BAREKENG: Jurnal Ilmu Matematika dan Terapan
March 2022 Volume 16 Issue 1 Page 137–146
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss1pp137-146

FORECASTING RAINFALL IN PANGKALPINANG CITY USING SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE WITH EXOGENOUS (SARIMAX)

Ririn Amelia[1*], Elyas Kustiawan[2], Ineu Sulistiana[3], Desy Yuliana Dalimunthe[4]

[1,2,3,4] Mathematics Department, Faculty of Technic, Universitas Bangka Belitung
Kampus Terpadu Universitas Bangka Belitung, Balunijuk, Bangka Regency, 33172, Indonesia

Corresponding author e-mail: [1] ririn-amelia@ubb.ac.id*

Abstract. *Changes in extreme rainfall can cause disasters or losses for the wider community, so information about future rainfall is also needed. Rainfall is included in the category of time series data. One of the time series methods that can be used is Autoregressive Integrated Moving Average (ARIMA) or Seasonal ARIMA (SARIMA). However, this model only involves one variable without involving its dependence on other variables. One of the factors that can affect rainfall is wind speed which can affect the formation of convective clouds. In this study, the ARIMA model was expanded by adding eXogen variables and seasonal elements, namely the SARIMAX model (Seasonal ARIMA with eXogenous input). Based on the analysis that has been carried out, the prediction of rainfall in Pangkalpinang City, Bangka Belitung Islands Province can be modeled with the SARIMAX model (0,1,3)(0,1,1){12} for monthly rainfall and SARIMAX (0,1,2)(0,1,3){12} for maximum daily rainfall. When compared with the actual data and previous studies using ARIMAX, the SARIMAX model is still better in the forecasting process when compared to the ARIMAX model. If viewed based on the AIC value of the SARIMA model, the SARIMAX model is also more suitable to be used to predict rainfall in Pangkalpinang City.*

Keywords: *forecasting, rainfall, sarima, sarimax*

Article info:

Submitted: 24th November 2021 *Accepted: 7th February 2022*

How to cite this article:
R. Amelia, E. Kustiawan, I. Sulistiana and D. Y. Dalimunthe, "FORECASTING RAINFALL IN PANGKALPINANG CITY USING SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE WITH EXOGENOUS (SARIMAX)", *BAREKENG: J. Il. Mat. & Ter.*, vol. 16, iss. 1, pp. 137-146, Mar 2022.

1. INTRODUCTION

Based on the results of monitoring and analysis of the Meteorology, Climatology and Geophysics Agency (BMKG) regarding climate prospects in Indonesia in 2020/2021, it shows that as many as 85 percent of season zones in Indonesia have entered the rainy season since the end of December 2020. Monthly rainfall in most areas of Indonesia is predicted to increase more than last year. Some areas also can experience high rainfall categories (300-500mm/month) in the January - April 2021 period and get an increase in rainfall of 40 percent to 80 percent higher than the rainfall in 2020 [1].

Furthermore, the BMKG of the Bangka Belitung Islands Province predicts a relatively high intensity of rain to occur in the Bangka Belitung area from March to April 2021. The rainfall is also predicted to exceed 350-400 mm. In 2018, the rainfall rate could reach 620 mm per month. There is a possibility of extreme weather conditions at the predicted rainfall figures in 2021 [2] .

On January 13, 2021, the Bangka Belitung Islands Province was determined by the BMKG to be in a state of alert. The determination of this alert condition after the occurrence of floods, tidal floods, and puddles of water in several areas in Bangka Belitung. Due to the rain falling with intensity and duration, it is pretty long. This condition is made worse by the high tide. As a result, there were tidal floods in several areas, including Pangkalpinang City. In the Pangkalpinang City area, tidal flooding occurred on Pasir Padi Beach, several parts of the Opas Indah Urban Village, Taman Sari Sub-District, and Selindung Bridge. Meanwhile, flooding occurred in Rejosari Urban Village, Pangkal Balam Sub-district, National Building Urban Village, Taman Sari Sub-district, Pasir Putih Urban Village, Bukit Intan Sub-District [3].

Many parties have realized that extreme changes in rainfall can cause disasters and losses for the wider community. For this reason, information about future rainfall is needed that can be used to estimate the worst event that can occur so that anticipatory steps can be taken. Information relating to conditions in the future cannot be determined with certainty but can be predicted.

Forecasting is predicting, projecting, or making an estimate/estimation of various possibilities that will occur. Forecasting is the art and science of predicting future events. Forecasting is an activity to predict the values of a variable based on the known value of the variable or related variables [4]. Rainfall is included in the category of periodic series data, so the forecasting method that can be used is the time series method, such as the Autoregressive Integrated Moving Average (ARIMA). Rainfall forecasting mostly used the ARIMA or Seasonal ARIMA (SARIMA) model. However, this model only involved one variable without its dependence on other variables. Several factors can affect rainfall, one of which is wind speed. According to [5], significant/intense zonal wind speeds can destroy the formation of convective clouds or rain-forming clouds that cause low rainfall. In addition, the rainfall pattern in Indonesia tends to form a seasonal pattern. Therefore, in this study, the ARIMA model was expanded by adding the eXogen variable and seasonal elements. The eXogen variable is a variable that affects the primary variable. Therefore, the model used was the SARIMAX model (Seasonal Autoregressive Integrated Moving Average with eXogenous input). Several studies use the ARIMAX method [6] to forecast currency netflow using the ARIMAX method and the Radial Basis Function Network. The next is forecasting motorcycle sales by type using the Autoregressive Integrated Moving Average With Exogenous Input (ARIMAX) approach in Banyuwangi Regency [7]. Then, the development of temporal modeling for forecasting and predicting malaria infections using time-series and ARIMAX analysis: a case study in endemic districts of Bhutan [8]. In addition, [9] also researched ARIMAX for export value in Thailand and showed that the MSFE value of the ARIMAX model was better than the ARIMA model. Furthermore, [10] also researched Rainfall Forecasting Using the Autoregressive Integrated Moving Average With Exogenous Input (ARIMAX) Method. However, this study focused on forecasting rainfall data in Pangkalpinang City using the SARIMAX model with seasonal data and maximum wind speed as exogenous variables.

2. RESEARCH METHODS

In this study, the data used was secondary data, namely daily rainfall data in Pangkalpinang City, Bangka Belitung Islands for the period January 2009 to December 2020 obtained from the Meteorology and Geophysics Agency (BMKG) of the Bangka Belitung Islands Province.

2.1 Data collection technique

The data collection technique was carried out by taking daily rainfall data and daily wind speed samples for the Pangkalpinang City area. The data was obtained from the Meteorology, Climatology and Geophysics Agency (BMKG) of the Bangka Belitung Islands Province through the Class I Meteorological Station of Pangkalpinang, Bangka Belitung Islands Province.

2.2 SARIMAX Model Analysis Stages

In the analysis phase of the SARIMAX model, what needed to be done was to identify the data, estimate the model's parameters, estimate the SARIMAX model, check diagnostics, and select the best model. Then, the best forecasting model was applied. The description of these stages was:

1. Data Sharing

The data was divided into two parts in this process, namely in-sample data and out-sample data. The in-sample data was used for the model estimation process, while the out-sample data was used for the forecasting process.

2. Identification Stage

In this identification stage, data stationarity was checked, and differencing processes and the transformation of existing data were carried out. Simple identification could be made by looking at the data plot to see trends, seasonal components, non-stationarity invariance, and others. Detection of non-stationary data in the mean could be seen by using plots from the original data, plots of the autocorrelation function (ACF), and plots of the partial autocorrelation function (PACF). The transformation process was carried out if the data was not stationary in the variance. The transformation that is often used is the Box-Cox transformation. One of the Box-Cox transformations that are often used in time series analysis is the logarithmic transformation, which is often combined with differencing the data resulting from the logarithmic transformation [11].

3. Parameter Estimator

The next step was to determine the non-seasonal (p, d, q) and seasonal (P, D, Q) values to obtain the ARIMA or SARIMA model from the dependent variable. Suppose the stationary data processing had been carried out on the existing data. In that case, the ARIMA and SARIMA model forms could be determined. Based on the differencing value, it is seen to determine the non-seasonal d value or seasonal D value. If there is no differencing, then d or D is 0.

If it only contained AR and differencing processes, the model followed the integrated autoregressive process and was denoted by ARIMA (p,d,0). If it only contained MA and differencing, the model followed the integrated moving average process and was denoted by ARIMA (0,d, q). Then, determination of the non-seasonal value of p and q or seasonal value P and Q was assisted by observing the autocorrelation function (ACF) and partial autocorrelation (PACF) pattern of the studied time series. Autocorrelation is the relationship between ordered values of the same variables. At the same time, partial autocorrelation is used to measure the level of association between Yt and Yt-k. If the effect of time lag 1,2,3, … and so on until k-1 is considered separately, it can also help determine the suitable forecasting model.

4. SARIMAX Model Estimation

After obtaining a tentative ARIMA or SARIMA model, the model was used to form a tentative SARIMAX model by adding independent variables to the SARIMA model. The ARIMAX model (p,d,q) in general can be written as follows [12]:

$$(1-B)^d \phi_p(B)Y_t = \theta_q(B)e_t + \alpha_1 X_{1t} + \ldots + \alpha_k X_{kt} \quad (1)$$

Then, the model of SARIMAX (p,d,q) (P,D,Q)S:

$$\begin{aligned}&(1-B)^d(1-B^s)^D(1-\phi_1 B-...-\phi_p B^p)(1-\theta_1 B^S-...-\theta_P B^{SP})Y_t = \\ &(1-\phi_1 B-...-\phi_q B^q)(1-\theta_1 B^S-...-\theta_Q B^{SQ})e_t+\alpha_1 X_{1,t}+...+\alpha_k X_{k,t}\end{aligned} \quad (2)$$

Where:

Y_t : the dependent variable is rainfall

$X_{1,t}$: exogen variable i.e. maximum wind speed
B : back shift operator
d : *difference* (non-seasonal)
D : *difference* (seasonal)
S : number of season

The value of $X_{k,t}$ is the independent variable or as the eXogen variable of k when t with $k = 1,2,3,….K$. In this model, Y_t and $X_{i,t}, i = 1,2,3…,k$ are time series data and are assumed to be stationary.

5. Diagnostic check and best model selection

The next step was to perform a diagnostic check of the estimated model to verify the model's suitability with the data properties. If the model were suitable, the data calculated by the model (fitted value) would have properties similar to the original data. The L-Jung Box Test was used for a diagnostic check in this study. L-Jung Box test also tested the assumption that the residual must not contain a correlation (Rosadi, 2011). The residual test was carried out using the Akaike Information Criterion (AIC) test and the Ljung-Box test or also known as the Q statistic. The AIC test formula was calculated as [13] :

$$AIC(p,q) = n\log(RSS/n) + 2(p+q) \tag{3}$$

Where n is the number of data points (observations), and RSS is the number of squared remainders. The model with the smaller AIC value is considered the best model. Meanwhile, Ljung-Box tested the residual autocorrelation for the significance of :

H_0: Independently distributed data without serial correlation
H_1: Data were not independently distributed and showed serial correlation.

The maximum likelihood method was applied for parameter estimation. The Ljung-Box Q test was performed to assess whether the residual series was *white noise* [14]. The best model was selected based on the smallest AIC value.

3. RESULTS AND DISCUSSION

This research is a continuation of the results of research published previously using the ARIMAX model in forecasting rainfall in Pangkalpinang City. This study's rainfall data were monthly and maximum daily rainfall taken from January 1, 2009, to June 30, 2021. The discussion this time included descriptive statistical analysis, SARIMA and SARIMAX modeling, and forecasting (forecasting) from rainfall data in Pangkalpinang City.

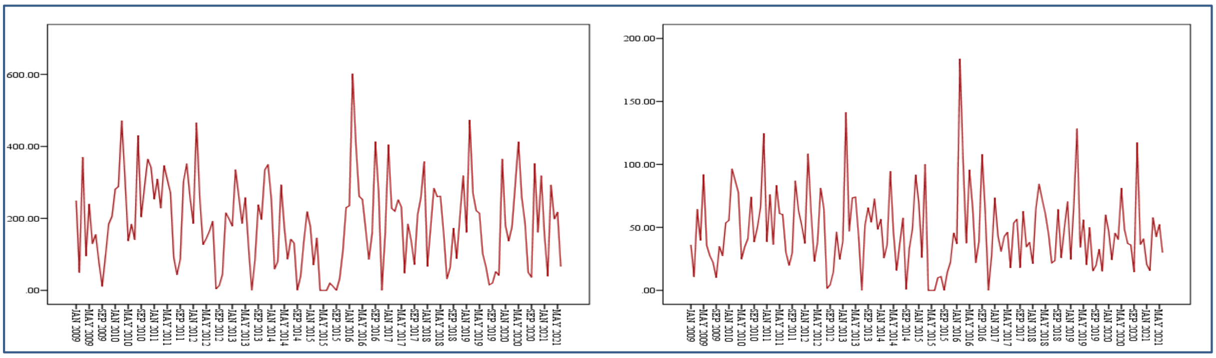
3.1. Descriptive Statistical Analysis

When viewed from descriptive statistical data, monthly and maximum daily rainfall have a high variance. It shows that the rainfall in Pangkalpinang City has high variability in the period 2009 - 2021. When viewed from the maximum value of each data, the highest monthly rainfall in the period January 2009 to June 2021 reached 602 mm, with a maximum daily rainfall of 183 mm.

Table 1. Descriptive Statistics of Monthly Rainfall and Maximum Daily Rainfall

Category	Monthly Rainfall	Rainfall Daily Maximum
N	150	150
Mean	189.19	47.58
Median	184.60	42.50
Mode	0.00	0.00
Std. Deviation	121.85	30.59
Variance	14846.72	935.82
Skewness	0.438	1.120
Kurtosis	-0.036	2.346
Range	602.00	183.90
Minimum	0.00	0.00
Maximum	602.00	183.00

If analyzed from the results of the data plot in Figure 1, it can be seen that in 2016 there was a reasonably high spike in rainfall. It can also be seen from the box-plot diagram of rainfall in Figure 2. Based on this figure, it was found that the rainfall data showed non-stationary data. The data must be transformed into a data transformation process or using a differencing process to proceed to the forecasting process.



(a) **(b)**

Figure 1. Graph of (a) monthly rainfall and (b) maximum rainfall

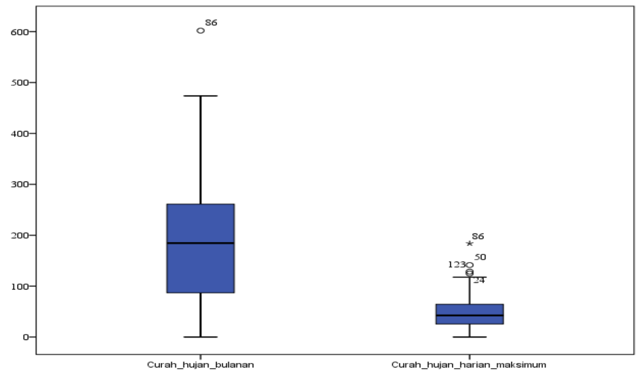
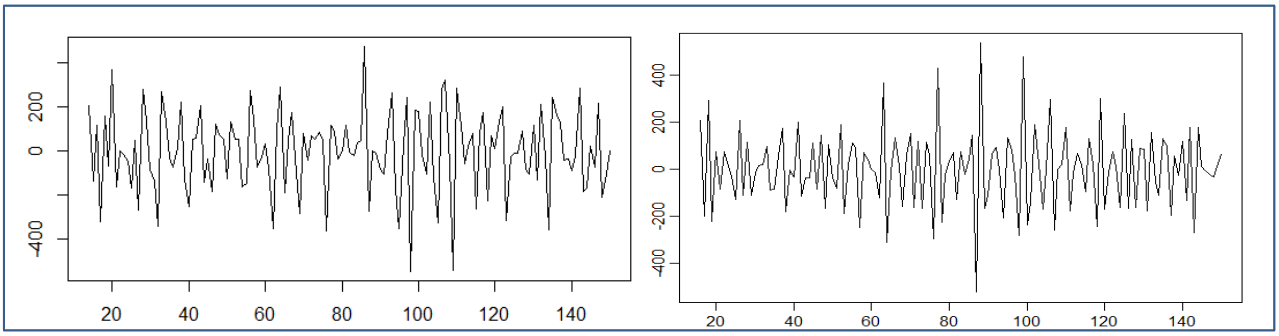


Figure 2. Box-plot of monthly rainfall and maximum rainfall

3.2. SARIMA Modeling

SARIMA modeling is a continuation of ARIMA modeling, which involves seasonal data. In the previous study, the ARIMA model obtained was ARIMA (0,1,3) for monthly rainfall and ARIMA (0,1,2) for maximum daily rainfall. Therefore, the following process was to model Seasonal ARIMA (SARIMA). In Figure 3, it can be seen that the monthly rainfall data and the maximum daily rainfall data can be assumed to be stationary after a one-time differentiation according to the existing data conditions. After that, the plot of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) from the SARIMA model would be seen to obtain a model fit.

In selecting the SARIMA model, a suitable model is selected by looking at the smallest number of parameters (Figure 4). Based on the smallest AIC value (see Table 2). The SARIMA model which has the smallest AIC for monthly rainfall data and maximum daily rainfall is SARIMA (0,1,3)(0,1,1){12} and SARIMA (0,1,2)(0,1,3){12}.



(a) **(b)**

Figure 3. Plots of monthly rainfall data and maximum daily rainfall for the period January 2009 – June 2021 which have undergone a differencing process.

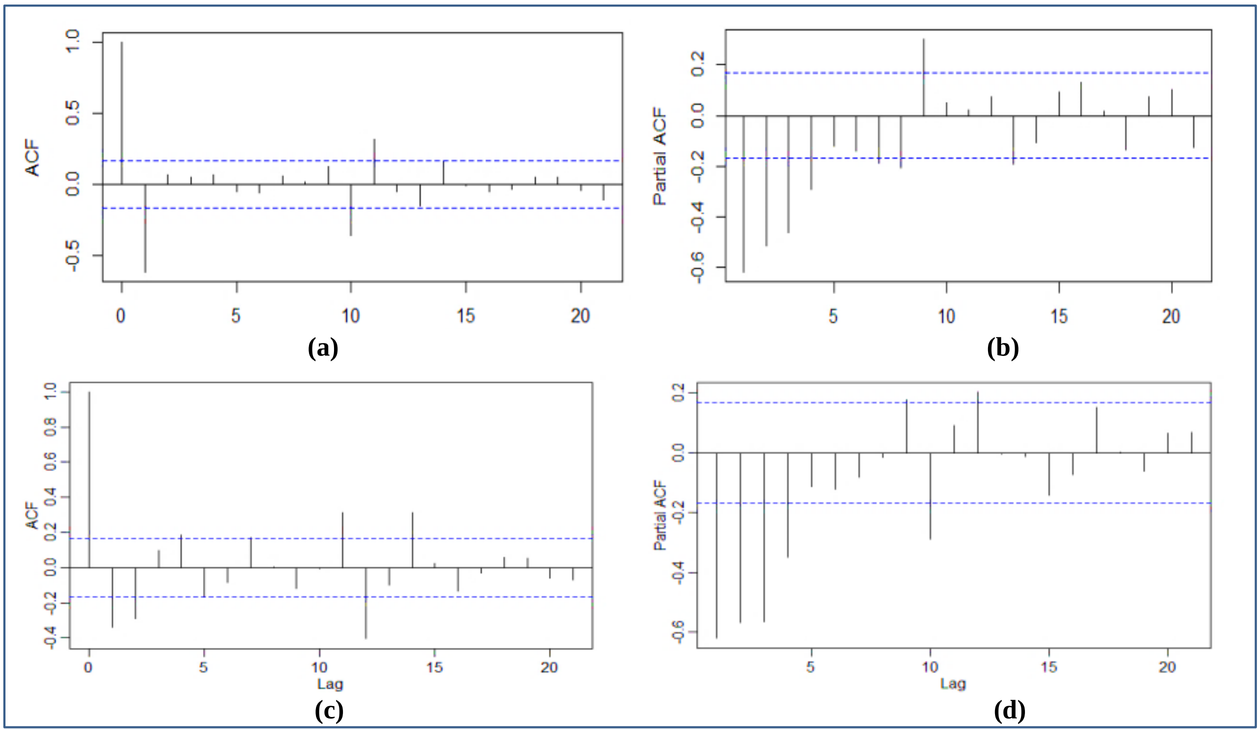


Figure 4. Plot (a) ACF and (b) PACF for monthly rainfall; (c) ACF and (d) PACF for maximum daily rainfall data.

Table 2. SARIMA model of monthly rainfall and maximum daily precipitation

Monthly rainfall		Maximum daily rainfall	
Model	**AIC**	**Model**	**AIC**
SARIMA (0,1,3)(0,1,1){12}	1709.78	SARIMA (0,1,2)(0,1,1){12}	1355,84
SARIMA (0,1,3)(0,1,2){12}	1711.23	SARIMA (0,1,2)(0,1,3){12}	1353,08

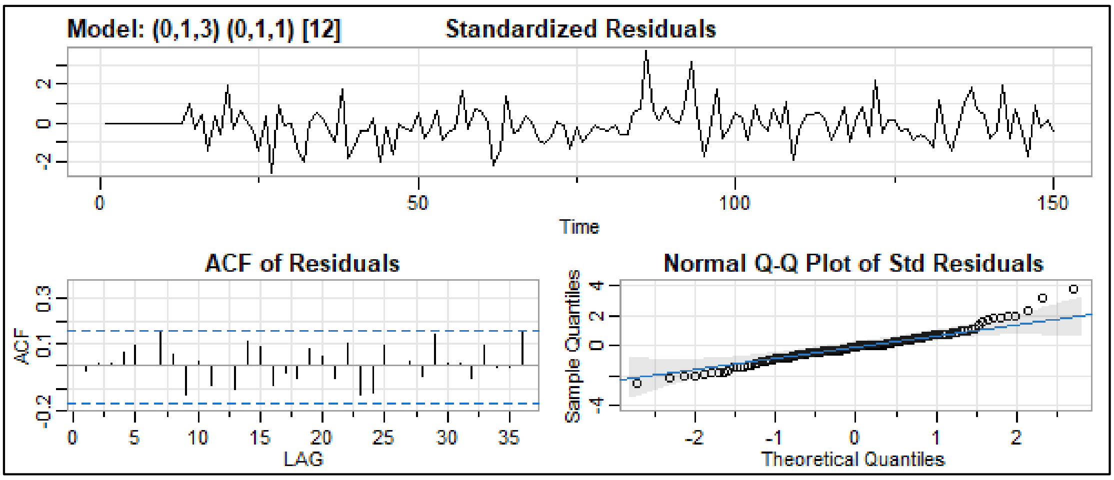


Figure 5. Residual test results of the SARIMA model for monthly rainfall

Then, the residual assumption test was carried out on the selected SARIMA model. Based on the QQ-plots (Figures 5 and 6) and the p-value of more than 5% (Table 3), it can be concluded that the SARIMA model chosen for the maximum monthly and daily rainfall has met the residual assumptions. After getting the best SARIMA model, SARIMAX modeling was carried out by simultaneously combining the SARIMA model that had been obtained with the eXogen variable taken, namely the maximum wind speed.

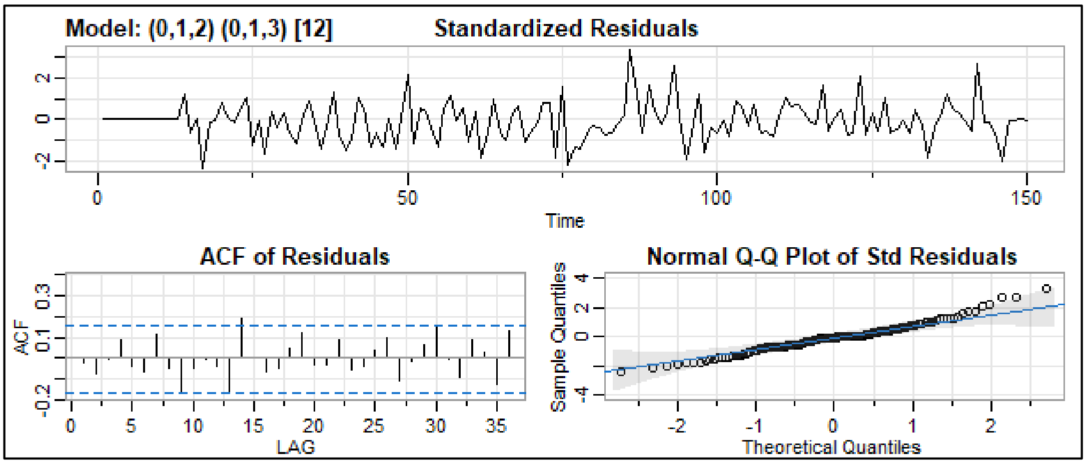


Figure 6. Residual test results from the SARIMA model for maximum daily rainfall

Table 3. Diagnostic check results using the Ljung-Box test

The Best Model	**Data**	***p-value***
SARIMA (0,1,3)(0,1,1){12}	Monthly Rainfall	0,7964
SARIMA (0,1,2)(0,1,3){12}	Maximum Daily Rainfall	0,2195

3.3. SARIMAX Modeling

In modeling SARIMAX, the best SARIMA model was used by involving maximum wind speed data as an exogenous variable. The results of the Ljung-Box test for the SARIMAX model can be seen in Table 4. At the AIC results in Table 2 and Table 4, it can be seen that the SARIMAX model has a smaller AIC value than the SARIMA model. Therefore, it can be said that the SARIMAX model was more suitable for predicting monthly rainfall data and maximum daily rainfall data.

Table 4. Ljung-Box test results for ARIMAX modeling

The Best Model	**Data**	**log likelihood**	**AIC**
SARIMAX (0,1,3)(0,1,1){12}	Monthly Rainfall	-848.23	1708.45
SARIMAX (0,1,2)(0,1,3){12}	Maximum Daily Rainfall	-669.39	1352.78

The SARIMAX equation based on Table 4 is:

1. SARIMAX (0,1,3) (0,1,1) {12}

$$(1-B)^1(1-B^{12})^1 Y_t = (1+0,7102B+0,2391B^2+0,0508B^3)(1+1,000B^{12})e_t + 8,8577X_{1,t} \tag{4}$$

2. SARIMAX (0,1,2) (0,1,3) {12}

$$(1-B)^1(1-B^{12})^1 Y_t = (1+0,8751B^2)(1+0,1249B^{12}+0,9359B^{12,2}+0,0848B^{12,3})e_t + 0,2692X_{1,t} \tag{5}$$

Furthermore, from this modeling, monthly rainfall forecasts and maximum daily rainfall were carried out.

3.4. Rainfall Prediction Using SARIMAX

The results of prediction (forecasting) of rainfall data in Pangkalpinang City using the SARIMAX model can be seen in Table 5. The highest monthly rainfall was predicted to occur in December 2021, with a total monthly rainfall of 293,079 mm. Then, in 2022, there will be the highest monthly rainfall in March 2022. The actual rainfall data from July 2021 - November 2021 was also taken in this rainfall prediction. Then, the actual data was compared with the forecasting results using the SARIMAX model. However, because this research was a follow-up study of the ARIMAX model, Figure 7 showed a graph of the actual rainfall data from July to November 2021 with a graph of forecasting results using the ARIMAX and SARIMAX models. Based on Figures 7. (a) and 7. (b), It can be seen that the rainfall forecasting using SARIMAX was closer to the actual rainfall data compared to the ARIMAX model.

Table 5. The results of forecasting monthly rainfall and daily rainfall maximum using the SARIMAX model

Year	Month	Monthly Rainfall	Maximum Daily Rainfall
2021	July	117.915	39.546
	August	111.125	39.073
	September	116.497	35.932
	October	184.977	55.571
	November	220.687	47.327
	December	293.079	62.145
2022	January	221.157	46.849
	February	258.836	72.710
	Maret	277.325	72.497
	April	244.623	58.294
	May	237.689	62.080
	June	154.991	44.564

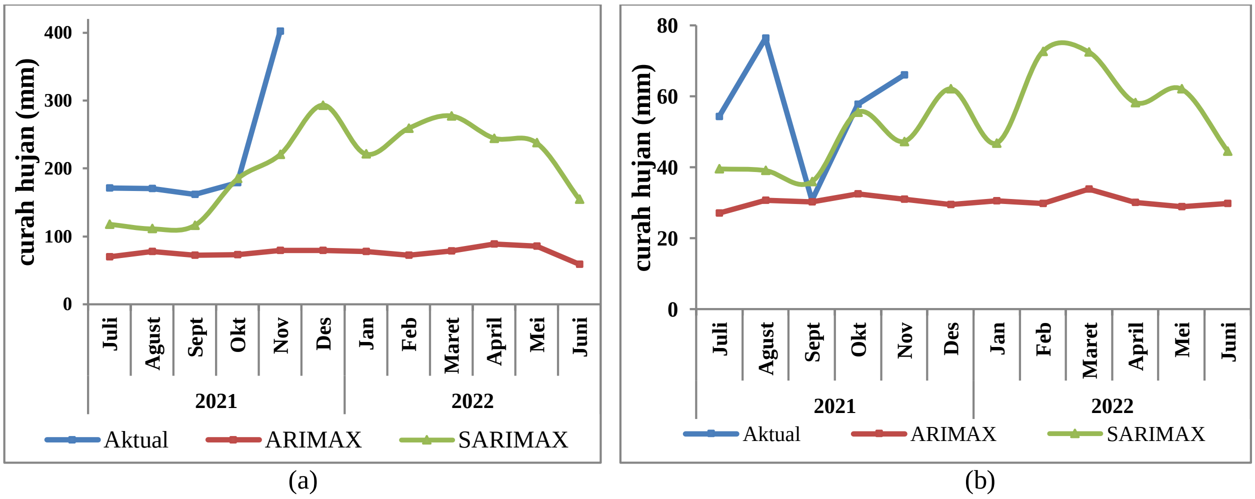


Figure 7. Graph of comparison of actual data, forecasting results of ARIMAX and SARIMAX for data (a) monthly rainfall and (b) maximum daily rainfall

4. CONCLUSIONS

Based on the results and discussions that have been presented, the prediction of rainfall in Pangkalpinang City, Bangka Belitung Islands Province can be modeled with the SARIMAX (0,1,3) (0,1,1) {12} model for monthly rainfall and SARIMAX (0,1,2) (0,1,3) {12} for maximum daily rainfall. Compared to the actual data and previous studies using ARIMAX, the SARIMAX model is still better in the forecasting process when compared to the ARIMAX model. If viewed based on the AIC value of the SARIMA model, the SARIMAX model is also more suitable to be used to predict rainfall in Pangkalpinang City.

AKNOWLEDGEMENT

Our Gratitude is to Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) of the Universitas Bangka Belitung for the contribution in financing the research carried out. Our Gratitude is also for the Class I Meteorological Station Office in Pangkalpinang, Bangka Belitung Islands Province.

REFERENCES

[1] BMKG, “BMKG Ingatkan Prospek Iklim 2021,” 2020. https://www.bmkg.go.id/press-release/?p=bmkg-ingatkan-prospek-iklim-2021&tag=press-release&lang=ID.
[2] Kompas.com, “Cuaca Ekstrem Diprediksi Landa Bangka Belitung Selama Tiga Hari ke Depan,” 2020. https://regional.kompas.com/read/2020/01/26/15181151/cuaca-ekstrem-diprediksi-landa-bangka-belitung-selama-tiga-hari-

ke-depan?page=all (accessed Apr. 01, 2021).

[3] Tempo.co, "Waspada! BMKG Tetpkan Babel Status Siaga," 2021. https://nasional.tempo.co/read/1422975/waspada-bmkg-tetapkan-babel-status-siaga (accessed Mar. 30, 2021).

[4] W. Teguh, "Peramalan iklim dan cuaca berbasis teknologi informasi," no. March, 2020, [Online]. Available: https://www.researchgate.net/publication/339602291_Peramalan_Iklim_dan_Cuaca_Berbasis_Teknologi_Informasi.

[5] S. Nurmita, Sugianto, and H. Wendi, "Analisa Arah Angin Terhadap Curah Hujan Menggunakan Equatorial Atmosphere Radar (EAR) dan Optical Rain Gauge (Org) Di Atas Kototabang Sumatera Barat," [Online]. Available: http://repository.unri.ac.id/xmlui/handle/123456789/7888.

[6] S. Renny Elfira Wulansari, "Peramalan Netflow Uang Kartal dengan Metode ARIMAX dan Radial Basis Function Network (Studi Kasus Di Bank Indonesia)," *J. Sains dan Seni POMITS*, vol. 3, no. 2, pp. 73–78, 2014.

[7] M. L. Izza, D. Susilaningrum, and Suhartono, "… Penjualan Sepeda Motor Menurut Tipe dengan Pendekatan Autoregressive Integrated Moving Average with Exogeneous Input (ARIMAX) di Kabupaten Banyuwangi," *J. Sains Dan Seni Pomits*, vol. 3, no. 2, pp. 176–181, 2014.

[8] K. Wangdi, P. Singhasivanon, T. Silawan, S. Lawpoolsri, N. J. White, and J. Kaewkungwal, "Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: A case study in endemic districts of Bhutan," *Malar. J.*, vol. 9, no. 1, p. 251, Dec. 2010, doi: 10.1186/1475-2875-9-251.

[9] Kongcharoen and Kruangpradit, "Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) Model for Nigerian Non Oil Export," *Eur. J. Bus. Manag.*, vol. 8, no. 36, pp. 2010–2015, 2016.

[10] A. R. Suryani, Sugiman, and P. Hendikawati, "Peramalan Curah Hujan Dengan Metode Autoregressive Integrated Moving Average With Exogenous Input (Arimax)," *Unnes J. Math.*, vol. 7, no. 1, pp. 120–129, 2018.

[11] D. Rosadi, *Analisis Ekonometrika & Runtun Waktu Terapan dengan R.* Yogyakarta: ANDI, 2011.

[12] J. D. Cryer and K.-S. Chan, *Time Series Analysis With Applications in R*. Springer, 2008.

[13] K. P. Chong, "Forecasting the Cocoa Black Pod Incidence in Sabah Using Arimax Model," *Malaysian Cocoa J.*, vol. 10, no. May 2018, pp. 39–48, 2018.

[14] M. S. Hossain, S. Ahmed, and M. J. Uddin, "Impact of weather on COVID-19 transmission in south Asian countries: An application of the ARIMAX model," *Sci. Total Environ.*, vol. 761, p. 143315, Mar. 2021, doi: 10.1016/j.scitotenv.2020.143315.