ARIMA Models of Dengue Cases in Kartamantul, Based on Area Risk Classification


* Correspondent Author: agus.rubaya@gmail.com
** Polytechnic of Health, Yogyakarta, Indonesia

INTRODUCTION

Dengue is a disease caused by infection of virus of Flavivirus genus and its transmission from a person to another is by Aedes mosquito bite. In the early of 2000, Dengue had been stated as the most important arthropod-borne viral diseases and as one of the main cause of high morbidity and mortality among 15-years old and younger children in tropical and sub-tropical countries. Southeast Asia is a region where the incidence of Dengue can be classified as hyper endemic, because epidemics repetitively occur in a cycle of almost three to five years. A study conducted in this region finds that the average number of cases per year between 2001 and 2010 reaches more than 2.9 millions with 5906 death tolls, and the estimation of economic burden hit 950 million USD.
Based on the number of cases between 2000 and 2011, the World Health Organization’s Office for Southeast Asia Region (SEARO) declared that the situation of Dengue in Indonesia was at A level, meaning that: the disease was a great problem of public health and main cause of inpatient and mortality among children. The four virus serotypes have been identified as being circulated in urban areas and tend to disperse slowly into rural ones.

A study conducted by Karyanti et.al. which analyzes the data collected by Dengue surveillance program in Indonesia since it was established in 1968 to 2013, finds that in Java Island, as the most populated island, the highest ranks of annual incidence rate (IR) per 100,000 inhabitants, after Jakarta the capital city of Indonesia, are West Java and Yogyakarta Provinces, consecutively. Even in 2013, Yogyakarta Province ranked at third place from all other provinces, with its IR reaching 95.99 per 100000 population, and was far above the national target.

Between 2006 and 2012 (except 2008), the number of Dengue cases in three out of five regencies/municipalities in Yogyakarta Province, i.e. Yogyakarta Municipality, Sleman and Bantul Regencies (acronym: Kartamantul), is always the highest, compared to the other two regencies.

One of key technical elements declared by WHO in global strategy for Dengue prevention and control in the 2012-2020 period11, is integrated surveillance and preparation to encounter the possibility of outbreak, with one of the goals is to monitor the trend of Dengue distribution and spread, based on time and geographical area.

One of the efforts for early detection against outbreak is by means of time-series analysis approach from previous Dengue cases and one of the commonly-used techniques is ARIMA (auto-regressive, integrated, moving average) model from Box-Jenkins. The auto-regressive element depicts the lingering effect of the previous values on a certain value; meanwhile, the integrated element represents the trend of the data; and the moving average element takes into account the lingering effect from one or some previous random shocks.

In relation to the outbreak alert, it is also important to map the high potential area for Dengue transmission accompanied by a good understanding of spatial and temporal distribution as a preventive measure. By doing so, the limited resources can then be effectively focused and prioritized to specific areas, and therefore surveillance function can run most effectively and will give the best effect.

A study conducted by Wen et.al. applies spatial mapping method of some new temporal risk indices to identify the risk of particular area on Dengue transmission. The indices comprise of: frequency, duration dan intensity indices; and their goals are to measure the occurrence of the disease, the duration of the disease, and the significance of the disease existing in a certain time and place. Those three temporal indices are developed to complement some limitations found in the use of spatial analysis only, time-series analysis, and the interpretation of incidence rate as comparison tool for areas.

This study aims at identifying the risk level of each village in Kartamantul towards Dengue transmission, and predicting the number of cases that will occur for each of the area risk level, by employing ARIMA model.

RESEARCH METHOD

Study Area

Kartamantul is located at the centre of Yogyakarta Province, lies from North to South between 110° 33’ 00” and 110° 12’ 34” longitude and between 07° 34’ 51” and 08° 00’ 27” latitude. Yogyakarta Municipality is the smallest area (32.50 km²), meanwhile for Sleman Regency and Bantul Regency, their areas are 574.82 km² and 508.85 km², respectively. Kartamantul has wet tropical climate with its rainy season usually coming between November-April and dry season between May and October.

In general, at the Southern part, Kartamantul has more sloping land contour because it is adjacent to Indonesia Ocean, meanwhile in the Northern part, the land altitude increases as it stepping up to Mount Merapi. The average population density in Yogyakarta Municipality reaches 15,000 lives per km² whereas in Sleman Regency and Bantul Regency they are only 1479 and 1910 lives per km², respectively.

206 villages are involved in this study, comprising of: 45 villages in Yogyakarta Municipality, 86 villages in Sleman Regency, and 75 villages in Bantul Regency.

This is an ecological study and its data are collected from serological confirmed Dengue cases registered at surveillance program of health offices in the three municipalities/regencies from 2010 to 2015. The data from 2010-2014 period are used as the source for calculating frequency, duration and intensity indices, as well as to construct the ARIMA model; meanwhile the 2015 data, are used to validate the accuracy of the risk levels and case prediction generated from the analysis. Two key
components are needed in the data, i.e. case’s address at village level and time (month and week) when the case acquired Dengue.

Data Analysis

The frequency index is calculated by comparing the number of weeks of each village where one or more cases were found in 2010-2014 period, with the total weeks in the same period. This index ranges between 0-1; where the closer the value to 1 the more frequent Dengue to occur in a particular area.

The duration index is derived using the same numerator as frequency index, and divided by the number of epidemic wave in 2010-2014 period. One epidemic wave is defined as one uninterrupted weeks where, at least, one Dengue case is found in Kartamantul\textsuperscript{15}. The bigger the value of duration index in a certain area the longer the Dengue emerging in that area to persist.

The intensity index is gained by dividing the incidence rate of Dengue of a certain area in 2010-2014 period with the number of epidemic wave in the same year period. This index measures the intensity of the disease, i.e. the bigger the value of this index the more temporally concentrated the Dengue cases in a particular area is. The data are processed using spreadsheet software.

The values of the three indices for each village are then analyzed with GeoDa of 1.8.14 version to determine their spatial association and to generate the risk map. Local indicator of spatial auto-correlation (LISA) technique is used as the tool for distinguishing the levels of the indices, spatially. The p-value derived from Monte Carlo simulation with 999 permutations is used to determine the significance of high-risk clusters.

Villages which have “high-high” spatial relationship with their neighbors for a certain index, and at least have a p-value of 0.05 is classified as "high" risk group for that index. On the other hand, areas which do not match that criteria, are then classified as “low” risk group. Subsequently, based on the composite risk level from the three indices, a village is classified as Dengue “high” risk area if all indices show “high” values; “medium” risk if one or two indices show "high" values; and “low” risk if none of the indices show "high" values.

The next analysis is ARIMA modelling by following the flow chart\textsuperscript{17} in Figure 1, by employing monthly cases data between January 2010 to December 2014 for the aggregation of villages according to their risk levels. Finally, the best models are then applied to predict the number of cases for each area risk classification in 2015.

![Flow-chart of ARIMA modelling with Box-Jenkins approach](Image)

RESULT AND DISCUSSION

The total number of Dengue cases between 2010 and 2014, and 2015, are 4515 in Yogyakarta Municipality, 2798 in Sleman Regency and 5332 in Bantul Regency. After detailed investigation to ensure the address and the time when the cases occurred, only a few cases are then considered as missing data because of their incomplete information. Those missing data are: five cases (0.11 %) in Yogyakarta Municipality, 10 cases (0.36 %) in Sleman Regency and 18 cases (0.34 %) in Bantul Regency.

Analysis of Three Temporal Indices

From the 206 villages in Kartamantul, the range of frequency index for 2010-2014 is between 0.000 and 0.529. It means from 261 weeks in that period, some villages were free from serologically positive Dengue infection, yet in contrast, there were also villages in which during the 138 weeks (0.529 * 261 weeks), at least, one case was always found. For 2015, the range is between 0-39 out of 52 weeks.

The range of duration index for 2010-2014 is between 0 and 69. It means for each epidemic wave, the number of consecutive weeks where Dengue were found, is from 0 up to 69 weeks. For 2015, the index ranges between 0-38 weeks for every epidemic wave. Meanwhile, the range of intensity index for 2010-2014 is between 0.0000 and 0.0015; which means that in a single epidemic wave, there are villages whose IR reach 15 cases per 10000 inhabitants. For 2015, the value of this index ranges between 0-42 cases per 10000 people.

To figure out the magnitude of the correlation of 2010-2014 indices and the same 2015 indices owned by
each village, Spearman correlation coefficient (\( \rho \)) is calculated (it is because the data do not follow the normal distribution assumption). The results are: the \( \rho \) for frequency index and duration index is similar, i.e. 0.855 (strong) with \( p \)-values < 0.001; while for intensity index it is 0.818 (strong) with \( p \)-value < 0.001. Therefore, it can be interpreted that the indices obtained from 2010-2014 can be used to predict the condition of 2015.

Based on the spatial analysis toward the three temporal indices for each area with 2010-2014 data, among all the 206 villages in Kartamantul, 31 (15.05 \%) are identified as “high” risk; 28 (13.59 \%) are identified as “medium” risk, and the rest 147 villages (71.36 \%) are classified as “low” risk (Figure 2). In terms of the municipality/regency where the villages are belong to, from 45 kelurahans in Yogyakarta City, 91.11 \% is categorized as “high” and “medium” risk based on 2010-2014 data, 20 (64.52 \%) are once again classified into same risk level in 2015; and among 28 villages which are classified as “medium” risk by 2010-2014 data, exactly 50.0\% is back into same risk category in 2015; and among the 147 villages which are classified as having “low” risk level, 135 or 91.84 \% are once again classified into the same group.

Table 1. Cross tabulation between risk categories of 2010-2014 and 2015 (f, \%)

<table>
<thead>
<tr>
<th>Risk category</th>
<th>2015</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Med</td>
</tr>
<tr>
<td>2010-2014</td>
<td>20 (64.52)</td>
<td>7 (22.58)</td>
</tr>
<tr>
<td>to</td>
<td>3 (10.71)</td>
<td>14 (50.00)</td>
</tr>
<tr>
<td>2014</td>
<td>6 (4.08)</td>
<td>6 (4.08)</td>
</tr>
<tr>
<td>Total</td>
<td>29 (14.08)</td>
<td>27 (13.11)</td>
</tr>
</tbody>
</table>

Time Series Analysis with ARIMA Modelling

From the analysis, it is found that the best ARIMA models constructed for each of the risk category follow a seasonal ARIMA pattern or SARIMA, i.e. for “high” and “medium” risk areas, the model is \((0, 1, 0)(1, 1, 0)_{12}\); and for “low” risk area, it is \((0, 1, 0)(0, 1, 0)_{12}\). The models are then applied to predict the number of Dengue cases for their corresponding risk areas in 2015 (Figure 4, Figure 5 and Figure 6).

Figure 4. Comparison between observed and predicted number of Dengue cases for 2015 in “high” risk areas with ARIMA model \((0, 1, 0)(1, 1, 0)_{12}\)
In the “high” risk group, up to June 2015, the actual or observed number of cases is close to the upper limit of the prediction rather than to their point estimates. Nonetheless, after that month, it is closer to the point prediction. In contrast, in the “medium” risk group, the observed case number is closer to the point prediction, throughout the year. Meanwhile, in the “low” risk group, for the first two months of 2015, the actual Dengue case number exceeds the upper limits, but for the following five month it is within the predicted range between point prediction and the upper limit.

Discussion

Combined with the spatial analysis, the application of the three temporal indices for the first time by Wen et al. aims to offer a complement for the following limitations: 1) the spatial analysis with mapping and just GIS only gives visual description and, thus, cannot confirm the clusters of the case occurrence or the emerged spatial correlation, 2) the use of spatial statistics can only offers a small portion of static condition of disease incidence and therefore neglects the dynamic of the disease under study by time, 3) the time series analysis approach cannot give clues for identifying the risk areas, spatially, and 4) the incidence rate used in an analysis to compare two or more areas has restrictive interpretation since IRs with same values may have different disease’s transmission dynamics.

The use of those temporal indices, collectively, can reveal how frequent, how persistent and how intense the Dengue problem in a certain area. Combining the indices with spatial analysis can discern whether among the areas with similar phenomenon will also have spatial association or not.

When introduced for the first time, the indices were implemented in epidemic situation. However, Dom et al. broaden the application by calculating them in a series, i.e. between 2006 and 2009, and compared the results, a technique which is more or less similar to that of this study. After that, Rasidi et al., not only make calculation of the indices annually in a sequence from 2003 to 2009, but also average the values. Now, this study is a further development of the implementation of the indices.

Theoretically, the result of each index calculation combination will give eight area categories that have different characteristic, depending on their formed indices. For example, an area with “high” frequency and duration indices, but “not high” intensity index, is expected to have conducted adult mosquitoes controlling. Nonetheless, some of the vector insects have become resistance and therefore disease incidence frequently occurs there for a long duration in a specific time period, even though not in a big number. Another example is: an area that only has “high” duration index is probably due to silent transmission.

However, the most significant category to take into account is still areas with all “high” index values, since they are the most severe epidemic areas where the transmission source might be originated. Therefore, this study considers villages which belong to this category as areas with “high” risk for Dengue transmission.

In 2010-2014 period, the highest proportion (71.36%) or 147 villages are classified as “low” risk category, and the majority of the remaining 59 villages falls to “high” risk. A similar pattern is also found for 2015. Based on the outcome map, it is visible that most of the “high” and “medium” risk areas are villages situated in Yogyakarta Municipality and those in Bantul Regency and Sleman Regency which are adjacent or close to Yogyakarta Municipality, mostly at Southern and Western/Southwestern parts. Nevertheless, in Yogyakarta Municipality itself, a few villages which are classified as cold-spot area can also be found.

Despite the dissimilarities of index values owned by each village every year in 2010-2014 period, the result of cross tabulation for risk level category yielded by 2010-
2014 data and those from 2015, give Kappa coefficient of 0.593 (p-value < 0.001). It give sufficient (and close to strong) evidence, that the classification method used in this study to determine the Dengue risk level for each village can be applied and utilized to complement the stratification method which is usually used to classify the level of Dengue endemcity of an area.

In its implementation, this new classification method needs no additional data, except those which are collected by health offices and community health centers, i.e. the time and place when and where diseases occur. However, to process the data, additional effort is needed to provide the information related to epidemic week and epidemic wave as components for the index calculation. In fact, it is not difficult to learn because basically it only uses simple formula that can be run by main spreadsheet software.

In the existing Dengue surveillance data processing and analysis, cases from previous years are used and considered for forecasting the transmission season. Those kinds of data furthermore can be better utilized to forecast the upcoming number of cases by the advantage of time series analysis.

Compared to other time series methods, the approach using ARIMA models is suitable for Dengue because this disease usually has high number of cases that will lead to better predictability. Other than that, ARIMA model is good as well for data that have autocorrelation feature as Dengue data has, and more appropriate for data that have linear correlation and stationary. Its statistical properties is also easy to understand and the model can be developed using familiar data processing software.

The technique to construct and separate ARIMA model for sub-level areas from their main areas and then make comparison is also used by Sitepu et.al. in Jakarta and Silawan et.al. in Thailand. Their studies also find that the best ARIMA model for each sub-area has different characteristics, one to another. However, this study differs from those two studies in that the previous studies use administrative areas rather than risk-level areas.

The prediction for 2015 based on those ARIMA models show that in “high” risk areas, the factual number of cases is closer to the upper limit of the estimation; meanwhile in “medium” risk areas, the observed number of cases is closer to its point estimation; and in “low” risk areas the prediction is missed for the first two months and becomes better in the following months.

However, the three models have similarities in that their predictions reach its ultimate only up to June and July, and after that the predictability decreases. This finding confirms the statement that the longer ARIMA model has to predict, the lower the predictability will be obtained.

Notation of the ARIMA models generated for “high” and “medium” risk areas explain that the number of cases in a certain month can predicted linearly based on the number of cases in the same month at previous years; and for “low” risk areas’ ARIMA model, the interpretation is the number of cases in a certain month can be predicted linearly by following a 12-month seasonal pattern.

In correlation with the predictability for Dengue case number as described above, Runge-ranzinger et.al. argue that routine Dengue surveillance with population basis, not only important for monitoring the trend of disease transmission, it can also provide baseline data for epidemic warning and preparedness.

Early Warning System (EWS) is a system that integrates: the analysis about risk, monitoring and prediction of a particular area and also the intensity of the hazard that will be faced; dissemination of information about alertness to important stakeholders and communities who are vulnerable to the danger; and provide adequate response for the danger itself. The main goal of this system is to collect information for punctual decision-making process in order to appropriately implement the strategical intervention for specific population. There are two objectives for the development of the EWS of infectious diseases, i.e. firstly, to identify whether or not an epidemic will take place, and secondly, to forecast the number of cases that will occur along with the outbreak.

CONCLUSION

Risk level categorization based on 2010-2014 Dengue data has concordance with that of 2015 Dengue data in Kartaman tul. The best ARIMA models for each Dengue risk category follow a seasonal pattern. Multi analysis approach used in this study can be applied to complement the analysis techniques and monitoring which are usually used in Dengue surveillance, because it can sharpen the preparedness of early alert for Dengue spread on the community. Therefore, it is suggested for the health offices of Yogyakarta Municipality, Sleman Regency and Bantul Regency to adopt this analysis to support the endemicity area stratification activities and the forecasting of transmission period, so that the stakeholders of Dengue controlling and monitoring program can focus more on high risk areas in order to be more effective. To gain more advantages, since it is known that Dengue spread has strong association with people mobility between adjacent areas,
the implementation of these analyses can be jointly conducted by cross administrative areas, and for that reason, uniformity in terms of data management and processing, analysis technique and software, need to be prepared.

ACKNOWLEDGMENT

The authors would like to sincerely thank the Health Offices of Yogyakarta City, Sleman Regency and Bantul Regency, for the permission to obtain the Dengue data for the analysis of this study. The Ministry of Health of Indonesia and the Polytechnic of Health of Yogyakarta are also duly acknowledged for the research funding and opportunity to conduct this study.

REFERENCE


