Assessing the temporal modelling for prediction of dengue infection in northern and northeastern, Thailand

Wongkoon, S., Jaroensutasinee, M.* and Jaroensutasinee, K.
Centre of Excellence for Ecoinformatics, School of Science, Walailak University, 222 Thaiburi, Thasala, Nakhon Si Thammarat 80161, Thailand
*Corresponding author email: jmullica@gmail.com
Received 20 December 2011; received in revised form 24 February 2012; accepted 19 March 2012

Abstract. This study aimed at developing a predicting model on the incidence rate of dengue fever in four locations of Thailand – i.e. the northern region, Chiang Rai province, the northeastern region and Sisaket province – using time series analysis. Seasonal Autoregressive Integrated Moving Average (SARIMA) model was performed using data on monthly incidence rate of dengue fever from 1981 to 2009, and validated using the monthly rate collected for the period January 2010 to October 2011. The results show that the SARIMA(1,0,1)(0,1,1)_{12} model is the most suitable model in all locations. The model for all locations indicated that the predicted dengue incidence rate and the actual dengue incidence rate matched reasonably well. The model was further validated by the Portmanteau test with no significant autocorrelation between residuals at different lag times. Our findings indicate that SARIMA model is a useful tool for monitoring dengue incidence in Thailand. Furthermore, this model can be applied to surveillance data for early warning systems for control and reduction of dengue transmission.

INTRODUCTION

Dengue is a serious public health problem in the tropical regions, particularly in Thailand. World Health Organisation (WHO) estimated that about 50-100 million cases of dengue are recorded from all over the world annually and two-fifths of the world population are at risk. More than one hundred countries have been affected by dengue or DHF/DSS epidemics (Sabin, 1959). In Thailand, dengue epidemics of increasing magnitude and severity occur every two to four years beyond the endemic levels (Strickman et al., 2000). The northern and north-eastern regions of Thailand have been most affected by dengue possibly due to the increase of a rare virus serotype against which there is limited immune protection, human ecology and behaviour (WHO, 1997; 2004; Thomas, 2004; Muttitanon et al., 2005; Bureau of Epidemiology, 2011). Although dengue occurs throughout the year, cases peak from June to August during the wet season (Strickman & Kittiyapong, 2003).

Neither a vaccine nor specific treatment for dengue fever is available (Morrison et al., 2004; Bruno et al., 2011). Vector control seems to be the most possible solution to prevent dengue transmission. Well-designed and reliable strategies for specific temporal patterns and a predictive model for the disease are needed. Dengue forecasting models have gained much interest in the past decades, but the temporal patterns and ability to develop a predictive forecasting model are still not reliable (Kanchanpairoj et al., 2000; Nagao et al., 2003; Muttitanon et al., 2005; Thammapalo et al., 2005). In addition, mosquito population dynamics are not the same in different geographical regions where dengue is transmitted suggesting that the influence of climate on dengue may be site specific (Scott & Morrison, 2003).
The Seasonal Autoregressive Integrated Moving Average (SARIMA) models are widely used to predict the incidence of various infectious diseases such as Ross River virus disease (Hu et al., 2004; 2006), malaria and hepatitis A (Nobre et al., 2001) and dengue fever (Silawan et al., 2008; Gharbi et al., 2011; Wongkoon et al., 2011). This research was carried out in an attempt to propose a forecasting model for dengue incidence in two regions of Thailand using SARIMA models.

MATERIALS AND METHODS

Study area

Thailand covers 518 000 km² (between latitude 5° and 21° N and longitude 97° and 107° E). The northern region, bordering on Laos and Myanmar, is mostly high heavily forested mountains with several flat river basins, (Fig. 1). The region, which is divided into 17 provinces, has a local population of 11 788 411 with average density of 69.7 people/km². Chiang Rai province is the northernmost province of Thailand with an average elevation of 580 m. Chiang Rai has an area of 11 678.37 km² with a population of 1 198 218. The province borders Myanmar on the north and Laos on the north and northeast (Fig. 1).

The north-eastern region occupies a large land area in a strategic location bordering on Laos and Cambodia to the east. The region is comprised of 19 provinces with a local population of 21 573 318 and average density of 124.31 people/km² (Fig. 1). The region usually conjures up the image of an arid land area. Sisaket, one of the north-eastern provinces, has a land area of 8 839.976 km² and a population of 1 446 345. The province borders Cambodia to the south (Fig. 1). The seasonal weather for the northern and north-eastern regions of Thailand consists of three seasons: summer season (February-May), rainy season (May-October) and winter season (October-February).

Data collection

We obtained monthly-notified dengue fever cases per 100 000 population in the northern region, Chiang Rai province, the north-eastern region and Sisaket province for the period January 1981 to October 2011 from the Bureau of Epidemiology, Department of Disease Control (DDC), Ministry of Public Health (MOPH) (Bureau of Epidemiology, 2010). Thailand has had a well-established surveillance system for dengue since 1967 (Tipayamongkholgul et al., 2009). All identified dengue cases are based on WHO clinical criteria (Tipayamongkholgul et al., 2009; WHO, 2009). A dengue case is clinically-diagnosed and laboratory-confirmed. The data were collected continuously and systematically from government (public) hospitals, provincial public health offices, and health centres by the National Disease Surveillance (Report Figure 1. Study areas in Thailand: northern region ( ), Chiang Rai province ( ), northeastern region ( ) and Sisaket province ( )
The 506 surveillance weekly summarised databases include information on the dengue cases such as gender, age, nationality and occupation. The data collection mechanism has been stable over time, and this routinely collected data can be used for analysing factors affecting the occurrence of dengue fever.

**Time series analysis**

We used the Box-Jenkins approach to SARIMA modelling of time series, which consists of a four-step process. First, we evaluated the need for variance-stabilising transformations using the mean-range plot. Second, SARIMA(p,d,q)(P,D,Q)s model was fitted, where p is the order of autoregression (AR), d is the order of integration, q is the order of moving average (MA), P is the order of seasonal autoregression (SAR), D is the order of seasonal integration, Q is the order of seasonal moving average (SMA) and s is the length of seasonal period, using the following five tools: 1) The plot of dengue incidence, which assists in the need for non-seasonal and seasonal differencing; 2) The auto-correlation (ACF) and partial auto-correlation (PACF) functions, which indicate the temporal dependence structure in the stationary time series; 3) Three measures, namely, Bayesian information criterion (BIC), Root Mean Square Error (RMSE) and Mean Absolute Percent Error (MAPE), based on information theory, were used to achieve a trade-off between an adequate prediction and a minimum of number of parameters (Weigend & Gershenfeld, 1994). Comparatively lower values of BIC, RMSE and MAPE were preferable; 4) The model diagnostic test was performed by examining the residual ACF and p-values for the Portmanteau test. The Portmanteau test is based on the statistic:

$$Q_h = n (n+2) \sum_{k=1}^{h} \hat{\rho}_z(k)^2 / (n-k),$$

where \( \hat{\rho}_z(k) \): correlation function of the residuals at each k

h: the first h correlation values

n: sample size

The Portmanteau statistic has an asymptotic \( \chi^2 \) distribution with h-p-q degree of freedom. If \( Q_h > \chi^2_{h-p-q} \), the adequacy of the model is rejected at level \( \alpha \). The mean errors for actual cases and predicted cases for the suitable models were calculated. The most suitable models were chosen based on their adequate predictions; and, 5) The significance of the parameters, which should be statistically different from zero (that is, the t statistic should exceed 2 in absolute value). Third, we estimated the parameters of the SARIMA model by maximum likelihood. Finally, we graphically compared the model’s fitted values with the observed data to check if it indeed models dengue incidence. In order to identify the best fit model, different models were examined including ACF and PACF. Possible SARIMA models, with seasonal differencing, were fitted to data series for each location from 1981 to 2009 (the training sets). Then the forecasting accuracy of this model was verified using the data between January 2010 and October 2011 (the testing sets). Twenty-two predicted months were used to verify the SARIMA model. All statistical analyses were conducted using Mathematica Software with Time Series package.

**RESULTS**

**Descriptive epidemiology**

During the January 1981 to October 2011 period, 377 213 dengue cases were reported in the northern region; 29 545 in Chiang Rai province; 671 219 in the north-eastern region; and 39 992 in Sisaket province. Dengue incidence rates differed among locations: the northern region, Chiang Rai, the north-eastern region and Sisaket (Kruskal-Wallis test: \( \chi^2_3 = 109.727, P < 0.001, \) Fig. 3a-d). The north-eastern region had higher dengue incidence rates than northern region (Mann-Whitney test: \( U = 238,338, P < 0.001, \) Fig. 3a, c). The northern region had higher dengue incidence rates than Chiang Rai (Mann-Whitney test: \( U = 41,244, P < 0.001, \) Fig. 3a, b). The north-eastern region had higher dengue incidence rates than Sisaket (Mann-Whitney test: \( U = 62,061, P < 0.001, \) Fig. 3c, d). There
were both seasonal and inter-annual fluctuations (Fig. 3a-d). The dengue season in all four locations started in April, peaked during June-August, declined in October, and reached the lowest level during December and January.

**Forecasting of dengue incidence**

The plots of the sample ACF and PACF describe the temporal dependence structure in dengue incidence and suggest that non-seasonal and seasonal parameters are needed in the model (Fig. 2a-h). The best fitted model in all locations was the SARIMA(1,0,1)(0,1,1)_{12} model with the lowest BIC, RMSE and MAPE (Table 1). The ACF of residuals for the model at different lag times were not significantly different from zero in all locations (Portmanteau test: the northern region: $Q_{20} = 28.125, P>0.05$;

![Figure 2. ACF and PACF of monthly dengue incidence rate from 1981 to 2009 in (a, b) northern, (c, d) Chiang Rai, (e, f) north-eastern and (g, h) Sisaket, Thailand](image_url)
Our results support the previous findings that the highest dengue incidence in Thailand usually occurs in June to August (Strickman & Kittiyapong, 2003; Wongkoon et al., 2002, 2008; Silawan et al., 2008; Thammapalo et al., 2008). This may be due to the patterns of vector occurrence and the high rates of dengue transmission in May and June (Strickman & Kittiyapong, 2003). The seasonal high peak of dengue incidence differs between countries within the Asian region, e.g. India (August and November), Indonesia (January to February) and Myanmar and Sri Lanka (May and August) (WHO, 2010). All these seasonal patterns of dengue outbreaks in the Asian region including Thailand coincide with the rainy season (e.g. Watts et al., 1987; Thongrungkiat et al., 2003; WHO, 2010; Wongkoon et al., 2011). The end of the rainy season leads to a return to lower transmission levels. This phenomenon is repeated every year and characterises the endemic mode of transmission (Barbazan et al., 2002). This may have implications for the health office in the area with regard to disease control and risk management planning before the start of the rainy season.

Our results show that the dengue incidence rates in the northern and northeastern regions from 1981 to 2011 had several outbreak peaks which were higher than the dengue control target set by the Ministry of Public Health of Thailand - i.e. 50 dengue cases per 100 000 population (Suwanchai-chinda, 2000; Juraserikul, 2002; Usaha et al., 2003; WHO, 2010; Wongkoon et al., 2011). The end of the rainy season leads to a return to lower transmission levels. This phenomenon is repeated every year and characterises the endemic mode of transmission (Barbazan et al., 2002). This may have implications for the health office in the area with regard to disease control and risk management planning before the start of the rainy season.

Our results show that the dengue incidence rates in the northern and northeastern regions from 1981 to 2011 had several outbreak peaks which were higher than the dengue control target set by the Ministry of Public Health of Thailand - i.e. 50 dengue cases per 100 000 population (Suwanchai-chinda, 2000; Juraserikul, 2002; Usaha et al., 2003; WHO, 2010; Wongkoon et al., 2011). The end of the rainy season leads to a return to lower transmission levels. This phenomenon is repeated every year and characterises the endemic mode of transmission (Barbazan et al., 2002). This may have implications for the health office in the area with regard to disease control and risk management planning before the start of the rainy season.

Our results show that the dengue incidence rates in the northern and northeastern regions from 1981 to 2011 had several outbreak peaks which were higher than the dengue control target set by the Ministry of Public Health of Thailand - i.e. 50 dengue cases per 100 000 population (Suwanchai-chinda, 2000; Juraserikul, 2002; Usaha et al., 2003; WHO, 2010; Wongkoon et al., 2011). The end of the rainy season leads to a return to lower transmission levels. This phenomenon is repeated every year and characterises the endemic mode of transmission (Barbazan et al., 2002). This may have implications for the health office in the area with regard to disease control and risk management planning before the start of the rainy season.
did not occur until late September. Piped water is not commonly available in most rural areas. Because of the unusually severe water shortages in 1987, the increase in the number of water storage containers kept in and around the house, which were potential *Aedes aegypti* breeding sites, might have contributed to the increased dengue incidence of that year (Emchan *et al.*, 1989). In addition, Vietnam, Malaysia, and Burma also reported record incidence in 1987 whereas Taiwan and the Philippines experienced major epidemics the following year (Halstead, 1990).
This study represents another attempt to develop an epidemic forecasting model for predicting dengue transmission in different locations of Thailand using SARIMA model. The results of this study suggest that the key determinants of the dengue fever transmission include autoregression, moving average and seasonal moving average. These variables may be used to assist in forecasting outbreaks of dengue incidence rate in Thailand. Cases of dengue typically vary throughout the year and assume a regular pattern, normally in association with changes of temperature and rainfall. This pattern of disease is described as seasonal (WHO, 2010). Our results indicate that the predicted values could follow the upturn and downturn of the observed data reasonably well, especially during the high peaks in each location in Thailand.

The SARIMA modelling is a useful tool for interpreting and applying surveillance data in disease control and prevention. Once a satisfactory model has been obtained, it can be used to forecast expected numbers of cases for a given number of future time intervals (Allard, 1998). Since the SARIMA model has the capacity to forecast when and where an outbreak is likely to occur, it therefore has great potential to be used as a decision supportive tool for planning public health interventions (Hu et al., 2004). This study has two strengths. Firstly, a sophisticated time series model was used in the attempt to develop an epidemic forecasting system for the control and prevention of dengue fever in the northern and north-eastern regions of Thailand. Secondly, the model developed in this study appears to have a high degree of accuracy. However, in this study we focused only on the dengue fever cases. For future work, other factors that are reported to have significant impacts on the transmission of dengue fever should been included in the models such as climatic (Thammapalo et al., 2005; Wu et al., 2007; Wongkoon et al., 2011), social (Wongkoon et al., 2005; Vanlerberghe et al., 2010), seasonal (Wongkoon et al., 2007), topographical (Wongkoon et al., 2005; 2007), biological (Tun Lin et al., 2000; Strickman & Kittayapong, 2003; Thongrungkit et al., 2003; Costa et al., 2010), and economic factors such as population immunity, housing conditions, mosquito control measures, local ecological environments (vegetation, irrigation system), and drug resistance (Eamchan et al., 1989; Hayes et al., 2003; Kittigul et al., 2003; Phuanukoonnon et al., 2005).

Acknowledgements. We thank Thana na Nagara and John Endler for constructive comments on previous versions of this manuscript. Financial support from the Thailand Research Fund through the Royal Golden Jubilee Ph.D. Program (Grant No. PHD/0201/2548) to Siriwan Wongkoon and Mullica Jaroensutasinee, Walailak University Fund 06/2552 and Centre of Excellence for Ecoinformatics, NECTEC/Walailak University are acknowledged. We also thank the Bureau of Epidemiology, Department of Disease Control, the Ministry of Public Health for the dengue incidence data.

REFERENCES


