Jurnal Ilmiah KURSOR
Menuju Solusi Teknologi Informasi
Vol. 12, No. 2, Desember 2023

ISSN 0216 – 0544
e-ISSN 2301– 6914

SHORT-TERM FORECASTING DAILY ELECTRICITY LOADS USING SEASONAL ARIMA PATTERNS OF GENERATION UNITS AT PT. PLN (PERSERO) TARAKAN CITY

[a]Ismit Mado, [b]Achmad Budiman, [c]Aris Triwiyatno

[a,b]Department of Electrical Engineering, Universitas Borneo Tarakan, Jl. Amal Lama No. 1, Tarakan 77115, Indonesia
[c]Department of Electrical Engineering, Universitas Diponegoro, Jl. Prof. Soedarto, SH, Tembalang, Semarang 50275, Indonesia
E-mail: ismitmado@borneo.ac.id, achmad1177@gmail.com, aristriwiyatno@undip.ac.id

Abstract

Electrical power requirements at load centers tend to change over time, so the State Electricity Company (PT. PLN) as a provider of electrical energy must be able to predict electrical load requirements every day. The city of Tarakan as a reference center in the northern region of Indonesia is developing rapidly. Along with this growth, the need for electric power is of course also increasing, so we must be able to provide an economical and reliable electric power supply system. This research aims to predict the electricity load at PT. PLN (Persero) Tarakan City. The author will carry out short-term forecasting using time series data in the form of daily electrical power usage data using the Autoregressive Integrated Moving Average (ARIMA) method. The ARIMA model or often called the Box-Jenkins technique is a method suitable for forecasting a number of variables quickly and simply because it only requires variable data. Analysis based on the Box-Jenkins time series taking into account the influence of seasonal patterns. Forecasting results show the data contains a daily seasonal pattern with a Mean Absolute Percentage Error (MAPE) of 3 percent.

Key words: electrical load, mape, sarima, time series.

INTRODUCTION

In general, all activities carried out by humans are based on forecasting and decision making in situations that contain uncertainty. Electrical load forecasting studies are needed in the operation of electric power systems, be it in operating systems, load growth, scheduling of generation systems and also maintaining the stability of the electric power system. Estimating the electric power generated so that it is the same as that consumed is also part of the forecasting study. An imbalance in electrical power will result in blackouts or otherwise a waste of electrical energy. Much literature describes the techniques and methods applied in forecasting studies. These include analytical approaches based on linear regression[1]–[3], time series approaches[1]–[5], and artificial intelligence such as artificial neural networks[4], [6] and fuzzy logic[7]. Likewise, this forecasting study is used to predict future electrical loads, to achieve a balance of electrical power in the generating system and load.

Time series forecasting studies use a quantitative approach with past data collected and used as a reference for the future. Time series forecasting techniques are divided into two parts. First, statistical mathematical models such as moving averages, exponential smoothing, regression and ARIMA[8], [9]. Second, artificial intelligence models such as neural networks, genetic algorithms, simulated annealing, and generic programming. However, the problems encountered in the real world are non-stationary or non-linear data. An accurate and effective tool is needed to predict the behavior of the data. Besides that, the problem of forecasting using neural networks that is often encountered is the problem of overfitting, where the model that is made only produces output for the data that is trained and not for the data that is validated.

Research that leads to the study of time series electric load forecasting is growing, especially the Box-Jenkins ARIMA model. Researchers have carried out short-term electricity load forecasting studies[5][10]–[12]. Time series modeling related to energy management and estimation of electricity load prices[13], [14].

In this study, short-term electrical load forecasting modeling will be carried out using time series ARIMA at the generation unit of PT. PLN (Persero) Tarakan City. The generating system runs continuously and works based on customer requests. The results of this study are expected to get the best data forecasting model for daily electricity loads. The data used in this study is univariate data taken from the electricity load production process during January 2019 from a total of eleven types of PLTU, PLTG, and PLTMG generation.

This study aims to determine the characteristics of the daily electricity load data for the Generation Unit of PT. PLN (Persero) Tarakan City in 2019. Next, make a model and predict the daily electricity load of PT. PLN (Persero) Tarakan City based on training data and daily electricity load testing in January 2019.

MATERIAL AND METHODS

Forecasting Short Term Electricity Load

Forecasting is the prediction of the values of a variable to a known value of that variable or a related variable. Forecasting can also be based on judgmental expertise, which in turn is based on historical data and experience. Short-term electric load forecasting is for a period of several hours to one week, taking into account various information that causes the electrical load to fluctuate in the system[15].

The load on the electric power system is the consumption of electric power from electricity customers. Therefore the size of the load and its changes depend on the needs of the customers for electricity. There is no exact calculation of how much the system load will be at any one time, what is usually done is to make an estimate of the load.

In the operation of the electric power system, efforts are made so that the generated power is not wasted or vice versa[16]. So the problem of load estimation is a very decisive issue for electricity companies, both from a managerial and operational perspective, and therefore needs special attention. To be able to make good load estimates, it is necessary to analyze the daily load data of electric power that has occurred in the past.

Time Series Modeling

The ARIMA method, also known as the Box-Jenkins method, is a model intensively developed by George Box and Gwilyn Jenkins in 1970. However, the Box-Jenkins ARIMA forecasting model still dominates many areas of research to date. ARIMA uses two algorithms, namely autoregressive (AR) and moving average (MA), and includes integrated elements to handle non-stationarity with the differencing method[17]. One of the stationary conditions is that the data pattern does not significantly contain a trend pattern[18]. ARIMA uses two algorithms, namely autoregressive (AR) and moving average (MA), and includes integrated elements to handle non-stationarity with the differencing method. To see the nature of MA, use the autocorrelation function (ACF) which represents the relationship between series of observations in a time series. MA is expressed as the number of ACF values from lag 1 to lag k sequentially which lie outside the confidence interval Z. Meanwhile, to see the nature of AR, we use the partial autocorrelation function (PACF). In general, the ACF and PACF values are 1 or 2, it is rare to find AR and MA properties with values greater than 2.

Generally, a non-seasonal time series can be modeled as a combination of past values and past errors, denoted as ARIMA(p, d, q) can be written as:

$$\phi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t \quad (1)$$

With $\emptyset_p(B) = \left(1 - \emptyset_1 B - \ldots - \emptyset_p B^p\right)$ is an AR(p) operator, $\theta_q(B) = \big(1 -$

$\theta_1 B - \ldots - \theta_q B^q)$ is an MA(q) operator. Where B is the backward shift operator and a_t is the purely random process.

In practice, time series contain seasonal periodicities. The ARIMA model is generalized so that it is defined as the Seasonal ARIMA model (SARIMA). The general ARIMA procedure is represented as follows[19]:

$$\emptyset_p(B)\Phi_P(B^s)(1-B^s)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (2)$$

With $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is a non-seasonal operator AR(p), $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_P B^{Ps}$ is a seasonal operator AR(P), $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q$ is a non-seasonal MA(q), $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_Q Q^{Qs}$ is a seasonal operator MA(Q). Where B is the backward shift operator and a_t is the purely random process.

The original series Z_t is differenced by appropriate differencing to remove non-stationary terms. While $(1-B)^d$ dan $(1-B)^D$ are the non-seasonal and seasonal differencing operators. If the integer D is not zero, then seasonal differencing is onvilved. The above model is called a SARIMA model of $(p, d, q)(P, D, Q)$. If d is not-zero, then there is a simple differencing to remove trend, while seasonal differencing $(1-B^s)^D$ may be use to remove seasonality. Basically, the value of d and D are usually zero or one and rarely two.

ARIMA Box-Jenkins Model Procedure

Data analysis was carried out using the ARIMA method with the help of statistical software, namely MINITAB 14. The steps for applying the ARIMA method successively are[17]:

a. preparation of data, including checking of data stationary.
b. model identification in ARIMA. Through the ACF and PACF plots we can determine the ARIMA model that can be used in predictions.
c. determination of parameters p, d and q in ARIMA.
d. determination of the ARIMA model equation. The coefficients used are generated from the results of the parameter analysis of the ARIMA model with the smallest MSE (mean square error).
e. predictive parameter validation.
f. prediction.

The next step is to use the best model for prediction. If the best model has been determined, the model is ready to be used for predicting the daily electricity load at the Generation Units of PT. PLN (Persero) Tarakan City due to usage on the load side.

RESULT AND DISCUSSION

The data used in this study are secondary data, namely electricity production data at the Generation Unit of PT. PLN (Persero) Tarakan City throughout January 2019. The data is data that is measured every one hour. The data used for training is 336 and the data used for testing is 24 data. The research variable used in this study is Z_t, which is the electrical load in MWH units.

Data Stationarity

The initial step is in the form of plotting the daily electricity load data at the PT. PLN (Persero) Tarakan City throughout January 2019 is shown in Figure 1.

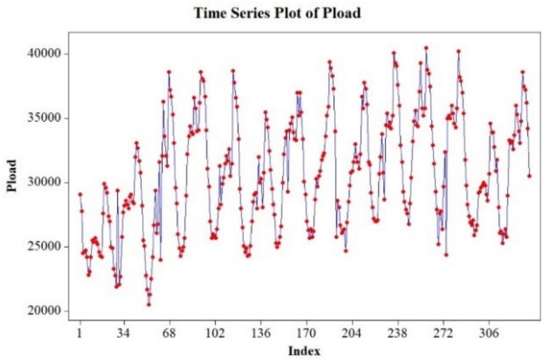

Fig 1. Time series plot of P_{load}.

Figure 2 shows the daily electricity load data at PT. PLN (Persero) Tarakan experiences fluctuations and tends to experience trends as in the data plot with the influence of the trend with the regression equation $Z_t = 27789 + 16.7t$.

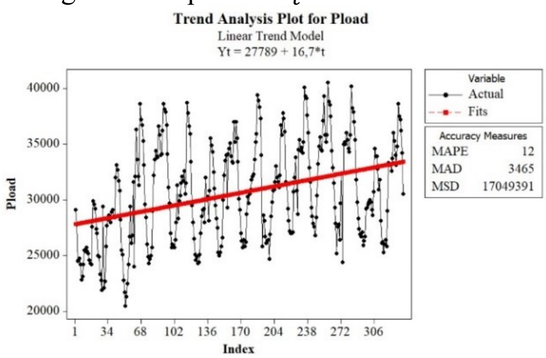

Fig 2. Trend analysis plot for P_{load} .

The temporary conclusion shows that the data is not stationary because it is still changing over time. Apart from using data plots to determine the characteristics of the data, data plots can also be observed through the coefficients of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF).

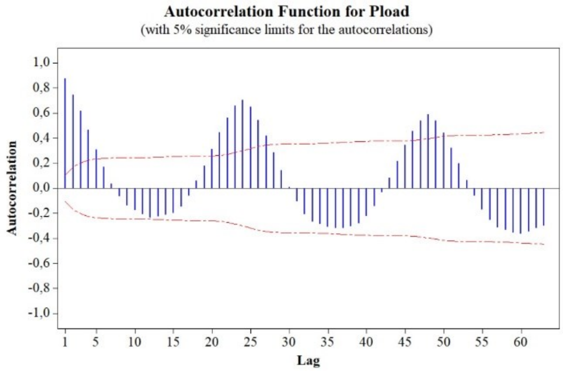

Fig 3. The ACF of Z_t.

From Figure 3 it can be seen that the autocorrelation coefficient differs significantly from zero and decreases slowly. Meanwhile, all partial autocorrelation coefficients are close to zero after the first lag. It can be seen in Figure 4 that the lag parameter MA cuts off but in fact it still indicates it dies down after the second lag. Referring to the $|T| > 2$ data for the initial lags, it indicates that the data is not stationary at the mean, even though the ARIMA method requires data to be stationary[19].

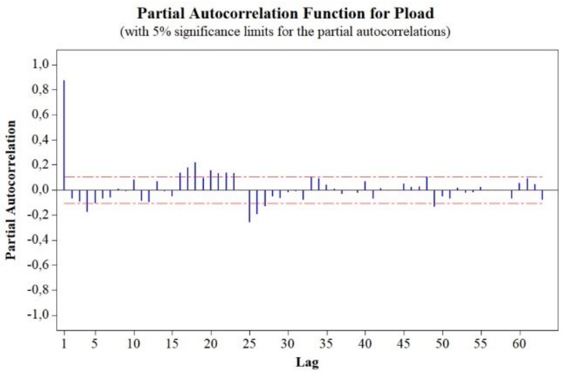

Fig 4. The PACF of Z_t.

Because of that it is necessary to do a differencing method. The results can be seen in Figure 5 below.

In Figure 5 it can be seen that the data has gone through a level 1 differentiation process. From this data it can be observed that there is data that is already stationary. The differentiation process that has been carried out indicates that the value of d that can be used is the value of $d = 1$.

After the data is stationary, a plot is carried out to see the ARIMA model to be used. By looking at the ACF and PACF plots as shown in Figure 6 and Figure 7.

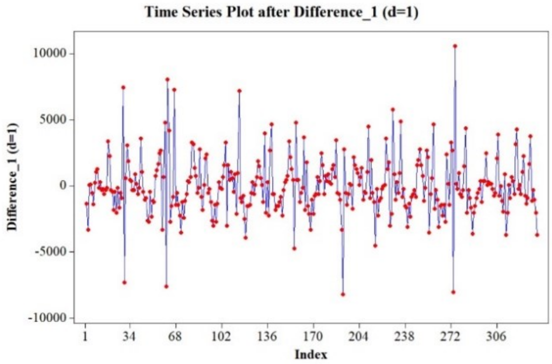

Fig 5. Time series plot after difference.

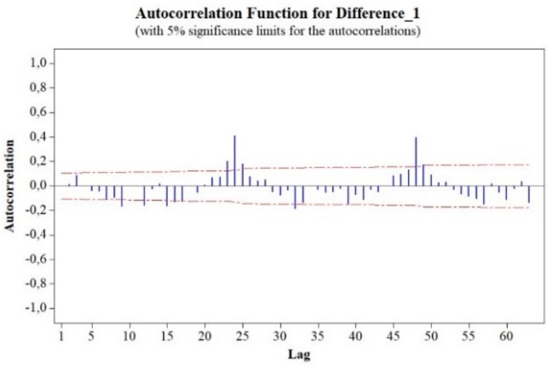

Fig 6. Plot ACF after difference $(d = 1)$.

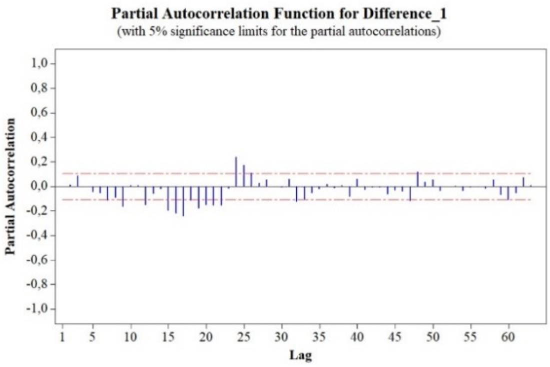

Fig 7. Plot PACF after difference $(d = 1)$.

It can be seen that the ACF data at the beginning of the lag dies down and cuts off at lag 5 in Figure 6. Likewise, the PACF data at the initial lag tends to die down and will cut off at lag 7 and then the data pattern indicates a cut off in Figure 7. The basis used to determine the model is as follows [8]:

- If the ACF shows a dying down pattern, and PACF shows a cut off, then it can be said that the ARIMA model is pure AR.
- If ACF shows a cut off pattern, and PACF shows dying down, then it can be said that the ARIMA model is a pure MA.
- If ACF and PACF show dying down, it can be said that the ARIMA model is a combination of AR and MA

From Figures 6 and 7 it can be seen that the ACF and PACF coefficients decrease exponentially. If ACF and PACF decrease exponentially or rapidly, this indicates that the ARIMA model that can be used is the ARIMA(p, d, q).

To find out more detailed ARIMA parameter values, the p, d and q parameters are determined. The level of prediction accuracy can be measured by the smallest MSE value so that a decision can be made if a model is feasible to use[20].

Because based on the correlation data after the differencing process, it is found that there are indications that both ACF and PACF die down, so the model is an ARIMA$(1,1,1)$. However, because of the cutoffs on the AR and MA parameters, there is a possibility that it will also be ARIMA$(0,1,1)$ or ARIMA$(1,1,0)$. To test the significance of determining the ARIMA model, it is necessary to test several models by looking at the lowest MSE value from the identification of the ARIMA stationary model.

Table 1. Determination of p, d and q with MSE

$d = 1$		$d = 2$	
Model	MSE	Model	MSE
(0,1,0)	-	(0,2,0)	-
(1,1,0)	5006097	(1,2,0)	7400374
(2,1,0)	5019925	(2,2,0)	6316277
(3,1,0)	4994964	(3,2,0)	6050771
(0,1,1)	5006097	(0,2,1)	-
(1,1,1)	5019487	(1,2,1)	-
(2,1,1)	5027700	(2,2,1)	-
(3,1,1)	5010097	(3,2,1)	-
(0,1,2)	5019855	(0,2,2)	-
(1,1,2)	5034259	(1,2,2)	5084202
(2,1,2)	-	(2,2,2)	-
(3,1,2)	5021423	(3,2,2)	-
(0,1,3)	4990631	(0,2,3)	-
(1,1,3)	-	(1,2,3)	-
(2,1,3)	-	(2,2,3)	-
(3,1,3)	-	(3,2,3)	-

From table 1 there are 32 models that can be used to predict electrical loads, but based on theory the smaller the MSE value produced by a model, the better the model. Unfilled MSE values indicate that the parameter cannot be estimated by Minitab. So it can be seen that the parameters $p = 0$, $d = 1$ and $q = 3$ or ARIMA$(0,1,3)$with the smallest MSE value of 4990631 can be used to predict PT. PLN (Persero) Tarakan City.

The parameter determination results were obtained by ARIMA$(0,1,3)$ with the following coefficients: MA$(1,1) = 0.5549$; MA$(1,2) = -0.021$; MA$(1,3 = -0.0963)$ and Constant$=$ 0.4, an equation analysis for ARIMA$(0,1,3)$ becomes:

$$Z_t = 0{,}4 + (1 - 0{,}5549B + 0.021B^2 + 0{,}0963B^3)a_t \quad (3)$$

Based on the results of forecasting data, the Mean Absolute Percentage Error (MAPE) is 7 percent. In this study it needs to be analyzed again. If you refer to Figure 1, you can see that the data contains seasonal patterns every week. And based on Figure 6, it can be seen that lag 12 and lag 24 in the ACF lag are still not significant. For this reason, it is necessary to re-analyze by including seasonal patterns.

The first step in building a seasonal ARIMA (SARIMA) model is to determine the amount of differentiation time to change the average data to stationary for both non-seasonal data and seasonal data[10].

The seasonal trend can be seen in the ACF coefficient of the results of the difference $(d = 1)$, where it can be seen that the data is not significant at lags 12 and 24 in Figure 6 with the value $|T|$ the dominance is < 2, but at lag 12, 24 has a value of > 2.

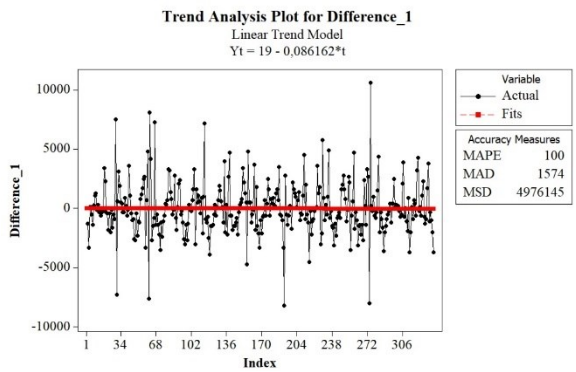

Fig 8. Trend analysis plot after $(d = 1)$.

Because there is a seasonal pattern every week and the ACF data in Figure 6 is still not significant at lags 12 and 24, it is necessary to difference the seasonal pattern again $(d = 1, D = 12)$.

Figures 9 and 10 show the ACF and PACF plots after seasonal difference.

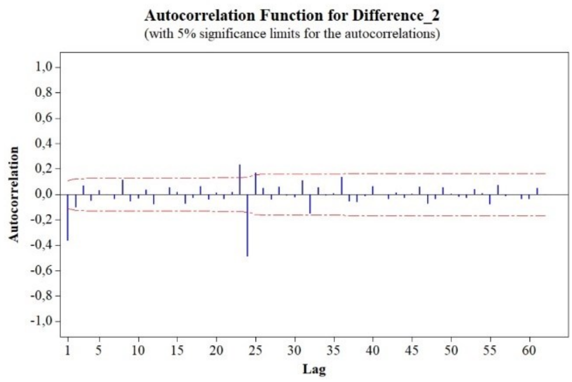

Fig 9. Plot ACF after $(d = 1, D = 12)$.

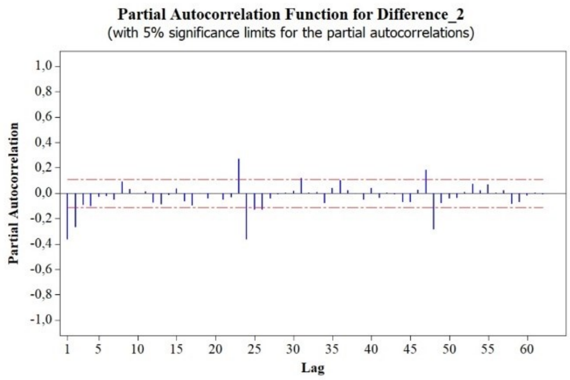

Fig 10. Plot PACF after $(d = 1, D = 12)$.

Determination of dying down or PACF for SARIMA if the pattern is said to be cut off, as follows:

a. the correlation coefficient is not significant at lag 2 or less for non-seasonal lag. It is said to be insignificant if $|T| < 2$ for non-seasonal, and
b. not significant at lag $+2$ or less for seasonal lag. It is said to be insignificant if $|T| < 1.25$.

If If seen from the plot criteria, the ACF pattern is subsiding which indicates that the AR element has seasonal value. Meanwhile, the PACF plot has a cut-off pattern because the lag value is $+2$ or less significant < 1.25.

For this reason, it can be assumed that the seasonal ARIMA model is $p = 1$, $d = 1$ and $q = 1$ or ARIMA$(1,1,1)$. Then by looking at the ACF and PACF plots for non-seasonal and seasonal data, respectively, it can be seen that the suggested models are $(0,1,3), (3,1,0), (0,1,1)$ and $(1,1,0)$ for non-seasonal and for seasonal are $(0,1,1), (1,1,0)$ and $(1,1,1)$. So there will be nine combinations of the SARIMA model with a periodicity of 12. MSE comparison for each SARIMA model

Table 2. Determination of p, d and q with MSE

Model	MSE
$(0,1,3)(0,1,1)^{12}$	2291026
$(0,1,3)(1,1,0)^{12}$	3455142
$(0,1,3)(1,1,1)^{12}$	2235663
$(3,1,0)(0,1,1)^{12}$	2337901
$(3,1,0)(1,1,0)^{12}$	3486753
$(3,1,0)(1,1,1)^{12}$	2256730
$(0,1,1)(0,1,1)^{12}$	2295398
$(0,1,1)(1,1,0)^{12}$	3487986
$(0,1,1)(1,1,1)^{12}$	2240484
$(1,1,0)(0,1,1)^{12}$	2546517
$(1,1,0)(1,1,0)^{12}$	3782306
$(1,1,0)(1,1,1)^{12}$	2475842

The results of determining the parameters obtained by the best model based on MSE values are SARIMA$(0,1,3)(1,1,1)^{12}$ with the following coefficients: AR(2,1) = -0,1561; MA(1,1) = 0.4984; MA(1,2) = 0.1021; MA(1,3) = -0.0899; MA(2,1) = 0.9150 and Constant = -1.40. The equation analysis for ARIMA$(0,1,3)(1,1,1)^{12}$ becomes:

$$(1 + 0{,}1561B)Z_t = (1 - 0{,}4984B - 0{,}1021B^2 + 0{,}0899B^3)(1 - 0{,}9150B^{12})a_t \quad (4)$$

Based on the prediction results, a comparison of the electrical load with actual data (testing) with a MAPE of 3.23 percent is obtained, as shown in Figure 11 below.

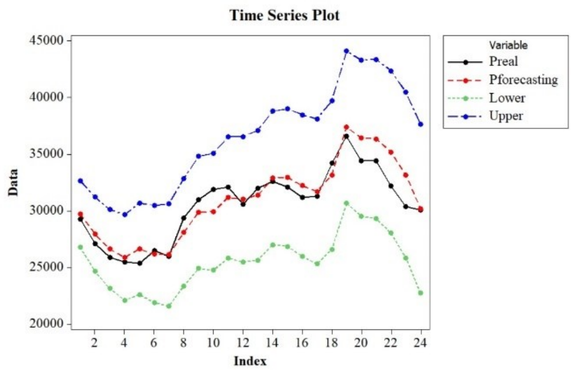

Fig 11. Plot the actual data model and the predicted results.

CONCLUSION

Based on the analysis and discussion that has been carried out, it is concluded that the prediction of the electrical load on the customer side of the Generation Unit of PT. PLN

(Persero) Tarakan City every hour with a daily peak load of 37397.4 MW and the lowest of 25907.6 MW. The highest average consumption occurred at 18:00 WIB and the lowest occurred at 02:00. In the analysis of electric load forecasting by taking into account seasonal patterns, namely model the $\mathrm{SARIMA}(0,1,3)(1,1,1)^{12}$.

ACKNOWLEDMENT

This research is an internal grant from the University of Borneo Tarakan with a Collaborative Research scheme in 2022. As an appreciation, gratitude is given to the Institute for Research and Community Service (LPPM), Universitas Borneo Tarakan for its cooperation so far.

REFERENCES

[1] S. Djanali and H. Studiawan, "REAL TIME ESTIMATION AND MONITORING OF POWER IN HOMOGENEOUS SERVER CLUSTER USING MULTIPLE LINEAR REGRESSION," *J. Ilm. Kursor*, 2012, [Online]. Available: http://kursorjournal.org/index.php/kursor/article/view/25%0Ahttps://kursorjournal.org/index.php/kursor/article/download/25/18

[2] A. Ubaidillah and S. I. Kholida, "PLANNING OF 5G NETWORK PATH LOSS IN GEOMETRY BASED STOCHASTIC CONCEPT BY USING LINEAR REGRESSION METHODS," *J. Ilm. Kursor*, vol. 10, no. 4, 2020, doi: 10.21107/kursor.v10i4.245.

[3] N. H. Harani, H. T. Siregar, and C. Prianto, "Implementation of Multiple Linear Regression Methods as Prediction of Village Spending on Village Financial Management System," *J. Ilm. Kursor*, vol. 10, no. 2, 2020, doi: 10.21107/kursor.v10i2.216.

[4] J. Buitrago and S. Asfour, "Short-term forecasting of electric loads using nonlinear autoregressive artificial neural networks with exogenous vector inputs," *Energies*, 2017, doi: 10.3390/en10010040.

[5] I. Mado, A. Soeprijanto, and Suhartono, "Applying of double seasonal arima model for electrical power demand forecasting at pt. pln gresik Indonesia," *Int. J. Electr. Comput. Eng.*, 2018, doi: 10.11591/ijece.v8i6.pp.4892-4901.

[6] S. Gregorius, R. Adipranata, and B. Setiahadi, "Solar Storm Type Classification Using Probabilistic Neural Network Compared With the Self-Organizing Map," *Kursor*, vol. 6, no. 4, pp. 211–220, 2012, [Online]. Available: https://journal.trunojoyo.ac.id/kursor/article/view/1070

[7] B. D. Saputra, "A FUZZY TIME SERIES-MARKOV CHAIN MODEL TO FORECAST FISH FARMING PRODUCT," *Kursor*, vol. 9, no. 4, 2019, doi: 10.28961/kursor.v9i4.167.

[8] R. T. Jurnal, "STUDI PERAMALAN BEBAN RATA – RATA JANGKA PENDEK MENGGUNAKAN METODA AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA," *Sutet*, vol. 7, no. 2, pp. 93–101, 2018, doi: 10.33322/sutet.v7i2.84.

[9] N. Pooniwala and R. Sutar, "Forecasting Short-Term Electric Load with a Hybrid of ARIMA Model and LSTM Network," in *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*, 2021. doi: 10.1109/ICCCI50826.2021.9402461.

[10] N. H. Miswan, R. M. Said, and S. H. H. Anuar, "ARIMA with regression model in modelling electricity load demand," *J. Telecommun. Electron. Comput. Eng.*, vol. 8, no. 12, pp. 113–116, 2016.

[11] Y. Ardilla and S. Suhartono, "Short Term Forecasting of Electricity Load: A Comparison of Methods to Paiton Subsystem East Java & Bali," *J. Inf. Technol. Comput. Sci.*, 2019, doi: 10.25126/jitecs.201943159.

[12] E. Chodakowska, J. Nazarko, and Ł. Nazarko, "Arima models in electrical load forecasting and their robustness to noise," *Energies*, vol. 14, no. 23, 2021, doi: 10.3390/en14237952.

[13] J. Contreras, R. Espínola, F. J. Nogales, and A. J. Conejo, "ARIMA models to

predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, 2003, doi: 10.1109/TPWRS.2002.804943.

[14] B. Nepal, M. Yamaha, A. Yokoe, and T. Yamaji, "Electricity load forecasting using clustering and ARIMA model for energy management in buildings," *Japan Archit. Rev.*, vol. 3, no. 1, pp. 62–76, 2020, doi: 10.1002/2475-8876.12135.

[15] E. N. Yaqin *et al.*, "Short Term Load Forecasting for Weekends in Indonesia: Comparison of Three Methods," in *IOP Conference Series: Materials Science and Engineering*, 2018. doi: 10.1088/1757-899X/288/1/012119.

[16] M. A. Hammad, B. Jereb, B. Rosi, and D. Dragan, "Methods and Models for Electric Load Forecasting: A Comprehensive Review," *Logist. Sustain. Transp.*, 2020, doi: 10.2478/jlst-2020-0004.

[17] E. R. Ziegel, G. Box, G. Jenkins, and G. Reinsel, "Time Series Analysis, Forecasting, and Control," *Technometrics*, 1995, doi: 10.2307/1269640.

[18] A. Azeem, I. Ismail, S. M. Jameel, and V. R. Harindran, "Electrical Load Forecasting Models for Different Generation Modalities: A Review," *IEEE Access*, vol. 9. pp. 142239–142263, 2021. doi: 10.1109/ACCESS.2021.3120731.

[19] S. C. Hillmer and W. W. S. Wei, "Time Series Analysis: Univariate and Multivariate Methods.," *J. Am. Stat. Assoc.*, 1991, doi: 10.2307/2289741.

[20] Apriliyah, W. F. Mahmudy, and A. W. Widodo, "Perkiraan Penjualan Beban Listrik Menggunakan Jaringan Syaraf Tiruan Resilent Backpropagation (Rprop)," *Kursor*, no. 2, pp. 41–47, 2007.

.