Airport Electricity Consumption Demand Forecasting Model using Seasonal Autoregressive Integrated Moving Average

Afwan Heru Cahya and Zulkarnain
Industrial Engineering Department, Faculty of Engineering
University of Indonesia, Depok, West Java, Indonesia
afwan.heru@ui.ac.id, zulkarnain17@ui.ac.id

Abstract

Electric load forecasting, also known as Probabilistic Load Forecasting (PLF), has played a role in the electric power industry. Forecasting the electricity consumption in business is necessary for planning power system operations, stability, and energy trading. Many business entities, such as commercial airports, require electric load forecasting to meet service and regulatory needs. Therefore, forecasting is needed to become a reference in determining strategic management energy. This research aims to forecast the electricity consumption of Soekarno-Hatta International Airport using the Seasonal Autoregressive Integrated Moving Average (SARIMA) model. The study uses daily historical data collected from airport operator companies from 01 January 2022 to 31 December 2022 to build and evaluate the model's performance. The findings show that the SARIMA model $(1,1,1)(0,1,1)^7$ has the best model accuracy with a MAPE of 4.62%. The study conclusions highlight the potential of the model to support energy management practices at Soekarno Hatta International Airport and other similar facilities.

Keywords
Forecasting, Airport, SARIMA, Electricity Demand, and Energy Management.

1. Introduction
Electrical load forecasting, or Probabilistic Load Forecasting (PLF), has played a role in the electric power industry in the last few decades. The need for load forecasting in business is used as a basis for planning power system operations, revenue projections, tariff design, the stability of the entire power system, and energy trading. Apart from electric utilities, various business entities such as regulatory bodies, large industrial companies, and commercial companies have a need for electrical load forecasting (Hong & Fan, 2016; Tang et al., 2022). Moreover, accurately forecasting electricity loads in electricity generation and distribution systems can help better balance electricity production and demand (Gunawan & Huang, 2021). The dynamic nature of generating capacity and electricity consumption poses challenges due to the inability to store electricity. Synchronizing power generation with changes in electricity demand is crucial for maintaining a dynamic equilibrium. Failure to do so can impact the quality of electricity supply and potentially jeopardize the safety and stability of the power system (Hu et al., 2017). (Kang et al., 2017) research developed a basic model of airport energy specifically by including flight information, time, and outside temperature using the Piecewise Linear Regression Technique. Nevertheless, this study lacks the ability to fully capture the actual influence of passengers on airport demand. The association between passenger volume and the power demand of airports is anticipated to exhibit non-linearity. Airports require substantial electricity for fundamental operational necessities, such as conveyors and lighting, which remain consistent regardless of flight schedules. In addition, airport power data includes electricity demand for air-side buildings and all facilities, including parking lots and hotels. Soekarno-Hatta Airport currently only has historical records regarding the use of electricity loads, but does not yet have forecasts of electricity loads. Based on this background, airport management companies need an electric load forecasting model as a reference in carrying out strategies related to electrical equipment and airport operations.

1.1 Objectives
With the background that has been discussed previously, the purpose of this research is as follows.
1. Using forecasting methods, create a SARIMA model to estimate the airport's electrical consumption.
2. Comparing the performance of the performance models used.

2. Literature Review

2.1 Electricity Load Forecasting

The current load on the power system is so dynamic that it requires sophisticated technology to monitor and control the power system, such as forecasting the electric load. Although demand forecasting was first discussed in 1925, it received a significant boost with the introduction of statistical methods (Yashwanth et al., 2021). However, there is no standard yet to classify the forecasting range of electrical loads. The classification of the forecasting process by (Hong & Fan, 2016) is structured into four distinct categories, determined by the time intervals and their respective applications. These categories include very short-term load forecasting (VSTLF), short-term load forecasting (STLF), medium-term load forecasting (MTLF), and long-term load forecasting (LTLF). The consumption of electricity is subject to a range of factors, comprising weather conditions (such as temperature, wind speed, and rainfall), the level of daily business activities (including peak and off-peak hours, weekdays and weekends, holidays, and approaching holidays), as well as seasonal patterns at the daily, weekly, and yearly levels. These factors can result in unforeseen spikes in electricity demand. Consequently, electric load forecasting has emerged as a crucial component for planning and operating energy systems (Abunofal et al., 2021).

Various electrical load forecasting methods, such as neural network-based forecasting, support vector machines, decision trees, time series analysis, and linear regression, have been introduced. While nonlinear models, such as neural networks and support vector machines, often yield superior predictive accuracy and can effectively capture non-linearity within a dataset, they present challenges in result interpretation and entail high computational burdens during model training and testing. On the other hand, linear regression models are user-friendly and provide easily interpretable results. In certain cases, they have demonstrated success in predicting building energy consumption (Kang et al., 2017). However, linear regression models are not suitable for datasets with nonlinear characteristics. Therefore, this research uses a time series analysis model, namely Seasonal ARIMA, because it is easier to interpret and can handle nonlinear and seasonal data types.

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model has gained significant traction in the field of forecasting. In a similar vein, Hasanah et al. (2021) employed the SARIMA model to predict loads of electricity on several national holidays, specifically Independence Day & New Year, during the period of 2015-2018. Their study aimed to identify suitable SARIMA models for accurately estimating loads of electricity in the future periods. The findings indicate that SARIMA (2,2,0) $(0,1,0)_{24}$ and SARIMA (1,1,0) $(1,1,0)_{24}$ are appropriate models for forecasting electricity loads on New Year and Independence Day, respectively. In the various models of forecasting, SARIMA models can be enhanced by incorporating exogenous factors to improve accuracy. As demonstrated by Abunofal et al. (2021), SARIMA models were employed to forecast future electricity prices for Germany by considering various input parameters and selecting the optimal modeling approach. The results indicate that the SARIMAX model outperformed other models, including SARIMA, multiple linear regression, and ARIMA, in terms of predictive performance. Apart from the demand side, SARIMA can also be used for forecasting on the energy production side, such as (Vagropoulos et al., 2016), which compares several methods for forecasting power generation from Photovoltaic (PV) generators connected to the grid (grid), including modeling Seasonal Autoregressive Integrated Moving Average (SARIMA). The results show that the SARIMA model is relatively accurate regarding future forecasting.

2.2 Seasonal Autoregressive Integrated Moving Average (SARIMA)

The SARIMA model is a method of modeling univariate time series with high seasonal patterns (Box et al., 2016). This model is adding seasonal effect as an extension of the ARIMA model, representing a generalization of the ARMA model. (Al-Shaikh et al., 2019) illustrated that a stationary time series can be represented by an Autoregressive Moving Average (ARMA) model with an order of (p, q). This model consists of two distinct polynomials: an Autoregressive (AR) polynomial of order p, as denoted by equation 1, and a Moving Average (MA) polynomial of order q, as represented by equation 2.

$$y(t) = c + \epsilon_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} \tag{1}$$

$$y(t) = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \tag{2}$$

In equation 3, the constant c represents a constant term, *y(t)* denotes the value of the time series at time t, ϵ_t represents a white noise error, y_{t-p} represents the value of the time series at lag p_{th}, and ϕ and θ are parameters of the ARMA model. The ARIMA model with an order of *(p, d, q)* can be represented by equation 3.

$$y'(t) = c + \epsilon_t + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \quad (3)$$

Here, $y'(t)$ is defined by differencing as $y'(t) = y(t) - y(t-1)$. Equation 4 can alternatively be expressed with the inclusion of first differencing.

$$y'(t) = y(t) - B^1(y(t)) \quad (4)$$

The operator B^1 in Equation 4 is referred to as the backward shift operator with an order of 1. The subsequent equations depict the SARIMA model with an order of *(p, d, q) x (P, D, Q)* $_L$.

$$\phi_{1_p}(B)\varphi_P(B^L)y(t) = \theta_q(B)\vartheta_Q(B^L)\epsilon_t \quad (5)$$

Here, L represents the seasonal period, and

$$\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \quad (6)$$
$$\theta_q(B) = 1 - \theta_q B - \cdots - \phi_q B^q, \quad (7)$$
$$\varphi_P(B^L) = 1 - \varphi_{1L} B^L - \cdots - \varphi_{PL} B^{PL} \quad (8)$$
$$\vartheta_Q(B^L) = 1 - \vartheta_{1L} B^L - \cdots - \vartheta_{QL} B^{QL} \quad (9)$$

Here, ϕ, θ, φ, and ϑ are parameters of the SARIMA model. ϕ represents the parameter for autoregressive, θ parameter for moving average, φ represents the parameter for seasonal autoregressive, and ϑ represents the parameter for the seasonal moving average.

3. Methods

In the forecasting research process using the SARIMA model, starting with determining objectives, data collection, data preprocessing, descriptive statistics, splitting data into train and test, stationarity test, model identification, model selection, model diagnostics, parameter evaluation, residual evaluation, model evaluation, and forecasting. The research process is described in Figure 1 as follows.

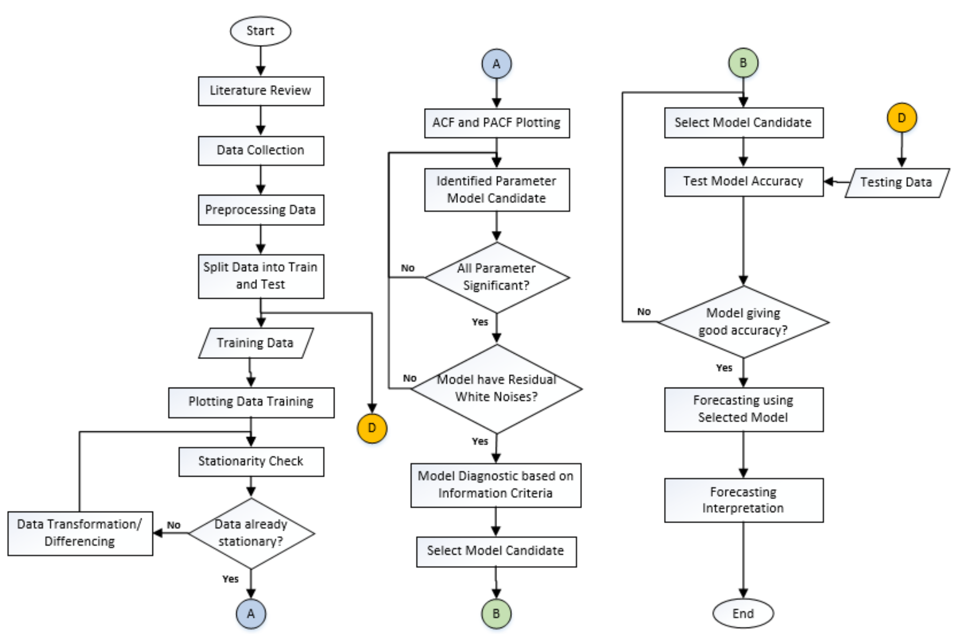


Figure 1. Research Flowchart

3.1 Model Identification

The preliminary specification stage is also referred to as model identification using the Box-Jenkins methodology that can only be applied effectively to stationary time series data. In these instances, the time series graph should ideally be analyzed, and data transformations should be made as needed. A time plot of the data should be first constructed, and any anomalies in the graph should be scrutinized. The variance would need to be stabilized if it was observed to increase over time (Chang, 2012; Damrongkulkamjorn & Churueang, 2005). Determining the ARIMA order (p,d,q) is typically based on the patterns observed in plots of ACF and PACF. The most critical elements are the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF). The Autocorrelation Function (ACF) quantifies the linear relationship between observations in a time series with a time lag of q. The Partial Autocorrelation Function (PACF) aids in determining the required number of autoregressive terms, p. The parameter d signifies the order of differencing needed to transform a non-stationary time series into a stationary one. Moreover, examining a time series plot and the ACF can provide insights into whether differencing is necessary. If differencing is utilized, the time plot will exhibit a discernible linear trend.

The determination of the appropriate values for p and q is based on the identification of significant lags in the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), as outlined by (Hasanah et al., 2021). A summary of this identification procedure is presented in Table 1.

Table 1. ACF and PACF Pattern Theoretically for Stationary Process

Models	ACF	PACF
AR (p)	Exponentially decreasing/sinusoidal	Cut off the significant level at lag p
MA (q)	Cut off the significant level at lag q	Exponentially decreasing/sinusoidal
ARMA (p,q)	Cut off the lag p and the vast majority decreasing after a lag	Cut off the lag q and the vast majority decreasing after lag q

In datasets that exhibit seasonality, it is common to employ ARIMA models that account for seasonal patterns. A seasonal ARIMA model extends the conventional ARIMA models by incorporating additional seasonal terms. This model is denoted as SARIMA *(p, d, q)(P, D, Q)m*, where the non-seasonal components are represented by the order (p, d, q), while the seasonal components are denoted by (P, D, Q), and *m* indicates the seasonality of the model. Once the initial values of D (seasonal difference) and d (non-seasonal difference) have been established, the subsequent step involves examining the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of seasonality to ascertain the appropriate values for Q and P. Subsequently, the selection of model parameters is guided by Akaike's Information Criterion (AIC), which aids in determining the optimal parameter values.

Box-Cox transformation

The initial step in forecasting involves data analysis to identify if the variances are stationary and to verify if the time series mean stationarity. When variances prove to be unstable, we can utilize the Box-Cox transformation. This traditional method aids in maintaining a stable residual variance (Halim et al., 2007). The formula for this transformation is as follows:

$$g(y) = \begin{cases} log\ (y),\ \lambda = 0 \\ \frac{y^{\lambda} - 1}{\lambda},\ \lambda \neq 0 \end{cases} \tag{10}$$

In the context of the analysis, y represents the original data, while λ is the selected parameter.

Unit Root Test

A time series is considered stationary if all the roots of its characteristic equation (such as equation (6)) have an absolute value greater than one, as stated by (Halim et al., 2007). In the case of an AR (1) model, the characteristic equation is given by $1-\phi B = 0$, which yields a root of $B = 1/\phi$. If the absolute value of this root exceeds one, the time series, denoted as y_t, is deemed stationary. Consequently, the AR (1) model demonstrates stationarity when $|\phi| < 1$. When the root equals one, it is referred to as a unit root.

1) Dickey-Fuller Test

Dickey and Fuller proposed three regression equations to identify the presence of unit roots, namely:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t \tag{11}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + \varepsilon_t \tag{12}$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_1 + \varepsilon_t \tag{13}$$

The Dickey-Fuller test is based on a set of hypotheses, which are:
$H_0: \gamma = 1\ (time\ series\ is\ not\ stationary)$
$H_0: \gamma < 1\ (time\ series\ is\ stationary)$

The primary version of the Dickey-Fuller (DF) test is expressed by Equation (11). This version does not consider the potential presence of drift and trend in the series. When Equation (11) matches the original data, but the value of y0 in the series is uncertain, it is recommended to introduce a constant, a0, into the regression model while testing for a unit root, as demonstrated in Equation (12). Additionally, if a time trend, t, is incorporated, the regression model specified in Equation (13) can be employed to test for a unit root. Testing continues toward more restricted specifications if the null hypothesis cannot be rejected using the most inclusive specification. The testing process ceases once we can reject the null hypothesis, which suggests the presence of a unit root.

2) Augmented Dickey-Fuller Test
When a basic AR (1) Dickey-Fuller (DF) model is used but the actual process followed by yt is an AR(p), the error term will exhibit autocorrelation to adjust for the incorrect representation dynamic structures of y_t. Therefore, if y_t follows an AR process of order p_{th}, the enhanced version of the Augmented Dickey-Fuller test can be applied, which is (Halim et al., 2007; Noureen et al., 2019):

$$y_t = \Psi_1 y_{t-1} + \Psi_2 y_{t-2} + \cdots + \Psi_p y_{t-p} + \varepsilon_t \qquad (14)$$
$$\Delta y_t = \Psi^* y_{t-1} + \Psi_1 \Delta y_{t-1} + \Psi_2 \Delta y_{t-2} + \cdots + \Psi_{p-1} \Delta y_{t-p+1} + \varepsilon_t \qquad (15)$$
Where, $\Psi^* = \left(\Psi_1 + \Psi_2 + \cdots + \Psi_p\right) - 1\ and\ \varepsilon_t \sim Independent\ and\ Identically\ Distributed\ (IID)(0, \sigma^2)$

The Augmented Dickey-Fuller test is based on a set of hypotheses, which are:
$H_0: \Psi^* = 0\ (there\ is\ a\ unit\ root)$
$H_1: \Psi^* < 0\ (there\ is\ no\ unit\ root)$

The p-value acquired from performing the ADF test determines the acceptance or rejection of the null hypothesis. For a confidence level of 95%, if the p-value is equal to or greater than 0.05, the null hypothesis is true. Conversely, if the p-value is less than 0.05, the value is significant enough to reject the null hypothesis, indicating that the time series is stationary.

Differencing
Differencing is a technique used to transform a time series into a stationary process by sequentially calculating the differences between consecutive observations. This approach aids in stabilizing the mean of the time series by removing variations in the time series' rate, thereby mitigating or eliminating trend and seasonality effects. ACF will decrease to zero relatively quickly for stationary time series, while non-stationary ACF will decrease slowly. The formula of differencing is shown as follows:

First order differencing:
$$y'_t = y_t - y_{t-1} \qquad (16)$$
Second-Order Differencing:
$$y''_t = y'_t - y'_{t-1} = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \qquad (17)$$
Seasonal Differencing:
$$y'_t = y_t - y_{t-m} \qquad (18)$$
Where y_t is the value observed, and m is seasonality.

Diagnostic Check
Akaike's Information Criterion (AIC) can be used to determine the order of the ARIMA model; this value evaluates the adequacy of a fitted statistical model by assessing its goodness of fit., which is obtained as follows. (Hyndman & Athanasopoulos, 2018)

$$AIC = -2\log(L) + 2(p + q + k + 1) \qquad (19)$$

In the above equation, represents the likelihood of the data, where k=1 if $c \neq 0$ and k = 0 if c = 0, and k is equal to 0 if c is equal to zero. The last term enclosed in parentheses denotes the count of parameters in the model, which encompasses the variance (σ^2) of the residuals. The model with the lowest Akaike Information Criterion (AIC) value is considered the most favorable. The approach of AIC aims to identify a model that provides the best representation of the data while utilizing the least number of free parameters.

Preliminary Parameter Estimation

Some of the model parameters do not exhibit statistical significance. The associated ratios are presented as

$$\left|\frac{Parameter}{1.96 \times Std.error}\right| > 1 \qquad (20)$$

The recommendation could be to experiment with a model where some parameters are assigned a value of zero [5]. After each parameter has been set to zero, it becomes necessary to re-evaluate the model.

Ljung-Box Statistic

The Ljung-Box statistic is used to evaluate whether the autocorrelations of a time series fall below zero (Halim et al., 2007). The computation of the test statistic proceeds as follows:

$$Q = T(T+2)\sum_{k=1}^{s}\frac{r_k^2}{T-k} \qquad (21)$$

Where,

- T: Number of observations
- s: length of coefficients to test autocorrelation
- r_k: autocorrelation coefficient (for lag k).

The Ljung-Box test involves the formulation of hypotheses, which include

- H_0: Residual is white noise
- H_1: Residual is not white noise

When the observed value of Q exceeds the critical value obtained from a chi-square distribution with s degrees of freedom, it can be concluded that, at the specified significance level, there exists at least one non-zero value of r that exhibits statistical significance.

3.2 Model Evaluation

The magnitude of the forecast error is aligned with the scale of the underlying data. Consequently, accuracy measures solely dependent on the data scale are inadequate for comparing series that involve varying units of measurement (Hyndman & Athanasopoulos, 2018). Fundamentally, the accuracy of prediction outcomes can be evaluated using several statistical analysis techniques (Mado et al., 2018), including root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The definition of each model evaluation is as follows:

$$Mean\ absolute\ error: MAE = mean(|e_t|) \qquad (22)$$

$$Root\ mean\ sqared\ error: RMSE = \sqrt{mean(e_t^2)} \qquad (23)$$

When evaluating forecasting methods on a single time series or multiple time series with consistent units, Mean Absolute Error (MAE) is widely recognized for its simplicity in interpretation and calculation. Optimal performance in terms of MAE corresponds to a median forecast, whereas minimizing Root Mean Square Error (RMSE) yields an average estimate. Consequently, the Root Mean Square. Error (RMSE) is frequently utilized, despite its comparatively more complex interpretation.

For a given percentage error is obtained by the following equation.

$$p_t = 100e_t/y_t \qquad (24)$$

The utilization of percentage error offers the benefit of being dimensionless, making it a commonly employed metric for comparing forecast performance across different datasets. The most commonly used measures include:

$$Mean\ absolute\ percentage\ error: MAPE = mean(|p_t|) \qquad (25)$$

4. Data Collection

The research data, comprising daily electricity consumption, was collected from the management company of Soekarno Hatta International Airport (CGK) for a year, from 01 January 2022 to 31 December 2022, as shown in Figure 2.

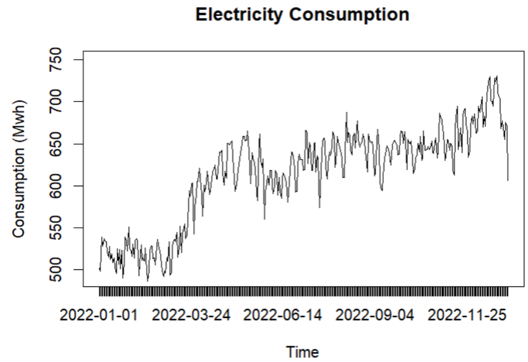


Figure 2. Electricity Consumption

Because these data tend to have trend, seasonality, and remainder characteristics, a time decomposition process is carried out to examine these characteristics. From Figure 3, it is shown that the data has regular seasonal characteristics at frequency = 7. Because the data is in the form of daily data, it can be assumed that the data has a weekly seasonal pattern.

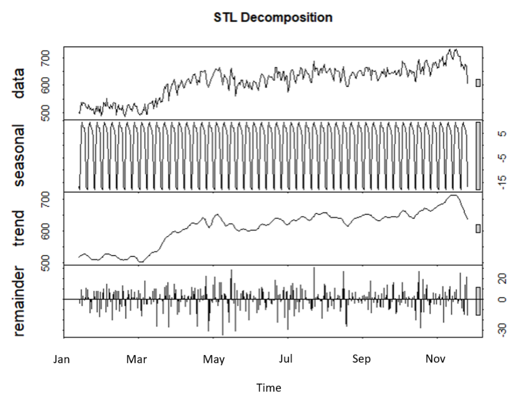
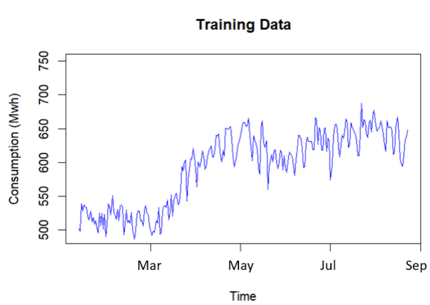
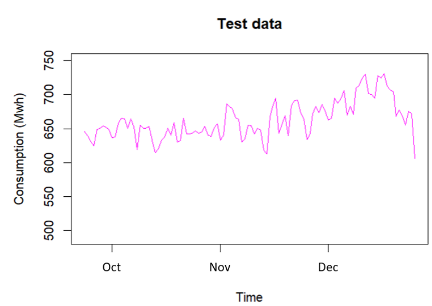


Figure 3. Data Time Series Decomposition

This dataset was subsequently divided into two segments: a training set and a test set shown in Figure 4. The training set, representing approximately 70% of the data, spans from 01 January 2022 to 12 September 2022. The remaining 30% of the data, running from 13 September 2022 to 31 December 2022, was reserved as the test set. The statistical description of the data is shown in Table 2.



(a) Training Data



(b) Test Data

Figure 4. Training Data and Test Data

Table 2. Statistical Descriptive

Parameter	Data Set	Training Data	Test Data
Number of Observation	365	255	110
Min	485.71	485.71	605.59
Max	730.45	687.20	730.45
Mean	613.39	592.88	662.20
Median	630.14	609.12	655.74
Variance	328.72	292.51	769.06
Std. Deviation	573.34	540.84	277.31

In the early stages of SARIMA modeling, stationary data to variance and mean is required. Where the stationarity test for variance is carried out on the training data, using the BoxCox function on RStudio, it is obtained that the Lambda value = -0.53 so that a BoxCox transformation is needed to change the training data to be stationary to variance. The results of the BoxCox transformation on the data are shown in Figure 5, which has the same shape as Figure 4 (a) but is stationary to variance.

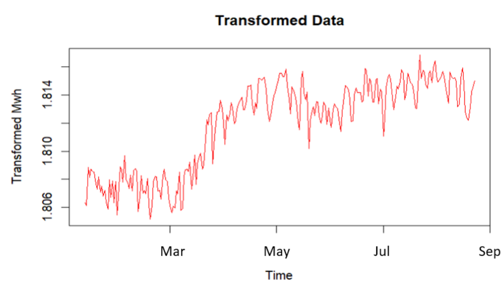
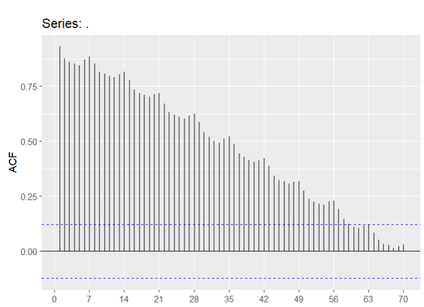
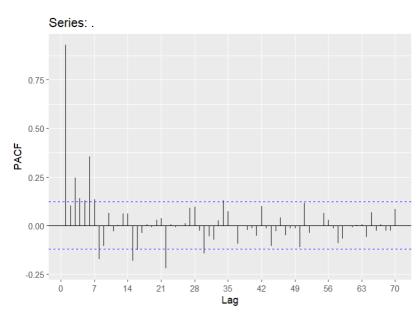


Figure 5. Transformed Training Data

The next step is to ensure that the data is stationary to the means by carrying out the ADF Test. The results of the ADF test on training have a p-value of 0.93 because the confidence interval is 5%, so it fails to evaluate H_0, and the data has a unit root or non-stationary. Before the differencing process, it is necessary to pay attention to the ACF plot and the PACF plot shown in Figure 6. The plot shows a seasonal effect for each lag 7, so seasonal differencing is needed, followed by differencing regular because it has a clear autocorrelation pattern, namely a slow decline.



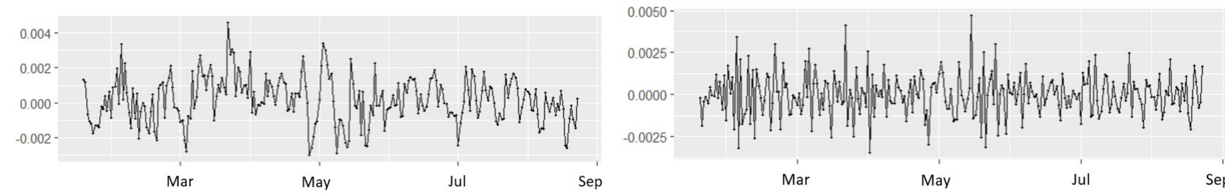
(a) ACF Plot



(b) PACF Plot

Figure 6. ACF Plot and PACF Plot of Transformed Data

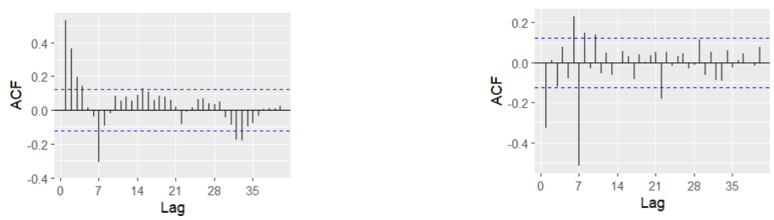
If seasonal differencing is only used, it will be seen in Figure 7 (a) that the plot still has a trend. After the first regular differencing is done at the continuation of the previous process, the data has no trend and is stationary to the mean, as shown in Figure 7 (b).



(a) Seasonal Differencing (b) First Regular Differencing

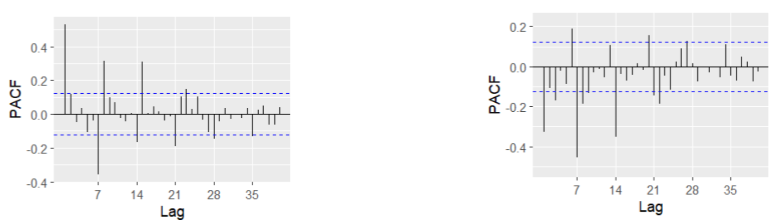
Figure 7. Seasonal Differencing and First Regular Differencing Data Plot

Thus, the ACF plot is shown in Figure 8, and the PACF plot is shown in Figure 9. From the plot in Figure 9, a candidate model can be made by looking at the lag that passes the confidence interval. Figure 8(a) and Figure 9(a) show the ACF plot and PACF plot respectively, where the seasonal differencing process is carried out. Meanwhile, Figure 8 (b) and Figure 9 (b) show the ACF plot and PACF plot which are then subjected to a regular differencing process after seasonal differencing.



(a) ACF Plot Seasonal Differencing (b) ACF Plot First Regular Differencing

Figure 8. ACF Plot: Seasonal Differencing and First Regular Differencing



(a) PACF Plot Seasonal Differencing (b) PACF Plot First Regular Differencing

Figure 9. PACF Plot: Seasonal Differencing and First Regular Differencing

For the non-seasonal model, it can be seen in Figure 8 (b) the ACF plot that the lag spikes at lag 1 and in Figure 9 (b) PACF is at lag 1,2,3, while for the seasonal model, there is a lag spike in Figure 8 (b) the ACF plot at lag 7, and PACF at lags 7 and 14, but at lag 8 in the Figure 9 (b) as PACF plot contains lag spikes so that the candidate models can be written as follows: MA (0,1), SMA (0,1), AR (0,1,2,3), and SAR (0,1,2).

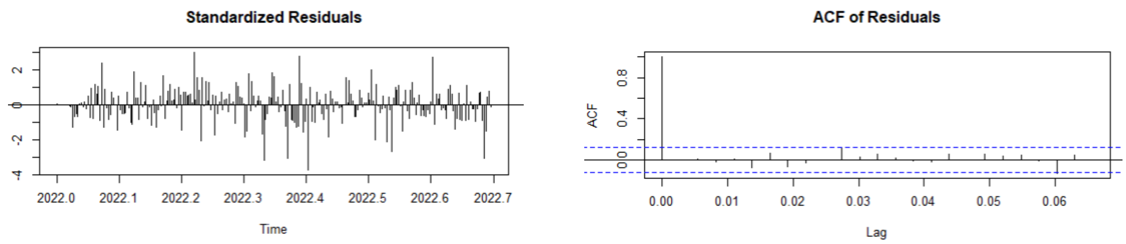
5. Results and Discussion

5.1 Numerical Results

This combination of parameters creates candidate models, which will be compared with the significance values of the parameters, residuals, and AIC values Each model has parameters d = 1 and D = 1 because they have experienced seasonal differencing with seasonal values = 7 and first regular differencing. Then a diagnosis of the model is carried out by looking at the significance of the parameters of each model, the absence of autocorrelation in the residuals (white noises), and the residuals are normally distributed. The model that has passed the diagnosis is then selected with the criterion of having the smallest AIC value. The best model chosen is SARIMA (1,1,1)(0,1,1)7 because it has the lowest AIC value of 2051, the parameters in this model are significant, then this model passes the diagnosis of residuals with residuals that do not have autocorrelation (white noise), has zero mean, and is normally distributed

5.2 Graphical Results

The SARIMA (1,1,1)(0,1,1)^7 model was chosen because it met the criteria, so the model will be used in the forecasting process later. The results of the diagnostic checking on the model are shown in Figure 10(a) for the residual model to have a zero mean, so the model residual are stationary. Figure 10 (b) shows that the residual model does not have a lag spike that exceeds the confidence interval, therfore the model residual have no autocorrelation, and Figure 11 shows that the model has residuals that follow a normal distribution.



(a) Model Residual Mean (b) Model Autocorrelation Residual

Figure 10. Model Residual Mean and Autocorrelation

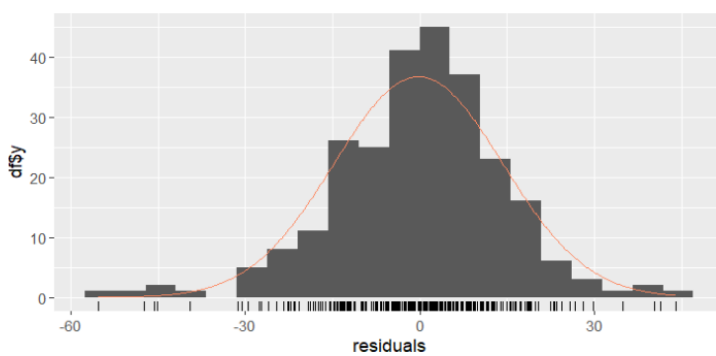


Figure 11. Histogram of Residual Distribution

The suitability of the model with the training data is shown in Figure 12. The figure shows that the model can capture patterns from data with an accuracy level of 1.83% MAPE compared with training data.

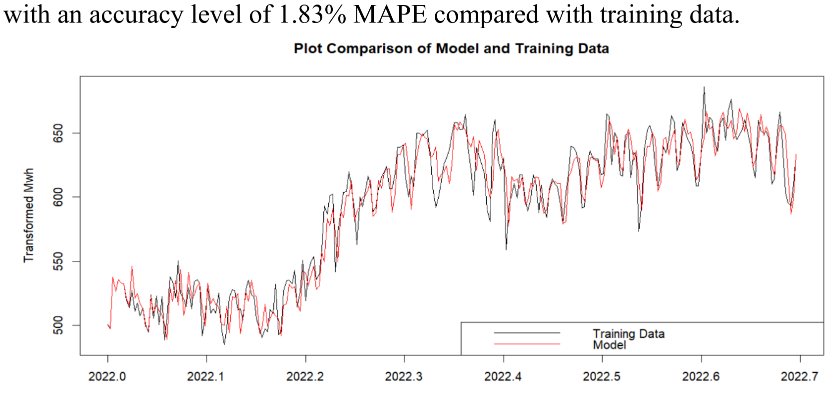


Figure 12. Plot Comparison of Model and Training Data

The following process is forecasting using the model for the following several periods, where the forecast results are compared with the previously prepared Test Data to measure the accuracy of the forecasting model. As a result, the model has an accuracy of MAE 31.37, RMSE 38.59, and MAPE 4.62%. A comparison plot of the forecasting model with Test Data is shown in Figure 13.

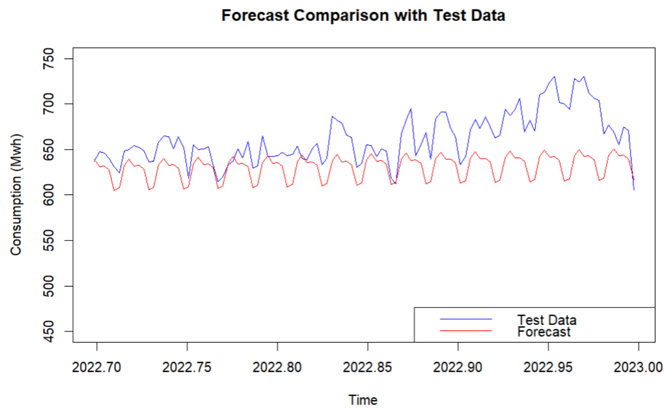


Figure 13. Model and Test Data Plot Comparison

5.3 Discussion

Because the data has a complex pattern, there may be more than one seasonal effect in the data used. Furthermore, there are various environmental factors, such as temperature, day type, and occupancy, which contribute to the dynamic nature of the airport electricity consumption model. The current time series model employed in this study exhibits limitations in effectively capturing extreme values. Future research endeavors could address this limitation by incorporating interventions in time series analysis, such as the inclusion of exogenous factors in Seasonal Auto-Regressive Integrated Moving Averages models.

6. Conclusion

This research introduces an ARIMA model that including the seasonal patterns of the time series called Seasonal Autoregressive Integrated Moving Average (SARIMA). The model was applied to analyze the daily electrical consumption data of Soekarno Hatta International Airport (CGK), a seasonal SARIMA (1, 1, 1) $(0, 1, 1)^7$ model is built with MAPE 4.62%. The findings demonstrate that the proposed model exhibited a good fit to the data, effectively capturing the stochastic seasonal fluctuations with the exception of a few extreme values. The predictions derived from this model suggest a gradual increase in electricity consumption for the upcoming month, aligning with the observed weekly seasonal patterns. This changing trend could be a recommendation for airport management to make proper energy and financial management strategies in response to electricity demand.

References

Abunofal, M., Poshiya, N., Qussous, R., and Weidlich, A. Comparative Analysis of Electricity Market Prices Based on Different Forecasting Methods. *2021 IEEE Madrid PowerTech*, 1–6, 2021. https://doi.org/10.1109/PowerTech46648.2021.9495034

Al-Shaikh, H., Rahman, Md. A., and Zubair, A. Short-Term Electric Demand Forecasting for Power Systems using Similar Months Approach based SARIMA. *2019 IEEE International Conference on Power, Electrical, and Electronics and Industrial Applications (PEEIACON)*, 122–126, 2019. https://doi.org/10.1109/PEEIACON48840.2019.9071939

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. *Time Series Analysis: Forecasting and Control* 5th Edition. Wiley. New Jersey. 2016

Chang, X., Gao, M., Wang, Y., and Hou, X. Seasonal Autoregressive Integrated Moving Average Model for Precipitation Time Series. *Journal of Mathematics and Statistics*, 8(4), 500–505, 2012. https://doi.org/10.3844/jmssp.2012.500.505

Damrongkulkamjorn, P., and Churueang, P. Monthly energy forecasting using decomposition method with application of seasonal ARIMA. *2005 International Power Engineering Conference*, 1–229, 2005. https://doi.org/10.1109/IPEC.2005.206911

Gunawan, J., and Huang, C.-Y. An Extensible Framework for Short-Term Holiday Load Forecasting Combining Dynamic Time Warping and LSTM Network. *IEEE Access*, 9, 106885–106894, 2021. https://doi.org/10.1109/ACCESS.2021.3099981

Halim, S., N. Bisono, I., Melissa, and Thia, C. Automatic seasonal auto regressive moving average models and unit root test detection. *2007 IEEE International Conference on Industrial Engineering and Engineering Management*, 1129–1133, 2007. https://doi.org/10.1109/IEEM.2007.4419368

Hasanah, P., Wiradinata, S., Azka, M., and Sari, C. Forecasting electricity loads on national holidays in East Kalimantan. *Journal of Physics: Conference Series*, 1821(1), 012009, 2021. https://doi.org/10.1088/1742-6596/1821/1/012009

Hong, T., and Fan, S. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3), 914–938, 2016. https://doi.org/10.1016/j.ijforecast.2015.11.011

Hyndman, R. J., and Athanasopoulos, G. *Forecasting Principles and Practice* 3rd Edition. OText.com, 2018.

Kang, M., Bergés, M., and Akinci, B. Forecasting airport building electricity demand on the basis of flight schedule information for demand response applications. *Transportation Research Record*, 2603, 29–38, 2017. https://doi.org/10.3141/2603-04

Mado, I., Soeprijanto, A., and Suhartono, S. Applying of Double Seasonal ARIMA Model for Electrical Power Demand Forecasting at PT. PLN Gresik Indonesia. *International Journal of Electrical and Computer Engineering (IJECE)*, 8(6), 4892, 2018. https://doi.org/10.11591/ijece.v8i6.pp4892-4901

Noureen, S., Atique, S., Roy, V., and Bayne, S. Analysis and application of seasonal ARIMA model in Energy Demand Forecasting: A Case Study of Small-Scale Agricultural Load. *2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS)*, 521–524, 2019. https://doi.org/10.1109/MWSCAS.2019.8885349

Tang, X., Chen, H., Xiang, W., Yang, J., and Zou, M. Short-Term Load Forecasting Using Channel and Temporal Attention Based Temporal Convolutional Network. *Electric Power Systems Research*, 205, 107761, 2022. https://doi.org/10.1016/j.epsr.2021.107761

Vagropoulos, S. I., Chouliaras, G. I., Kardakos, E. G., Simoglou, C. K., and Bakirtzis, A. G. Comparison of SARIMAX, SARIMA, modified SARIMA, and ANN-based models for short-term PV generation forecasting. *2016 IEEE International Energy Conference (ENERGYCON)*, pp. 1–6, 2016. https://doi.org/10.1109/ENERGYCON.2016.7514029

Biographies

Afwan H. Cahya is currently a Master student in Industrial Engineering at the University of Indonesia. He also holds a degree in Mechanical Engineering from Trisakti University and a diploma in Mechatronics Engineering from the Bandung State Manufacturing Polytechnic. Currently, he works as an Electrical Engineer in an airport operator company and has experience as a Senior Aircraft Component Technician in an aircraft MRO company.

Zulkarnain is a lecturer in the Industrial Engineering Department, Faculty of the Engineering University of Indonesia. He earned his Bachelor of Engineering degree in Industrial Engineering from University of Indonesia, Master degree in Industrial Engineering from University of Indonesia, and Doctor of Science (Technology) degree in Industrial Engineering and Management from University of Oulu. His industrial engineering topics and fields of interest are operations research; statistics; transport, logistics, and supply chain management; quality engineering; ergonomics; product design and development; marketing; operation management; and industrial management.