International Journal of Advances in Computer Science & Its Applications
Volume 6 : Issue 3 [ISSN 2250-3765]
Publication Date : 30 December. 2016

Analysis of Performance Prediction Models in Predicting Dengue Fever Patients Number in Each Group of Malang, Indonesia

Wiwik Anggraeni, Febriliyan Samopa, Edwin Riksakomara, Radityo P.W., Eka Mulya A. Lulus Condro T., Pujiadi

Abstract - **Dengue Fever is one of acute and deadly diseases that commonly happens in tropical area. The spread of it is also influenced by geographical condition. Indonesia, Particularly in Malang that is a tropical area with a geographical condition supports the development of this disease. It needs a fat-moving action to the early step precaution so that the number of patients can be reduced. As the primary decision for an early prevention, it needs predictions about several cases of dengue fever of some period in the future. The result of this prediction is needed by Public Health Office of Malang as one of instances that responsible of dengue fever cases.**

This research analyses performance as a prediction model in getting the predictions in a number of dengue fever cases in Malang, Indonesia for some different group of data. The models suggested are Multiplicative Holt-Winters, Additive Holt-Winters, Multiplicative Decomposition and Autoregressive Moving Average (ARIMA). Those models are applied in special data to some cases in Malang which is categorized in 3 groups, namely Lowlands (Malang Rendah), Mediumlands (Malang Sedang), Highlands (Malang Tinggi). The result shows that Multiplicative Holt-Winters model is the best model for lowlands, and mediumlands. Meanwhile, for highlands is best-used with Multiplicative Decomposition

Keywords - **Performance, Model Predictions, Holt-Winters, Decomposition, ARIMA, Early Detection, Dengue**

I. Introduction

The dengue Fever is one of the acute and deadly diseases. It is spread to humans from mosquitos named Aedes Aegypti and Aedes Albopictus [1][2]. The dengue fever is the fastest spreading disease in the world and predicted that every year, there are 50 million infections [2][3].

The dengue fever develops in tropical area [6]. It consists of 2.5 billion people living in an endemic area [2][3][4]. Indonesia, as one of the tropical country surely becomes an endemic area for this disease. East Java, mainly in Malang has a supportive climate to the development of Aedes mosquito. Moreover, the society keeps water in bathtub in couple of days. Water storage in household can be a place for mosquitos to breed [5].

W Anggraeni, F Samopa, E Riksakomara, Radityo P.W., Eka Mulya A.
Department of Information Systems, Institut Teknologi Sepuluh Nopember Surabaya (ITS)
Indonesia

Lulus Condro T., Pujiadi
Department of Public Health Malang Regency
Indonesia

The East Java Public Health Office records that in 2015, there are 19.942 cases of dengue fever with 277 numbers of dead victims. So the level of Case fatality Rate (CFR) reaches to 1,4 percent. On the other hand, for January 2016 to January 18 reported that dengue fever cases in East Java are 213 and 7 people died [6].

During January 2016, Malang as one of districts in East Java decided as one of five districts as endemic area of dengue fever. In 2015, Malang is also one of 27 districts in East Java that has KLB dengue fever status [6]. Over all, during January to February, cases of dengue fever in Malang attacked 517 people. Meanwhile, in 2015 in the same period, the number of dengue fever patients are 654 [7]. For positive dengue fever patients known as 184 people and 4 of them are dead. Last year, the number of positive dengue fever are 205 and 7 of them are dead. The endemic dengue fever has happened for years [7].

The number of patients followed by the significance number of death, one of them is caused by the late handling [8]. So, most of the patients are brought to the hospital in an unusual condition. For reducing that, The Public Health Office is responsible to take preventive actions. Benefits of preventive actions will feel if it is done in the right time and place. That is why, it needs predictions of the number of dengue fever patients in the future. An early warning system of predictive modelling can be effective tools to determine preparation step and control [9][10].

Because Malang has varied height, therefore the discovery of predictive models is done by dividing Malang into some groups. It is important because predictive models of the disease can be different in a different location. Other than that, predictive models are not easily used for predicting diseases in other locations [11].

So, this research is made in some predictive models that can be used for predicting number of patients of dengue fever and testing how those models work in predicting the number of dengue fever in district of Malang for lowlands, mediumlands, and highlands group.

II. Literature Review

A. *Related Work*

Models development for predicting a disease has been done by Thailand in 2012 [12]. The model that has been used is Autoregressive Integrated Moving Average (ARIMA) by the using of monthly data from January 1981 to December 2006 and validated with January 2007 data to April 2010. The result shows that the best model used is ARIMA (3, 1, 4) based on small MAPE value.

SEEK DIGITAL LIBRARY

Proc. of The Fifth Intl. Conf. On Advances In Computing, Control And Networking - ACCN 2016
Copyright © Institute of Research Engineers and Doctors, USA .All rights reserved.
ISBN: 978-1-63248-104-7 doi: 10.15224/ 978-1-63248-104-7-16

In 2015, Dung Pung et al also found predictive models incidents of diseases in Can To City, Vietnam [13]. This research used prediction period 3, 6, 9 and 12 months. It is decided that for that periods, it is found different models.

Another work is done by Siriyasatien et al in 2016 [14]. Tis research says that the power of predictive models can be seen from dome factors such as Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC) and Mean Absolute Percentage Error (MAPE). The best model is the one who has the lowest AIC, BIC and smallest MAPE score. In 2015, Malaysia has ever done modelling and predicting disease incidents [15]. Even if SARIMA used, however in this research it didn't find a seasonal significance patterns. Next, there is also another disease prediction incident in Taiwan [16].

ARIMA and SARIMA are also mostly used in predicting other subjects such as energy consumption predictions in China [17], forecasting heat demand [18], analysis of aerosol optical [19], and electricity demand [20].

Besides ARIMA, in some subjects it often uses Exponential smoothing and decomposition of predicting. For example, global horizontal prediction [21], energy consumption prediction [22], forecasting for interval-valued time series [23], inventory [24], and pig price forecasting [25].

From the best model that has been used in previous research, it is not always suitable for predicting patients of dengue fever in Malang, Indonesia [11]. It is because Malang has a different type with the previous research. The focus of this research is the making of some predictive models and testing those performance models in predicting patients of dengue fever in the group of low, average and high Malang area.

III. Research Methods

This research is done to numbers of dengue fever patients in Malang from 2009-2014. It's monthly data. The division data of training and testing is 70:30. Data is grouped into 3, Malang low, medium and high group. The method used is Multiplicative Holt-Winter's, Additive Holt-Winter's, Multiplicative decomposition and ARIMA. Steps used in this research is explained in details is every chapter of methodology

- **Holt-Winter's Additive**

Holt-Winter's method based on 3 smoothing elements which are stasioner, trend and seasonal. Here are the steps used in Holt-Winter's Additive [26][27].

1. Get the (Initial value) for ℓ0 level, development rate b0 and seasonal factor sn-3, sn-2, sn-1, and sn0.
2. Get the initial value for seasonal.
3. Count prediction value of training data by using regression equation least square.
4. Detrend data by reducing actual data value with prediction value that is obtained from previous step every period.
5. Count seasonal average value data for every seasonal period.
6. Count average seasonal factor every period. The average is 0 (zero).
7. Count predicting value using initial value.
8. Do value update ℓT, bT, and snT using different values.
9. Find combinationα, γ, dan δ that can minimize SSE or MSE.
10. Count predicting value for some periods in the future by using combination of optimumα, γ, and δ.

- **Holt-Winter's Multiplicative**

Steps that are used by this research are [26][27]:

1. Get the (Initial value) for level (ℓ), development rate (b), and seasonal factor (s).
2. Get the initial value for seasonal by using regression equation least square.
3. Count predicting and training value.
4. Detrend data by dividing actual data value with predicting value that is gotten in previous step in every period.
5. Count seasonal average value data for every seasonal period
6. Multiply seasonal average value with one constant value so, seasonal average factor is equal to 1.
7. Count predicting value.
8. Do value update ℓT, bT, and snT using different values.
9. Find combinationα, γ, dan δ that can minimize SSE or MSE.
10. Count predicting value for some periods in the future by using combination of optimumα, γ, and δ.

- **Decomposition Multiplicative**

Decomposition method consists of elements which are trend, seasonal and others (cyclical and other random influences). Steps that are used are [26][27]:

1. Seasonally adjust the data. It can be done as follows:
    - Count average value per season.
    - Count seasonal index value for 1 season.
    - Count seasonal index value for twice seasonal.
    - Doing adjust data by dividing actual value with seasonal index value for twice seasonal.
2. Extract trend value. In here, we can use the least adjust data in I time seasonal. Next, we can get linear regression or forecast function for extracting trend value.
3. Count predicting by doing multiple combination between trend and seasonal.

- **Autoregressive Moving Average (ARIMA)**

Steps used in ARIMA consist of [26][27]:

1. Checking data stasionarity and making stationer data.
2. ARIMA model components Identification
3. Stasioner data then identified with its ARIMA model components.
4. Data parameter estimation
5. Model parameter estimation is done by getting coefficient from each component AR and MA.
6. Model Diagnose Test

SEEK DIGITAL LIBRARY

Proc. of The Fifth Intl. Conf. On Advances In Computing, Control And Networking - ACCN 2016
Copyright © Institute of Research Engineers and Doctors, USA .All rights reserved.
ISBN: 978-1-63248-104-7 doi: 10.15224/ 978-1-63248-104-7-16

IV. Result and Discussion

The data from number of dengue fever cases in Malang is generally seasonal. In a half of area, there is a trend and a half of it the trend is so small or it can be said there is no trend anymore. Here is pattern data plot in each group of Malang on Fig. 1.

Figure 1 shows that number of dengue fever cases in lowlands Malang area is seasonal pattern with decreasing trend. It also happens in mediumlands Malang area. It is also seasonal and trend. In highlands Malang area, data shows that it is seasonal and less trend.

From that pattern, finally it is used for training process to get the best model. The next training model will be used for predicting testing data. Comparison plot result between actual data and predicting result data from each group in Malang by using Multiplicative Holt-Winter's method can be seen in Fig. 2.

However, comparison result plot between actual data and predicting data result in every group area of Malang by using Additive Holt-Winter's can be seen on Fig. 3.

Comparison result plot between actual data and predicting data result in every group area of Malang by using Multiplicative Decomposition can be seen in Fig. 4.

Comparison result plot between actual data and predicting result data in each group of area in Malang by using ARIMA method can be seen on Fig. 5.

From all of figures, it is shown that numbers of patients for lowlands group area less than medium and high area. It is based on a fact that average and high area group have more supportive climate of the development of dengue fever vector.

However predicting models used in this research is like on Table 1 and Table 2.

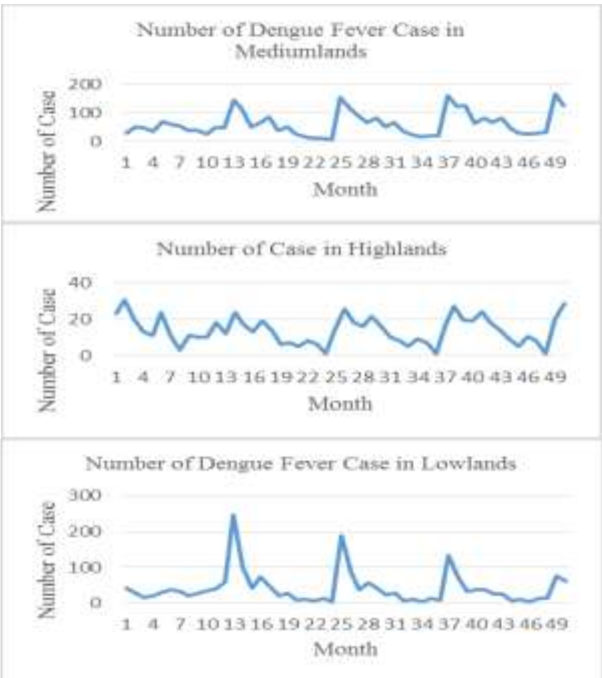


Figure 1. Actual data plot in each group area.

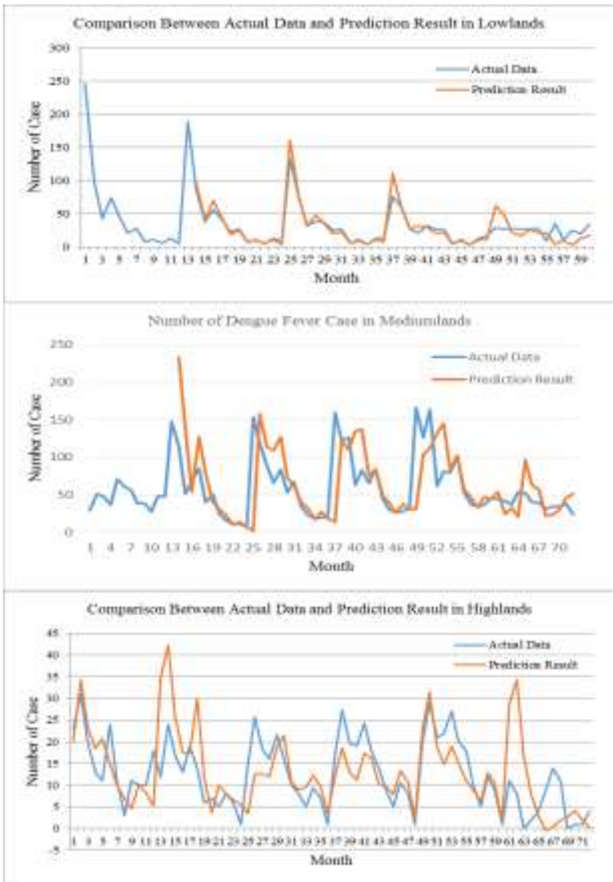


Figure 2. Actual data comparison plot and predicting data in every group of area Multiplicative Holt-Winter's method.

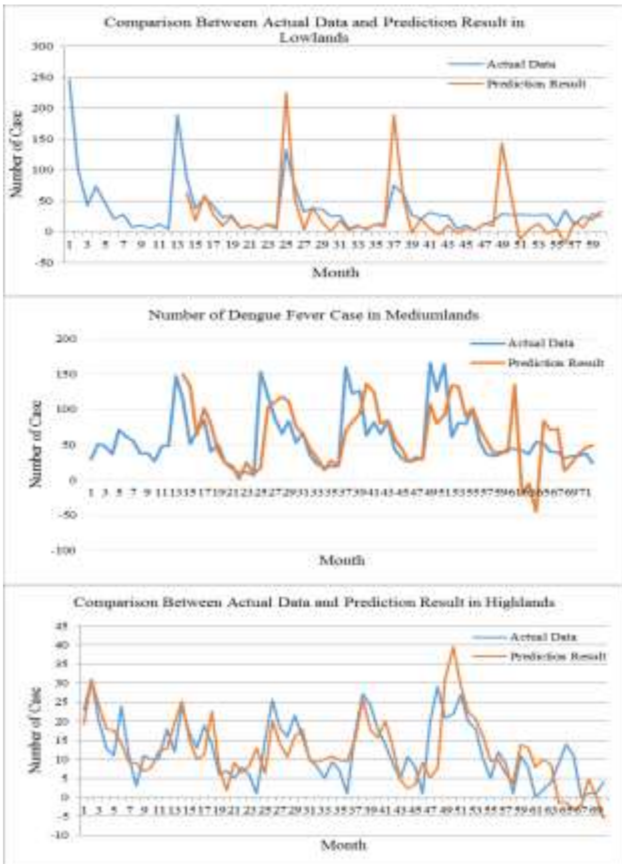


Figure 3. Actual data comparison plot and predicting data in every group area in Malang by using Additive Holt-Winter's.

SEEK DIGITAL LIBRARY

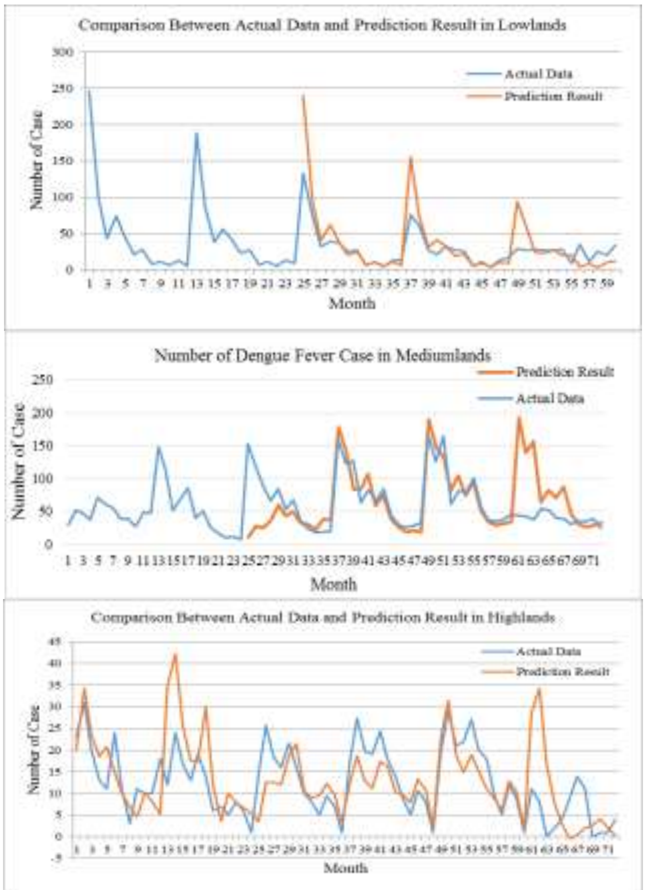
Copyright © Institute of Research Engineers and Doctors, USA .All rights reserved.
ISBN: 978-1-63248-104-7 doi: 10.15224/ 978-1-63248-104-7-16

TABLE 1. PREDICTING MODEL FOR HOLT-WINTERS AND DECOMPOSITION.

Model	ℓ_T, b_t, Sn_T	$\hat{y}_{T+p}(T)$
Multiplicative Decomposition	-	$\hat{y}_{T+1} = \hat{T}_{T+1} S_{T-p+1}$
Multiplicative Holt-Winter's	$\ell_T = \alpha(y_T / sn_{T-L}) + (1-\alpha)(\ell_{T-1} + b_{T-1})$ $b_T = \gamma(\ell_T - \ell_{T-1}) + (1-\gamma) b_{T-1}$ $sn_T = \delta(y_T / \ell_T) + (1-\delta) sn_{T-L}$	$\hat{y}_{T+p}(T) = (\ell_T + p b_T) sn_{T+p-L} \quad (p = 1, 2, 3, \ldots)$
Additive Holt-Winter's	$\ell_T = \alpha(y_T - sn_{T-L}) + (1-\alpha)(\ell_{T-1} + b_{T-1})$ $b_T = \gamma(\ell_T - \ell_{T-1}) + (1-\gamma) b_{T-1}$ $sn_T = \delta(y_T - \ell_T) + (1-\delta) sn_{T-L}$	$\hat{y}_{T+p}(T) = \ell_T + p b_T + sn_{T+p-L} \quad (p = 1, 2, 3, \ldots)$

TABLE 2. PREDICTING MODEL FOR ARIMA

Group	The Best Model	Equation of The Best Model
Lowlands	ARIMA(1,0,5)+c	$Y_t = 3.39919 + 1.45977 + 0.5575\,(Y_{t-1} - Y_{t-2}) + 0.1822\, e_{t-1} - 0.2297\, e_{t-2} + 0.3897\, e_{t-3} + 0.0287\, e_{t-4} + 0.9116\, e_{t-5}$
Mediumlands	ARIMA(1,0,5)+c	$Y_t = 3.85393 + 2.17194 + 0.4364(Y_{t-1} - Y_{t-2}) - 0.1181\, e_{t-1} - 0.0348\, e_{t-2} + 0.275\, e_{t-3} + 0.0175\, e_{t-4} + 0.7797\, e_{t-5}$
Highlands	ARIMA(4,0,3)+c	$Y_t = 3.52458 + 1.62838 + 1.5169(Y_{t-1} - Y_{t-2}) - 1.2155(Y_{t-2} - Y_{t-3}) + 0.8681(Y_{t-3} - Y_{t-4}) - 0.6315(Y_{t-4} - Y_{t-5}) + 1.4776\, e_{t-1} - 0.6355\, e_{t-2} - 0.1393\, e_{t-3}$



Figures 4. Actual data comparison plot and predicting data in every group of area by using Multiplicative Decomposition.

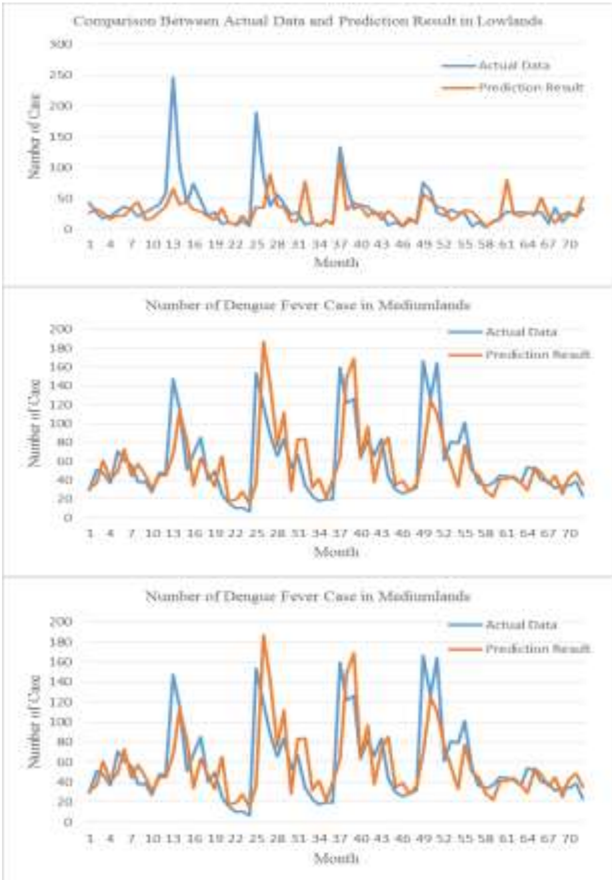


Figure 5. Actual data comparison plot and predicting data in each group of area by using ARIMA method.

SEEK DIGITAL LIBRARY

Copyright © Institute of Research Engineers and Doctors, USA .All rights reserved.
ISBN: 978-1-63248-104-7 doi: 10.15224/ 978-1-63248-104-7-16

Predicting model by using Holt-Winter's involved 3 parameters, namely alpha, beta and gamma which the value can be different. The parameter values can be gotten by doing trial so that the smallest error value is gotten. In this research, parameter value determines automatically by helping of solver. For optimum parameter value for each model can be seen on Table 3.

Performance results of forecasting for each model can be measured from several grades, including Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE), Mean Absolute Deviation (MAD), and so forth. The performance of the proposed model in this study can be seen in Table 4.

Performance in Table 4 is presented in training and testing process. From that table, it is shown that model which are from each method has different performance in each area. It is because the data in every group has different pattern, so every model is not always compatible with a certain data pattern.

v. Conclusion

Each model has a different performance in predicting the number of dengue fever cases in the group of Low Malang, Medium Malang, and High Malang. The performance is strongly influenced by the pattern of the existing data. For the case in a group of low Malang, the model Multiplicative Holt-Winter's works the best performance. While in the case of groups of medium Malang, Multiplicative Holt-Winter's method also works the best. And in the case of the High Malang, models of Multiplicative Decomposition has the most excellent performance. If noted, the accuracy of each model in each region is still relatively large group even though it can be said enough. For future research, this method can be combined with other methods to obtain a smaller degree of accuracy.

TABLE 3. OPTIMUM PARAMETER IN HOLT-WINTER'S MODEL IN EVERY GROUP OF AREA.

Group	Holt-Winter's					
	Multiplicative			Additive		
	Alpha	Beta	Gamma	Alpha	Beta	Gamma
Lowlands	0.013	0.064	1	0.583	0.09	1
Mediumlands	0.788	0.113	0.113	0.714	0.166	0.166
Highlands	0.047	0.016	0.059	0.3	0.2	0.8

References

[1] Hay S.I., Myers M.F., Burke D.S., Vaughn D.W., Endy T., Ananda N., "Ethology of interepidemic periods of mosquito-borne disease", Proc Nat Acad Sci, USA, 2000, pp. 9335-9339.

[2] WHO, "Dengue: Guidelines for diagnosis, treatment, prevention and control", WHO and TDR Publication, 2009.

[3] Farrar J., Focks D., Gubler D., Barrera R., Guzman M.G., Simmons C., Towards a global dengue research agenda, Trop Med Int Health, 2007, pp. 695-699

[4] WHO, 2009, " Dengue Guidelines for Diagnosis, Treatment, Prevention and Control", WHO Press, World Health Organization and the Special Pro- gramme for Research and Training in Tropical Diseases, France, [Online]. http://whqlibdoc.who.int/publications/2009/9789241 547871 engl.pdf [accessed 30 June, 2016]

[5] Nagao Y., Svasti P., Tawatsin A., Thavara U., Geographical structure of dengue transmission and its determinants in Thailand, Epidemiol Infect, 2008, pp. 843-851.

TABLE 4. MODEL TRAINING PERFORMANCE AND MODEL TESTING IN EVERY GROUP OF AREA

Group	Model	Performance					
		MAPE		MSE		MAD	
		Training	Testing	Training	Testing	Training	Testing
Lowlands	Multiplicative Decomposition	33.25%	50.15%	1377.08	363.86	19.54	11.88
	Multiplicative Holt-Winter's	14.92%	19.81%	103.69	178.98	5.49	9.32
	Additive Holt-Winter's	37.71%	86.31%	1005.28	1075.18	0.58	21.46
	ARIMA	58.76%	34.39%	1384.58	60.38	19.05	6.09
Mediumlands	Multiplicative Decomposition	35.41%	66.97%	1602.30	2387.96	25.34	29.24
	Multiplicative Holt-Winter's	19.44%	27.69%	1804.41	794.08	31.31	19.41
	Additive Holt-Winter's	40.49%	66.38%	1559.57	1919.59	25.81	32.49
	ARIMA	40.27%	12.05%	1141.61	249.40	21.07	7.44
Highlands	Multiplicative Decomposition	22.64%	39.85%	11.35	79.21	0.78	3.77
	Multiplicative Holt-Winter's	431.08%	460.83%	32262.53	1099.56	5.28	6.57
	Additive Holt-Winter's	69.76%	180.57%	64.99	74.25	4.94	6.23
	ARIMA	60.22%	114.16%	42.49	25.35	4.90	3.85

SEEK DIGITAL LIBRARY

Copyright © Institute of Research Engineers and Doctors, USA .All rights reserved.
ISBN: 978-1-63248-104-7 doi: 10.15224/ 978-1-63248-104-7-16

[6] "Lima Kabupaten Kota di Jatim Endemik DBD," Januari 2016. [Online]. http://www.malang-post.com/kota-malang/98588-masyarakat-was-was-demam berdarah. [accessed 5 June, 2016].

[7] "Kabupaten Malang Belum Klb DBD", Maret 2016. [Online]. http://radarmalang.co.id/kabupaten-malang-belum-klb-dbd-32834.htm [accessed 2 March, 2016].

[8] "Empat Nyawa Pasien DBD Meninggal Ini Langkah Dinas Kesehatan Kabupaten Malang," Juni 2016. [Online]. http://suryamalang.tribunnews.com/2016/02/26/empat-nyawa-pasien-dbd-meninggal-ini-langkah-dinas-kesehatan-kabupaten-malang. [accessed 16 June, 2016].

[9] Banu S., Hu W., Guo Y., Hurst C., Tong S., Projecting the impact of climate change on dengue transmission in Dhaka, Bangladesh. Environ. Int, 2014, pp. 137–142.

[10] Colon-Gonzales F.J., Fezzi, C, Lake I.R., Hunter P.R., The effects of weather and climate change on dengue, PLoS Negl. Trop. Dis. 7, e2503, 2013.

[11] Racloz V., Ramsey R., Tong S., Hu W., Surveillance of dengue fever virus: a review of epidemiological models and early warning systems, PLoS Negl. Trop. Dis. 6, e1648, 2012

[12] M.J Siriwan Wongkoon, Development of temporal modeling for prediction of dengue infection in Northeastern Thailand, Asian Pacific Journal of Tropical Medicine, 2012.

[13] Dung Pung, Cunrui Huang, Shannon Rutherford, Cordia Chu, Xiaoming Wabg, Minh Nguyen, Nga Huy Nguyen, Cuong Do Manh, Identification of The Prediction Model for Dengue Incidence in Cao Tho City, a Mekong Delta Area in Vietnam, Acta Tropika, 2015.

[14] Padet Siriyasatien, Atchara Phumee, Phatsavee Ongruk, Katechan Jampachaisri, and Kraisak Kresorn, Analysis of Significant Factors for Dengue Fever Incidence Prediction, BMC Bioinformatics, Apr 2016.

[15] B. Gill, J. Xiao, G. Yun, S. Carter, A. Imrie and G. Shellam, Modelling and Forecasting Dengue Fever Incidence in Malaysia By Interpolating Environmental Data, Health Geographics, 2015.

[16] T.-H. H. a. J.-S. H. Ta-Chien Chan, Daily forecast of dengue fever incidents for urban villages in a city, International Journal of Health Geographics, 2015

[17] Chaoqing Yuan, Sifeng Liu, Zhigeng Fang, Comparison of China's Primary Energy Consumption Forecasting by Using ARIMA (The Autoregressive Integrated Moving Average) Model and GM(1,1) Model, Energy Journal, 2016.

[18] Tingting Fang, Risto Lahdelma, Evaluation of a Multiple Linear Regression Model and SARIMA Model in Forecasting Heat Demand For District Heating System, Applied Energy Journal, 2016

[19] Kanika Taneja, Shamshad Ahmad, Kafeel Ahmad, S.D. Attri, Time Series Analysis of Aerosol Optical Depth Over New Delhi Using Box-Jenkins ARIMA Modelling Approach, Atmospheric Pollution Research Journal, 2016

[20] Yuanyuan Wang, Jianzhou Wang, Ge Zhao, Yao Dong, Application of Residual Modification Approach in Seasonal ARIMA for Electricity Demand Forecasting: A Case Study of China, Energy Policy Journal, 2012

[21] Dazhi Yang, Vishal Sharma, Zhen Ye, Lihong Idris Lim, Lu Zhao, "Forecasting of Global Horizontal Irradiance by Exponential Smoothing, Using Decompositions, Energy Journal, 2015

[22] P.M. Macaira, R.C. Souza, F.L. Cyrino Oliveira, Modelling and Forecasting The Residential Electricity Consumption in Brazil with Pegels Exponential Smooting Techniques, Procedia Computer Science, 2016

[23] Andre Luis Santiago Maia, Francisco de A.T. de Carvalho, Holt's Exponential Smoothing and Neural Network Models For Forecasting Interval-Valued Time Series, International Journal of Forecasting, 2016

[24] Giacomo Sbrana, Andrea Silvestrini, Random Switching Exponential Smoothing and Inventory Forecasting, International Journal Production Economics, 2016

[25] Lifeng Wu, Sifeng Liu, Yingjie Yang, Grey Double Exponential Smoothing Model and Its Application on Pig Price Forecasting in China, Applied Soft Computing, 2016

[26] Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman, Forecasting: Methods and Applications, 3rd ed.: John Wiley & Sons, 1998.

[27] O'connell Bowerman and Koehler , Forecasting, Time Series, and Regression, 4th ed. United States of America: Curt Hinrichs, 2005.

SEEK DIGITAL LIBRARY