MJFAS MALAYSIAN JOURNAL OF FUNDAMENTAL AND APPLIED SCIENCES
PRINT ISSN: 2289-5981 | ONLINE ISSN: 2289-599X

RESEARCH ARTICLE

Estimating relative risk for dengue disease in Peninsular Malaysia using INLA

Nurul Syafiah Abd Naeeim*, Nuzlinda Abdul Rahman

School of Mathematical Science, Universiti Sains Malaysia, Penang, Malaysia.

* *Corresponding author: nurulsyafiahabdnaeeim@yahoo.com*

Article history
Received 18 February 2017
Accepted 4 October 2017

Abstract

Study in spatio-temporal disease mapping models give a great worth in epidemiology, in describing the pattern of disease incidence across geographical space and time. This paper studies generalized linear mixed models (GLMM) for the analysis of spatial and temporal variability of dengue disease rates. For spatio-temporal study, the models accommodate spatially correlated random effects as well as temporal effects together with the space time interaction. The space time interaction is used to capture any additional effects that are not explained by the main factors of space and time. However, as study including time dimension is quite complex for disease mapping, the temporal effects that only relate to structured and unstructured time pattern are considered in these models as initial screening in studying disease pattern and time trend. The models are fitted within a hierarchical Bayesian framework using Integrated Nested Laplace Approximation (INLA) methodology. For this study, there are three main objectives. First, to choose the best model that represent the disease phenomenon. Second, to estimate the relative risk of disease based on the model selected and lastly, to visualize the risk spatial pattern and temporal trend using graphical representation. The models are applied to monthly dengue fever data in Peninsular Malaysia reported to Ministry of Health Malaysia for year 2015 by district level.

Keywords: Spatio-temporal analysis, Disease mapping, Bayesian estimation, GLMM, INLA

© 2017 Penerbit UTM Press. All rights reserved

INTRODUCTION

Disease mapping is a very active research field in investigating the spatial relationship of disease burden with the geographical distribution, risk factors in the environment and human populations. Year by year, numerous methods for disease mapping have been expanded in accordance with the growing amount of routinely collected health information worldwide. Recently, the combination study of spatial and temporal effects in disease mapping is being emphasizes compared to study on spatial effects only. This is because ignoring temporal evolution by assuming static in time is not always realistic for every disease especially for infectious diseases.

In disease mapping, disease rates in a population are regularly been used to visualize on a map rather than the number of disease count. Traditionally, standardized incidence ratio (SIR) was calculated as improvement for crude rate by dividing the observed disease case count with the expected number of observations, where the expected number of observations is very sensitive to the number of population. Uneven number of population in different areas leads to large variances in rates across the map and this results to unstable patterns of disease distribution. Besides that, a small change in the number of observations, would cause abrupt changes in the expected number of observations, which is known as the small number problems. Hence, to address the drawbacks of SIR, smoothing techniques for relative risk estimation have been developed by borrowing strength or information across neighbouring spatial units (Clayton and Kaldor, 1987). Smoothing techniques help to remove the variation in disease rates, reduce the effect of measurement errors, and better smoothing leads to less bias results in further analyses (Kang *et al.*, 2016).

More interestingly, for spatio-temporal disease mapping studies, smoothing technique is not only relating to the use of information from spatial neighbours but the context is lengthened to borrowing temporal neighbours too. Information is shared in time in a similar manner as sharing information in space. Hence, risks are smoother and more reliable because now the model is based on greater amount of information.

Bayesian models have become a familiar approach for smoothing purposes, both empirical Bayes (EB) and fully Bayes (FB). Several studies like Clayton and Kaldor (1987), Marshall (1991) and Lahiri and Maiti (2000) used EB method for smoothing in their disease mapping studies while Besag *et al.* (1991), MacNab *et al.* (2004), and Wakefield (2007) applied FB. However, with the aid of modern programming, FB is preferably being used to solve the posterior distribution especially in a complex model that requires many parameters to be estimated. Commonly, estimating parameter with FB in most studies will involve Markov Chain Monte Carlo (MCMC) algorithm computation but this method requires a huge computation time and may lead to large Monte Carlo errors especially for a big data set. Furthermore, according to Schrödle and Held (2011), specific block-sampling algorithms have to be applied in order to get reliable estimates if complex spatial and spatio-temporal models are to be fitted.

In order to circumvent these drawbacks, an alternative method using integrated nested Laplace approximation (INLA) has been proposed by Rue *et al.* (2009) to compute the posterior marginals of all parameters of interest. This method is believed to provide precise parameter estimates in shorter time and more practical to use. Recently, the INLA method becomes an active research development in disease mapping and has been shown to work well with generalized linear

mixed models, Bayesian quantile additive mixed models, survival analysis, and spatio-temporal models (Martins *et al.*, 2013).

Variant models in spatio-temporal disease mapping have been described in the literature including parametric and non-parametric time trend and space-time interactions, most of them based on conditional autoregressions (CAR) models with extension to Besag, York and Mollié (BYM) models. For instance, Bernardinelli *et al.* (1995b) proposed a linear time trend model which an additional linear and a differential time trend. Then, Assunção *et al.* (2001) modified the study by applying second degree polynomial. Knorr-Held (2000) focuses on the inclusion of space-time interactions using non-parametric time models by proposing four types of interactions together with four different prior distributions for the interactions. Ugarte *et al.* (2009) compare the performance of six space-time disease mapping models by adopt the extension proposed by Bernardinelli *et al.* (1995b) and Knorr-Held (2000). In other study, Martínez-Beneito *et al.* (2008) link spatio-temporal study with autoregressive approach.

The combination study of spatio-temporal model is very interesting to explore as this study is not yet in abundance as spatial model study. This may be due to difficulties of linking both spatial and temporal dependence on a single model. In this study, we will focus on general time trend model as a linear time trend model may be unrealistic for some case of study. It is common to observe some changes in time trends due to the changes of the environment effects, the improvement in treatments and the advancement in research. The objectives of this study are to choose the best model that represent the disease phenomenon, to estimate the relative risk of disease based on the model selected and to visualize the spatial pattern and temporal trend using graphical representation. The models are fitted using INLA methodology and dengue data will be applied in this study.

MATERIALS AND METHODOLOGY

Case study: Dengue in Peninsular Malaysia

Dengue is the most ordinary mosquito-borne viral disease of humans, mainly transmitted by the female mosquitos *Aedes aegypti* and *Aedes albopictus*. Due to the dramatically increase of dengue cases day by day, dengue has become a major world-wide public health concern especially for regions near in tropical and sub-tropical climate. According to Mia *et al.* (2013), the factors that contribute to the spread of dengue in Malaysia are the changes in climate factors such as warm temperature, increased rainfall, and relative humidity that indirectly serve a conducive condition for mosquito breeding. Besides that, factors such as unstopping urbanization, massive infrastructure development, change in population number, deforestation, poor waste sanitation, inadequate domestic water supplies, faster modes of transportation, increased migration and internationally travel are also contributing to the dramatically increase of dengue (Shafie *et al.*, 2015; Pang and Loh, 2016). Presently, dengue outbreak can only be controlled using vector-controlled methods as licensed vaccine still in the developmental stage. In Malaysia, the vector control strategies are adulticiding, larviciding, personal protection, environmental management, community participation, legislation, and integrated control. Besides that, personal protection such as the use of mosquito coils, insecticide mats, aerosols, and bed-nets are also give benefits in controlling dengue incidence.

With a deep concern on dengue incidence in Malaysia and in line with government's effort to control this problem, statistically research on dengue with inclusion of spatial and temporal trend are considered because it is believed that there is a relationship between geographical areas and time points on dengue risk. It is hope that this research may help the target audiences in planning more systematic vector-control prevention programmes especially for hotspots areas and also areas that has a tendency to become hotspots area.

Areal data on dengue incidence in 86 districts in Peninsular Malaysia for year 2015 recorded by monthly is used in this study. This data is obtained from Vector Borne Disease Sector, Ministry of Health Malaysia. In Malaysia, every single dengue case is recorded properly to Vector Borne Disease Sector. It is compulsory to health officer to notify any suspected or confirmed dengue case to the nearest district health office via online notification system within 24 hours of diagnosis.

Spatio-temporal models for disease mapping

The relative risk estimation typically involves generalized linear mixed models (GLMM). Let the study region, Peninsular Malaysia be divided into n districts. Data for each area, i ($i = 1, 2, ..., n$) with time points, t ($t = 1, 2, ..., T$) are available for study. The number of dengue cases, O_{it} conditional to the relative risk, r_{it} is assumed to have Poisson distribution with mean, $\mu_{it} = E_{it} r_{it}$.

$$O_{it} \mid r_{it} \sim Poisson(\mu_{it} = E_{it} r_{it}) \tag{1}$$

where E_{it} is an expected number of cases,

$$E_{it} = N_{it} \times \left(\sum_{i=1}^{n} \sum_{t=1}^{T} O_{it} \Big/ \sum_{i=1}^{n} \sum_{t=1}^{T} N_{it} \right) \tag{2}$$

and N_{it} refers to number of population.

Taking logarithm to the both side of the mean, μ_{it} leads to

$$\log(\mu_{it}) = \log(E_{it}) + \log(r_{it}). \tag{3}$$

Then, different models are defined depending on the specification of $\log(r_{it})$.

General time trend models

The general time trend models describe here are similar to those proposed by Knorr-Held (2000). The log-risk is modelled as

$$\log r_{it} = \alpha + \xi_i + \varphi_t + \gamma_t + \delta_{it} \tag{4}$$

where α is logarithm of the global risk, ξ_i is refer to spatial effect, φ_t is defined as unstructured temporal effect, γ_t is structured temporal effect, and δ_{it} is space-time interaction effect.

From Eq. (4) $\boldsymbol{\xi}$, $\boldsymbol{\varphi}$, and $\boldsymbol{\gamma}$ represent main effects and $\boldsymbol{\delta}$ represents interaction effect. The prior distribution used for $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_n)$ follows the Leroux *et al.* (2000) CAR prior, defined as $\boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{Q}^{-1})$, where $\mathbf{Q}^{-1}$ is the inverse of precision matrix.

$$\mathbf{Q}^{-1} = \sigma_s^2 (\lambda_s \boldsymbol{R}_s + (1 - \lambda_s) \boldsymbol{I}_s)^{-1} \tag{5}$$

Here, the subscript s refers to spatial elements where σ_s^2 is a spatial variance component, λ_s is a spatial smoothing parameter taking values between 0 and 1, I_s is an $n \times n$ identity matrix and $\boldsymbol{R}_s$ is a spatial neighbourhood matrix. The entries of matrix $\boldsymbol{R}_s$ follow the following rules.

$$(R_s)_{ij} = \begin{cases} w_i, & i = j \\ -1, & i \sim j \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

OPEN ACCESS Freely available online

where w_i is the number of neighbours of the ith area and $i \sim j$ indicates the areas i and j are neighbours. In our study, the areas are neighbours if they share common boundary. Corresponding to the prior distribution for $\boldsymbol{\xi}$, the univariate full conditional distribution model is expressed as

$$\xi_i \mid \xi_{j \neq i} \sim N\left(\frac{\lambda_s}{1 - \lambda_s + \lambda_s w_i} \sum_{i \sim j} \xi_i, \frac{\sigma_s^2}{1 - \lambda_s + \lambda_s w_i} \right). \quad (7)$$

Next, an independent and identically distributed normal prior with mean zero and unknown variance s_φ^2 is used for $\boldsymbol{\varphi}$. That is $\boldsymbol{\varphi} \sim \boldsymbol{N}(\mathbf{0}, \sigma_{\boldsymbol{\varphi}}^2 \boldsymbol{I}_t)$, where $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, ..., \varphi_T)'$ and I_t is the $T \times T$ identity matrix. For $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, ..., \gamma_T)'$, a random walk of first order (RW1) is considered. That is $\boldsymbol{\gamma} \sim \boldsymbol{N}(\mathbf{0}, \sigma_{\boldsymbol{\gamma}}^2 \boldsymbol{R}_t^-)$. The structure for $\boldsymbol{R}_t$ is in the following form.

$$\boldsymbol{R}_t = \begin{pmatrix} 1 & -1 & & & & & & \\ -1 & 2 & -1 & & & & & \\ & -1 & 2 & -1 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & -1 & 2 & -1 & & \\ & & & & -1 & 2 & -1 & \\ & & & & & -1 & 1 \end{pmatrix} \quad (8)$$

Lastly, similar for $\boldsymbol{\delta} = (\delta_{11}, \delta_{12}, ..., \delta_{nT})'$, it is assumed to be normally distributed as $\boldsymbol{\delta} \sim N(\mathbf{0}, \sigma_\delta^2 \boldsymbol{R}_\delta)$ where $\boldsymbol{R}_\delta$ is the structure matrix given by the Kronecker product, corresponds to the structure matrices of the main effects. Similar to Knorr-Held (2000), four different types of interactions are considered and can be interpreted in a different way (see Table 1). In Type I interaction, all δ_{it} are independent (do not have any structure in space and time). In Type II interaction, each $\delta_{i\cdot}$ follows time trends (in this study, we use a RW1), independently to all other areas. In Type III interaction, each $\delta_{\cdot t}$ follows spatial pattern, without any temporal structure. Lastly, for Type IV, δ_{it} are completely dependent over space and time.

Table 1 Four different types of space-time interaction terms.

Space-time interaction	$\boldsymbol{R}_\delta$	RW1 for $\boldsymbol{\gamma}$
Type I	$\boldsymbol{I_s} \otimes \boldsymbol{I_t}$	$I \times T$
Type II	$\boldsymbol{I_s} \otimes \boldsymbol{R_t}$	$I \times (T-1)$
Type III	$\boldsymbol{R_s} \otimes \boldsymbol{I_t}$	$(I-1) \times T$
Type IV	$\boldsymbol{R_s} \otimes \boldsymbol{R_t}$	$(I-1) \times (T-1)$

Note: Table followed Ugarte *et al.* (2014)

Note that keeping the main effects only and dropping the space-time interaction effect leads to the additive model. The additive model will be represented by Model 1. For the next models, models with different interaction types are run. Model 2, 3, 4 and 5 follow Type I, II, III and IV interactions respectively with the unstructured and the structured time effects. In addition, models without the unstructured time effects are considered in which Model 6 represents the additive models, Model 7 and 8 represent Type II and IV interactions respectively.

Integrated nested Laplace approximations: INLA

The spatio-temporal models are built as Bayesian model in the form of three stages hierarchy with latent Gaussian model.

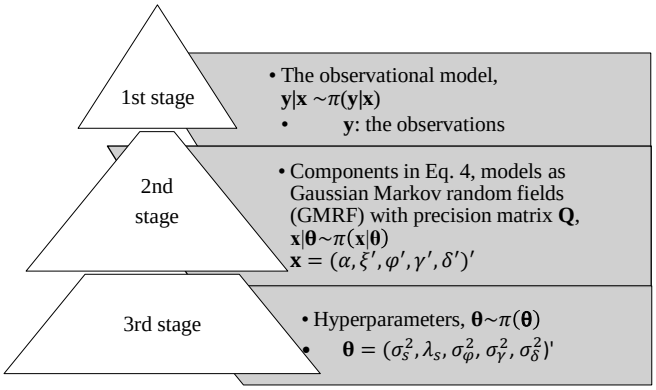

Fig. 1 The Bayesian hierarchical model.

The main goal in INLA is to estimate the marginal posterior distribution of all components of GMRF. Briefly, GMRF is a multivariate Gaussian distribution with a sparse precision matrix.

$$\pi(x_i \mid \mathbf{y}) = \int_\theta \pi(x_i \mid \boldsymbol{\theta}, \mathbf{y}) \pi(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta}. \quad (9)$$

The first component of the integral $\pi(x_i \mid \boldsymbol{\theta}, \mathbf{y})$ can be approximated using three different approaches: a Gaussian approximation, a simplified Laplace approximation and a full Laplace approximation. The Gaussian approximation is the simplest and the fastest approximation. However, inaccurate result might be obtained due to numerical error in location, error due to the lack of its skewness or both (Rue and Martino, 2007). The most accurate approximation is the full Laplace, but the computation is too long. As an alternative, the simplified Laplace approximation is used which is less time consuming and only bring a slight loss of accuracy.

Meanwhile, the second component can be approximated using a Laplace approximation where $\tilde{\pi}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})$ denotes the Gaussian approximation to the full conditional distribution, and $\mathbf{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional of $\mathbf{x}$ for a given $\boldsymbol{\theta}$.

$$\tilde{\pi}(\boldsymbol{\theta} \mid \mathbf{y}) \propto \frac{\pi(\mathbf{x}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{\pi}_G(\mathbf{x} \mid \boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x} = \mathbf{x}^*(\boldsymbol{\theta})}. \quad (10)$$

Finally, an approximation of the posterior marginal density in Eq. (9) is given by

$$\pi(x_i \mid \mathbf{y}) = \sum_k \tilde{\pi}(x_i \mid \boldsymbol{\theta}_k, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}_k \mid \mathbf{y}) \Delta_k. \quad (11)$$

An area weight, Δ_k has to be assigned to each θ_k for substitution of the integral in Eq. (9). For more details, see Rue *et al.* (2009) and Martins *et al.* (2013).

Next, the hierarchical model is completed by assigning an appropriate prior distribution for the hyperparameters of the model. The prior distributions used for the hyperparameters may influence the results (posterior distributions) and hence, should be carefully considered and compared. Usually, using the prior distributions based on literature review to determine the prior distributions is helpful in Bayesian models. For details, see papers by Bernardinelli *et al.* (1995a) and Wakefield (2007). For this study model, the only priors that should

be specified correspond to the precision parameters which are the inverse of the variance components; $t_s = 1/s_s^2$, $t_j = 1/s_j^2$, $t_g = 1/s_g^2$, and $t_d = 1/s_d^2$. The hyperpriors distributions used in this paper are similar to what has been proposed by Ugarte *et al.* (2014).

Besides that, to guarantee identifiability of the interaction term $\boldsymbol{\delta}$, specific sum-to-zero constraints have to be used except for Type I interaction. The vector $\boldsymbol{\delta}$ follows an intrinsic Gaussian Markov random field (IGMRF). An IGMRF is improper and its structure matrix, $\boldsymbol{R}_\delta$ is not of full rank. The improper density, $\boldsymbol{\pi^*(\delta)}$ can be written as

$$\boldsymbol{\pi^*(\delta) = \pi(\delta \mid A\delta = e)} \quad (12)$$

where $\mathbf{A\delta = e}$ are linear constraints $\boldsymbol{\delta}$, $\mathbf{A}$ is a matrix consists of $\boldsymbol{R}_\delta$ eigenvectors which span the null space, and $\mathbf{e}$ is a vector of zeros. The number of necessary linear constraints is always equal to the rank deficiency of $\boldsymbol{R}_\delta$.

Interestingly, in the INLA approach, the deviance information criterion (DIC) can be computed for selecting the best model. According to Spiegelhalter *et al.* (2002), DIC is the sum of the deviance posterior mean, $\bar{D}$ (a measure for model fit) and the effective parameters number, p_D (a measure for model complexity). The lowest DIC values provides the best trade-off between model fit and model complexity.

$$\mathrm{DIC} = \bar{D} + p_D. \quad (13)$$

In this study, R-programming via the R-INLA package is used for completing the methodology that has briefly described above.

RESULTS AND DISCUSSION

Table 2 DIC values for the study models.

Model	Space-time interaction	$\bar{D}$	p_D	DIC
1	Additive model	19172.65	97.19382	19269.84
2	Type I	5990.825	765.7633	6756.588
3	Type II	6003.618	641.8277	6645.446
4	Type III	6172.708	696.5397	6869.248
5	Type IV	6196.147	581.0601	6777.207
6	Additive model	19172.55	97.19417	19269.75
7	Type II	6003.29	641.8658	6645.156
8	Type IV	6196.199	580.9653	6777.164

Table 2 shows the result for the eight fitted models that has been described previously. This result is based on a simplified Laplace approximation. The additive models exhibit the highest values of DIC and show the worst fit although their estimated model complexity is lower. This result implies the importance of including space-time interaction in the study model. The models without unstructured temporal component seem slightly better although the difference of the DIC value between model with and without unstructured temporal component is too small. This may imply the inclusion or exclusion of the unstructured temporal component for this study model is not too important. Among the eight models proposed, Model 7 has the smallest DIC values. Hence, Model 7 is chosen as the best model in terms of model fit and complexity. This model consists of the spatial effect with a Leroux CAR prior, a structured temporal effect with a RW1 prior and a type II interaction. Then after choosing the best model, this selective model has been fitted again using the 'full Laplace' approximation and the result is used for relative risk computation.

The estimated log-relative risks obtained with Model 7 can be separated into individual components: an overall global risk ($\hat{a}$), a risk related to the spatial location ($\hat{x}$), a temporal risk trend ($\hat{g}$) common to all areas, and an area specific temporal risk trend ($\hat{d}$) for each district. These are useful as the spatial and temporal effects can be varied across time and districts respectively.

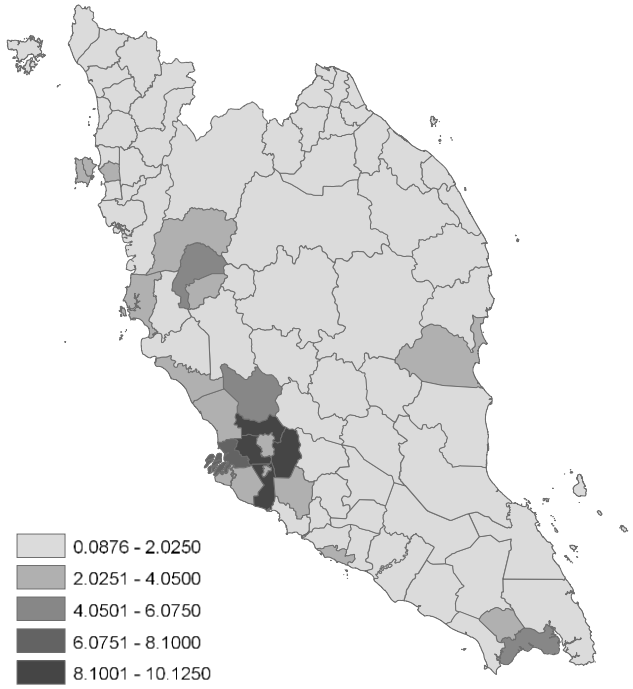

Fig. 2 The spatial pattern of dengue risk map, $\hat{z}_i = e^{\hat{x}_i}$.

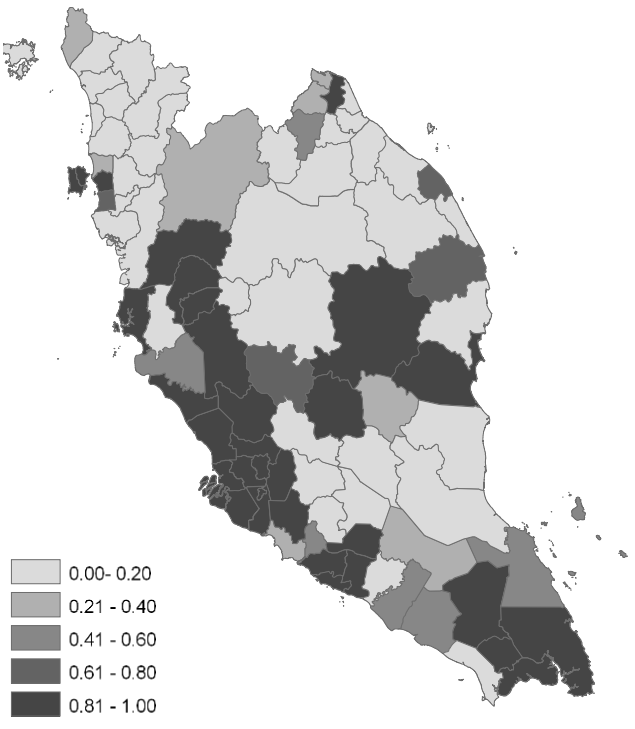

Fig. 3 The posterior probabilities map, $P(\hat{z}_i > 1 \mid \mathbf{O})$.

OPEN ACCESS Freely available online

Fig. 2 presents the map of spatial dengue risk, $\hat{z}_i = e^{\hat{x}_i}$ associated to each district and constant along the year. Meanwhile, Fig. 3 presents the posterior probability that the spatial risk is greater than 1, $P(\hat{z}_i > 1 \mid \mathbf{O})$. The degree of spatial risk is differentiated with different shades. The darker the region indicates the region's risk is higher. For simplicity, usually in the disease mapping studies, the regions with probabilities above 0.8 and 0.9 are considered as high risk regions, similar to what Richardson *et al.* (2004) has suggested. In this study, a reference threshold equal to 1 and cut-off value of 0.8 is used to detect high risk regions in all the time periods. Hence, from the both figure, Fig. 2 and Fig. 3, it is clearly seen that the districts in Selangor, Kuala Lumpur, Penang, Melaka, Southern part of Johor, and some districts in Perak and Pahang are high risk areas of dengue in year 2015.

Fig. 4 presents the line graph of the general temporal trend of dengue risk common to all districts in Peninsular Malaysia. The line graph shows a non-linear trend with a decreasing trend for the first four months and a drastic increasing pattern to the seventh month as well as a fluctuating pattern until the end of the year. This trend indicates that there might be some factors that are affecting the dengue outbreaks in Peninsular Malaysia along this period such as the climate changes or the prevention schedule.

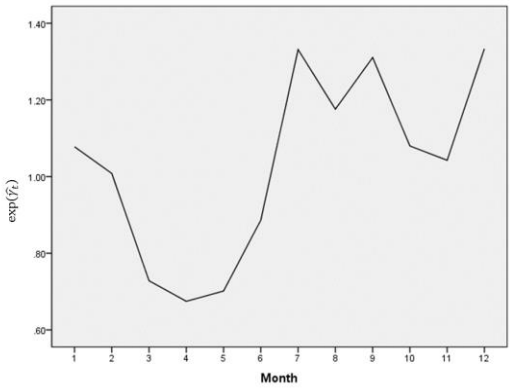

Fig. 4 The general temporal trend of dengue risk.

For the specific temporal trends, five selected districts are chosen for representing the northern, eastern, southern, western, and middle part of Peninsular Malaysia in order to see the temporal trend of different districts. These specific temporal trends (in log scale) are represented by Fig. 5 until Fig. 9. These line graphs are not converted to actual values because we are only interested to see their pattern as using the actual values will reveal the same pattern too. Kota Bharu, Kota Setar and Kuantan show slightly similar pattern which is a decreasing trend towards the middle of the year and increasing trend after that. However, Kuala Lumpur and Johor Bahru display a different trend. The temporal trend for Kuala Lumpur is quite steady along the year while for Johor Bahru, the time effect is higher in the middle of the year.

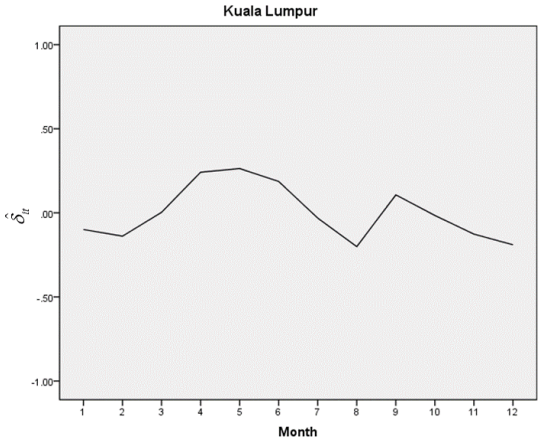

Fig. 5 Specific temporal trend for Kuala Lumpur.

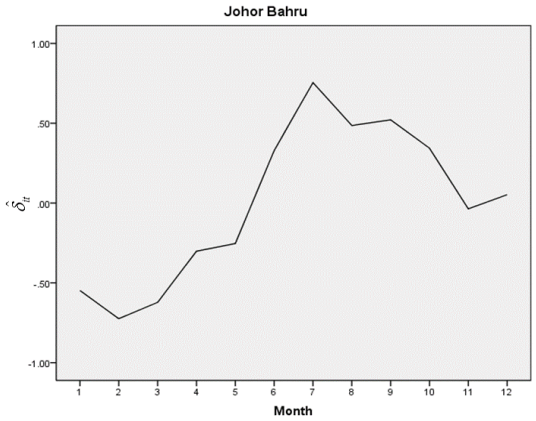

Fig. 6 Specific temporal trend for Johor Bahru.

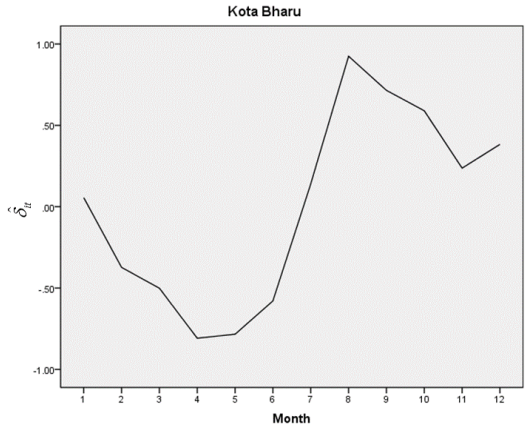

Fig. 7 Specific temporal trend for Kota Bharu.

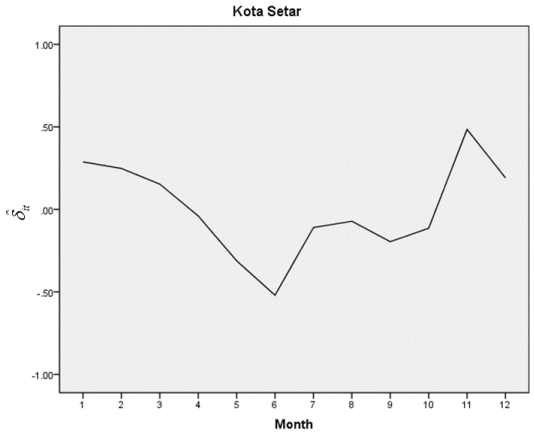

Fig. 8 Specific temporal trend for Kota Setar.

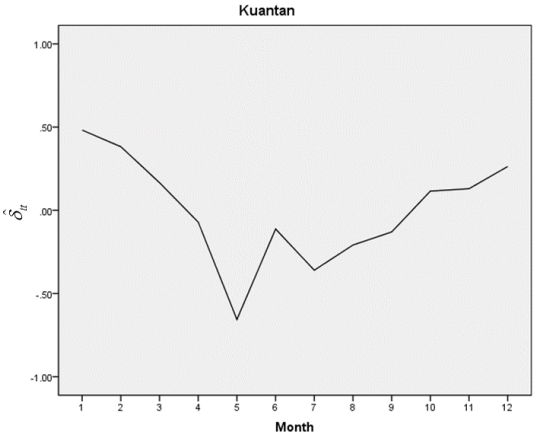

Fig. 9 Specific temporal trend for Kuantan.

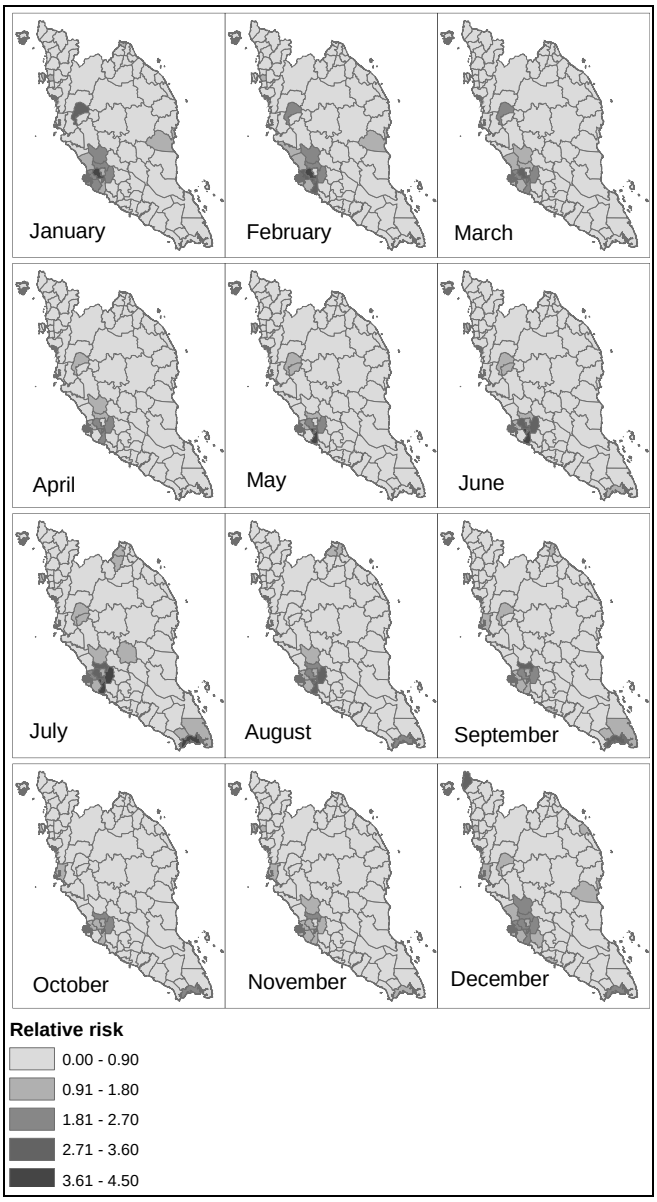

Fig. 10 Relative dengue risk distribution.

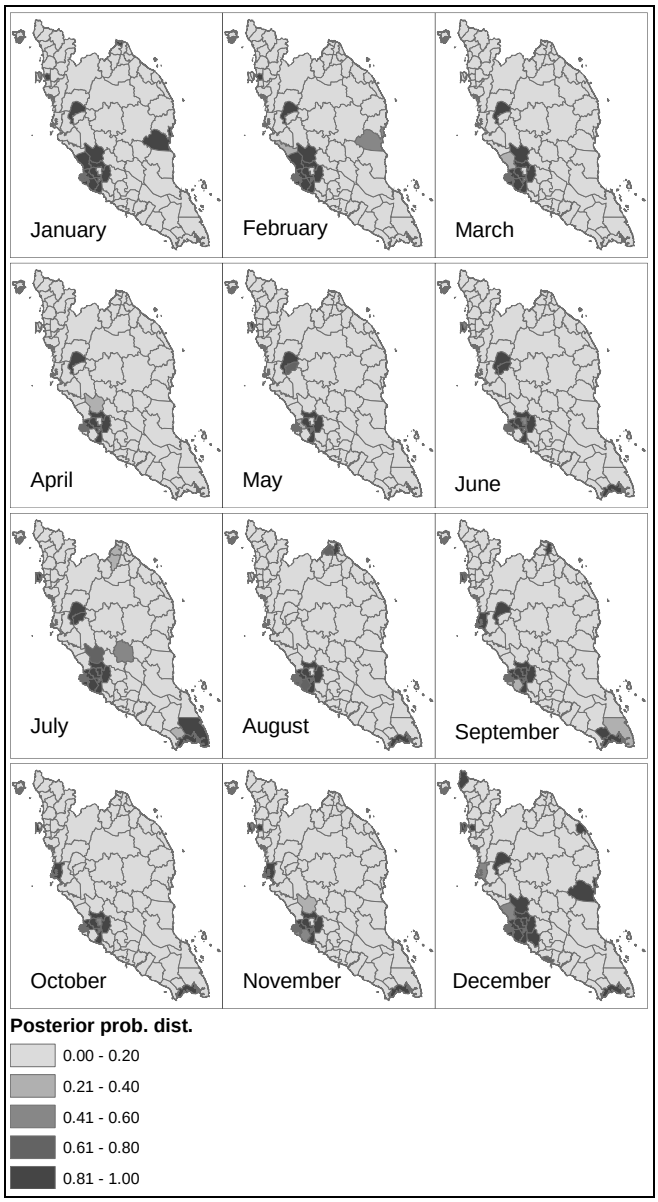

Fig. 11 Posterior probability distribution, $P(\hat{r}_{it} > 1 \mid \mathbf{O})$ by districts.

Fig. 10 and Fig. 11 display the maps of relative risk of dengue and the posterior probabilities that the relative risk greater than 1, $P(\hat{r}_{it} > 1 \mid \mathbf{O})$ for each district in Peninsular Malaysia from January to December of 2015 respectively. These figures give a clearer graphical representation in visualizing dengue disease phenomenon for all areas throughout the study period. Based on information from both figures, there are several districts that show high significant risk of dengue disease in Peninsular Malaysia for year 2015. A group of districts near to the most urbanized area in Peninsular Malaysia, Kuala Lumpur shows high significant risk throughout the year. Other than that, Kinta also shows high dengue risk for almost every month. Meanwhile, the most southern district, Johor Bahru exhibits high dengue risk during June until December with its neighbours Kota Tinggi and Kulai Jaya are also slightly affected during July and September. The rest of the areas do not give any clearer pattern of relative risk and just report high case for certain month.

CONCLUSION

Our analysis of the dengue case in Peninsular Malaysia for year 2015 shows that the gap of the relative risks of dengue between the districts under study is big. There are areas that show significant high risk throughout the year such as areas near to the capital city of Malaysia, Kuala Lumpur which are Petaling, Sepang, Hulu Langat, Gombak, Klang and Hulu Selangor. Meanwhile, Seberang Perai Tengah, Kinta, Kuantan and Johor Bahru have a tendency to become high dengue risk area. Hence, the authorities in charge should give prioritized to these areas in planning intervention strategies to reduce the dengue cases. The rest of the districts are still under control. However, precaution must be continued and monitor from time to time. Besides that, the results obtained also shows that some areas have different temporal trends in dengue outbreak compare to other areas. Hence, we can get some ideas on the effective time for vector control activities for each district in Peninsular Malaysia. For example, areas like Kota Setar, Kota Bharu and Kuantan, the prevention control should focus more on the earlier and at the end of the year, Johor Bahru in the middle of the year while Kuala Lumpur, the prevention should be for the overall months in the year.

In general, this study provides a useful starting point for spatio-temporal dengue analysis. The result in this study can be used in clustering analysis, hotspot identification and also spatial regression. As dengue is a major infectious disease in Malaysia and believe to have strong relationship with environmental factors, adding seasonal effects in the spatio-temporal model might give more appropriate model for dengue study. Besides that, instead of using monthly data, this model can be rerun using weekly data with other suitable time series model such as autoregressive first order model (AR1).

OPEN ACCESS Freely available online

ACKNOWLEDGEMENT

Special thanks to Vector Borne Disease Sector, Ministry of Health Malaysia for providing dengue data to this study. This research was financially supported by Short-Term Grant offered by Universiti Sains Malaysia. (304/PMATH/6313154)

REFERENCES

Assunção, R. M., Reis, I. A., Oliveira, C. D. L. 2001. Diffusion and prediction of Leishmaniasis in a large metropolitan area in Brazil with a Bayesian space–time model. *Statistics in Medicine*. 20, 2319-2335.

Bernardinelli, L., Clayton, D., Montomoli, C. 1995a. Bayesian estimates of disease maps: How important are priors? Statistics in Medicine 14, 2411-2431.

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M. 1995b. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*. 14, 2433-2443.

Besag, J., York, J., Mollié, A. 1991. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*. 43, 1-20.

Clayton, D., Kaldor, J. 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*. 43, 671-681.

Kang, S. Y., Cramb, S. M., White, N. M., Ball, S. J., Mengersen, K. L. 2016. Making the most of spatial information in health: a tutorial in Bayesian disease mapping for areal data. *Geospatial Health*. 11, 190-198.

Knorr-Held, L. 2000. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*. 19, 2555-2567.

Lahiri, P., Maiti, T. 2000. Empirical Bayes estimation of relative risks in disease mapping. University of Nebraska Technical Report.

Leroux, B. G., Lei, X., Breslow, N.. Estimation of disease rates in small areas: a new mixed model for spatial dependence. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Springer New York, 2000. 179-191.

MacNab, Y. C., Farrel, P. J., Gustafson, P., Wen, S. 2004. Estimation in Bayesian disease mapping. *Biometrics*. 60, 865-873.

Marshall, R. J. 1991. Mapping disease and mortality rates using empirical Bayes estimators. *Applied Statistics*. 40, 283–294.

Martínez-Beneito, M. A., López-Quilez, A., Botella-Rocamora, P. 2008. An autoregressive approach to spatio-temporal disease mapping. *Statistics in medicine*. 27, 2874-2889.

Martins, T. G., Simpson, D., Lindgren, F., Rue, H. 2013. Bayesian computing with INLA: New features. *Computational Statistics and Data Analysis*. 67, 68-83.

Mia, M. S., Begum, R. A., Er, A., Abidin, R. D. Z. R. Z., Pereira, J. J. 2013. Trends of dengue infections in Malaysia, 2000-2010. *Asian Pacific Journal of Tropical Medicine*. 6, 462-466.

Pang, E., Loh, H. 2016. Current perspectives on dengue episode in Malaysia. *Asian Pacific Journal of Tropical Medicine*. 9, 395-401.

Richardson, S., Thomson, A., Best, N., Elliot, P. 2004. Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*. 112, 1016-1025.

Rue, H., Martino, S. 2007. Approximate Bayesian inference for hierarchical Gaussian Markov random field models. *Journal of statistical planning and inference*. 137, 3177-3192.

Rue, H., Martino, S., Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of Royal Stat Society B*. 71, 319-392.

Schrödle, B., Held, L. 2011. Spatio temporal disease mapping using INLA. *Environmetrics*. 22, 725-734.

Shafie, A., Rainis, R., Ahmad, A. H. *Sistem maklumat geografi dalam kajian demam denggi di Malaysia*. Penerbit Universiti Malaya, 2015.

Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*. 64, 583-639.

Ugarte, M. D., Adin, A., Goicoa, T., Militino, A. F. 2014. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Statistical Methods in Medical Research*. 23, 507-530.

Ugarte, M. D., Goicoa, T., Ibáñez, B., Militino, A. F. 2014. Evaluating the performance of spatio-temporal Bayesian models in disease mapping. *Environmetrics*. 20, 647-665.

Wakefield, J., 2007. Disease mapping and spatial regression with count data. *Biostatistics*. 8, 158-183.