Bayesian Spatial-Temporal Autologistic Regression Model on Dengue Hemorrhagic Fever in East Java, Indonesia

Suci Astutik\textsuperscript{1}, Bayu Rahayudi\textsuperscript{2}, Agustin Iskandar\textsuperscript{3} and Rahma Fitriani\textsuperscript{1}

\textsuperscript{1}Statistics Study Program, Mathematics Department
Brawijaya University, Malang, Indonesia
suci_sp@yahoo.com, suci_sp@ub.ac.id

\textsuperscript{2}Computer Science Study Program, Mathematics Department
Brawijaya University, Malang, Indonesia
ubay1@ub.ac.id, ubay3@yahoo.com

\textsuperscript{3} Medical Faculty, Brawijaya University, Malang, Indonesia
agustin_almi@yahoo.co.id

Abstract. The purpose of this study is to discuss and develop Spatial-Temporal Autologistic Regression Model (STARM) to represent spreading of the \textit{Aedes aegypti} which is indicated by the endemic level of DHF (Dengue Hemorrhagic Fever) in East Java. The method which is used to estimate STARM parameter is Bayesian method with Markov Chain Monte Carlo (MCMC) and Gibbs Sampler simulation. This study observed 38 districts as spatial lattice units, meanwhile temporal unit is represented by monthly period of evidence (January-December) in 2002-2008. Result of the research was obtained STARM model that indicate the spreading pattern of the \textit{Aedes aegypti} that is indicated by the endemic level of DHF incidence in East Java have spatially and temporally positive correlation. Model validation using 95\% credible interval shows that all estimators are significant. This is also supported by a MAE value 0.09 and the percentage of correctly classified predicted data 90\%, which means there are 90 correctly classified data of 100 prediction data.

Keywords: Dengue Hemorrhagic Fever (DHF), Spatial Temporal Autologistic Regression Model (STARM), Bayesian methods, Markov Chain Monte Carlo (MCMC)
Introduction

Dengue Hemorrhagic Fever (DHF) is a contagious infectious disease vectors that often cause Unusual Incident, or epidemic, and also cause many case of death. This is a seasonal disease that usually happens in the rainy season, which allows the vector lives and spreads out at stagnant water [7]. The spread of dengue in East Java province usually spread from cities to villages because the virus that causes and also the vector, *Aedes aegypti*, are live widespread both in homes and in public places, except for areas with an altitude of more than 1000 meters above the sea [19,21].

Factors that influence the increase and spread of dengue cases are very complex, such as geographic and climatic conditions in the local area [12]. WHO (2005) state that has historically, outbreaks of DHF affected directly or indirectly by high rainfall. This statement is supported by Hales research in 1999 (WHO, 2005) that showed a positive relationship between rainfall against time series of monthly periods DHF incidence in Bangkok. Other research on Dengue epidemic in Nicaragua in 1998 concluded that the epidemiology of dengue can vary depending on geographic region and virus serotypes [13-18,6,8].

Based on the facts, it is possible to form a model to predict the presence of this mosquito species based on the state of the factors that influence it in any location, so that immediate prevention action can be optimal done by eradicating it without having to wait for the fall of victim. Obviously, the resulting predictions will be accurate if the relationship between climatic factors and the presence of mosquito species can be described through an appropriate model.

DHF incidence can be described as a binary response (there is an endemic or no endemic in a region) with including factors that cause it. Model that can describe this relationship is known as a logistic regression model [1,10]. In fact, the geographical conditions (proximity of the location area / spatial) also affects the incidence of dengue. So, it required a little modification to the logistic regression model in order that can involve spatial effects, i.e the existence of the mosquito species in the surrounding locations become an additional factor. This model was introduced as an autologistic regression by Besag [3-4] which is then much more detail studied by many researchers [5,9,23,25,28].

In the subsequent research [26] proposed a spatial-temporal autologistic regression model (STARM) which is an extension of autologistic regression. This model was applied to observe the pine beetle species outbreak. In addition to the factor that involving the spatial, STARM also enter temporal factors as explanatory variables. STARM is a very useful model for dealing with repeated measurements of binary observations taken on spatial lattices, instead of spatial data at one point in time. One major weakness of [27] is that Zhu’s STARM regression model parameters estimation is based on pseudo-maximum
Bayesian spatial-temporal autologistic regression model

likelihood, which is not statistically efficient. To overcome this, [28] proposed an MCMC algorithm that made STARM model parameter estimation easier to understand, and more useful to practitioners. In particular, the idea is used in the MCMC maximum likelihood [9,23].

This research aims to develop STARM parameter estimation methods that have been proposed by Zheng and Zhu [28], i.e Bayesian method with the MCMC algorithm on the incidence of DHF in East Java province to predict the incidence of dengue endemic level. This study is a continuation of research conducted by Astutik [2], which only studied simulation autologistic regression model on the incidence of DHF in East Java. In Indonesia, especially East Java, the factors that cause dengue fever incidence not only involves the geographical factor (spatial) but also time (temporal). Furthermore, the STARM model will be validated using a credible interval and Mean Absolute Error (MAE).

2 Materials and Methods

2.1. Data
The data in this study consider the spatial and temporal information. Unit of spatial/geographic observed in this study is the districts in East Java, which consists of 38 cities/counties. Unit time/temporal observed is monthly period (January-December) in 2002-2008 (Figure 1).

![Figure. 1 districts/counties map in East Java province](image)

The research variables are data on the number of cases of DHF and rainfall index (mm). The number of dengue cases in every city/county in East Java is obtained from the Province and City/District Health Office (Dinkes) in East Java in monthly period from 2002 to 2008. Furthermore, the data is expressed
in IR (Incidence Rate) i.e. the number of DHF incidence per 100,000 population and then scaled to the binary as a representation of the spread of the *Aedes aegypti*, i.e 1 if there is endemic (IR is higher than 10) and 0 if there is no endemic [19]. Rainfall index data (mm) in the monthly period 2002-2008 is obtained from irrigation department district office and BMKG Karangploso Malang, East Java.

2.2. Procedure in the STARM

In general, the stages of STARM model are as follows:

1) Digitize East Java map using GIS and determine the spatial weights matrix (adjacent / contiguity matrix)

2) Detect the presence or absence of spatial-temporal relationship to the incidence of DHF using the Moran autocorrelation multivariate method and LISA multivariate.

3) Define the STARM model as [28].

4) Estimate STARM model parameters using the MCMC algorithm [28]. Endemic level as the dependent variable. Rainfall index as independent variable.

5) Test of significance of model parameters. The significance of parameter estimation by using Bayesian can be made through Credible Interval (CI) of model parameters. If CI doesn’t include the zero value, it is said that the estimated parameters of the model was significant.

6) Perform validation of models using MAE. STARM is said as well data representation if it produces a minimum MAE. MAE is formulated as

\[
\text{MAE} = \frac{\sum_{i=1}^{T} \sum_{t=1}^{N} |e_{i,t}|}{NT}
\]

where \( e_{i,t} \) is error on the \( i \)-th location and the \( t \)-th time, \( Z(s_{i,t}) \) is the actual value at the \( i \)-th location and the \( t \)-th time, \( \hat{Z}(s_{i,t}) \) is the predicted value at the \( i \)-th location and the \( t \)-th time, \( N \) \( N \) is the number of area / location, \( T \) is the number of observation

7) Interpret STARM models, include studied the relationship between the independent variable and dependent variable, the significance of model parameters, odds ratios, and percentage correctly classified data.

Stages of analysis are done by software R, WinBUGS.

3. Results and Discussion

3.1. STARM on the DHF endemic level

STARM is an extension of autologistic regression model that include the temporal factor which is developed by Zhu et al. (2005) [28]. It is used for binary data that are measured repeatedly over time in spatial lattices.
involves the calculation of covariates, spatial dependence and temporal dependence simultaneously.

This study developed a STARM model as a model to approach DHF endemic level relationship (0 means there is not a DHF endemic; 1 means there is a DHF endemic) with climatic factors (rainfall) as a predictor variable that considers spatial and temporal factors. Furthermore, the model forms is used to study the patterns of *Aedes aegypti* mosquito distribution that cause of DHF.

STARM of DHF endemic level with climatic factors (rainfall) as a predictor variable considering spatial and temporal factors in the city/district in East Java province in the month period (January-December) during the years 2002-2008 is defined as follows:

\[
p(Y_{iT,i} | Y_{uj,t}; (i,t) = (i,t)) = p(Y_{iT,i} | Y_{uj,t}; (i,t) \in N_{iT}) = \frac{\exp[\theta_0 Y_{iT,i} + \theta_2 X_{iT,i} + \sum_{j \in N_{iT}} \theta_3 Y_{uj,t} + \theta_4 Y_{uj,t-1} + \theta_5 Y_{uj,t+1}]}{1 + \exp[\theta_0 Y_{iT,i} + \theta_2 X_{iT,i} + \sum_{j \in N_{iT}} \theta_3 Y_{uj,t} + \theta_4 Y_{uj,t-1} + \theta_5 Y_{uj,t+1}]} \tag{2}
\]

where \(Y_{iT,i}\) is endemic level (0 means no endemic dengue; 1 means there is endemic dengue) in the \(i\)-th location \((i = 1, \ldots, 38)\), at the \(t\)-th time \((t = 2002, 2003, \ldots, 2008)\), \(Y_{uj,t}\) is the dengue endemic level (0 means no endemic dengue; 1 means there is endemic dengue) in the \(j\)-th location in the \(t\)-th time; \(N_{iT}\) is sites adjacent to the \(i\)-th location at the \(t\)-th time, \(X_{iT,i}\) is rainfall predictor variable (rainfall index) at the \(i\)-th location (mm), \(\theta_0\) is intercept, \(\theta_1\) is the slope of the predictor variable, \(\theta_2\) is the spatial autoregressive coefficient, \(\theta_3\) is the temporal autoregressive coefficient. Based on the model STARM (2), there are 4 model parameters that must be estimated.

### 3.2. STARM Parameter Estimation on the DHF endemic level

STARM parameter estimation method is done by Bayesian inference approach using Markov Chain Monte Carlo algorithms (MCMC) combined with the Gibbs Sampler algorithm. Gibbs sampler algorithm is a special case of the Metropolis Hasting algorithm. We compute the posterior distribution model parameters with Gibbs sampler algorithm with burn in. STARM model parameter estimation algorithm with MCMC algorithms approach and Gibbs sampler are presented in WinBugs14 software. MCMC algorithm implementation details as follows: Determining the initial value of the parameter vector, it is selected \(\psi = (-1, -1, -1, -1)\). Prior distribution of \(\theta_0, \theta_1, \theta_2, \theta_3\) is the inverse Gaussian in the interval \((0, 0.00001)\) and \((-0.00001, 1)\). Then generated 100.000 Monte Carlo samples with 500 first samples is discarded (burn in).

Results of STARM model parameter estimation by MCMC approach is presented in Table 1.
Table 1. Results of STARM model parameters estimated by MCMC Method

<table>
<thead>
<tr>
<th>node</th>
<th>mean</th>
<th>sd</th>
<th>MC error</th>
<th>2.50%</th>
<th>median</th>
<th>97.50%</th>
<th>start</th>
<th>sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.098</td>
<td>0.001</td>
<td>0.000</td>
<td>0.096</td>
<td>0.097</td>
<td>0.101</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>A</td>
<td>-0.006</td>
<td>0.007</td>
<td>0.001</td>
<td>-0.021</td>
<td>-0.004</td>
<td>0.000</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>$\theta[1]$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>$\theta[2]$</td>
<td>-0.004</td>
<td>0.004</td>
<td>0.001</td>
<td>-0.016</td>
<td>-0.003</td>
<td>0.000</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>$\theta[3]$</td>
<td>-0.006</td>
<td>0.007</td>
<td>0.001</td>
<td>-0.031</td>
<td>-0.004</td>
<td>0.000</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>$\theta_\alpha$</td>
<td>0.994</td>
<td>0.007</td>
<td>0.001</td>
<td>0.979</td>
<td>0.996</td>
<td>1.000</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>0.996</td>
<td>0.004</td>
<td>0.001</td>
<td>0.984</td>
<td>0.997</td>
<td>1.000</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>0.994</td>
<td>0.007</td>
<td>0.001</td>
<td>0.970</td>
<td>0.996</td>
<td>1.000</td>
<td>1</td>
<td>100</td>
</tr>
</tbody>
</table>

3.3. Model Validation

STARM model validation is done by two methods, the test of significance of model parameters and testing models simultaneously. STARM model parameters test of significance are based on 95% of CI that the estimated parameters are in that interval. If the value of zero is not in the 95% of CI means that the estimated parameters of the model are significant. Table 1 shows that at 95% of CI, the estimated STARM model parameters are significant.

The MAE value of STARM model obtained was 9% (Table 1). This model can predict well level of endemic DHF in East Java. This is also indicated by percentage of STARM model predicted results which is consistent with the results of observation i.e 90%. Thus, this model can be used to predict the level of endemic districts in East Java.

3.4. STARM Interpretation

Interpretation of STARM model can be interpreted directly from Table 1 that there is a positive relationship between rainfall index and dengue endemic level ($\theta_1$). However, for spatial and temporal factors, it is showed that there are spatial autocorrelation among regions ($\theta_2$) and temporal autocorrelation of the time ($\theta_3$) and rather strong, which is consistent with Multivariate Moran and LISA autocorrelation test. Contributions or the role of predictor variables, spatial factors, temporal factors on the response variable can also be known from STARM model. This can be done by inserting STARM model parameter estimation from Table 1. ($\alpha$, $\theta[1]$, $\theta[2]$, $\theta[3]$) into the equation (2), and the result is:

$$p(Y_i,t|Y_{i,t};(i',t') \neq (i,t)) = p(Y_i,t|Y_{i,t};(i',t') \in N_{i,t}) =$$
Bayesian spatial-temporal autologistic regression model

\[ \exp \left[ -0.0062 \gamma_{i,t} + 3.67 \gamma_{j,t} + \sum_{j \neq i} -0.0041 \gamma_{i,t} + 0.00564 \gamma_{i,t-1} + \gamma_{i,t+1} \right] \]

Model (3) can be interpreted as follows: for the slope parameters: if other variables are considered constant then the increase 1 mm rainfall index will increase the probability of DHF endemic as height as \( \exp(0.0000367) = 1 \). For spatial autocorrelation coefficient: if other variables are considered constant then the closer a region to the other regions, it will increase the probability of DHF endemic as many as \( \exp(-0.00415) = 0.9959 \). Meanwhile, the temporal autocorrelation coefficient: if other variables are considered constant then the nearly time will increase the probability of DHF endemic as many as \( \exp(-0.00564) = 0.9944 \). The time required to calculate the results of statistical inference of STARM model parameter estimation by MCMC approach is 986 seconds.

4 Conclusion

Based on the goodness of fit test, STARM can predict the relationship between DHF endemic level which reflects the pattern of spread of the mosquito *Aedes aegypti* and rainfall index that involve spatial and temporal factors. This is demonstrated by 95% of CI, MAE value and the percentage of correctly classified predicted data 90%.

Future research: develop STARM models for endemic level of DHF prediction (spread of the mosquito *Aedes aegypti*), which involves more predictor variables include: host mosquito variable (vector mosquito density), demographic variables, and behavioral society variables. It is also necessary to develop another model that can predict the incidence of DHF in a longer period of time.

Acknowledgments

We would like to express our sincere gratitude to Directorate General for Higher Education (DIKTI) of National Education Ministry funded the research and Districts Health Department in East Java Province as providing DHF data.

References


Received: October, 2012