Spatio-temporal modelling of dengue fever incidence in Malaysia

Norziha Che-Him¹, MGhazali Kamardan¹, Mohd Saifullah Rusiman¹, Suliadi Sufahani¹, Mahathir Mohamad¹ and Nafisah @ Kamariah Kamaruddin²

¹ Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia (UTHM), Pagoh Educational Hub, 84600 Pagoh, Johor, Malaysia
² Centre of Diploma Studies, Universiti Tun Hussein Onn Malaysia (UTHM), Pagoh Educational Hub, 84600 Pagoh, Johor, Malaysia

E-mail: norziha@uthm.edu.my, mghazali@uthm.edu.my, saifulah@uthm.edu.my, suliadi@uthm.edu.my, mahathir@uthm.edu.my, nafisah@uthm.edu.my

Abstract. Previous studies reported significant relationship between dengue incidence rate (DIR) and both climatic and non-climatic factors. Therefore, this study proposes a generalised additive model (GAM) framework for dengue risk in Malaysia by using both climatic and non-climatic factors. The data used is monthly DIR for 12 states of Malaysia from 2001 to 2009. In this study, we considered an annual trend, seasonal effects, population, population density and lagged DIR, rainfall, temperature, number of rainy days and El Niño-Southern Oscillation (ENSO). The population density is found to be positively related to monthly DIR. There are generally weak relationships between monthly DIR and climate variables. A negative binomial GAM shows that there are statistically significant relationships between DIR with climatic and non-climatic factors. These include mean rainfall and temperature, the number of rainy days, sea surface temperature and the interaction between mean temperature (lag 1 month) and sea surface temperature (lag 6 months). These also apply to DIR (lag 3 months) and population density.

1. Introduction
Dengue fever (DF) is a disease where its virus transmitted by the bite of female Aedes aegypti mosquitoes [1] and about 40% of the population in worldwide reported is in the risk situation of DF [2]. DIR reported that influenced by weather and climate variability with the evidence of studies from China, Singapore, Brazil and Venezuela [3]-[6].

In Malaysia, dengue fever first time reported after an outbreak in Penang around December 1901 [7] and the dengue incidence has steadily increased then become a significant public health concern [8]. In the last 5 years, dengue cases continuously rise annually and has been recognised as a significant public health problem [9]-[10], therefore many efforts were arranged to help Malaysia in reducing the DIR.

There are four serotypes of dengue viruses in Malaysia known as DENV1 to DENV4 being the predominant virus in different periods, however DENV4 reported less influenced compared to other three [11].
In addition [12], reviewed status of scientific studies in Malaysia in terms of climate and dengue, and the challenges of such research on a climate-based dengue early warning system in a dengue-endemic country. This paper divided into Section 2 which presents the related dataset which is monthly from 2001 to 2009 dataset meanwhile Section 3 describes the research methodology in analysing the gathered data. Lastly, Section 4 shows the results and analysis of this study and Section 5 concludes the study.

2. Data
The data for this study comprises monthly dengue cases during the 108 months period between January 2001 and December 2009 for each of the 12 states of Peninsular Malaysia along with the total population of those states. North East region refers to most of the East of Peninsular Malaysia and includes the states of Kelantan, Terengganu and Pahang; the South East region consists of the states of Johor, Melaka and Negeri Sembilan which are actually located more in the South of the Peninsular; North West region refers to the Northern part of Peninsular Malaysia which contains Kedah, Penang and Perlis. Finally, the South West region includes the capital of Malaysia, Kuala Lumpur along with Selangor and Perak (see also Figure 1).

3. Research Methodology
The study started with exploratory data analysis using R. The dengue incidence rate (DIR) was calculated and embedded into dataset by using formula in Equation (1). The relevant plots and tables in R were created to help this analysis completed. Then related modelling framework with appropriate results present at the end of this study.

4. Results and Analysis
We found the population number of Malaysia increased double time for the last 30 years. Figure 1 is clearly marked the difference between the populations in the South West region as compared to the other three regions for the 108 months.

Figure 1. Annual population estimates of Malaysia (4 regions) from 2001 to 2009.
The dengue incidence rate (DIR) is defined by [13] as the number of new confirmed cases of DF and DHF, $y_{st}$, diagnosed in state, $s$ ($s=1,\ldots,12$) in month, $t$ ($t=1,\ldots,108$) divided by the total estimated population of the state, $p_{sj}$ (in 100,000s) for the year, $j$ ($j=1,\ldots,9$) in which month, $t$ falls. So the DIR is the monthly incidence per 100,000 persons at risk could be simplify as in Equation (1).

$$\text{DIR} = \frac{y_{st}}{p_{sj}} \times 100,000 \quad (1)$$

In addition, Figure 2 presents January and July become two months in the year which DIR peaks. This is due to the monsoon seasons where from May to September recognised as the Southwest monsoon and November to March known as Northeast monsoon [14].

Figure 3 shows monthly DIR per 100,000 populations for these four regions over the 108 months period. The most obvious aspect of Figure 3 is the generally higher level of monthly DIR in the South West. In addition, the average monthly DIR in each region over the nine years as shown in Figure 4 which displays the pattern from June through to May in the subsequent year.
Figure 3. Monthly DIR for 4 Regions in Malaysia (2001 to 2009).

Figure 4 similarly indicates the notable difference between the South West and the other regions in that monthly DIR remains consistently high throughout the year. Besides, there are similar features in the seasonal cycle in the North East, South East and South West with an annual peak in January and another in July. This pattern is to be expected due to the similar impact of the South West Monsoon on these three regions [15].

Figure 4. Mean Monthly DIR per 100,000 population for North East, South East, North West and South West in Malaysia (2001 to 2009).
A generalised additive models (GAM) framework was adopted [16] to capture the various influences such as global trend and the impact of population density also investigating potential association with climate and lagged climate variables.

A negative binomial GAM with the dengue incidence rate $\rho_{st}$ modelled as in Equation (2). The observed dengue counts $y_{st}$, for state $s$ ($s = 1,...,12$) and month $t$ ($t = 1,...,108$) are assumed to follow a negative binomial distribution to allow for overdispersion with mean value, $\mu_{st} = p_{st}\rho_{st}$ and scale parameter, $\theta$ where $p_{st}$ is the known population offset. The parameter estimates together with variables could be seen in Table 1.

$$\log \rho_{st} = \alpha + \sum_{j=1}^{7} \beta_{j} x_{jt} + \beta_{67} x_{6t} x_{7t} + \gamma_{1} z_{1st} + \gamma_{2} z_{2st} + \gamma_{3} z_{3st} + f_{r(s)}(z_{4st}) + \delta_{r(t)}$$

(2)

Population was included as an offset in the model to account for variations in population between state and over time. Smooth functions of year and of calendar month were included to capture global trend and seasonal cycle, while a 4 levels factor was incorporated to reflect regional variations.

Interactions between the regional factor and the year trend and seasonal cycles were also considered. Population density and the climate variables: mean monthly rainfall and temperature and number of rainy days, along with the Niño4 SST index. Lagged values of all climate variables (from one to six months previously) were also considered.

The final climate variables to emerge as most significant from the model selection process were rainfall (current and lag 3 months), number of rainy days (current and lag 3 months), temperature (current month) and sea surface temperature (lag 6 months).

The seasonal cycle is represented by a smooth function of calendar month; the interaction between this and region was also significant indicating that the apparent regional differences in seasonal cycle noted earlier do appear to be important even when other variables are allowed for. Lastly, population density was highly significant.

### Table 1. Parameter Estimates, standard errors and $p$-values for negative binomial GAM.

| Variables                          | Coefficients | Estimates   | Std. Error | Z -value | Pr(>|Z|) |
|------------------------------------|--------------|-------------|------------|----------|----------|
| Intercept                          | $\alpha$     | 4.110e+01  | 2.476e+01 | 1.660    | < 0.1    |
| Rainfall                           | $\beta_1$    | -3.201e-04 | 1.133e-04 | -2.826   | < 0.01   |
| Rainfall Lag 3 Months              | $\beta_2$    | 2.149e-04  | 9.507e-05 | 2.261    | < 0.05   |
| Rainy Days                         | $\beta_3$    | 1.846e-02  | 4.607e-03 | 4.008    | < 0.001  |
| Rainy Days Lag 3 Months            | $\beta_4$    | 8.081e-03  | 4.195e-03 | 1.927    | < 0.1    |
| Temperature                        | $\beta_5$    | 1.184e+00  | 3.472e-02 | 3.410    | < 0.001  |
| Temp. Lag 1 Month                  | $\beta_6$    | -1.661e+00 | 9.147e-01 | -1.815   | < 0.1    |
| Nino4 Lag 6 Months                 | $\beta_7$    | -1.530e+00 | 8.577e-01 | -1.784   | < 0.1    |
| Temp. Lag 1 and SST Lag 6 Months   | $\beta_{67}$ | 5.927e-02  | 3.176e-02 | 1.866    | < 0.1    |
| Population Density                 | $\gamma_1$   | 1.256e-04  | 1.213e-05 | 10.361   | < 0.001  |
| Year                               | $\gamma_2$   | 5.50e-02   | 7.003e-03 | 7.861    | < 0.001  |
| DIR Lag 3 Months                   | $\gamma_3$   | 8.198e-03  | 1.792e-03 | 4.575    | < 0.001  |
5. Conclusion

Overall, the value of DIR in the South West region is significantly higher compared to the other regions and the South West region also shows a significant difference in DIR patterns compared to the other three regions for the 108 months. The population density is found to be positively related to monthly DIR. There are generally weak relationships between monthly DIR and climate variables. A negative binomial GAM shows that there are statistically significant relationships between DIR with climatic and non-climatic factors. These include mean rainfall and temperature, the number of rainy days, sea surface temperature and the interaction between mean temperature (lag 1 month) and sea surface temperature (lag 6 months). These also apply to DIR (lag 3 months) and population density.

Acknowledgement

We would like to say thank you to the Universiti Tun Hussein Onn Malaysia (UTHM) and the Office for Research, Innovation, Commercialization and Consultancy Management (ORICC), UTHM for kindly providing us with the internal funding for publishing this paper.

References

[1] Gubler D 1998 Dengue and dengue hemorrhagic fever Clinical Microbiol Rev, Vol. 11, pp. 480 - 496
[2] Guzman M and Kouri G 2002 Dengue: An update The Lancet Infection Diseases, Vol. 2, pp. 33 - 42
[3] Lu L, Lin H, Tian L, Yang W, Sun J and Liu Q 2009 Time series analysis of dengue fever and weather in Guangzhou, China BMC Public Health, 9(395)
[4] Hii Y L 2009 Climate variability and increase in intensity and magnitude of dengue incidence in Singapore Global Health Action
[5] Lowe R 2011 Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil Computers and Geosciences, 37(3), 371 - 381
[6] Aura D H M and Alfonso J R M 2010 Potential influence of climate variability on dengue incidence registered in a western pediatric Hospital of Venezuela Tropical Biomedicine, 27(2), 280 - 286
[7] Skae F M T 1902 Dengue fever in Penang, The British Medical Journal, 1581 - 1582
[8] Abu Bakar S and Shafee N 2002 Outlook of dengue in Malaysia: a century later Malaysian Journal Pathology, 24(1), 23 - 27
[9] Aziz S, Ngui R, Lim Y, Solehah I, Nur Farhana J, Azizan A and Wan Yusoff W 2012 Spatial pattern of 2009 dengue distribution in Kuala Lumpur using GIS application Tropical Biomedicine Vol. 29 No. 1 pp. 113-120
[10] Lam S K 1993 Strategies for dengue control in Malaysia Tropical Medicine, 35(4), 303 - 307
[11] Rose Nani M 2015 Dengue incidence and the prevention and control program in Malaysia. IMJM Vol. 14 No. 1 pp. 05-10
[12] Hii Y L, Ahmad Zaki R, Aghamohammadi N and Rocklov J 2016 Research on Climate and Dengue in Malaysia: A Systematic Review Curr Environ Health Rep 2: 81-90
[13] Hafiz H, Shamrina S and Nor R 2012 Risk mapping of dengue in Selangor and Kuala Lumpur, Malaysia, Geospatial Health, Vol. 7, No. 1, pp. 21 - 25
[14] Suhaila J, Deni S, Wan Zin W and Jemain A A 2010 Spatial patterns and trends of daily rainfall
regime in Peninsular Malaysia during the southwest and northeast monsoon *Meteorol Atmos Phys*, Vol. 110, pp. 1 - 18

[15] Che Him N Bailey T C and Stephenson D B, 2012 Climate variability and dengue incidence in Malaysia *Proceedings of the 27th International Workshop on Statistical Modelling* Prague, 2, 435 - 440

[16] Hastie T and Tibshirani R 1986 Generalized Additive Models *Statistical Science*, 1(3):297 - 318