Performance optimization through hybrid modelling: An application to dengue disease

K. WIJESINGHE, S. PERERA, C. ATTANAYAKE

khisalwijesinghe1998@gmail.com, ssnp@maths.cmb.ac.lk, succ@kln.ac.lk

Department of Mathematics, University of Colombo,
Department of Statistics, University of Kelaniya

Although Dengue virus could be prevented through responsible human actions, it has become a serious threat to mankind. This study was intended to increase the prediction accuracy of dengue transmission using hybryd models. After forecasting with Grey Forecasting Model, Growth Curve Model, Alpha Sutte Indicator and Generalized Additive Model, the models with the best prediction accuracy were determined through lowest Mean Absolute Percentage Error (MAPE) recorded in error calculation. Accordingly, a hybrid model was developed, by using a weighted average method as a coupling technique. Through the calculations and the analysis carried out, Alpha Sutte Indicator and the Generalized Additive Model were chosen to develop the Hybrid Model. The model enhances the prediction accuracy for most of the regions in Sri Lanka. Forecasting dengue transmission accurately is important to allocate medical personnel and equipment, conduct effective environmental management and awareness programs and chemical vector controlling in correspondence to the rising figures of dengue patients.

1. Introduction

Dengue is a mosquito-borne viral disease, transmitted by female mosquitoes *Aedes aegypti* and *Ae. albopictus* in the tropical regions According to the WHO website, the scientists have identified four serotypes of the Dengue Virus (DENV1, DENV2, DENV3 and DENV4) [1].

Dengue is mostly found in tropical climates worldwide and about 100-400 million dengue infections are occurring each year worldwide. In Sri Lanka, Dengue fever results in an unfortunate figure of around 30,000 cases and 100 deaths annually. The highest number of deaths in the recent past; 440 were recorded in 2017, which has made it a crucial public concern. Thus, forecasting in this field are vital, which May otherwise result in loss of people's lives and collapse of economic security of the country [2].

The objective of this study was to minimize the error in predicting dengue transmissions by using hybrid models rather than the individual models.

2. Methodology

Forecasting of dengue disease transmission was done using Grey Forecasting Model, Growth Curve Model, Alpha Sutte Indicator and Generalized Additive Model. Accordingly, dengue cases were forecasted monthly, for each district. The best models were identified by Rusing the MAPE of the testing dataset. Details of the four models used are found below.

2.1. Data: The dengue case data in Sri Lanka was obtained by the Epidemiology Unit, monthly, for each district from 2010 to 2019. Rainfall (mm) and Mean Temperature (ºC) data were used as predictors for the Generalized Additive Model which were obtained from the Department of Meteorology, Sri Lanka.

2.2. Grey Forecasting Model: The Grey Forecasting models are developed based on the grey system theory, which was introduced by Deng in 1982. The grey models have effectively tackled uncertain problems with small samples with insufficient information. These models are extensively used in fields as science and technology, energy consumption and economy [3].

2.2.1. Standard Grey Forecasting Model:
This is the most commonly used grey model and this model uses at least four observations. The prediction values of $X^{(1)}$ is calculated as

$$\hat{x}^{(1)}(k) = \left(x^{(0)} - \frac{b}{a} \right) e^{-a(k-1)} + \frac{b}{a}$$
$$k = 1, 2, 3, \dots \qquad (1)$$

then the prediction values of $X^{(0)}$ can be calculated by

$$\hat{x}^{(0)}(1) = x^{(0)}(1)$$
$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1)$$
$$k = 1, 2, 3, \dots n \qquad (2)$$

2.2.2. Exponential Grey Forecasting Model:
GM(1,1) is typically appropriate for data that satisfies exponential growth. But when the data decreases it can lead to decreased prediction accuracy. To enhance the prediction accuracy the model incorporates a decreasing term (e^{-1}) in the whitenization equation [4]. Then the grey prediction model is

$$\hat{x}^{(1)}(k) = \left(x^{(0)}(k) - \frac{b}{a} - \frac{c}{a-1} e^{-t} \right) e^{a(1-t)} +$$
$$+ \frac{b}{a} + \frac{c}{a-1} e^{-1}$$
$$(3)$$

then the predicted values of the original series, $\hat{x}^{(0)}(k)$ can be obtained as

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1)$$
$$k = 1, 2, 3, \dots n \qquad (4)$$

2.3. Growth Curve Model:
Growth Curve models record the growth in variables over time and these models are widely used to model cumulative growth processes.

2.3.1. Exponential Growth Curve Model:
This model is mainly used to fit an early phase of an epidemic.

$$N(t) = N_0 e^{rt} \qquad (5)$$

2.3.2. Logistic Growth Curve Model:
This model shows exponential growth in the first phase and then starts to grow at a reduced rate till it reaches the maximum in the end.

$$N(t) = \frac{K}{1 + ae^{-rt}} \qquad (6)$$

2.3.3. Gompertz Growth Curve Model:
This model shows the slowest growth at the beginning and at the end.

$$N(t) = N(0) e^{c\left(e^{at-1}\right)} \qquad (7)$$

2.4. Alpha Sutte Indicator:
The Alpha Sutte Indicator has been originated from the Sutte Indicator which was proposed by Ahmar in 2015. This indicator is primarily used to predict stock market trends [5]. The Alpha Sutte Indicator uses only four observations to predict the fifth observation. Hence it is mainly used for short term forecasting. Also it has no assumptions to validate [6].

$$a_t = \frac{\alpha\left(\dfrac{\dfrac{\Delta x}{\alpha+\delta}}{2}\right) + \beta\left(\dfrac{\dfrac{\Delta y}{\beta+\alpha}}{2}\right) + \gamma\left(\dfrac{\dfrac{\Delta z}{\gamma+\beta}}{2}\right)}{3}$$
$$(8)$$

2.5. Generalized Additive Model:
Generalized Additive Models (GAMs) are a type of regression model that allows nonlinear relationships between the predictors and the response variable.

$$z = s_1 x_1 + s_2 x_2 + \dots + s_n x_n \qquad (9)$$

where s_n are smooth functions used to denote the non-linear combination of variables.

2.6. Hybrid Model:
Optimizing the forecasting models is carried out by developing a hybrid model by coupling two or more individual models. By optimizing the models we can reduce the prediction error of the model to increase the accuracy. The coupling technique used to develop the hybrid model was a weighted average method.

2.7. Evaluation of accuracy:
2.7.1. Mean Absolute Percentage Error:
The Relative Percentage Error (RPE) and the Mean Absolute Percentage Error (MAPE) are used toevaluate the performance of the models

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}\left|\frac{\hat{x}(k)-x(k)}{x(k)}\right|*100$$

(10)

where $\hat{x}(k)$ is the predicted value, $x(k)$ is the actual value and n is the number of observations.

2.7.2. Coefficient of Determination: Coefficient of Determination (R^2) is a measure of the goodness of fit of a regression model. It measures the portion of variability of the response variable that can be explained by the predictors in the model

$$R^2 = 1 - \frac{RSS}{TSS}$$

(11)

where total sum of squares (TSS) is the total variability in the response variable and sum of squared residuals (RSS) is the difference between the actual values of the response variable and the predicted values from the model.

2.7.3. Residual Mean Squared Error: Residual Mean Squared Error (RMSE) is used to evaluate the performance of regression models. It measures the average of the squared differences between the actual values and the predicted values of the model

$$RMSE = \sqrt{\frac{\left(\hat{x}(k)-x(k)\right)^2}{n}}$$

(12)

where $\hat{x}(k)$ is the predicted value, $x(k)$ is the actual value and n is the number of observations.

3. Results

3.1. Grey Forecasting Model: The Mean Absolute Percentage Error of the data validation set (year 2018 and year 2019) for the three models of the Grey Forecasting Model is plotted in Figure 1. Here:
- Model 1 – GM(1,1) – The traditional grey forecasting model which uses the

data from 2010–2017 for the model development.
- Model 2 – GM(1,1) – 3 years – The traditional grey forecasting model which uses the data of three prior years to forecast for a given year.
- Model 3 – EXGM – The exponential grey forecasting model which uses data of three prior years to forecast for a given year.
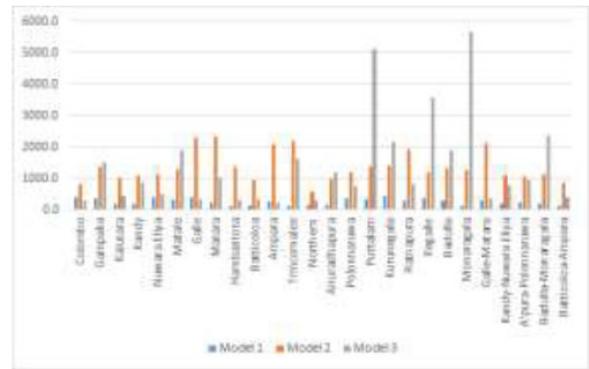

Fig. 1. MAPE: Grey Models

3.2. Growth Curve Model: The Mean Absolute Percentage Error of the data validation set (year 2018 and year 2019) for the two most accurate growth curve models are plotted in Figure 2.
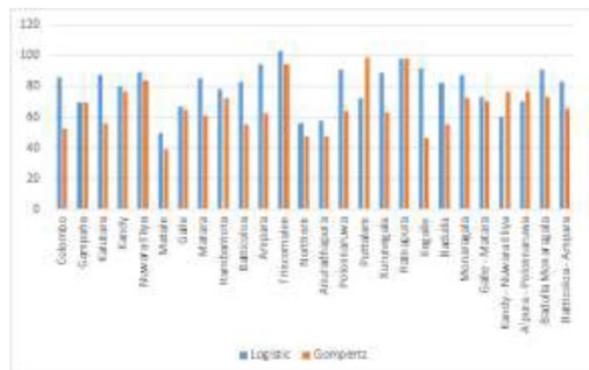

Fig. 2. MAPE: Growth Models

3.3. Alpha Sutte Indicator: Since the predicted cases from the Alpha Sutte indicator were significantly lower than that of the actual cases, the model was optimized in a way which reduced the MAPE of the model of the data development set. The optimization was done by assigning an adding variable to each of the regions.

3.4. Generalized Additive Model Since the temperature and rainfall data could not be obtained for each of the districts in Sri Lanka, several districts were combined in

order to forecast the dengue transmission using the limited data resources.

Figure 3 shows the MAPE of the Gompertz Growth Curve Model, Alpha Sutte Indicator (after optimizing) and the Generalized Additive Model for each of the regions.
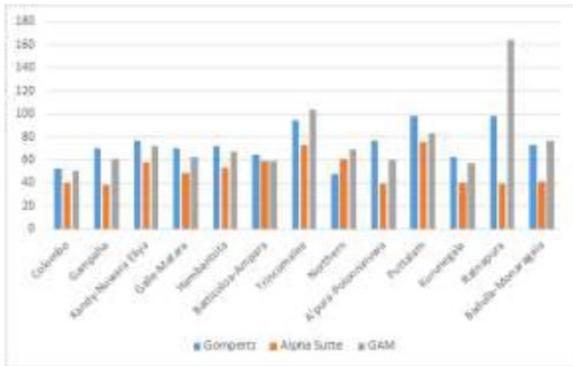

Fig. 3. MAPE: Accurate Models

Evidently, the lowest MAPE is recorded for the Alpha Sutte Indicator and the Generalized Additive Models. Hence, these two models are used to generate the hybrid model to further increase the prediction accuracy. Since the hybrid model is generated from a weighted average method, the weights for each district are calculated so that the MAPE of the model is reduced for the model development set.

Figure 4 compares the MAPE of the Alpha Sutte indicator, Generalized Additive model and the hybrid model. We can see that the accuracy of the hybrid model is better than the Alpha Sutte indicator and the Generalized Additive model for most of the regions.
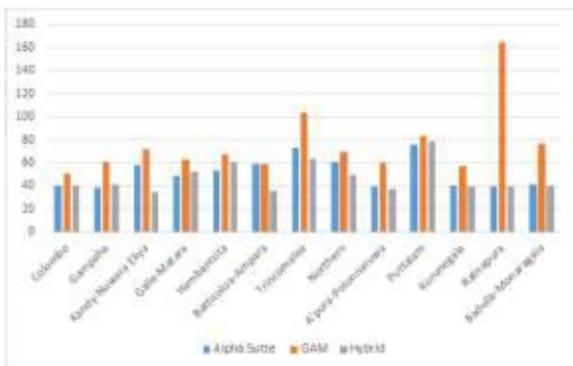

Fig. 4. Comparison of MAPE

4. Conclusion

It is intended to propose a hybrid dengue forecasting model which will increase the accuracy of predicting the dengue incidence in Sri Lanka in a monthly basis. The optimization of the model was done by developing a hybrid model by coupling two most precise individual models; Alpha Sutte Indicator and Generalized Additive Model. The proposed hybrid model provides better forecasts than the individual models.

However, the non-availability of data in certain districts and contribution factors like COVID-19 and non-climatic factors, which cannot be incorporated in the models, may challenge the accuracy of the model.

5. Bibliography

[1] Dengue fever – Sri Lanka. (n.d.), www.who.int. https://www.who.int/emergencies/diseaseoutbreak-news/item/19-july-2017-dengue-srilanka-en.

[2] Tissera H.A. et al., "Severe Dengue Epidemic, Sri Lanka, 2017", *Emerging Infectious Diseases*, vol. 26, no. 4 (2020), https://doi.org/10.3201/eid2604.190435.

[3] Shodiq M. et al., "Grey forecasting model implementation for forecast of captured fisheries production", *Kursor*, vol. 9(4), 169 (2019), https://doi.org/10.28961/kursor.v9i4.170.

[4] Bilgil H., Bilgil H., "New grey forecasting model with its application and computer code", *AIMS Mathematics*, vol. 6(2), 1497–1514 (2021), https://doi.org/10.3934/math.2021091.

[5] Ahmar A.S., Rahman A., Mulbar U., α-Sutte Indicator: a new method for time series forecasting", *Journal of Physics, Conference Series*, vol. 1040, 012018 (2018), https://doi.org/10.1088/17426596/1040/1/012018.

[6] Attanayake A.M.C.H., Perera S.S.N., "Forecasting COVID-19 Cases Using Alpha-Sutte Indicator: A Comparison with Autoregressive Integrated Moving Average (ARIMA) Method", *BioMed Research International*, 1–11 (2020), https://doi.org/10.1155/2020/8850199.

Optymalizacja wydajności przez modelowanie hybrydowe: aplikacja do choroby Denga

K. WIJESINGHE, S. PERERA, C. ATTANAYAKE

Chociaż wirusowi dengi można zapobiec przez odpowiedzialne działania człowieka, stał się on poważnym zagrożeniem dla ludzkości. Celem pracy było zwiększenie dokładności przewidywania transmisji dengi z wykorzystaniem modeli hybrydowych. Po prognozowaniu za pomocą modelu prognozowania Greya, modelu krzywej wzrostu, wskaźnika Alpha Sutte i uogólnionego modelu addytywnego, określono modele o najlepszej dokładności przewidywania poprzez najniższy średni bezwzględny błąd procentowy (MAPE) zarejestrowany w obliczeniach błędu. W związku z tym opracowano model hybrydowy, wykorzystując metodę średniej ważonej jako technikę łączenia. Na podstawie przeprowadzonych obliczeń i analiz wybrano wskaźnik Alpha Sutte i uogólniony model addytywny do opracowania modelu hybrydowego. Model zwiększa dokładność przewidywań dla większości regionów Sri Lanki. Dokładne prognozowanie przenoszenia się dengi jest ważne, aby przydzielić personel i sprzęt medyczny, przeprowadzić skuteczne programy zarządzania środowiskowego i podnoszenia świadomości, a także kontrolować wektory chemiczne w odpowiedzi na rosnącą liczbę pacjentów z dengą.

Słowa kluczowe: transmisja dengi, uogólniony model addytywny, model hybrydowy.