Jurnal Mekintek, 15 (1) (2024) pp. 27-33
Published by:IHSA Institute
Mekintek Journal: Journal of Mechanical, Energy, Industry and Technology
Homepage journals:www.ejournal.isha.or.id/index.php/Mekintek

Rainfall Data Modeling in Simalungun Regency Using the Arima Box-Jenkins Method

Desi Fransiska D
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, Indonesia

ARTICLEINFO

Article history:

Received Mar 28, 2024
Revised Apr 15, 2024
Accepted Apr 29, 2024

Keywords:

Forecasting;
ARIMA Box – Jenkins;
Rainfall.

ABSTRACT

One of the components of the environment that determines the success of plant cultivation is climate. To predict rainfall, the author uses the ARIMA Box Jenkins method, which is a quantitative forecasting method. The data used are data for the period July 2012 to June 2017. In this study, the right model is the ARIMA model (2,0,2) with Xt = 4.05668 + 0.9416Xt-1 - 1.0039Xt-2 - 0, 8558et-1 + 0.9617et-2 + et which is used to forecast rainfall for the next 12 periods. The selection is based on the smallest MSE (average error squared) value of 0.033401954 and the smallest RMSE (root mean square error value), which is 0.001115691 and the smallest MAPE (absolute average error percentage) is -0 , 00801773.

Corresponding Author:

Desi Fransiska D,
Department of Mathematics,
North Sumatra University Medan,
Jl. Dr. T. Mansur No.9, Padang Bulan, Medan
E-mail: desifransiska@gmail.com

1. INTRODUCTION

One of the components of the environment that determines the success of plant cultivation is climate(Shaw, 1988). Extreme climates can be bad for the growth and quality of cultivated crops, especially seasonal crops such as food crops(Shaw, 1988). One of the climate indicators is rainfall, rainfall is defined as the amount of water that falls on a flat land surface during a certain period which is measured in units of height (mm) above the horizontal surface if there is no evaporation, runoof, and infiltration. The rainfall is one millimeter (1 mm), meaning that it is in an area of one square meter on a flat place where one millimeter of water is collected(Bertrand, 1965). The amount of rainfall is very important in determining the yield of crop cultivation(Lobell et al., 2009). Increased rainfall in an area creates the potential for flooding(Qin et al., 2013).

Simalungun Regency, as one of the rice producing districts in North Sumatra Province, has a land area of 17.244 kw / ha and has a productivity of 6.5 kw / ha to 6.7 kw / ha. This figure even exceeds the national rice productivity, namely 5.9 kw / ha to 6.0 kw / ha(Chauhan et al., 2006). Rice production in Simalungun had dropped in 2011 due to pest attacks, but in 2012 it rose again(Fadhliani, 2016).

Given the important role of rainfall in determining the planting season to achieve maximum results in order to meet national food needs, it is necessary to predict future rainfall. The Center for Research and Development of the Meteorology, Climatology and Geophysics Agency (BMKG) explained that BMKG forecasts rainfall using the Ensemble Meandan Ensemble Bayesian Model Averaging (EBMA) method. The results of forecasting with this method are validated by the Taylor

ISSN 2086-7026 (Print), 2808-7372 (Online)

diagram to see the goodness of the forecasting results(Anderson, 1996). The results of the Taylor diagram show that the ensemble-me-ensemble BMA technique does not always provide the best accuracy.

ARIMA (Autoregressive Integrated Moving Average) method is a periodic series data modeling method(Nelson, 1998). Forecasting using the ARIMA model assumes that the data used are linearly related(Fattah et al., 2018). The data assumption is linearly related to the ARIMA model characterized by residual values that are normally distributed and white noise. In fact, there are extreme weather conditions in Simalungun. Data modeling that contains extreme values using the ARIMA model causes the residual value to not be normally distributed and white noise(Mahan et al., 2015).

In this study, the Autoregressive Integrated Moving Average (ARIMA) method is used for forecasting rainfall in Simalungun Regency. The use of the ARIMA method is used for a linear approach to rainfall data(Somvanshi et al., 2006). The results of this study are expected to be a more accurate approach to forecasting rainfall in Simalungun Regency.

2. RESEARCH METHOD

2.1 Climate

Climate is a state of the average weather in an area in a certain period. Rainfall is the amount of rain that falls in an area at a certain time(Dore, 2005). To find out the amount of rainfall a rainfall gauge is used, called a Rain Gauge(Wood et al., 2000).

2.2 Data

Information or illustrations about something can be in the form of categories, for example: damaged, good, happy, satisfied, successful, failed and so on, or it can be in the form of numbers(Mumford, 2006). All of these are called data or statistical data completeness.

According to its nature, data can be classified into two, namely:

a. Qualitative data, namely data that is presented in the form of words that contain meaning (not in the form of numbers)(Leech & Onwuegbuzie, 2007).
b. Quantitative data, namely data that is presented in the form of numbers.Based on how to obtain it, data can be divided into two, namely(Firdaus et al., 2020):
    - Primary data, namely data that is collected by individuals or an organization directly from the object under study and for the purposes of the study concerned, which can be in the form of interviews or observations(Swanborn, 2010).
    - Secondary data, namely data obtained / collected and compiled by previous studies data published by other agencies(Leech & Onwuegbuzie, 2008).
    - Secondary data, namely data obtained indirectly from the object under study, usually the data is obtained from third parties either from the object individually (respondent) or from an entity that deliberately discloses facts to a second party so that the second party exploits the facts in question. in mass media or other media, for later the data (facts) are reused by the researcher as a reference in writing.

Based on the collection technique, the data are classified into two, namely:

a. Cross section data, is an event that is collected at a certain time to describe the situation and activities at that time(Beck et al., 1998). For example, research data using a questionnaire.
b. Time series / periodic data is data that is collected from time to time to see the development of an event / activity during that period, for example, developments in circulation(Nerlove et al., 2014).

The data according to the source is divided into two namely:

a. Internal data, namely data that describes the situation in an organization (an agency) which is used for its own purposes(Khan et al., 2014).
b. External data, namely data obtained from outside for the purposes of an agency (institution)(Jubb, 1999).

2.3 Forecasting

Forecasting is a technique for predicting conditions in the future based on both past and present conditions. Based on the period, forecasting is divided into 3 forms:

a. Short Term (Short Term)

Short-term forecasting includes a period of time from days, weeks, to months. Historical data is very relevant in this forecast, because the forecast period is very short(Fan & Chen, 2006). An example of short-term forecasting is forecasting product sales for the next month.

b. Medium Term
Medium-term forecasting covers the period from one season (quarter, quarter, or other) to the next two years. In medium-term forecasting, historical data are still considered relevant(Kurawarwala & Matsuo, 1998). One example of medium-term forecasting is forecasting the production budget.

c. Long Term
Past data are less relevant in long-term forecasting(Ardakani & Ardehali, 2014). This is due to long-term forecasting covering the next two years or more. In stock price forecasting, long-term forecasting usually uses fundamental analysis and intuition.

2.4 The Box – Jenkins Classification Model

The Box – Jenkins model is grouped into three groups, namely:

a. Autoregressive Model (AR)
The Autoregressive (AR) model was first introduced by Yule (1926) and later developed by Walker (1931) The AR (Autoregressive) model in the order p states that observations at time t are linearly related to previous observations of time (t - 1), (t - 2),…, (t - p). The form of the equation function for the AR model in the order p is stated as follows:
The general equation of the ARIMA model (p, 0, 0):

$$X_t = \mu' + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \ldots(1)$$

with:
= VX_te of data in a period t
= Cμ'stant Value
= 1;Φ_1Autoregressive parameter
= Ee_tr value at time t

b. Moving Average (MA) Model
The Moving Average (MA) model was first used by Slutzky (1937). However, it was Wold (1938) who produced the theoretical underpinnings and combined process of ARMA. Wold established an ARMA model that was developed in three directions - efficient identification and assessment procedures (for mixed AR, MA and ARMA processes), expanding these results to include seasonal time series and a simple development that includes non-stationary processes[24]. (non-stationary processes) and is useful for explaining an observation when t is expressed as a linear combination of a number of residuals. The form of the equation function for the MA model in the order q is stated as follows:

$$\text{ARIMA } (0, 0, q):$$
$$X_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \ldots(2)$$

with:
: M$\theta_1, \theta_2 \ldots \theta_q$/erage Parameters
: ee_{t-q} e at time of tq

c. Mixed Model Autoregressive Moving Average (ARMA)
ARMA model is a combined model between AR (Autoregressive) and MA (Moving Average) models which are sometimes written in ARMA (p, q) notation. The form of the ARMA model function in the order p and q is denoted as follows:
ARIMA (1, 0, 1)

$$X_t = \mu' + \Phi_1 X_{t-1} + e_t - \theta_1 e_{t-1} \quad (3)$$

d. Autoregressive Model Integrate Moving Average (ARIMA)

ISSN 2086-7026 (Print), 2808-7372 (Online)

The ARIMA model (p, d, q) was introduced by Box and Jenkins (1976) where p is the operator order of AR, d is the differencing order and q is the operator order of MA. This model is used for stationary time series data after d times of differencing, namely by calculating the difference between observations and previous observations where the form of the equation for the ARIMA model is as follows:

ARIMA (1, 1, 1)

$$(1-B)(1-\Phi_1 B)X_t = \mu' + (1-\theta_1 B) \quad \ldots(4)$$

with:

: F$_{(1-B)}$ erence

: A$_{(1-\Phi_1 B)X_t}$

: M$_{(1-\theta_1 B)e_t}$

3. RESULTS AND DISCUSSIONS

The data analyzed in this study is rainfall data in Simalungun Regency in July 2012-June 2017, can be seen in Table 1 as follows:

Table 1. Rainfall Data in Simalungun Regency

	RAINFALL (MM)				
MONTH	2012 - 2013	2013 - 2014	2014 - 2015	2015 - 2016	2016 - 2017
July	291	133	159	84	204
August	143	235	186	204	197
September	340	221	235	236	248
October	204	427	401	211	312
November	285	392	194	403	220
December	230	560	266	102	145
January	480	57	148	90	103
February	367	119	56	237	183
March	208	115	139	274	198
April	386	309	211	341	324
May	246	347	339	221	265
June	93	115	132	153	130

Source: Statistics Indonesia (BPS) North Sumatra Province

To see the seasonal effect, a seasonal test is carried out according to the theoretical basis as follows:

$$R_y = \frac{\left(\sum_{i=1}^{k} J_i\right)^2}{\sum_{i=1}^{k} n_i}$$

with:

Ry = Sum of squares (JK) for mean

Ji = The number of observed values

ni = trial sample size

obtained:

$$R_y = \frac{(3.273 + 3.030 + 2.466 + 2.556 + 2.529)^2}{60}$$

$$R_y = \frac{(13.854)^2}{60}$$

$$R_y = 3.198.889{,}6$$

3.1 Rainfall Data Analysis

The ARIMA model assumes that the data used is stationary with respect to variance and means, therefore the initial stage of forming the ARIMA model is to check the stationarity of the data against

variance and means. The time series plot of rainfall data for Simalungun Regency is shown in the following figure:

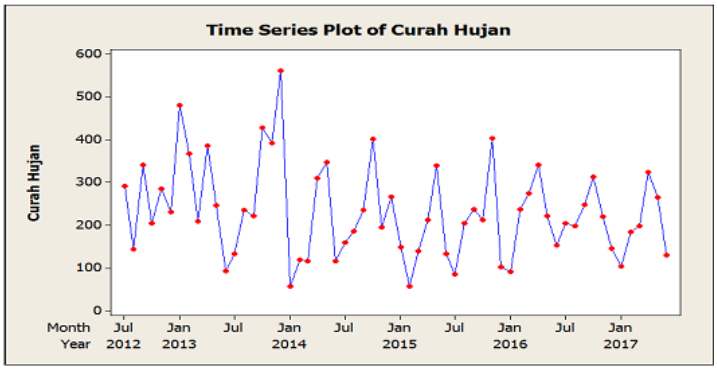

Figure 1. Time series plot of Rainfall Data in Simalungun Regency

Time series plot of rainfall data in Simalungun Regency is not stationary to variance. This can be seen from the volatile data fluctuation. For more details, the box-cox parameter will be estimated.

3.2 Selection of the Best ARIMA Model

The selection of the best model in this study uses the MSE, RMSE and MAPE criteria. The best model is the model with the smallest MSE, RMSE, and MAE values calculated from the out-sample data. From the comparison of the MSE values of the three models, the best model is ARIMA (2, 0.2) because it has the smallest MSE value in the out-sample data. The comparison of the MSE, RMSE, and MAPE values for the outsample data from the three models is described in Table 1

Table 1. Comparison of the Goodness of ARIMA Models

ARIMA Model	MSE	RMSE	MAPE
(2, 0, 2)	0.0334	0.1828	-0.0080
(3, 0, 3)	0.0433	0.2081	-0.0106
(4, 0, 1)	0.0645	0.2540	-0,0210

3.2 Forecasting Using the Box-Jenkins ARIMA Model

The mathematical equations built from the ARIMA (2, 0, 2) model are as follows:

$$X_t = \mu + \Phi_1 X_{t-1} + \Phi_1 X_{t-2} - \theta_1 e_{t-1} - \theta_1 e_{t-2} + e_t$$

Based on the calculation of the formula, the predicted value of the amount of rainfall in Simalungun Regency in the 61st period or in July 2017 is 141 mm. In the same way, the calculation of the forecast value of the amount of rainfall for the period 62 to 72 or until June 2018 is also carried out. The results obtained are shown in Table 2.

Table 2. Values of Rainfall Forecast in Simalungun Regency for Model (2, 0, 2) for the period July 2017 to June 2018

Year	T	Month	Forecast
2017 - 2018	61	July	141
	62	August	216
	63	September	312
	64	October	302
	65	November	200
	66	December	136
	67	January	149
	68	February	236
	69	March	323
	70	April	286
	71	May	182
	72	June	132

ISSN 2086-7026 (Print), 2808-7372 (Online)

4. CONCLUSION

From the description in the previous discussion chapter, the use of the Box-Jenkins method in predicting rainfall in Simalungun Regency can be concluded as follows: a) Rainfall data in Simalungan district from July 2012 to June 2017 which is used to forecast rainfall for the next 12 periods shows the following characteristics: The data plot shows that the data is not stationary, The partial autocorrelation plot shows that the data is stationary with a rapidly decreasing trend towards zero, The time series plot shows that the data has no seasonality, as seen from the graph that does not have any lag period. b) The forecasting model chosen to predict rainfall in Simalungun Regency for the next 12 periods is the ARIMA (2, 0, 2) model which has the following forecasting equation:

$$X_t = \mu + \Phi_1 X_{t-1} + \Phi_2 X_{t-2} - \theta_1 e_{t-1} - \theta_2 e_{t-2} + e_t$$

$$X_t = 4{,}05668 + (0{,}9416) X_{t-1} + (-1{,}0039) X_{t-2} - (0{,}8558) e_{t-1} - (-0{,}9617) e_{t-2} + e_t$$

$$X_t = 4{,}05668 + 0{,}9416\, X_{t-1} - 1{,}0039 X_{t-2} - 0{,}8558 e_{t-1} + 0{,}9617 e_{t-2} + e_t$$

REFERENCES

Anderson, J. L. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, *9*(7), 1518–1530.

Ardakani, F. J., & Ardehali, M. M. (2014). Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types. *Energy*, *65*, 452–461.

Beck, N., Katz, J. N., & Tucker, R. (1998). Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. *American Journal of Political Science*, *42*(4), 1260–1288.

Bertrand, A. R. (1965). Rate of water intake in the field. *Methods of Soil Analysis: Part 1 Physical and Mineralogical Properties, Including Statistics of Measurement and Sampling*, *9*, 197–209.

Chauhan, N. S., Mohapatra, P. K. J., & Pandey, K. P. (2006). Improving energy productivity in paddy production through benchmarking—An application of data envelopment analysis. *Energy Conversion and Management*, *47*(9–10), 1063–1085.

Dore, M. H. I. (2005). Climate change and changes in global precipitation patterns: what do we know? *Environment International*, *31*(8), 1167–1181.

Fadhliani, Z. (2016). *The Impact of Crop Insurance on Indonesian Rice Production*. University of Arkansas.

Fan, S., & Chen, L. (2006). Short-term load forecasting based on an adaptive hybrid method. *IEEE Transactions on Power Systems*, *21*(1), 392–401.

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management*, *10*, 1847979018808673.

Firdaus, A. M., Juniati, D., & Wijayanti, P. (2020). Number pattern generalization process by provincial mathematics olympiad winner students. *Journal for the Education of Gifted Young Scientists*, *8*(3), 991–1003.

Jubb, P. B. (1999). Whistleblowing: A restrictive definition and interpretation. *Journal of Business Ethics*, *21*, 77–94.

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Mahmoud Ali, W. K., Alam, M., Shiraz, M., & Gani, A. (2014). Big data: survey, technologies, opportunities, and challenges. *The Scientific World Journal*, *2014*.

Kurawarwala, A. A., & Matsuo, H. (1998). Product growth models for medium-term forecasting of short life cycle products. *Technological Forecasting and Social Change*, *57*(3), 169–196.

Leech, N. L., & Onwuegbuzie, A. J. (2007). An array of qualitative data analysis tools: A call for data analysis triangulation. *School Psychology Quarterly*, *22*(4), 557.

Leech, N. L., & Onwuegbuzie, A. J. (2008). Qualitative data analysis: A compendium of techniques and a framework for selection for school psychology research and beyond. *School Psychology Quarterly*, *23*(4), 587.

Lobell, D. B., Cassman, K. G., & Field, C. B. (2009). Crop yield gaps: their importance, magnitudes, and causes. *Annual Review of Environment and Resources*, *34*, 179–204.

Mahan, M. Y., Chorn, C. R., & Georgopoulos, A. P. (2015). White Noise Test: detecting autocorrelation and nonstationarities in long time series after ARIMA modeling. *SciPy*, 97–104.

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential.

Information Systems Journal, *16*(4), 317–342.
Nelson, B. K. (1998). Time series analysis using autoregressive integrated moving average (ARIMA) models. *Academic Emergency Medicine*, *5*(7), 739–744.
Nerlove, M., Grether, D. M., & Carvalho, J. L. (2014). *Analysis of economic time series: a synthesis*. Academic Press.
Qin, H., Li, Z., & Fu, G. (2013). The effects of low impact development on urban flooding under different rainfall characteristics. *Journal of Environmental Management*, *129*, 577–585.
Shaw, R. H. (1988). Climate requirement. *Corn and Corn Improvement*, *18*, 609–638.
Somvanshi, V. K., Pandey, O. P., Agrawal, P. K., Kalanker, N. V, Prakash, M. R., & Chand, R. (2006). Modeling and prediction of rainfall using artificial neural network and ARIMA techniques. *J. Ind. Geophys. Union*, *10*(2), 141–151.
Swanborn, P. (2010). Case study research: What, why and how? *Case Study Research*, 1–192.
Wood, S. J., Jones, D. A., & Moore, R. J. (2000). Accuracy of rainfall measurement for scales of hydrological interest. *Hydrology and Earth System Sciences*, *4*(4), 531–543.