International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:8, No:7, 2014

Spatio-Temporal Analysis and Mapping of Malaria in Thailand

Krisada Lekdee, Sunee Sammatat, Nittaya Boonsit

***Abstract*—**This paper proposes a GLMM with spatial and temporal effects for malaria data in Thailand. A Bayesian method is used for parameter estimation via Gibbs sampling MCMC. A conditional autoregressive (CAR) model is assumed to present the spatial effects. The temporal correlation is presented through the covariance matrix of the random effects. The malaria quarterly data have been extracted from the Bureau of Epidemiology, Ministry of Public Health of Thailand. The factors considered are rainfall and temperature. The result shows that rainfall and temperature are positively related to the malaria morbidity rate. The posterior means of the estimated morbidity rates are used to construct the malaria maps. The top 5 highest morbidity rates (per 100,000 population) are in Trat (Q3, 111.70), Chiang Mai (Q3, 104.70), Narathiwat (Q4, 97.69), Chiang Mai (Q2, 88.51), and Chanthaburi (Q3, 86.82). According to the DIC criterion, the proposed model has a better performance than the GLMM with spatial effects but without temporal terms.

***Keywords*—**Bayesian method, generalized linear mixed model (GLMM), malaria, spatial effects, temporal correlation.

I. Introduction

MALARIA has been a leading cause of morbidity and mortality in Thailand for many decades. Around 32 million people are at risk of malaria. All the four types of malaria are prevalent in the country [1]. For the effective control utilization of resources is required. The type and degree of interventions need to be based on epidemiological patterns of malaria risk. Malaria risk varies in space and time [2]. It is important to describe the spatio-temporal variability of malaria risk to guide control programs [3]-[5]. Disease mapping defined as a method for displaying the spatial distribution of disease occurrence, or exposure occurrence, on a map is the most basic way of visualizing the spatial distribution of the disease of interest in a defined area. The production of malaria maps relies on modeling to predict the risk for most of the map.

Accurate prediction of risk is dependent on knowledge of a number of environmental and climatic factors that are related to malaria transmission [6], [7]. Routine hospital malaria data provide a proxy for the incidence or severe malaria and for the crudely estimating morbidity rate or equivalent indicators. Analysis of these data may allow to access, compare and improve the care provided at all levels of health care. It assists in monitoring and planning resource needs in a health system and designing appropriate interventions, tailored towards communities at high risk or lead to further investigations to identify important risk factors [8]. This paper was motivated by a yearly report of the malaria data collected by the Bureau of Epidemiology, Ministry of Public Heath, Thailand, in which descriptive statistics such as percentages and charts are commonly presented [9]. An alternative and useful tool is presenting in disease maps. Therefore, we seek for a proper model for disease mapping. Reference [10] studies factors related to malaria incidence rates in Thailand using generalized linear mixed models (GLMM). Reference [11] extends their model [10] by adding spatial random effects in the GLMM under a hierarchical Bayesian framework. Their model is applied to malaria data in Thailand and malaria maps are constructed.

This paper, extending the model of [11], proposes a GLMM with spatial effects and random effects that account for the temporal correlation. Hierarchical Bayesian inference via Gibbs sampling Markov Chain Mote Carlo (MCMC) is used for parameter estimation. The posterior means of the estimated mortality rates are used for producing malaria maps. The response variables are the repeated count data, the number of malaria patients collected quarterly in each province of Thailand in 2012. The factors considered are rainfall and temperature. This paper is organized as follows. Section II briefly describes the methodology. The application is illustrated in Section III. In Section IV, the result of the study is presented. Lastly, in Sections V and VI the discussion and conclusion are drawn.

II. Methodology

A GLMM with spatial effects and temporal effects are employed. The spatial effects are assumed to arise from the conditional autoregressive (CAR) model originally proposed by [12]. The temporal correlation is specified in the covariance matrix of the random effects. The model is expressed as:

For $i = 1,...,m$, $j = 1,...,q$,

Assume $y_{ij} \mid \mathbf{b}_i, v_i \overset{iid}{\sim} Pois(\mu_{ij})$, $\mathbf{y}_i = (y_{i1},...,y_{iq})^T$ and $\mathbf{b}_i = (\mathrm{b}_{i1},...,\mathrm{b}_{iq})^T$, the proposed model is:

$$\log(\mu_{ij}) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + \sum_{k=1}^{q} b_{ki} + v_i, \quad (1)$$

where $\mathbf{x}_{ij}$ are vectors of fixed effects, $\boldsymbol{\beta} = (\beta_0,...,\beta_k)^T$ is a vector of fixed effect coefficients, $\mathbf{b}_i$ are vectors of random

L. Krisada, S. Sunee, and B. Nittaya are with the Department of Mathematics and Statistics, Faculty of Science and Technology, Rajamangala University of Technology Phra Na Khon, Bangkok, Thailand (phone: 662-9132424 ext. 197 ; fax: 662-9132424 ext. 156; e-mail: krisada.l@rmutp.ac.th, Sunee.sa@rmutp.ac.th, nittaya.bo@rmutp.ac.th).

effects, and v_i are spatial effects.

$$v_i \mid v_{i\neq j} \sim \text{CAR.Normal}\left(\sum_{j=1}^{m}\frac{w_{ij}v_j}{w_{i+}}, \frac{\tau_v^{\,2}}{w_{i+}}\right), \quad (2)$$

where $w_{ij}=1$ if i, j are adjacent (or 0 if they are not), $w_{i+}=\sum_j w_{ij}$, and $\tau_v^{\,2}$ is a conditional variance of v_i.

Model estimation was achieved using a hierarchical Bayesian approach [13]. Bayesian Inference is based on estimating the probability density of the parameters $\boldsymbol{\theta}$ in the model after observing the data, their posterior distributions, $\mathrm{p}(\boldsymbol{\theta} \mid \mathbf{Y})$.

$$\mathrm{p}(\boldsymbol{\theta} \mid \mathbf{Y}) = \frac{\mathrm{p}(\boldsymbol{\theta}, \mathbf{Y})}{\mathrm{p}(\mathbf{Y})} = \frac{\mathrm{p}(\mathbf{Y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(\mathbf{Y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta}} \quad (3)$$

$$\mathrm{p}(\boldsymbol{\theta} \mid \mathbf{Y}) \propto \mathrm{f}(\mathbf{Y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) \quad (4)$$

$\mathrm{f}(\mathbf{Y} \mid \boldsymbol{\theta})$ is the likelihood of the model, which reflects the relationship between the data and the parameters. $\pi(\boldsymbol{\theta})$ is the prior distribution of the parameters, which reflects the initial information on the parameters. Usually, $\mathrm{p}(\boldsymbol{\theta} \mid \mathbf{Y})$ is computed by simulation using Markov Chain Monte Carlo techniques. MCMC aims at simulating a series of values for the parameters in the model, so that, in the end, these values will be drawn from the posterior distribution. The most common hierarchical Bayesian model is the case that there are 3 stages. At the first stage, a distribution for the data given parameters is specified. At the second stage, prior distributions for parameters given hyper-parameters are specified, and distributions for hyper-parameters are specified at the third stage.

III. Application

Data: Malaria quarterly data have been extracted from the Bureau of Epidemiology, Ministry of Public Health, Thailand [9]. Factors considered in this study are rainfall and temperature collected for the Thai Meteorological Department [14].

For province i and quarter j, $i = 1,...,76$, $j = 1,...,4$,

let y_{ij} = the number of the malaria patients.

Assume $y_{ij} \mid \mathbf{b}_i, v_i \overset{iid}{\sim} Pois(\mu_{ij})$, $\mathbf{y}_i = (y_{i1}, y_{i2}, y_{i3}, y_{i4})^T$ and $\mathbf{b}_i = (\mathrm{b_{i1}}, \mathrm{b_{i2}}, \mathrm{b_{i3}}, \mathrm{b_{i4}})^T$, the proposed model is:

$$\log(\mu_{ij}) = \log(pop_i) + \beta_0 + \beta_1 * rain_{ij} + \beta_2 temp_{ij} + b_{1i} + b_{2i} + b_{3i} + b_{4i} + v_i \quad (5)$$

where $\beta_0, \beta_1, \beta_2$ are fixed effect coefficients, $rain_{ij}$ are the amounts of rainfall, and $temp_{ij}$ are the average temperatures, $\mathbf{b}_i$ are vectors of random effects that take into account the temporal correlation via its covariance matrix, and v_i are spatial effects that account for the spatial dependence. We add the offsets, $\log(pop_i)$, where pop_i are mid-year population, so that we can get the morbidity rates (MR).

$$\mathrm{MR} = \exp(\beta_0 + \beta_1 * rain_{ij} + \beta_2 temp_{ij} + b_{1i} + b_{2i} + b_{3i} + b_{4i} + v_i) \quad (6)$$

Diffuse priors are assigned to the fixed effect coefficients.

$$\beta_0, \beta_1, \beta_2 \sim \mathrm{N}(0.0, 10^4),\ \mathbf{b}_i \sim \mathrm{MN}(\mathbf{0}, \mathbf{D}).$$

The covariance structure of $\mathbf{D}$ is assumed to be exchangeable,

$$\mathbf{D} = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & 0 & 1 & \rho \\ \rho & \rho & 0 & \rho \end{bmatrix} \quad (7)$$

$$\sigma^2 \sim \mathrm{IG}(0.5, 0.005),\ \rho \sim \mathrm{Unif}(0,1)$$

$v_i \mid v_{i\neq j} \sim \text{CAR.Normal}\left(\sum_{j=1}^{m}\frac{w_{ij}v_j}{w_{i+}}, \frac{\tau_v^{\,2}}{w_{i+}}\right)$ as defined in (2). We assume $\tau_v^{\,2} \sim \mathrm{IG}(0.5, 0.005)$.

The model was implemented using WinBugs [15]. The Gibbs sampling MCMC are run for 25,000 iterations, with burn-in of 5,000. We assess MCMC convergence of all model parameters by visual analysis of history and Kernel density plots. The performance of the proposed model is compared with the GLMM with spatial effects but without temporal correlation:

$$\log(\mu_{ij}) = \log(pop_i) + \beta_0 + \beta_1 * rain_{ij} + \beta_2 temp_{ij} + b_i + v_i. \quad (8)$$

The deviance information criterion (DIC) is used [16]. Smaller DIC is preferable to larger DIC.

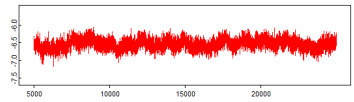

Fig. 1 History plot of β_0 showing convergence

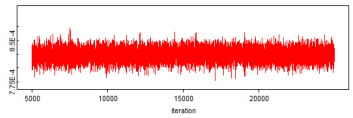

Fig. 2 History plot of β_1 showing convergence

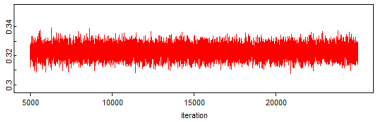

Fig. 3 History plot of β_2 showing convergence

IV. Result

As an example, for β_0, β_1 and β_2, the history plots in Figs. 1-3 show no trend and snake around the mean and the Kernel density plots in Figs. 4-6 look more bell-shape or not multi-modal. These suggest Gibbs sampling MCMC convergence. Table I presents the posterior means, the corresponding 95% credible intervals (CI) for the parameters, and morbidity rates. If the amount of rainfall increases by 1 mm, the malaria morbidity rate (per 100,000 population) will increase by 0.08%. If the temperature increases by 1°C, the morbidity rate will increase by 38.14. The spatial variation across the areas is not large ($\tau_v^2 = 0.0146$). The temporal correlation ($\rho = 0.4677$) is quite large, as well as the variability in each area is quite large ($\sigma^2 = 2.4470$). Since the DIC of the proposed model (21,000) is less than the DIC of the GLMM with spatial effects but without temporal correlation (DIC =21,700), it is evident that the proposed model has a better performance. The province and quarter (Q) whose morbidity rate (per 100,000 population) is higher than 80 is presented in Table II. The malaria maps showing the distribution of the mortality rates in each province are presented in Figs. 7–10.

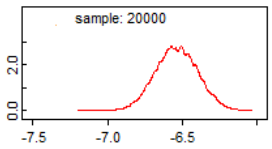

Fig. 4 Kernel density plot of β_0 showing convergence

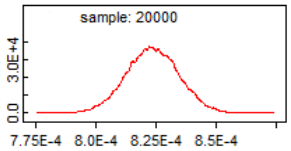

Fig. 5 Kernel density plot of β_1 showing convergence

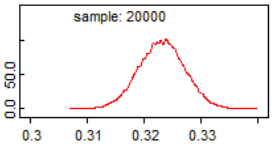

Fig. 6 Kernel density plot of β_2 showing convergence

TABLE I
Posterior Means, the Corresponding 95% Credible Intervals (CI) for the Parameters, and Morbidity Rates

Param.	Mean	Stand. Error	95% Credible Interval		Morbid. Rate
β_0	-6.5740	0.1413	-6.8460	-6.2950	0.0014
β_1 (rain)	0.0008	0.0000	0.0008	0.0008	1.0008
β_2 (temp)	0.3231	0.0040	0.3153	0.3309	1.3814
ρ	0.4677	0.3034	0.0190	0.9819	-
σ^2	2.4470	1.1100	1.0920	5.0400	-
τ_v^2	0.0146	0.0221	0.0015	0.0711	-

TABLE II
Province and Quarter (Q) in Which the Posterior Mean of the Estimated Morbidity Rate (per 100,000 Population) is Higher than 80

Province	Q	Mean	Stand. Error	95% Credible Interval	
Trat	3	111.70	5.50	101.20	122.70
Chiang Mai	3	104.70	1.75	101.30	108.20
Narathiwat	4	97.69	2.64	92.61	102.90
Chiang Mai	2	88.51	1.52	85.57	91.54
Chanthaburi	3	86.82	2.58	81.87	91.95
Nakhon Sawan	4	81.09	1.68	77.79	84.39
Surin	2	80.79	1.54	77.83	83.85

V. Discussion

The malaria maps cluster and identify the regions which have high risk of this disease. Such identification is important because policy makers may wish to target regions associated with such extreme risks for financial assistance while epidemiologist may wish to target such regions for further study. Rainfall and temperature are related to the malaria morbidity rates. This result supports the finding of [17] indicating that the temperature and rainfall are strong positive predictors of increased annual malaria incidence in Zimbabwe. Our proposed model has a better performance than the GLMM with spatial effects but without temporal correlation because the temporal correlation usually occurs in the spatio-temporal data. We select exchangeable covariance structure because our data are quarterly collected, so each period of time is not far from each other. Moreover, the data are usually spatially correlated according to the first law of geography stating that "everything is related to everything else, but near things are more related than distant things." [18]. Our proposed model can be applied to other kinds of spatio-temporal data.

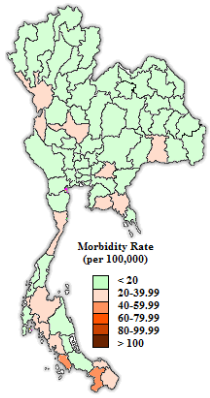

Fig. 7 Malaria morbidity rate in Q1

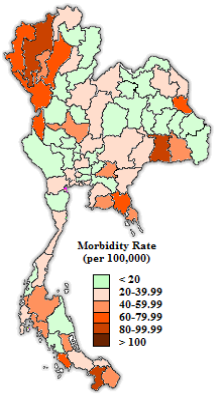

Fig. 8 Malaria morbidity rate in Q2

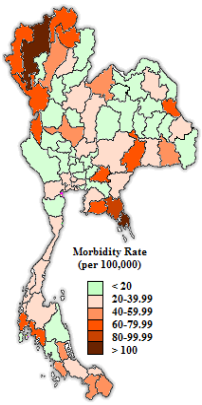

Fig. 9 Malaria morbidity rate in Q3

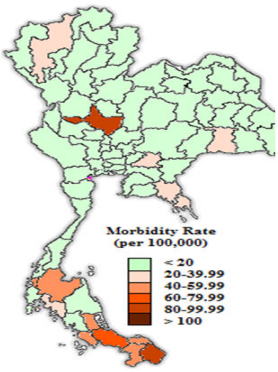

Fig. 10 Malaria morbidity rate in Q4

VI. Conclusion

This paper proposes a GLMM with spatial and temporal effects for malaria data in Thailand. The malaria quarterly data have been extracted from the Bureau of Epidemiology, Ministry of Public Health of Thailand [9]. The related factors are rainfall and temperature. A Bayesian method is used for parameter estimation via Gibbs sampling MCMC in WinBUGS. The spatial dependence is presented in a CAR model and the temporal correlation is presented through the covariance matrix of the random effects. The result shows that rainfall and temperature are positively related to the malaria morbidity rate. The posterior means of the estimated morbidity rated are used to construct the malaria maps. The malaria maps show the distribution of the morbidity rate (per 100,000 population) in each province. The top 5 highest morbidity rates are in Trat (Q3, 111.70), Chiang Mai (Q3, 104.70), Narathiwat (Q4, 97.69), Chiang Mai (Q2, 88.51), and Chanthaburi (Q3, 86.82). In epidemiology and surveillance, the map is a useful tool for gathering some information. Using DIC criterion, the proposed model has a better performance than the GLMM with spatial effects but without temporal terms.

Acknowledgment

We gratefully thank Assist. Prof. Dr. Amara Amornkaew, Dean of the Faculty of Science and Technology, Assist. Prof. Jutamas Peerapatchara, Director of the institute of research and development, and Assoc. Prof. Dr. Wirote Impitak , Rector of Rajamangala University of Technology Phra Nakhon for their financial support.

References

[1] WHO. (2013, Jan 11). "Malaria Situation in SEAR Countries: Thailand." Available: http://www.searo.who.int/en/Section10/Section21/Section340_4027.htm.
[2] I. Kleinschmidt, B. Sharp, I. Mueller, and P. Vounatsou, "Rise in malaria incidence rates in South Africa: small area spatial analysis of variation in time trends," Am J Epidemiol, vol. 155, 2002, pp. 257–264.
[3] R. Carter, K.N. Mendis, and D. Roberts, "Spatial targeting of interventions against malaria," Bull WHO, vol. 78, 2000, pp 1401–1411.
[4] D. Le Sueur, F. Binka, C. Lengeler, D. De Savigny, B. Snow, T. Teuscher, and Y. Toure, " An atlas of malaria in Africa," Africa Health. Vol. 19, 1997, pp. 23–24.
[5] R.W. Snow, K. Marsh, D. Le Sueur, "The need for maps of transmission intensity to guide malaria control in Africa," Parasitol Today, vol. 12, 1996, pp. 455–457.
[6] R.W. Snow, E. Gouws, and J.A. Omumbo, "Models to predict the intensity of Plasmodium falciparum transmission: applications to the burden of disease in Kenya," Trans R Soc Trop Med Hyg, vol. 92, 1998, pp. 601-606.
[7] L.R. Beck, M.H. Rodriguez, and S.W. Dister, "Remote sensing as a land-scape epidemiologic tool to identify villages at high risk for malaria transmission," Am J Trop Med Hyg, vol. 51, 1994, pp. 271–80.
[8] L. N. Kazembe, I. Kleinschmidt, T.H. Holtz, and B. L. Sharp, "Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data," Int J Health Geogr, vol. 5, 2006, pp. 41.
[9] Bureau of Epidemiology. (2012, Jan 20). "Table of notifiable diseases," Available: http://www.boe.moph.go.th/Annual/AESR2012/index.html.
[10] K. Lekdee and L. Ingsrisawang, "Risk factors for malaria in Thailand using generalized estimating equations (GEE) and genalized linear mixed model (GLMM)," Journal of Health Science, vol. 19, 2010, pp. 364-373.
[11] K. Lekdee, S. Sammatat, N. Boonsit, and L. Ingsrisawang, "Spatial Analysis and Malaria Mapping in Thailand," The World Congress on Engineering and Technology (CET2011), Oct. 28 to Nov. 2, 2011, Shanghai, China,pp.445-448.
[12] J. Besag, J.York, and A. Mollie, "Bayesian image restoration, with two applications in spatialstatistics," Ann. Inst. Stat. Mat, vol. 43, 1991, pp. 1-59.
[13] P. Congdon, *Bayesian Statistical Modelling*, 2nd ed., John Wiley & Sons: New York, pp. 1-56, 2006.
[14] TMD. (2011, May 10). "Weather." Available: URL: http://www.tmd.go.th/en/.
[15] The BUGS Project. (2011, Jan 30). Available: http://www.mrcbsu.com.ac.uk .

[16] D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn, D. *WinBUGS Version 1.4 User Manual*, MRC Biostatistics Unit, Institute of PublicHealth: London, pp. 15, 2003.
[17] M. LH, Mabaso, P.Vounatsou, S. Midzi, J. Da Silva, and T. Smith, "Spatio-temporal analysis of the role of climate in inter-annual variation of malaria incidence in Zimbabwe," International Journal of Health Geographics, 2006, pp. 5-20.
[18] W.R.Tobler, "A computer movie simulating urban growth in the Detroit region," Economic Geography, vol. 46, 1970, pp. 234-240.