RAINFALL FORECASTING WITH TIME SERIES MODEL IN ALOR SETAR, KEDAH

FAIQAH MOHAMAD FUDZI*, ZAHAYU MD YUSOF AND MASNITA MISIRAN

School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia[a]
Institute of Strategic Industrial Decision Modelling, School of Quantitative Sciences, Universiti Utara Malaysia, 06010 UUM Sintok Kedah, Malaysia[b]
Centre for Testing, Measurement and Appraisal, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia[c]

* Corresponding author: faiqahfudzi@gmail.com

Abstract: The prediction of rainfall on monthly and seasonal time scales is not only scientifically challenging but is also important for planning and devising agricultural strategies. In this paper, the study is conducted to examine the pattern of monthly rainfall in Alor Setar, Kedah within ten years which is from 2008 to 2018. This paper considered a model based on real data that obtained from Department of Meteorology Malaysia. This study indicates that the monthly rainfall in Alor Setar has a seasonal and trend pattern based on y_t vs t plotting, autocorrelation function and Kruskal Wallis Test for seasonality. The examined rainfall time-series modelling approaches include Naïve Model, Decomposition Method, Holt-Winter's and Box-Jenkins ARIMA. Multiplicative Decomposition Method was identified as the best model to forecast rainfall for the year of 2019 by analysing the previous ten-year's data (2008-2018). As a result from the forecast of 2019, October is the wettest month with highest forecasted rainfall of 276.15mm while the driest month is in February with lowest forecasted rainfall of 50.55mm. The model is therefore adequate and appropriate to forecast future monthly rainfall values in the catchment which can help farmers to plan their farming activities ahead of time.

Keywords: rainfall, Naïve model, decomposition method, Holt-Winter's, ARIMA

Introduction

Rain water, also called precipitation is a natural occurrence of nature. It is very integral part of sustenance for living beings on earth including humans, animals, and plants. Rain is formed through the process of condensation of water vapor, which then rises in the atmosphere. When cloud becomes saturated and too heavy to stay up in the air, it will then fall under gravity in form of droplets of water. In Malaysia, the threats of unpredictable rainfalls can be in a form of flood disaster (Azis *et al.,* 2019), agricultural activities such as large-scale crop planting (Oettli, behera & Yamagata, 2018; Alam *et al,* 2019; Lai, 2019) and shrimp farming (Bir *et al.,* 2017; Aziz *et al.*, 2018), and health complication (Ahmad *et al.,* 2018; Tokachil & Yusoff, 2018; Alhoot *et al.,* 2019), among others.

Among sector that heavily dependent in rainfall as a source of water is the agricultural production. In northern Malaysia in particular, this scenario is more prominent as this region is the main producer of rice in the country (Yusof *et al.,* 2019). Nonetheless, too little, and too much rainfall can be harmful in paddy plantation (Gumel *et al.,* 2017; Agnese & Othman, 2019; Herath *et al.,* 2019). The most optimized is the regular rainfall pattern that can produce optimum rice production (Ahmad *et al.,* 2017).

In the year 2017 for example, a rainfall crisis had occurred in Alor Setar, Kedah where water levels at major rivers rose dramatically, and compounded by the high-tide phenomena. These unforeseen events resulted in heavy flood caused by unusually heavy downpour which lasted for

several days. The Drainage and Irrigation Department (DID) and Meteorological Department also reported that this event caused major rivers to overflow which severely affected paddy fields being flooded by sea water. Hence, the need to understand more thoroughly the rainfall characteristics in this region as it spans large paddy granary areas. By better forecasting the volume of water from the rainfall, we are being equipped to prepare the countermeasures to overcome future disaster.

Many of research groups have carried out the analysis and forecasting of the rainfall based on different approaches. In this article, we make a comparison several methods of forecasting monthly rainfall in Alor Setar, Kedah and all the methods are Naïve Model, Decomposition Method, Holt-Winter's and Box-Jenkins ARIMA. Based on Hyndman and Athanpolous (2016), Naïve Model is some simple methods of forecasting and considered as benchmarks for forecast model comparison. Decomposition Method consists of two types in this study which is Additive Decomposition and Multiplicative Decomposition. Decomposition method is one of the oldest forecasting methods and it usually will be used for short-term forecasting and easy to understand. This method can be applied to any kind of pattern. Tirkes e*t al.* (2017) examined on a comparison between the holt-winters, trend analysis and decomposition models result that Holt-Winter's and Decomposition Method were both successful in a trendy-seasonal-cyclic data.

Holt-Winter's method can be classified into Additive and Multiplicative effect assumption. Manideep *et al.* (2018) made a prediction using different methods of Holt-Winters algorithm with big data approach and they said Multiplicative Method works better than Additive Method in most of the cases because it widely used for forecasting purposes. Box-Jenkins ARIMA has been widely been over the years to predict the rainfall trend (Mahsin *et al.,* 2012) because it eases the forecasting process by used only single variable time data series while also allow multiple for more complex cases (Bari *et al.,* 2015). Bari *et al.* (2015) with their study of forecasting monthly precipitation in Sylhet City by using ARIMA model was found that the ARIMA(0,0,1) (1,1,1) the most effective to predict future precipitation with a 95% confidence interval.

In this study, we will investigate monthly rainfall's pattern and characteristic in Alor Setar, Kedah, and further make a forecast to the rainfall volume into the future. Once forecasted values is obtained, the result can be used by the State Department of Agriculture Kedah to improve strategic plan in farming activities through incorporation of these inputs. Better policies can also be formulated (Twenefour, *et al.,* 2018).

Methodology

Data Collection

In this research, secondary data has been used for data collection from Department of Meteorology Malaysia. We will consider samples from city of Alor Setar which the dataset consists of monthly rainfall data for 10 years from 2008 to 2018 with a total of 132 readings. Large sample data is taken to determine the accuracy of the performance and pattern of the rainfall.

Phases

Phase 1: Partitioning

The second forecasting approaches are used, where data is divided into modelling and evaluation part, before transferred to the forecasting part. In modelling part, 90% of the dataset was used which is nine consecutive years is considered (year 2008 to 2017), while the remaining 10% was used for the evaluation part (year 2018).

Phase 2: Determine the pattern

In this study, time series analysis will be used and the components in time series need to be identified first before the study can be proceed to the next step.

i.   Plotting *t* versus

ii. Autocorrelation function

iii. Kruskal Wallis test for seasonality

Phase 3: List down possible techniques

Based on the pattern discovered, there are a few techniques which are suitable to use which involve the Naïve Model, Decomposition Method (Additive and Multiplicative Model), Holt-Winter's (Additive and Multiplicative Effect Assumption) and Box-Jenkins ARIMA.

Phase 4 and 5: Modelling and Evaluation

In this phase, calculations for all techniques are performed to obtain the error measurements for modelling and evaluation part.

Phase 6: Choose the best technique

i.   Error measurement within family method

The best method among all the techniques within its family that has been compute must be decided by using the error measurement. The error measurement consists of MAD (Mean Absolute Deviation), MSE (Mean Squared Error), MAPE (Mean Absolute Percentage Error) and MPE (Mean Percentage Error). The error measurement is calculated using the formula as shown in Table 1:

Table 1: Error measurement formula

MAD	MSE	MAPE	MPE
$\dfrac{\sum |e|}{n}$	$\dfrac{\sum e^2}{n}$	$\dfrac{\sum \frac{|e|}{y_t}}{n}$	$\dfrac{\sum \frac{e}{y_t}}{n}$

There will be two type of error measurement, one is for modelling part, and another one will be for evaluation part. As a result, the smallest error and the least error method in the evaluation part will be chosen as the best method to proceed for the next step.

ii. Error Measurement across family method

After error measurement is obtained within family method, all the best method of their family method will be put together to observe the best of the best method.

Phase 7: Forecast

The best method will be used in forecasting the future. The forecast of the data will be done in simple approach.

Results And Discussion

Data Pattern

By using SPSS, plot y_t vs t are used to determine the pattern of the data which y_t represents as actual rainfall while t represent as time. As shown in Figure 1, the time plot shows the stationary pattern exists in the data. Next, this data needs to proceed to autocorrelation function.

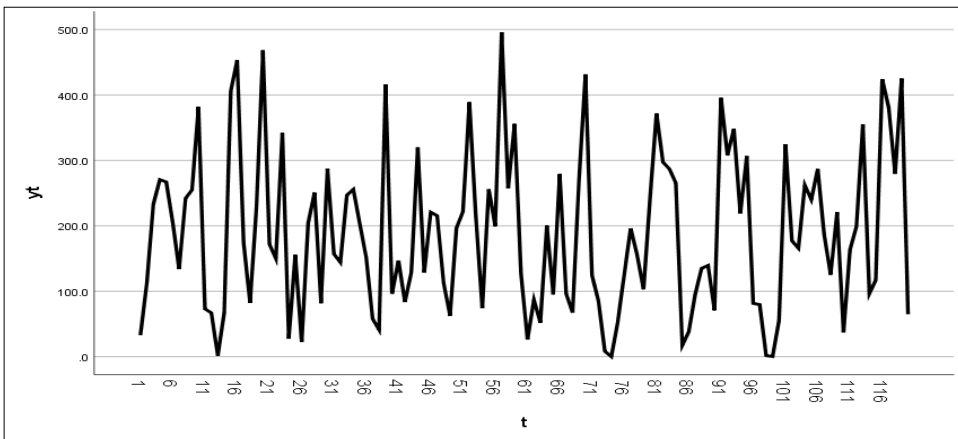Plot Actual Rainfall (y_t) versus time (t)

Figure 1: Line graph of actual rainfall
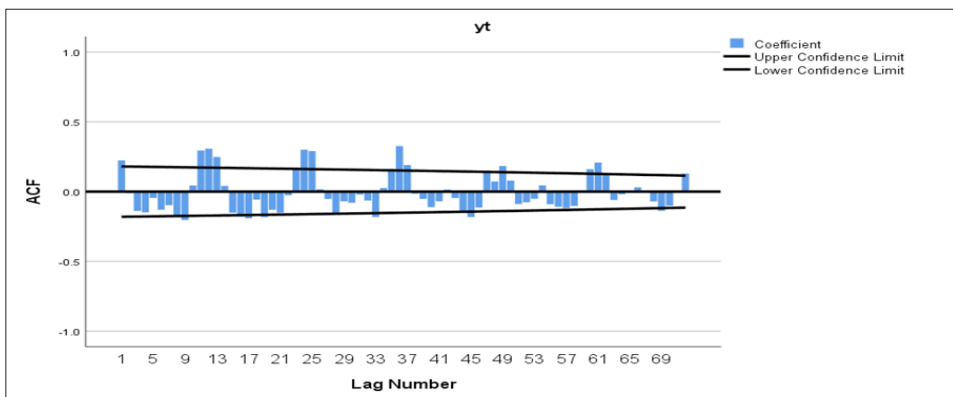
Autocorrelation

Figure 2: Autocorrelation plot

Based on Figure 2, some of the number of lagged is repeated and the pattern is slowly going to zero. It means that there is existing of seasonal and trend component in the data. For the confirmation of seasonal pattern exist, this data needs to proceed to the next step through hypothesis testing.

Kruskal Wallis Test for Seasonality

1. Hypothesis Testing:

H_0 :
$Sn_1 = Sn_2 = Sn_3 = Sn_4 = Sn_5 = Sn_6 = Sn_7 = Sn_8 = Sn_9 = Sn_{10} = Sn_{11} = Sn_{12} = 1$ (*There is no seasons*)

H_1: $Sn_1 \neq 1$ *for some seasons (There is seasonality in the data)*

where, $Sn_t =$ *Seasonal component* ($t = 1,2,3,...,12$)

2. Test Statistics: H = 45.1028429

3. Critical Value: $X^2_{(0.05,11)} = 19.675$

4. Decision: H (45.1028429) > X^2 (19.675), so reject H0.

5. Conclusion: There is seasonality in the data.

From the Kruskal Wallis test we can conclude that there is seasonal pattern exist in this data. As a conclusion the pattern of monthly rainfall in Alor Setar, Kedah from 2008 to 2018 has a seasonal and trend pattern.

Techniques for Seasonal and Trend Pattern

The performance of the models was evaluated based on four error measurement; Mean Absolute Deviation (MAD), Mean Squares Error (MSE), Mean Absolute Deviation (MAD), Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE). The best method can be identified by the lowest evaluation of the error measurement. The result is presented in Table 2 for modelling part and Table 3 for evaluation part.

Table 2: Error measurement for modelling part

Techniques (Modelling)	MAD	MSE	MPE	MAPE
Naive Model	115.6505	19831.42	-4.79956	5.279042
Decomposition Method (Multiplicative)	74.77597	9041.946	-8.21816	8.462945
Holt-Winter (Additive)	93.69879	13013.21	-7.81963	8.115413
Box-Jenkins ARIMA (0,0,1)	97.69	13853.08	-8.81648	9.110534

As shown in Table 2, Multiplicative Decomposition Method has the lowest value of error measurements for MAD (74.77597) and MSE (9041.946), while Naïve Model has the lowest value error measurements for MPE (4.79956) and MAPE (5.279042). As conclusion, it can be said that both Naïve Model and Multiplicative Decomposition Method are representing the best method for modelling part since both have the lowest value of error measurement.

Table 3: Error measurement for evaluation part.

Techniques (Evaluation)	MAD	MSE	MPE	MAPE
Naive Model	108.0222	19166.8	-0.44384	0.842861
Decomposition Method (Multiplicative)	75.16673	9786.889	-0.2628	0.548641
Holt-Winter (Additive)	84.03839	10781.64	-0.45099	0.61865
Box-Jenkins ARIMA (0,0,1)	74.89167	9211.791	-0.42995	0.631517

Based on Table 3, shown that Box-Jenkins ARIMA (0,0,1) recorded the lowest value of error measurement for MAD (74.89167) and MSE (9211.791), while Multiplicative Decomposition Method recorded the lowest error measurement for MPE (0.2628) and MAPE (0.548641). So, it can be concluded that the best method for evaluation part is Box-Jenkin ARIMA(0,0,1) and Multiplicative Decomposition Method since both have the lowest error measurement.

As referred from the Table 2 and 3, every technique is preselected as the best method from all their corresponding family methods due to the smallest error measurements. All these four techniques' errors are then compared among themselves. For the modelling part, Naïve Model and Multiplicative Decomposition Method are represent the best method, while Box-Jenkins ARIMA (0,0,1) and Multiplicative Decomposition Method are representing the best method for evaluation part. As

the result, Multiplicative Decomposition Method is chosen as the best technique between all the methods after the comparison of the evaluation performance in Table 3 with the smallest error measurement of MPE (0.2628) and MAPE (0.548641). In addition, this decision is supported when Multiplicative Decomposition Method is also has the smallest error measurement of MAD (74.77597) and MSE (9041.946) for modelling performance in Table 2. Thus, since Multiplication Decomposition Method has been chosen as the best in both modelling and evaluation part, so it can be used to forecast the future monthly rainfall with the least error among all, as well as to help farmers to plan their farming activities ahead of time in Alor Setar, Kedah.

Forecast

Based on the result, Multiplicative Decomposition Method is the best forecasting model to forecast the future monthly rainfall in 2019. Table 6 showed the result of monthly rainfall in 2019.

Table 4: Forecast

Year	Month	t	y_t
2019	1	133	66.27892
	2	134	50.54992
	3	135	166.9501
	4	136	196.3546
	5	137	183.8462
	6	138	148.4112
	7	139	176.0894
	8	140	236.3197
	9	141	271.3511
	10	142	276.1552
	11	143	225.8478
	12	144	113.6289

Based on Table 4, the monthly rainfall is forecasted for one year ahead which is 2019. As a result, October is the wettest month with highest forecasted rainfall of 276.15mm while the driest month is February with lowest forecasted rainfall of 50.55mm.

Conclusion

In conclusion, this study analyzes seasonal and trend pattern in monthly rainfall in Alor Setar, Kedah. Multiplicative decomposition method was identified as the best model to forecast rainfall for the year of 2019 by analysing ten year's data (2008-2018). As a result, from the forecast of 2019, October is the wettest month with highest forecasted rainfall of 276.15mm while the driest month is in February with lowest forecasted rainfall of 50.55mm. The model is therefore adequate and appropriate to forecast future monthly rainfall values in the catchment which can help farmers to plan their farming activities ahead of time in area Alor Setar, Kedah.

Acknowledgements

References

Agnese, F., & Othman, Z. (2019). Response to climate change impact on paddy farming: the system of rice intensification (SRI). *Journal of Technology and Operations Management, 14*(1), 43-53.

Ahmad, A. A., Yusof, F., Mispan, M. R. & Kamaruddin, H. (2017). Rainfall, Evapotranspiration and Rainfall Deficit Trend in Alor Setar, Malaysia. *Malaysian Journal of Fundamental and Applied Sciences*, 400-404.

Ahmad, R., Suzilah, I., Wan, W. N., Topek, O., Mustafakamal, I., & Lee, H. L. (2018). Factors determining dengue outbreak in Malaysia. *PloS one, 13*(2), e0193326-e0193326.

Alam, M. M., Siwar, C., Molla, R. I., & Toriman, M. E. (2019). Climate change and vulnerability of paddy cultivation in North-West Selangor, Malaysia: a survey of farmers' assessment.

Alhoot, M. A., Nagarajan, R., Abdulmajid, M. S., Alabed, A. A. A., Al-Majeed, L. R. A., Bisri, N. M., & Al-Maleki, A. R. (2019). Effects of global warming on the incidence of infectious diseases in Malaysia. *International Journal of Medical Toxicology & Legal Medicine, 22*(1and2), 111-117.

Azis, T. M. F., Sum, L. W., Adnan, A. A., Sapiri, H., & Misiran, M. (2019) The Use of Correspondence Analysis on the Visualization of Locality and Seasonal Behaviour on the Flood Pattern in Malaysia.

Aziz, N., Misiran, M., Yin, T. S., Yee, T. H., Hui, N. J., Pei, L. Y., ... & Ruddin, A. S. (2018). The Effect of Climate Change to the Farm Shrimp Growth and Production: An Empirical Analysis. *International Journal of Engineering & Technology, 7*(4.1), 138-141.

Bari, S. H., Rahman, M. T., Hussain, M. M., & Ray, S. (2015). Forecasting Monthly Precipitation in Sylhet City Using ARIMA Model. *Civil and Environmental Research, 7*(1), 69-77.

Bir, J., Howlader, P., Ray, S., Sultana, S., Khalil, S. I., & Banu, G. R. (2017). A critical review on white spot syndrome virus (WSSV): A potential threat to shrimp farming in Bangladesh and some Asian countries. *International Journal of Microbiology and Mycology, 6*(1), 39-48.

Drainage and Irrigation Department (DID) and the Meteorological Department. 2017. Report on unusual rainfall, not delay in flood mitigation project which caused flood: Kedah MB.

Gumel, D. Y., Abdullah, A. M., Sood, A. M., Elhadia, R. E., Jamalani, M. A., & Youssefa, K. A. A. B. (2017). Assessing Paddy Rice Yield Sensitivity to Temperature and Rainfall Variability in Peninsular Malaysia Using DSSAT Model. *International Journal of Applied Environmental Sciences, 12*(8), 1521-1545.

Herath, G., Hasanov, A., & Park, J. (2019, August). Impact of Climate Change on Paddy Production in Malaysia: Empirical Analysis at the National and State Level Experience. In International Conference on Management Science and Engineering Management (pp. 656-664). Springer, Cham.

Hyndman, R .J., & Athanasopoulos, G. (2016). Forecasting: Principles and practice, a online text book. Retrieved from https://www.otexts.org/fpp/2/3.

Lai, C. H., Settinayake, A. R. H., Yeo, W. S., Lau, S. W., & Jong, T. K. (2019). Crop Nutrients Review and the Impact of Fertilizer on the Plantation in Malaysia: A Mini Review. *Communications in Soil Science and Plant Analysis, 50*(17), 2089-2105.

Mahsin, M., Akhter, Y. and Begum, M. 2012. Modeling rainfall in Dhaka division of Bangladesh using time series analysis, *Journal of Mathematical Modelling and Application, 1*(5), 67-73

Manideep, K., & Sekar, K. R. (2018). Rainfall Prediction Using Different Methods of Holt Winters Algorithm: A Big Data Approach. *International Journal of Pure and Applied Mathematics, 119*(15), 379-386.

Oettli, P., Behera, S. K., & Yamagata, T. (2018). Climate based predictability of oil palm tree yield in Malaysia. *Scientific reports, 8*(1), 1-13.

Tokachil, N., & Yusoff, N. (2018, June). Effect of rainfall duration on Aedes aegypti population. In *AIP Conference Proceedings, 1974* (1), 020081. AIP Publishing LLC.

Twenefour, F. B. K., Quaicoe, M.T. & Baah, E.M. (2018). Analysis of Rainfall Pattern in The Western Region of Ghana. *Asian Journal of Probability and Statistics, 1*(30), 1-12.

Yusof, Z. M., Misiran, M., Baharin, N. F., Yaacob, M. F., Aziz, N. A. B. A., & Sanan, N. H. B. (2019). Projection of Paddy Production in Kedah Malaysia: A Case Study. *Asian Journal of Advances in Agricultural Research*, 1-6.