{ex} according to Model 4 (in red).

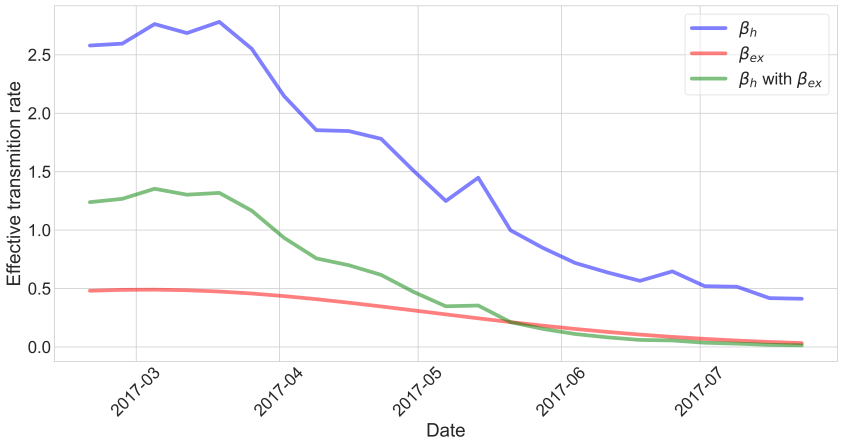

Figure 3: Plot of the β_{ex} component, β_h based only on temperature and with the added β_{ex} component for the year 2017.

Models 2 and 3 set a parameter at 1, k and σ, respectively, while Model 1 only adjusts u. These three models have similar performance, which tells us that setting a parameter prevents our proposed component β_{ex} from fitting correctly.

Although Model 4 is selected as the best in terms of the metrics chosen, Table 2 shows that there is a consistency between the values taken by the variables. This Table shows the values taken by the adjusted parameters for each model and allows us to analyze and interpret them in the context of the epidemiological outbreak. Since $\beta_{h,v}$ represents the transmission rate, the parameters u, σ and k of β_{ex} have a direct bearing on it. The mean value u indicates the week when the effective transmission rate is higher, the variance σ^2 is relate to the duration of the outbreak and k represents the importance of exogenous factor against the temperature in the adjustment. Therefore the model can extract new information from the data.

Table 2: Values obtained in the adjusted parameters in each experiment.

Model	Parameter	2017	2019	2020
Model 1	u	5.4881	6.9874	5.2181
Model 2	u	6.1528	7.3931	5.1911
	σ	1.0719	0.7461	0.9843
Model 3	u	5.4863	6.8787	5.1141
	k	1.0033	0.7371	0.9744
Model 4	u	2.3703	7.9240	0.0507
	σ	7.8431	4.7796	5.7511
	k	0.3833	0.2249	0.4785

These parameters can be used to characterize the outbreaks, in addition to the information already obtained with the SIR-SI model and the information from the climate-dependent variables. Therefore the model can provide new information from the data.

4 Conclusions

The experiments indicate that Gaussian function is effective to overcome the limitation of the reference model [6]. The model selection experiment indicates that is better to have the three parameters of the Gaussian function as free parameters.

The amount of information obtained from the phenomenon is even richer than that the reference model. The parameters are capable to extracting the duration of the epidemic (determined by σ^2), the peak of the transmission rate (determined by u) and the importance of the exogenous or hidden factors in the model (determined by k).

The model can capture the epidemic dynamic process. However, the question that arises is the understanding of the process itself (in terms of physical, biological, and entomological meanings). An entomological or transport-based explanation, human mobility, socio-cultural behavior such as accumulating water in a reservoir are viable options to be explored in future works.

Acknowledgments

This work is granted by the P-2020-LIM-01 project of the Technological University of Peru. Christian E. Schaerer and Diego H. Stalder thanks FEEI-PROCIENCIA-CONACYT-PRONII.

References

[1] H. Akaike. “Information theory and an extension of the maximum likelihood principle”. In: **Selected papers of hirotugu akaike**. Springer, 1998, pp. 199–213.

[2] J. V. Bogado. “Time series clustering and data augmentation techniques to improve the forecast of Dengue cases in Paraguay with deep learning”. Master dissertation. Universidad Nacional de Asunción, Facultad Politécnica, 2021.

[3] J. V. Bogado et al. “Time Series Clustering to Improve Dengue Cases Forecasting with Deep Learning”. In: **2021 XLVII Latin American Computing Conference (CLEI)**. IEEE. 2021, pp. 1–10.

[4] M. Derouich, A. Boutayeb, and E. H. Twizell. “A model of dengue fever”. In: **Biomedical engineering online** 2.1 (2003), pp. 1–10.

[5] S. Gómez-Guerrero et al. “Measuring Interactions in Categorical Datasets Using Multivariate Symmetrical Uncertainty”. In: **Entropy** 24.1 (2021), p. 64.

[6] H. Lee et al. “Potential effects of climate change on dengue transmission dynamics in Korea”. In: **PLoS One** 13.6 (2018), e0199205.

[7] J. M. Reinhold, C. R. Lazzari, and C. Lahondère. “Effects of the environmental temperature on Aedes aegypti and Aedes albopictus mosquitoes: a review”. In: **Insects** 9.4 (2018), p. 158.

[8] C. Rodrıguez. “The ABC of model selection: AIC, BIC and the new CIC”. In: **AIP Conference Proceedings**. Vol. 803. 1. American Institute of Physics. 2005, pp. 80–87.

[9] P. Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: **Nature methods** 17.3 (2020), pp. 261–272.

