

# Forecasting Monthly Rainfall Using Arima Seasonal Method: Case Study: Rainfall in West Lampung Regency

Nusyirwan<sup>1,\*</sup>, Lusmeilia Afriani<sup>2</sup>, Faturrahman<sup>1</sup>

<sup>1</sup>Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Lampung, Bandar Lampung, Indonesia

<sup>2</sup>Department of Civil Engineering, Faculty of Engineering, University of Lampung, Bandar Lampung, Indonesia

**Abstract** Rain comes from rain clouds that condense and fall to the earth as a result of natural phenomena so that the occurrence is so random in terms of time, location and magnitude, as a result the actual value is difficult to estimate. West Lampung Regency has two seasons (Rain and Dry) and also has very fast weather changes due to the geographical location of West Lampung Regency which borders the Indian Ocean which results in low pressure in the Indian Ocean, high rainfall will occur. Rainfall consists of two seasons and occurs repeatedly, so that rainfall is included in the monsoon pattern which is characterized by the type of unimodal rainfall (one peak of the rainy season). If using data on the amount of past rainfall, the appropriate method for modeling the amount of rainfall is the Seasonal Autoregressive Integrated Moving Average (SARIMA) method. This study aims to determine the SARIMA model that is formed on the data on the amount of monthly rainfall. The best SARIMA model for monthly rainfall data in West Lampung Regency is the SARIMA (1,0,0)(1,0,1)<sup>12</sup> model with an AIC value of 887.40. Forecasting results for the next 12 months are dominated by high and very high rainfall.

**Keywords** Total Rainfall, West Lampung Regency, SARIMA

## 1. Introduction

Rain comes from clouds that condense rain\_ and fall to the ground due to natural phenomena so that it occurs so randomly in terms of time, location, and magnitude, as a result, the actual value is difficult to estimate. Rainfall is the amount of water that falls on flat ground during a certain period measured in millimeters (mm) above the horizontal surface. Rain can also be interpreted as the height of rainwater that collects in a flat place, does not evaporate, does not seep and does not flow [1]. Likewise, it is very difficult to prove a series of recordings from rain data, so sometimes to predict or add rain recording data it is necessary to simulate synthetic rain data. Various methods have been developed by researchers in engineering and science to prove this information. The most widely used methods today are the deterministic method and the stochastic method.

The intensity of heavy rain is very dangerous because it has a negative impact on people's life activities such as floods and landslides. This will have a negative impact on various sectors such as agriculture, fisheries, tourism and transportation. In Indonesia there are two seasons, namely the rainy season and the dry season. West Lampung Regency

is a district that has two seasons but also has very fast weather changes, because of the geographical location of West Lampung which borders the Indian Ocean. This causes if there is low pressure in the Indian Ocean then West Lampung will experience high rain. So that West Lampung Regency becomes a district with high rainfall and very wet nature. This makes information on rainfall in West Lampung Regency very important to know.

Monthly rainfall data modeling in West Lampung Regency can be done using the Seasonal Autoregressive Integrated Moving Average (SARIMA) method if the modeling uses univariate time series data and has a seasonal pattern. According to [7], forecasting using univariate time series data which was first introduced by [2] is the Autoregressive Integrated Moving Average (ARIMA) and is still the most popular forecasting model. This model is derived from the Autoregressive (AR) model, the Moving Average (MA) model and the combination of AR and MA, the ARMA model. In the event that there is a seasonal component in the model, this model is referred to as the SARIMA model [7].

Several studies on rainfall, namely time series modeling and forecasting of rainfall at twelve stations in Bogor by [6] obtained the best time series models for 12 stations in Bogor: ARIMA(1,1,1), ARIMA(1,1,0), and ARIMA(3,1,0) as well as rainfall forecasting data with Seasonal Autoregressive Integrated Moving Average (SARIMA) with outlier detection as an effort to optimize agricultural production in Mojokerto Regency by [5] found this method able to

\* Corresponding author:

nusyirwan.1966@fmipa.unila.ac.id (Nusyirwan)

Received: Oct. 31, 2022; Accepted: Nov. 19, 2022; Published: Nov. 24, 2022

Published online at <http://journal.sapub.org/statistics>

accommodate rainfall data rain that has high fluctuations and extreme values. This study aims to obtain the best model and predict the amount of rainfall in West Lampung Regency based on the best model that has been obtained using the Seasonal ARIMA method.

## 2. Materials and Methods

The object of this research is West Lampung Regency. The variable used is the amount of daily rainfall in West Lampung Regency. The stages of this research include data collection activities. In this study, daily rainfall data (mm) for the period 1 January 2017–31 May 2022 was used in West Lampung Regency. The data source comes from the Meteorology, Climatology, and Geophysics Agency (BMKG) Climatology Station Masgar Lampung. Data from BMKG is in the form of daily data which is then converted into monthly data by adding up the daily rainfall data for West Lampung Regency. The second is data exploration, the purpose of data exploration is to provide an overview of monthly rainfall data in West Lampung Regency. Third, modeling using the Box-Jenkins method. Building the SARIMA model, the modeling stages will be carried out, namely testing and handling stationarity, then model identification, model parameter estimation, and model diagnostics. The four best model choices, from several SARIMA models that meet the parameter significance test, will be selected the best model based on the Akaike Information Criterion (AIC) criteria. Finally, forecasting uses the best selected model.

The general form of the SARIMA model  $(p,d,q)(P,D,Q)_s$  is [3]:

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_p(B)\Theta_p(B^s)a_t \quad (1)$$

where:

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$$

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}$$

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

$$\Theta_q(B^s) = 1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_q B^{qs}$$

Where  $\phi$  and  $\Phi$  are the autoregressive (AR) parameters of the non-seasonal and seasonal components, respectively;  $\theta$  and  $\Theta$  is the parameter of the non-seasonal and seasonal component moving average (MA);  $B$  is the Backward operator,  $B(Z_t) = Z_{t-1}$ ,  $(1-B)^d$  is the number of non-seasonal differences and  $(1-B^s)^D$  the number of seasonal differences;  $a_t$  is an independently distributed random variable;  $P$  and  $p$  order of AR components;  $Q$  and  $q$  order MA components;  $D$  and  $d$  are the sum of the differences, and are usually abbreviated as SARIMA  $(p,d,q)(P,D,Q)_s$  [8].

The stages of modeling using the Box-Jenkins method are:

1. Create ACF and PACF plots.
2. Test the stationarity of the data by performing a unit root test using the ADF test.

3. Perform data transformation using the Box-Cox transformation method if the data used is not stationary on the variant. Differentiate the data if the data used is not stationary. Then examine the transformed or differentiated stationary data. This stage is carried out until the data is stationary.
4. Establish a provisional conjecture model. Determine the order of the model by looking at the stationary ACF plot and PACF data.
5. Estimating the parameters then testing the significance of the model parameters by using the t-test or looking at the p-value.
6. The selection of the best model uses several criteria for selecting the best model by looking at the AIC value. The criteria for selecting the best model are carried out if there is more than one model that meets the requirements, namely those that meet the significance of the parameters (step 5).
7. Performing overfitting by increasing the order of  $p$ ,  $q$ ,  $P$ , and  $Q$  from the temporary conjecture model so that several models are obtained while still looking at the models that meet the significance of the parameters.
8. Choose the best model based on the smallest AIC value.
9. Doing forecasting using the best model.
10. Categorize the results of forecasting the amount of rainfall into 3 parts, namely low (0-100 mm), medium (100-300 mm), high (300-500 mm), and very high (>500 mm).

In determining the forecasting model to be used, an evaluation of the model is carried out. How to evaluate the forecasting model that can be done is by using information criteria. Information criteria are criteria for assessing the quality of the model by using a statistical model of a parameter. Akaike Information Criteria (AIC) is defined as:  $AIC(M) = 2k - \log L$ , where:  $k$ =number of parameters;  $L$  = maximum likelihood [8].

## 3. Results and Discussion

The first step in modeling time series data is the stationarity test of the data. The ADF test showed that the data was stationary and found a significance value smaller than the 5% significance level ( $p$ -value  $\leq 0.05$ ) which was 0.01. After that, the ACF and PACF plots were made with the following results:

The ACF plot (a) shows that for non-seasonal significance at lag-1, lag-6 and for seasonality at lag-12. Based on the PACF plot (b) shows that for non-seasonal significant at lag1, lag-6. In determining the initial model, apart from seeing the significant lag, one must also pay attention to the parsimony principle, which is to load as few parameters as possible so that the model is more stable. Thus, the identification of the provisional predictive model for this data is SARIMA  $(1,0,0)(1,0,0)^{12}$ , SARIMA  $(1,0,0)(1,0,1)^{12}$ , SARIMA

$(1,0,0)(0,0,1)^{12}$ , SARIMA  $(1,0,1)(1,0,0)^{12}$ , SARIMA  $(0,0,1)(0,0,1)^{12}$ . Furthermore, to estimate the initial model parameters obtained as follows:

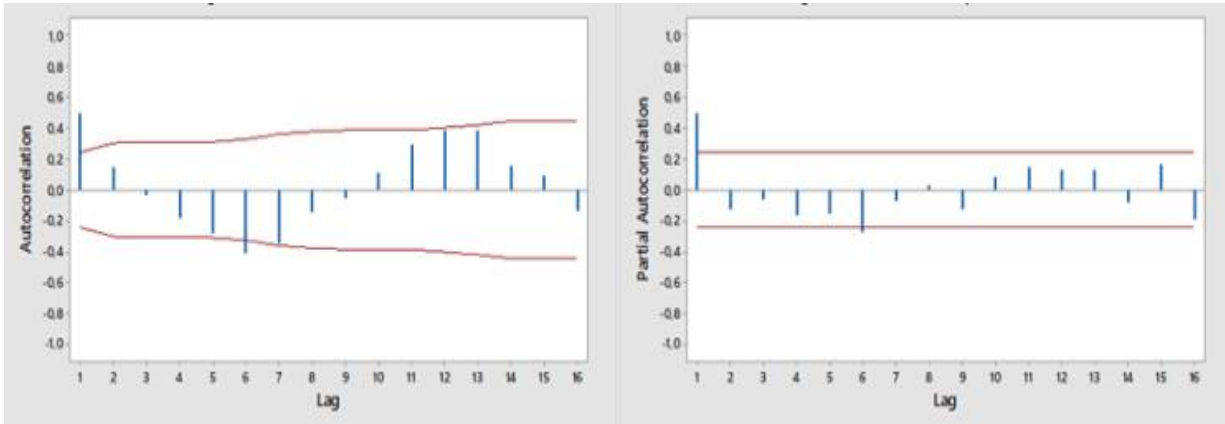


Figure 1. ACF and PACF plots Monthly Rainfall Data in West Lampung Regency

Table 1. Interim Model Parameter Estimation

Model	Parameter	Estimate	p-value	AIC	Information
SARIMA $(1,0,0)(1,0,0)^{12}$	AR	0.7723	<0.001	887.47	Significant Model
	SAR	0.3303	0.015		
SARIMA $(1,0,0)(1,0,1)^{12}$	AR	0.2794	0.025	887.40	Significant Model
	SAR	0.9990	<0.001		
	SMA	0.8611	<0.001		
SARIMA $(1,0,0)(0,0,1)^{12}$	AR	0.8187	<0.001	888.06	Model No Significant
	SMA	-0.2295	0.090		
SARIMA $(1,0,1)(1,0,0)^{12}$	AR	0.9093	<0.001	888.89	Significant Model
	SAR	0.5656	<0.001		
	MA	0.5191	0.001		
SARIMA $(1,0,1)(1,0,1)^{12}$	AR	0.0976	0.828	889.05	Model No Significant
	SAR	0.9991	<0.001		
	MA	-0.1954	0.659		
	SMA	0.8619	<0.001		
SARIMA $(1,0,1)(0,0,1)^{12}$	AR	0.8762	<0.001	889.30	Model No Significant
	MA	0.1997	0.200		
	SMA	-0.2758	0.043		
SARIMA $(0,0,1)(1,0,0)^{12}$	SAR	0.9634	<0.001	887.58	Model No Significant
	MA	-0.2042	0.103		
SARIMA $(0,0,1)(1,0,1)^{12}$	SAR	0.9992	<0.001	887.64	Significant Model
	MA	-0.2906	0.020		
	SMA	0.8625	<0.001		
SARIMA $(0,0,1)(0,0,1)^{12}$	MA	-0.5785	<0.001	888.29	Significant Model
	SMA	-0.6025	<0.001		

Based on Table 1, obtained five (5) models whose all model parameters significant namely the model SARIMA  $(1,0,0)(1,0,0)^{12}$ , SARIMA  $(1,0,0)(1,0,1)^{12}$ , SARIMA  $(1,0,1)(1,0,0)^{12}$ , SARIMA  $(0,0,1)(0,0,1)^{12}$  and SARIMA  $(0,0,1)(0,0,1)^{12}$ .

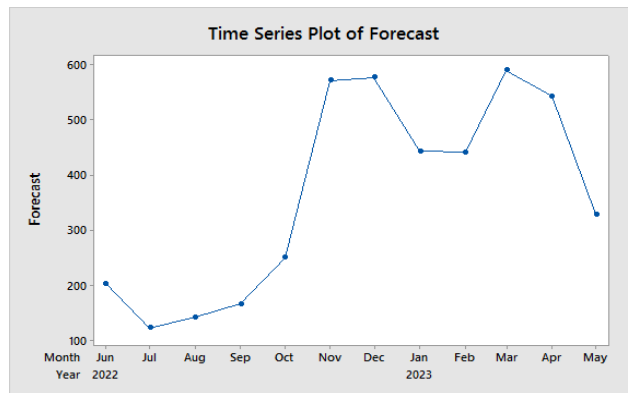
The five models above meet the parameter significance test, then overfitting is carried out, it is hoped that there will

be a final model that meets and is able to model better but in fact there is no model with significant parameters. So that only five models are feasible to use. To obtain the best model of the five models, the value of the AIC criteria can be seen in Table 1. The smallest AIC criterion value is the SARIMA model  $(1,0,0)(1,0,1)^{12}$  with AIC 887.40. Thus, the SARIMA  $(1,0,0)(1,0,1)^{12}$  model is the best model that can

be used for forecasting rainfall. The forecast results are obtained as follows:

**Table 2.** Monthly Rainfall Forecast in West Lampung Regency

Period	Amount of rainfall (mm)	Category
June 2022	203.742	Intermediate
July 2022	123.405	Intermediate
August 2022	142,577	Intermediate
September 2022	166,229	Intermediate
October 2022	250,388	Intermediate
November 2022	570,953	Very high
December 2022	576,410	Very high
January 2023	442,956	Tall
February 2023	440,779	Tall
March 2023	589,505	Very high
April 2023	542,717	Very high
May 2023	329,050	Tall



**Figure 2.** Graph of monthly rainfall forecasting data in West Lampung Regency

## 4. Conclusions

Modeling amount bulk rain monthly West Lampung Regency uses data from the West Lampung BMKG. The best SARIMA models for quantity data bulk rain monthly

in West Lampung Regency is the SARIMA model  $(1,0,0)(1,0,1)^{12}$  with AIC value of 887.40. Forecasting results for the next 12 months are dominated by high and very high rainfall.

## REFERENCES

- [1] Suroso. 2006. Rainfall Analysis to Create Intensity Duration Curves Frequency (IDF) in Flood Prone Areas, Banyumas Regency. *Journal Civil Engineering* Vol. 3.
- [2] Box, G.E. & Jenkins, G.M., 1976. *Time series analysis: Forecasting and control*. California: Holden-Day, San Francisco.
- [3] Geurts, M., Box, G.E.P. & Jenkins, G.M., 2006. *Time Series Analysis: Forecasting and Control*. *Journal of Marketing Research*. 14(2): 269.
- [4] Haryanto, S.A.F., Ernawati, E., & Puspitaningrum, D., 2015. Implementation of Backpropagation Imitation Neural Networks to Predict Weather (Case Study: Bengkulu City). *Recursive: Journal of Informatics*, 3(2).
- [5] Huda, A.M., Choiruddin, A., Budiarto, O., & Sutikno, S., 2012. Forecasting Rainfall Data with Seasonal Autoregressive Integrated Moving Average (SARIMA) with Outlier Detection as an Effort to Optimize Agricultural Production in Mojokerto Regency. In the National Seminar: Food and Energy Sovereignty, Faculty of Agriculture, Trunojoyo University, Madura.
- [6] Maulana, H.A. 2018. Row time modeling and forecasting of rainfall at two dukun substations in Bogor. *Journal of Mathematics, Statistics and Computing*. 15(1): 50-63.
- [7] Suhartono, S. 2011. Time series forecasting using autoregressive seasonal integrated moving averages: Subset, multiplicative or additive models. *J. Mathematics. statistics*. 7: 20-27.
- [8] Tadesse, K. & Dinka, M., 2017. Application of the SARIMA model for monthly flow forecasting in Waterval River, South Africa. *Journal of water and land development*.