Res. Math. Nat. Sci. vol. 5, no. 2, pp. 101–113, 2026, doi: doi:10.55657/rmns.v5i2.379

Research in the Mathematical and Natural Sciences
Journal Homepage: https://journal.scimadly.com/index.php/rmns
eISSN: 2828-6804

Original Research

Application of the SARIMAX Model for Forecasting the Number of Domestic Ship Departing Passengers at Soekarno-Hatta Port, Makassar

Wahidah Sanusi[1], Irwan[1,*], and Alni Safitri[1]

[1]*Department of Mathematics, Universitas Negeri Makassar, Makassar 90224, Indonesia*

Article information

***Corresponding author:**
Irwan
irwanthaha@unm.ac.id

Received: 1 September 2026
Revised: 8 November 2026
Accepted: 9 November 2026
Published: 13 November 2026

Licensed under CC BY-NC-SA 4.0

Published by:
scimadly PUBLISHING

Abstract

The number of domestic ship departure passengers at Soekarno-Hatta Port, Makassar, fluctuates seasonally and is affected by weather conditions. This study aims to forecast the number of domestic ship departure passengers at Soekarno-Hatta Port, Makassar, for the next 12 months as supporting information for port operational planning, using a SARIMAX model with monthly rainfall as an exogenous variable. The data are secondary data on the number of passengers and monthly rainfall from January 2017 to December 2024; the 2017–2023 data are used for training and the 2024 data for testing. The research stages include stationarity testing, Box-Cox transformation, model identification using ACF and PACF plots, parameter estimation and significance testing, residual diagnostics, and accuracy evaluation using MAPE. The best model is selected based on the significance of all parameters and the white-noise assumption of the residuals. The results show that the best model is SARIMAX$(2,1,2)(0,2,1)_{12}$, with a negative and significant rainfall coefficient. The MAPE of 29.62% falls into the reasonable category but indicates that the forecast error is still considerable, so the forecasts should be interpreted as an indication of the demand pattern. The model projects the lowest number of passengers in March 2025 (39,630 passengers) and the highest in July 2025 (68,318 passengers).

Keywords: Number of Departure Passengers; Forecasting; SARIMAX; MAPE

1. Introduction

A time series is a sequence of observations arranged chronologically according to a time index at equal intervals. Time series modeling utilizes the temporal dependence patterns in historical data to forecast values for future periods [1]. Such forecasting is widely used in operational planning, including capacity planning for sea passenger transportation.

One of the widely used time series methods is the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA is a forecasting technique based on a time series approach that utilizes correlations within

a time series. ARIMA is divided into two types, namely non-seasonal ARIMA and seasonal ARIMA [2, 3]. Seasonal ARIMA is used when a series is influenced by seasonal factors, such as daily, weekly, or monthly patterns. Seasonal ARIMA is also known as Seasonal Autoregressive Integrated Moving Average (SARIMA) [4].

A limitation of the SARIMA model is that it can only be used to forecast a single time series and does not account for the influence of other variables. To overcome this limitation, an extension of SARIMA called Seasonal Autoregressive Integrated Moving Average with eXogenous variables (SARIMAX) was developed by adding another time series as an exogenous variable [3, 4]. The SARIMAX model aims to determine the relationship between the variable being forecast and the exogenous variable based on historical data, thereby providing more comprehensive forecasts.

Forecasting the number of public transportation passengers in Indonesia has been investigated in several studies. Khoiriyah et al. [5] applied SARIMA to ship passengers at Balikpapan Port, Putri and Sofro [6] compared ARIMA and SARIMA for domestic shipping passengers at Tanjung Perak Port, Riung et al. [7] used SARIMA at Melonguane Port, and Sigalingging et al. [8] applied ARIMA to train passengers in Sumatra. Fernando and Mailoa [9] used the Holt–Winters method for passengers at Tanjung Perak Port, while at the same location as the present study, Abdy et al. [10] forecast the number of departing ship passengers at Soekarno-Hatta Port, Makassar, using the Holt–Winters triple exponential smoothing method. All these studies are univariate, meaning that forecasts are constructed solely from the historical patterns of the passenger series itself (level, trend, and seasonality), so the influence of external factors cannot be explicitly incorporated and tested.

Domestic shipping is influenced by weather conditions. Heavy rainfall during the rainy season is generally accompanied by strong winds and high waves, which can delay or cancel ship departures and reduce the demand for sea travel. Therefore, rainfall is a substantively relevant candidate for an exogenous variable in modeling the number of ship passengers. Unlike the study by Abdy et al. [10], this study models the number of departing passengers at Soekarno-Hatta Port, Makassar, using SARIMAX with monthly rainfall as an exogenous variable. The inclusion of this variable provides two benefits: (i) the direction and significance of the influence of rainfall can be statistically tested through its regression coefficient, and (ii) forecasts can be constructed conditional on a particular rainfall scenario. The novelty of this study lies in the use of rainfall as an exogenous variable in forecasting the number of domestic ship passengers at Soekarno-Hatta Port, Makassar.

Based on the above description, this study aims to (i) develop the best SARIMAX model for the number of domestic ship departure passengers at Soekarno-Hatta Port, Makassar, with rainfall as an exogenous variable, (ii) interpret the influence of rainfall on the number of passengers, and (iii) produce a 12-month forecast for 2025 as supporting information for port operational planning.

2. Method

2.1 ARIMA

The ARIMA model involves the AR process, differencing (I), and the MA process. When nonstationary data are incorporated into a combination of ARMA processes, an ARIMA(p, d, q) model is obtained. For example, the ARIMA$(1, 1, 1)$ model is written as in equation (1) [6].

$$(1 - \phi_1 B)(1 - B)Y_t = \mu + (1 + \theta_1 B)e_t, \tag{1}$$

where:

Y_t : value of the response variable at period t,
$(1 - \phi_1 B)$: AR operator,
$(1 + \theta_1 B)$: MA operator,
e_t : forecast error at time t.

2.2 SARIMA

SARIMA is an extension of the ARIMA model that can analyze recurring or seasonal data patterns at fixed time intervals, such as quarterly, semiannual, and annual intervals. The SARIMA model is denoted by ARIMA$(p,d,q)(P,D,Q)^S$, with the following general form:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)\Theta_Q(B^S)a_t, \tag{2}$$

where:

Y_t : observation at time t,
p, d, q : orders of nonseasonal AR, differencing, and MA, respectively,
P, D, Q : orders of seasonal AR, differencing, and MA, respectively,
$\phi_p(B)$: nonseasonal autoregressive operator $= 1 - \phi_1 B - \cdots - \phi_p B^p$,
$\Phi_P(B^S)$: seasonal autoregressive operator $= 1 - \Phi_1 B^S - \cdots - \Phi_P B^{PS}$,
$\theta_q(B)$: nonseasonal moving average operator $= 1 + \theta_1 B + \cdots + \theta_q B^q$,
$\Theta_Q(B^S)$: seasonal moving average operator $= 1 + \Theta_1 B^S + \cdots + \Theta_Q B^{QS}$,
$(1-B)^d$: nonseasonal differencing of order d,
$(1-B^S)^D$: seasonal differencing of order D,
a_t : residual at time t,
S : seasonal period,
B : backshift operator.

The parameterization of the moving average operator with positive signs follows the `arima` function in the R software used for estimation. Therefore, the signs of the MA coefficients in Table 2 can be directly interpreted in equation (2).

2.3 SARIMAX

The SARIMAX model is an extension of SARIMA that incorporates exogenous variables that are assumed to influence the response series [3, 4]. In this study, SARIMAX is formulated as a regression model with SARIMA errors, which is the form estimated by the `arima` function in R:

$$\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D(Y_t - \beta x_t) = \theta_q(B)\Theta_Q(B^S)a_t, \tag{3}$$

where the symbols are as defined in equation (2), x_t is the exogenous variable (rainfall) at time t, and β is its regression coefficient. The differencing operators in equation (3) are applied to both Y_t and x_t, so that β represents the average change in Y_t resulting from a one-unit increase in x_t, with the SARIMA dynamics remaining unchanged. If Y_t is transformed, Y_t in equation (3) is replaced by the resulting transformed series.

2.4 Stationarity

A time series $\{Y_t\}$ is said to be (weakly) stationary if its mean $E(Y_t) = \mu$ and variance $\mathrm{Var}(Y_t) = \sigma^2$ are constant over time, and its covariance $\mathrm{Cov}(Y_t, Y_{t+k})$ depends only on the lag k [3, 7].

2.5 Differencing

A time series is said to be stationary in the mean if its fluctuations occur around a constant (horizontal) mean, do not depend on time, and have a consistent variance. The first-order differencing equation is as follows:

$$\Delta Y_t = Y_t - Y_{t-1}. \tag{4}$$

2.6 Box–Cox Transformation

The Box–Cox transformation is a power transformation applied to a dependent variable with positive values. This transformation is useful for stabilizing nonconstant data variance. The Box–Cox transformation uses the parameter λ and is defined as follows [11]:

$$Y_t^{(\lambda)} = \begin{cases} \dfrac{Y_t^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \ln(Y_t), & \lambda = 0. \end{cases} \tag{5}$$

2.7 ACF and PACF

2.7.1 Autocorrelation Function (ACF)

The autocorrelation function (ACF) indicates the magnitude of the correlation between an observation at time t and observations in previous periods. Autocorrelation values can be used to determine the model order [3].

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}. \tag{6}$$

2.7.2 Partial Autocorrelation Function (PACF)

The partial autocorrelation function (PACF) represents a set of partial autocorrelations for various lags k. PACF measures the strength of the relationship between Y_t and Y_{t+k} after removing the effects of intermediate lags $1, 2, \ldots, k-1$. In general, $\hat{\phi}_{kk}$ is obtained recursively as follows:

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} \hat{\rho}_j} \tag{7}$$

2.8 Best Model Selection

According to Khoiriyah et al. [5], several adequate models can be obtained in time series modeling, namely models in which all parameters are significant and the residuals satisfy the white-noise assumption. Residual normality is examined as an additional diagnostic because it is related to the validity of prediction intervals.

If more than one adequate model is obtained, the model with the smallest Akaike's Information Criterion (AIC) value is selected as the best model [3]. AIC is formulated as follows:

$$\text{AIC} = n \log\left(\frac{\text{RSS}}{n}\right) + 2k, \tag{8}$$

where n is the number of observations, RSS is the residual sum of squares, and k is the number of estimated parameters. Thus, the best model is selected through the following stages: (i) tentative models containing insignificant parameters at the $\alpha = 0.05$ significance level are eliminated; (ii) the remaining models must have white-noise residuals; (iii) if more than one model satisfies both criteria, the model with the smallest AIC is selected; and (iv) the selected model is validated using MAPE on the test data.

2.9 Forecasting Model Accuracy

Several factors can cause deviations in forecasting results, such as outliers, trend components, seasonal components, or cyclical components in the data series, which may indicate that the method used is less appropriate [12]. The accuracy measure used in this study is MAPE.

2.9.1 Mean Absolute Percentage Error (MAPE)

MAPE is the average relative absolute error between actual and forecast values, expressed as a percentage [13]. The interpretation of MAPE follows the criteria proposed by Lewis [14]: MAPE < 10% indicates highly accurate forecasting, 10–20% indicates good forecasting, 20–50% indicates reasonable forecasting, and > 50% indicates inaccurate forecasting. MAPE is formulated as follows:

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|. \tag{9}$$

where Y_t is the actual value, $\hat{Y}_t$ is the forecast value at period t, and n is the number of periods in the test data.

2.10 Research Procedure

This study is an applied study that uses the SARIMAX method with the assistance of RStudio software. The study was conducted from January to May 2025. The data used are secondary data consisting of monthly data on the number of domestic ship departure passengers at Soekarno-Hatta Port, Makassar, from 2017 to 2024, obtained from the website of the Central Statistics Agency (BPS) of South Sulawesi Province, as well as monthly rainfall data for Makassar City over the same period, obtained from BMKG Region IV Makassar, measured in mm. The data analysis procedures in this study are as follows:

1. Collect historical data on the number of domestic ship departure passengers at Soekarno-Hatta Port, Makassar, from January 2017 to December 2024.
2. Divide the data into two parts, namely data from January 2017 to December 2023 as training data and data from January to December 2024 as testing data [15].
3. Construct time series plots as well as ACF and PACF plots of the actual data.
4. Calculate descriptive statistics for the passenger number data.
5. Identify data stationarity. If the data are nonstationary in the mean, differencing is performed; if the data are nonstationary in variance, the Box–Cox transformation is applied.
6. Construct ACF and PACF plots of the differenced and transformed data. If the data are stationary, tentative models are determined based on the ACF and PACF plots to identify the orders $(p, d, q)(P, D, Q)^S$.
7. Estimate the parameters of the selected models.
8. Perform residual diagnostic tests, namely the white-noise test (Ljung–Box) [16] and the normality test (Shapiro–Wilk) [17].
9. Select the best model based on parameter significance, the white-noise assumption, and AIC.
10. Validate the best model using the 2024 test data with MAPE.
11. Forecast the number of passengers for the next 12 months (January–December 2025) using 2025 rainfall values approximated by the corresponding monthly average rainfall over the period 2017–2024.

3. Results and Discussion

The results and discussion include the data stationarization process and model estimation to obtain the best SARIMAX model, which is subsequently used to forecast the number of domestic ship departure passengers at Soekarno-Hatta Port for the next 12 months.

3.1 Data Description

The data used in this study are secondary data from the Central Statistics Agency for the period January 2017–December 2023 (in-sample data) and January–December 2024 (out-of-sample data). The descriptive statistics of the in-sample data on the number of departing passengers are presented in Table 1.

Table 1 shows that the highest number of departing ship passengers occurred in June 2019, with 70,160 passengers, whereas the lowest number occurred in May 2020, with only 118 passengers. The average monthly

Table 1: Descriptive statistics of the number of departing passengers from 2017 to 2023.

Data	*N*	Mean	Std. Dev.	Max.	Min.
Number of passengers	84	26,894.26	14,092.49	70,160	118

number of departing passengers on the sea transportation route was 26,894.26, with a standard deviation of 14,092.49. The number of departing passengers at Soekarno-Hatta Port can be observed through the time series plot in Figure 1.

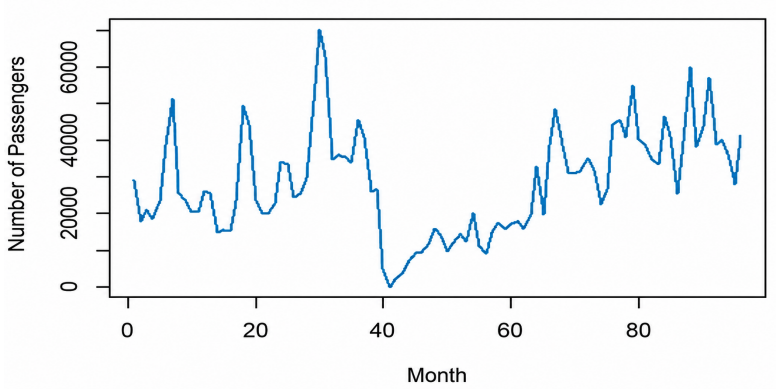


Figure 1: Time series plot of the actual number of departing passengers from 2017 to 2023.

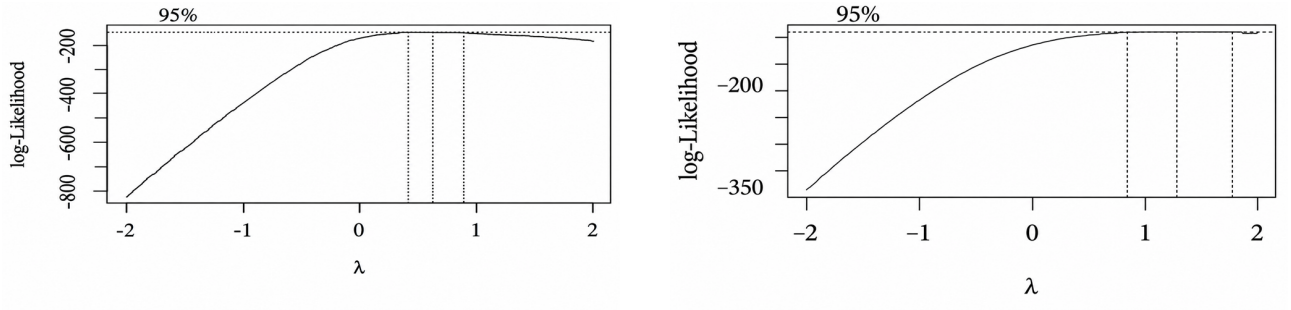
Figure 1 shows that the series experiences increases and decreases in certain months. A significant increase in the number of departing passengers can be observed in June 2019. The graph indicates that the data on the number of departing passengers are nonstationary because the data fluctuate upward and downward throughout the period from 2017 to 2023.

3.2 Data Stationarity Identification

The stationarity of the data in the mean was examined using the Augmented Dickey–Fuller (ADF) test. The ADF test yielded a p-value of $0.01 < \alpha = 0.05$, indicating that the data are stationary in the mean.

However, the ADF test does not examine the presence of seasonal unit roots. Therefore, the need for non-seasonal and seasonal differencing is further determined through the ACF and PACF plots (Subsection 3.3).

Next, stationarity in variance was examined. The Box–Cox plot yielded $\hat{\lambda} = 0.6359676$, indicating that the data on the number of departing passengers are not yet stationary in variance. Therefore, the data cannot be directly used for modeling and need to undergo a Box–Cox transformation. The results of the identification of variance stationarity using the Box–Cox transformation are presented in Figure 2.



(a) Box–Cox plot of the in-sample data

(b) Box–Cox plot after transformation

Figure 2: Box–Cox transformation.

Figure 2a shows that the data are not stationary in variance, with a lambda value of 0.6359676 in the Box–Cox plot. Since this value is far from 1 and the value of 1 is not included in the 95% confidence interval, a square-root transformation ($\lambda = 0.5$) was performed. After the square-root transformation, the Box–Cox analysis was repeated on the transformed data, yielding a lambda value of 1.271932, which falls within the 95% confidence interval, as shown in Figure 2b. Thus, the transformed data are stationary in variance.

After the square-root transformation, the series to be modeled is $W_t = \sqrt{Y_t}$. Figure 3 presents the monthly plot of W_t.

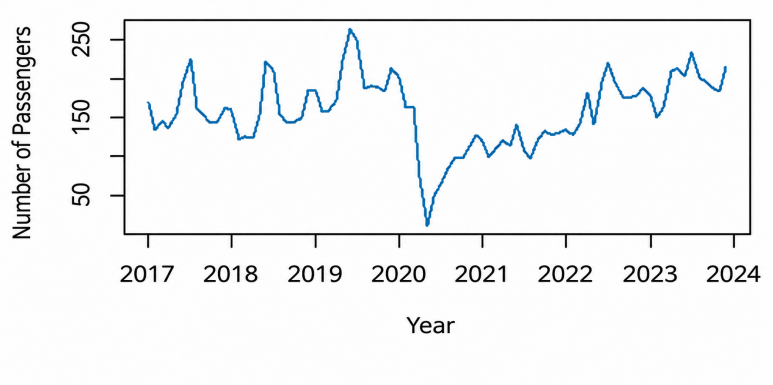
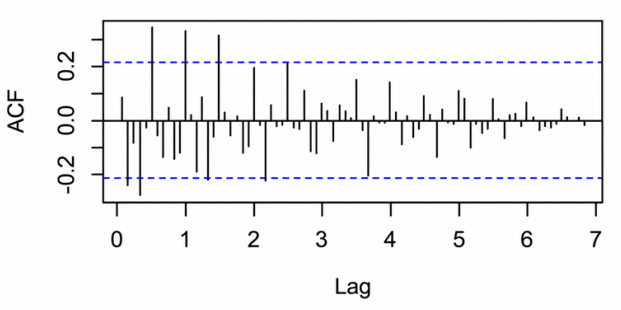


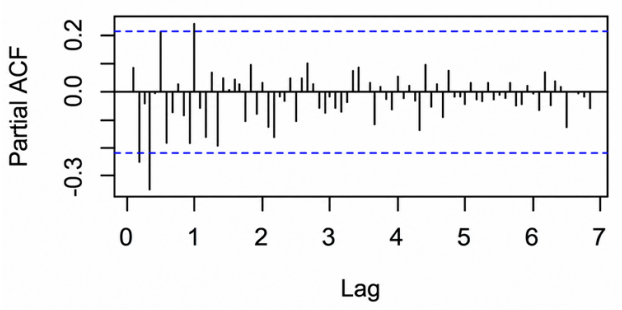
Figure 3: Time series plot of the number of passengers after the Box–Cox transformation

3.3 Model Identification

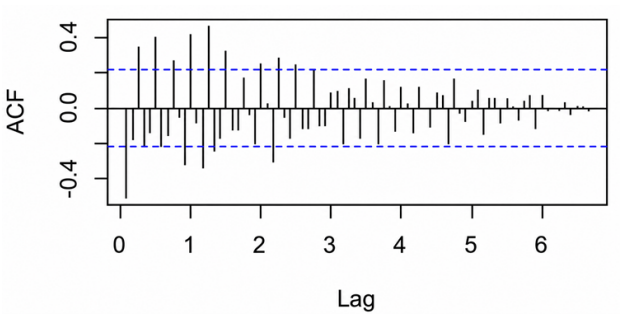
Model identification was performed by examining the ACF and PACF plots of the number of domestic sea transportation departure passengers at Soekarno-Hatta Port, for which the data had been made stationary in both the mean and variance.



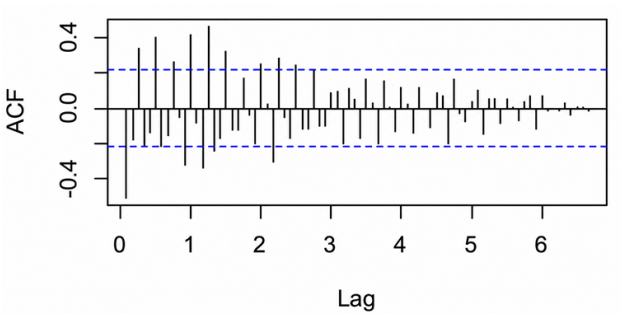
(a) ACF plot after nonseasonal differencing



(b) PACF plot after nonseasonal differencing



(c) ACF plot after seasonal differencing



(d) PACF plot after seasonal differencing

Figure 4: ACF and PACF plots after transformation.

Based on the ACF and PACF plots in Figure 4a–Figure 4d, several tentative SARIMAX models were obtained, namely SARIMAX$(2,1,2)(0,2,1)_{12}$, SARIMAX$(2,1,1)(0,2,2)_{12}$, SARIMAX$(1,1,2)(0,2,1)_{12}$, SARIMAX$(2,1,1)(0,2,1)_{12}$, SARIMAX$(1,1,2)(0,2,2)_{12}$, and SARIMAX$(2,1,2)(0,2,2)_{12}$.

3.4 Parameter Estimation and Model Parameter Significance Tests

Based on the identified models, the parameters of each model were estimated using the maximum likelihood method, as presented in Table 2. The significance of each parameter γ was tested using the hypotheses $H_0 : \gamma = 0$ and $H_1 : \gamma \neq 0$, with the test statistic

$$Z = \frac{\hat{\gamma}}{SE(\hat{\gamma})};$$

H_0 is rejected if the p-value is less than $\alpha = 0.05$. The xreg coefficient in Table 2 represents the estimate of β (rainfall) in equation (3).

Table 2: Parameter estimation and significance tests of the tentative models.

Model	Coefficient	Estimate	Std. Error	Z-value	P-value
$\text{SARIMAX}(2,1,2)(0,2,1)_{12}$	AR1	0.38054	0.06715	5.66674	< 0.0001
	AR2	−0.90684	0.07015	−12.92667	< 0.0001
	MA1	−0.24692	0.05344	−4.62021	< 0.0001
	MA2	0.99808	0.06203	16.08910	< 0.0001
	SMA1	−0.94737	0.29828	−3.17608	0.00149
	xreg	−8.98291	3.88010	−2.31512	0.02061
$\text{SARIMAX}(2,1,1)(0,2,2)_{12}$	AR1	1.0309	0.1312	7.8585	< 0.0001
	AR2	−0.1695	0.1328	−1.2763	0.2017
	MA1	−0.9920	0.1288	−7.7006	< 0.0001
	SMA1	−1.0075	0.3167	−3.1807	0.0015
	SMA2	0.0923	0.1820	0.5071	0.6123
	xreg	−9.0055	4.1343	−2.1781	0.0294
$\text{SARIMAX}(1,1,2)(0,2,1)_{12}$	AR1	−0.3952	0.5361	−0.7374	0.4607
	MA1	0.5441	0.1574	3.4565	0.00055
	MA2	0.1596	0.1594	1.0013	0.3168
	SMA1	−0.9615	0.2610	−3.6832	0.00023
	xreg	−9.0121	3.8316	−2.3514	0.0187
$\text{SARIMAX}(2,1,1)(0,2,1)_{12}$	AR1	1.0459	0.1273	8.2161	< 0.0001
	AR2	−0.1862	0.1283	−1.4514	0.1468
	MA1	−0.9940	0.1092	−9.1007	< 0.0001
	SMA1	−0.9573	0.2726	−3.5124	0.0004
	xreg	−9.0704	4.0308	−2.2503	0.0244
$\text{SARIMAX}(1,1,2)(0,2,2)_{12}$	AR1	0.8526	0.0893	9.5466	< 0.0001
	MA1	−0.8394	0.1757	−4.7764	< 0.0001
	MA2	−0.1535	0.1310	−1.1725	0.2411
	SMA1	−1.0579	0.4946	−2.1394	0.0325
	SMA2	0.0914	0.2010	0.4549	0.6488
	xreg	−9.0281	4.2857	−2.1066	0.0351
$\text{SARIMAX}(2,1,2)(0,2,2)_{12}$	AR1	0.3674	0.0758	4.8468	< 0.0001
	AR2	−0.9010	0.0797	−11.3004	< 0.0001
	MA1	−0.2545	0.0564	−4.5124	< 0.0001
	MA2	0.9985	0.0606	16.4777	< 0.0001
	SMA1	−1.1181	1.2096	−0.9245	0.3552
	SMA2	0.1907	0.2460	0.7745	0.4379
	xreg	−9.0345	4.1504	−2.1766	0.0295

Based on Table 2, five of the six tentative models contain parameters that are not significant at the $\alpha = 0.05$ level. These are AR2 and SMA2 in SARIMAX $(2,1,1)(0,2,2)_{12}$; AR1 and MA2 in SARIMAX $(1,1,2)(0,2,1)_{12}$; AR2 in SARIMAX $(2,1,1)(0,2,1)_{12}$; MA2 and SMA2 in SARIMAX $(1,1,2)\ (0,2,2)_{12}$; and SMA1 and SMA2 in SARIMAX $(2,1,2)\ (0,2,2)_{12}$. These five models were not retained because they contain parameters that do not make a statistically significant contribution and may potentially lead to overfitting. Only SARIMAX $(2,1,2)(0,2,1)_{12}$ has all parameters statistically significant, including the rainfall coefficient. Since only one model satisfies the first-stage criteria, an AIC comparison is not necessary to determine the best

model; AIC serves as a selection criterion when more than one model is eligible. This model was subsequently subjected to residual diagnostic tests.

3.5 Residual Assumption Tests

3.5.1 White Noise Test

The white noise assumption was tested using the Ljung–Box test [16], with the hypotheses H_0: the residuals satisfy the white noise assumption, and H_1: the residuals do not satisfy the white noise assumption.

Table 3: Ljung–Box test for the residual white noise assumption.

Model	X^2	P-value
SARIMAX$(2,1,2)(0,2,1)_{12}$	16.606	0.165

Based on Table 3, the p-value is greater than 0.05; therefore, H_0 is not rejected. This indicates that the SARIMAX$(2,1,2)(0,2,1)_{12}$ model satisfies the white noise assumption.

3.5.2 Normality Test

Residual normality was tested using the Shapiro–Wilk test [17], with the hypotheses H_0: the residuals are normally distributed, and H_1: the residuals are not normally distributed. The test results are presented in Table 4.

Table 4: Shapiro–Wilk normality test for the residuals.

Model	W	P-value
SARIMAX$(2,1,2)(0,2,1)_{12}$	0.92612	0.0001285

Based on Table 4, the p-value of 0.0001285 is less than $\alpha = 0.05$; therefore, H_0 is rejected. This means that the residuals of the SARIMAX$(2,1,2)(0,2,1)_{12}$ model do not satisfy the normality assumption.

This residual non-normality should be interpreted with caution. The most important assumption for point forecasting is that the residuals are uncorrelated (white noise), and this assumption has been satisfied (Table 3). If the errors are independent and identically distributed with finite variance, the Gaussian maximum likelihood estimator remains consistent and asymptotically normally distributed even when the errors are not normally distributed [18]. Therefore, the parameter significance tests in Table 2 remain asymptotically valid. Residual normality is primarily required for constructing prediction intervals based on the normal distribution [12]. Consequently, this study reports only point forecasts and does not use prediction intervals based on the normal distribution. The residual non-normality is suspected to be caused by extreme values during the COVID-19 pandemic, such as in May 2020, when only 118 passengers were recorded.

3.6 Forecasting

Based on the parameter significance tests and the white noise test, the SARIMAX$(2,1,2)(0,2,1)_{12}$ model was selected for forecasting. With $W_t = \sqrt{Y_t}$ and the parameter estimates in Table 2, namely $\hat{\phi}_1 = 0.3805$, $\hat{\phi}_2 = -0.9068$, $\hat{\theta}_1 = -0.2469$, $\hat{\theta}_2 = 0.9981$, $\hat{\Theta}_1 = -0.9474$, and $\hat{\beta} = -8.9829$, the estimated model can be expressed as follows:

$$\begin{aligned}(1 - 0.3805B + 0.9068B^2)(1 - B)(1 - B^{12})^2(W_t + 8.9829x_t)\\ = (1 - 0.2469B + 0.9981B^2)(1 - 0.9474B^{12})a_t.\end{aligned} \tag{10}$$

The rainfall coefficient is negative ($\hat{\beta} = -8.9829$; p-value = 0.0206). This indicates that, with the SARIMA dynamics held constant, a one-unit increase in rainfall is associated with a decrease of 8.98 in $W_t = \sqrt{Y_t}$. Since $Y_t = W_t^2$, the marginal effect on the original scale is

$$\frac{\partial Y_t}{\partial x_t} = 2W_t\hat{\beta}.$$

Thus, at the mean of the training data ($\sqrt{26,894.26} \approx 164$), a one-unit increase in rainfall is associated with a decrease of approximately 2,946 passengers per month. This direction of association is consistent with maritime

transportation conditions, where high rainfall is generally accompanied by adverse weather and high waves, which may delay or cancel ship departures and reduce the demand for sea travel. This coefficient indicates a statistical association and is not intended to represent a causal relationship.

Forecasting for the 2024 test data was performed using equation (10), with the actual rainfall data for 2024 used as the values of x_t. Thus, this evaluation is ex-post, meaning that the values of the exogenous variable are assumed to be known. The forecasting results are presented in Table 5.

Table 5: Forecasted number of departing passengers in 2024.

Month	Forecast for 2024
January	42,628
February	45,213
March	39,716
April	55,808
May	56,924
June	52,958
July	70,055
August	53,587
September	47,928
October	45,749
November	45,988
December	53,595

The plot of the actual data and the forecasted number of departing passengers using the SARIMAX$(2,1,2)(0,2,1)_{12}$ model is shown in Figure 5.

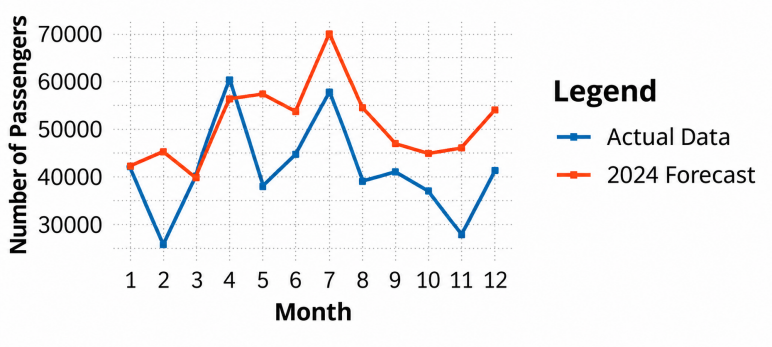


Figure 5: Forecast plot using the SARIMAX$(2,1,2)(0,2,1)_{12}$ model for 2024.

Next, the accuracy of the SARIMAX$(2,1,2)(0,2,1)_{12}$ model on the test data was evaluated using MAPE in equation (9):

$$\begin{aligned}\text{MAPE} &= \frac{\left|\frac{41{,}168-42{,}628}{41{,}168}\right| + \cdots + \left|\frac{41{,}276-53{,}595}{41{,}276}\right|}{12} \times 100\% \\ &= \frac{3.5549}{12} \times 100\% \\ &= 0.2962 \times 100\% = 29.62\%.\end{aligned}$$

The resulting MAPE value is 29.62%. Based on the Lewis criteria [14], this value falls into the reasonable category, but does not reach the good (10–20%) or highly accurate (< 10%) categories. This means that the average forecasting error is still approximately 30% of the actual values. Therefore, the forecasting results are more appropriately used to describe passenger demand patterns and levels rather than as exact numerical predictions. In several months, the model produced forecasts higher than the actual values, such as January 2024 (forecast: 42,628; actual: 41,168) and December 2024 (forecast: 53,595; actual: 41,276). Furthermore, this MAPE was obtained from ex-post forecasts using actual rainfall data. Therefore, the ex-ante forecasting error for 2025 may be larger because rainfall in 2025 must also be estimated.

For forecasting in 2025, actual rainfall data are not yet available. Therefore, the monthly rainfall values for 2025 were approximated using the average rainfall for the corresponding months over the 2017–2024 period (monthly climatological averages), which were then substituted into equation (10). The forecasting results for 2025 are presented in Table 6.

Table 6: Forecasted number of departing passengers in 2025.

Month	Forecast for 2025
January	45,181
February	46,954
March	39,630
April	54,327
May	54,625
June	50,944
July	68,318
August	51,147
September	45,599
October	43,727
November	46,518
December	61,083

The plot of the actual data and the forecasted number of domestic sea transportation departure passengers at Soekarno-Hatta Port in 2025 using the SARIMAX$(2,1,2)(0,2,1)_{12}$ model is shown in Figure 6.

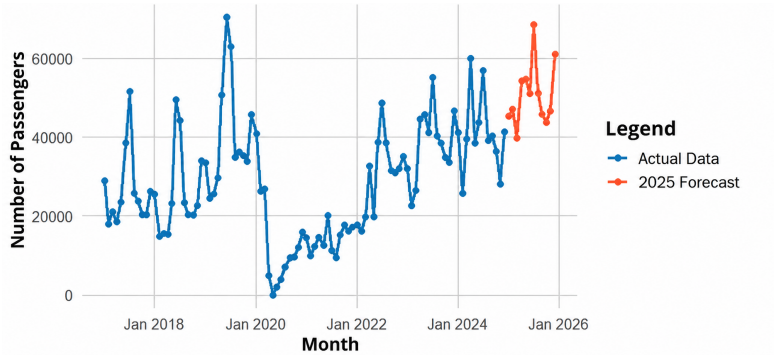


Figure 6: Forecast plot using the SARIMAX$(2,1,2)(0,2,1)_{12}$ model for 2025.

Based on Table 6 and Figure 6, the forecasts for the next 12 periods indicate that the number of departing passengers fluctuates throughout the year. The highest number of departing passengers is forecasted to occur in July 2025, at 68,318 passengers, while the lowest is forecasted for March 2025, at 39,630 passengers. These forecasts are conditional point forecasts based on the average rainfall scenario. Therefore, they are best used as supporting information for port operational planning, such as adjusting ship schedules and capacities during months with the highest forecasted passenger numbers, rather than as predictions of outcomes that will certainly occur.

4. Conclusion

The best model for forecasting the number of domestic ship departing passengers at Soekarno-Hatta Port, Makassar, based on data from January 2017 to December 2024, is the SARIMAX$(2,1,2)(0,2,1)_{12}$ model with rainfall as an exogenous variable. Rainfall has a significant negative effect, indicating that an increase in rainfall is associated with a decrease in the number of departing passengers. The model produces a MAPE of 29.62% on the 2024 test data, which falls into the reasonable forecasting accuracy category but still indicates a relatively large forecasting error. The 2025 forecasts show that the highest number of passengers is expected in July 2025 (68,318 passengers), while the lowest is expected in March 2025 (39,630 passengers). These results should be used as supporting information for port operational planning rather than as predictions of certain future outcomes.

This study has several limitations. First, the model residuals are not normally distributed; therefore, normality-based prediction intervals are not reported. Second, the model accuracy was evaluated *ex-post* using actual rainfall data, whereas the 2025 forecasts rely on average rainfall as a proxy. Third, the COVID-19 pandemic period was not explicitly modeled. Future research is recommended to compare the SARIMAX model with SARIMA without exogenous variables and the Holt–Winters method, incorporate intervention variables to account for the pandemic period, and employ bootstrap prediction intervals.

Declaration of Artificial Intelligence (AI) Use

The authors used Claude (Anthropic) to assist with language editing, translation, and formatting the manuscript according to the journal template. The AI tool was not used to generate data, perform the statistical analysis, or draw the conclusions of this study. The authors reviewed, verified, and edited all AI-assisted content and take full responsibility for the content of this manuscript.

Author Contributions

Wahidah Sanusi: Conceptualization, Methodology, Validation, Supervision, Writing – review and editing. **Irwan**: Conceptualization, Methodology, Validation, Supervision, Project administration, Writing – review and editing. **Alni Safitri**: Data curation, Investigation, Formal analysis, Software, Visualization, Writing – original draft.

Acknowledgments

The authors thank Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan and the institution that provided the rainfall data for making the data used in this study available.

Funding

This research received no external funding.

Data Availability

The monthly passenger data used in this study are publicly available from Badan Pusat Statistik (BPS) Provinsi Sulawesi Selatan. The processed dataset and the R code used for the analysis are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] L. Yuliana, "Sales planning analysis using time series method (case study at pd. sumber jaya aluminium)," *Jurnal Mitra Manajemen*, vol. 3, no. 7, pp. 780–789, 2019, doi:10.52160/ejmm.v3i7.255.

[2] S. Deviana, Nusyirwan, D. Azis, and P. Ferdias, "Analisis model autoregressive integrated moving average data deret waktu dengan metode momen sebagai estimasi parameter," *Jurnal Siger Matematika*, vol. 2, no. 2, pp. 57–67, 2021.

[3] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2015.

[4] 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: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 1, pp. 137–146, 2022, doi:10.30598/barekengvol16iss1pp137-146.

[5] N. S. Khoiriyah, M. Silfiani, R. Novelinda, and M. Rezki, "Peramalan jumlah penumpang kapal di pelabuhan balikpapan dengan sarima," *Jurnal Statistika dan Komputasi (STATKOM)*, vol. 2, no. 2, pp. 76–82, 2023, doi:10.32665/statkom.v2i2.2303.

[6] S. Putri and A. Sofro, "Peramalan jumlah keberangkatan penumpang pelayaran dalam negeri di pelabuhan tanjung perak menggunakan metode arima dan sarima," *MATHunesa: Jurnal Ilmiah Matematika*, vol. 10, no. 1, pp. 61–67, 2022, doi:10.26740/mathunesa.v10n1.p61-67.

[7] R. Riung, N. Nainggolan, and Y. A. R. Langi, "Peramalan jumlah penumpang kapal laut di pelabuhan melonguane kabupaten kepulauan talaud menggunakan metode sarima," *Jurnal MIPA*, vol. 13, no. 2, pp. 62–66, 2024.

[8] F. L. Sigalingging, M. Y. Irsan, and J. Sesa, "Forecasting the number of train passenger in sumatra using arima models," in *Proceeding of the 2nd First Symposium on Actuarial Science*, 2022, pp. 8–13.

[9] Z. Fernando and E. Mailoa, "Implementasi time series dalam prediksi jumlah keberangkatan penumpang di pelabuhan tanjung perak," *Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi*, vol. 13, no. 2, pp. 1296–1303, 2024, doi:10.35889/jutisi.v13i2.1998.

[10] M. Abdy, Irwan, and M. Mayangsari, "Penerapan metode triple exponential smoothing holt-winters dalam meramalkan jumlah keberangkatan penumpang kapal di pelabuhan soekarno-hatta," *Saintifik*, vol. 10, no. 2, pp. 182–190, 2024, doi:10.31605/saintifik.v10i2.533.

[11] G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 26, no. 2, pp. 211–252, 1964, doi:10.1111/j.2517-6161.1964.tb00553.x.

[12] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021, Available online: https://otexts.com/fpp3/.

[13] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, 2006, doi:10.1016/j.ijforecast.2006.03.001.

[14] C. D. Lewis, *Industrial and Business Forecasting Methods*. London, U.K.: Butterworth Scientific, 1982.

[15] H. Bichri, A. Chergui, and M. Hain, "Investigating the impact of train/test split ratio on the performance of pre-trained models with custom datasets," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 2, pp. 331–339, 2024, doi:10.14569/IJACSA.2024.0150235.

[16] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978, doi:10.1093/biomet/65.2.297.

[17] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3–4, pp. 591–611, 1965, doi:10.1093/biomet/52.3-4.591.

[18] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 3rd ed. Cham, Switzerland: Springer, 2016, doi:10.1007/978-3-319-29854-2.