Short-Term Electricity Load Forecasting Model Based DSARIMA

Ismit Mado*
Dept. of Electrical Engineering
Universitas Borneo Tarakan
Tarakan, Indonesia
ismitmado@borneo.ac.id

Antonius Rajagukguk
Dept. of Electrical Engineering
Universitas Riau
Pekanbaru, Indonesia
antoniusrajagukguk@gmail.com

Aris Triwiyatno
Dept. of Electrical Engineering
Universitas Diponegoro
Semarang, Indonesia
aristriwiyatno@undip.ac.id

Arif Fadllullah
Dept. of Computer Engineering
Universitas Borneo Tarakan
Tarakan, Indonesia
arif.fadl@borneo.ac.id

Corresponding author: ismit mado, ismitmado@borneo@ac.id

Abstract—**Forecasting short-term electrical load is very important so that the quality of the electrical power supplied can be maintained properly. The study was conducted to measure the results of electrical load forecasting based on parameter estimates and the presentation of time series data. It is important to manage stationary data, both in terms of mean and variance. Data presentation is done by determining the value of variance through the Box-Cox transformation method and the mean value based on the ACF and PACF plots. This study considers the pattern of electricity consumption which contains double seasonal patterns. The results of previous studies show the electric power prediction model, the DSARIMA model $([1, 2, 7, 16, 18, 35, 46], 1, [1, 3, 13, 21, 27, 46]) (1, 1, 1)^{48} (0, 0, 1)^{336}$ with a MAPE of 2.06%. The condition of the model used to predict the electrical load still has a tendency not to be normally distributed and it is estimated that there are outliers. Improvements to the AR and MA parameters that meet the standard error tolerance value of 5 percent are increased in this study. The results showed improvement of parameters to predict electrical load with DSARIMA model $([1, 2, 5, 6, 7, 11, 16, 18, 35, 46], 1, [1, 3, 13, 21 27, 46]) (1, 1, 1)^{48} (0, 0, 1)^{336}$. The significance of this study was obtained by the MAPE value of 1.56 percent when compared to the actual data.**

Keywords—*double seasonal ARIMA, electricity consumption, least squares estimation, time series*

I. INTRODUCTION

It is very important to maintain the quality of electrical power in the operation of the generation system. Electrical load patterns tend to change all the time. The electric power generated must always be the same as the customer's electrical load. If the power generated is greater than required, there will be a waste of energy. And if the power generated is less than required, a blackout will occur. Electrical energy providers must estimate changes in electrical load at any time in order to achieve a balance of electrical power in the generation system and load.

Forecasting electrical power demand is an important step in the planning and operation of the generating system[1]. Forecasting results are used to prepare the planning and operation of the generating system. This step is to ensure the economic value, reliability of the generating system, and quality of service are maintained. Electrical load forecasting is grouped into four categories, namely long-term, medium-term, short-term and very short-term forecasting[2]. So that one of the researchers' attention is the operating system and scheduling of power generation systems. This scheme is part of a short-term forecasting study of the generation system[2].

Research on load forecasting through a statistical approach has begun in the last decade. Time series analysis has become a research trend, where the results of this study are very effective with a very high level of accuracy in predicting future electrical loads. Forecasting short-term electrical load based on time series has been carried out by several Indonesian researchers[3]–[5]. This time series forecasting method is then also compared with artificial intelligence methods, as has been done previously[6], [7]. Other researchers, such as Jaime Buitrago and Shihab Asfor develop forecasting studies using time series-based ANN methods[8]. Ali Azadeh et al., developed a more complex ANN method based on analytical time series[9]. Meanwhile, the seasonal time series model has been developed by Norizan Mohamed et al., by comparing the SARIMA and DSARIMA models for short-term forecasting of electrical loads in Malaysia[10].

This study is a development of previous research in applying the double seasonal ARIMA method to predict electrical load time series data[11]. The steps taken in this research are in the form of smoothing the presentation of the data so as to produce the best data pattern. This research obtained the best model based on the percentage of MAPE from previous studies.

II. Typical Daily Electrical Load

Distribution of electrical loads from the generating system to customers is very unique because the load data used is data every half hour or an hour. Characteristics of daily electrical loads have a seasonal tendency. In each hour, the electrical load data form the daily characteristics of each week. The second seasonal trend, every week will form a monthly data pattern. The characteristics of this daily data are usually called double seasonal, namely the tendency of patterns that are almost the same every day and every week. Daily usage activities have definitely experienced their own trends, be it in the morning, afternoon, evening or the characteristics of the day each week. Usage during working hours, peak hours as well as national and religious holidays and holidays has unique data characteristics. This daily load forecasting study is a short-term forecasting[2]. This daily electrical load data is used for day-to-day utility operations, generator system scheduling, and distribution of electrical power.

The seasonal pattern is indicated by the emergence of high correlation coefficient repetitions at certain lags[12]. Times series is a quantitative approach based on time. There are two purposes of time series analysis, namely to model a stochastic mechanism of data observation based on time and to predict the value of future observations based on existing data.

Stochastic processes as statistical phenomena are arranged in chronological order based on the law of probability. The data can be seen as the reality of the stochastic process[13]. According to Wei et al.[14], the stochastic process is a set of data based on time which is composed of random variables, namely the sample space and t is the time index. The distribution function of the random variable is as follows:

$$F(z_{t1}, .., z_{tn}) = p\{\omega: z(\omega, t) \leq z_{t1}, .., z(\omega, t_n) \leq z_{tn}\} \ (1)$$

Observation $z_1, z_2, ..., z_n$ is stochastic process, so the random variable $z_{t1}, z_{t2}, ..., z_{tn}$ is said to be stationary in the distribution if:

$$F(z_{t1}, z_{t2}, ..., z_{tn}) = F(z_{t1+k}, z_{t2+k}, ..., z_{tn+k}) \ (2)$$

DSARIMA is written in ARIMA (p, d, q) $(P_1, D_1, Q_1)^{s1}(P_2, D_2, Q_2)^{s2}$ general form and has the following general form[13].

$$\emptyset_p(B)\Phi_{P_1}(B^{s_1})\Phi_{P_2}(B^{s_2})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}\dot{Z}_t = \theta_q(B)\Theta_{Q_1}(B^{s_1})\Theta_{Q_2}(B^{s_2})a_t \ (3)$$

With s_1 and s_2 is different seasonal periods.

III. Time Series Analysis Procedure

The steps for applying the Box-Jenkins ARIMA method are as follows[13]:

a) Prepare electrical load data as well as check stationary data.

b) Model identification in ARIMA through ACF and PACF plots.

c) Determination of parameters p, d, and q in ARIMA.

d) Determination of the ARIMA model equation. The coefficients used are generated from the analysis of the ARIMA model parameters with the smallest MSE.

e) Parameter estimation.

f) Validation or diagnostic test

g) Forecasting

IV. Result And Discussion

A. Dataset

This study aims to refine the data used for training. Previous research data still contain a trend pattern and a downward trend in the period range of November 20 - December 31, 2011[8]. So that the data used in the study is data that has been smoothed from previous research. This study uses electrical power consumption data every thirty minutes during the period from January 2, 2009 to November 19, 2011 at the Generation Unit (UP) of PT. PLN Gresik Indonesia. The data is distributed on: 1. Data for training with a period of January 2, 2009 – November 12, 2011, 2. Data for testing with the assumption that real data is compared to training data from forecasting results for the period from November 13 to 19, 2011. Stationary data analysis and parameter identification using Minitab software. Furthermore, in estimating parameters and forecasting using SAS software, Statistical Analysis System.
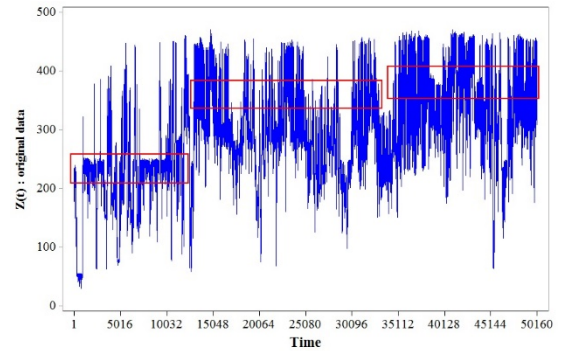

Fig. 1. Plot of load data with seasonal pattern (red box)

Graphically, it can be seen that the data is highly volatile and does not meet the stationary requirements in terms of variance and mean. For this reason, a stationary test was carried out. Stationary in variance is done by using Box-Cox transformation method[15]. The mathematical formula for the transformation is shown as follows[15]:

$$TZ_t = \frac{z_t^\lambda - 1}{\lambda} \ (4)$$

Where, TZ_t is Box-Cox transformation and λ is the value of the Box-Cox transformation coefficient. Based on the stationary test in variance as shown in Figure 2. Stationarity test in variance with a value of $\lambda = -0,13$ and after going through the data transformation process it becomes significant with a value of $\lambda = 1$. So we get the reverse transformation $Z_t = (Z_t^*)^{-\frac{100}{13}}$. The results of this transformation have not shown stationary data in the mean because it has not shown a constant value in the middle of the data plot (Figure 2b). For this reason, it is necessary to do differencing once $(d = 1)$. Based on the ACF plot, it can be seen that the non-seasonal data is stationary, while the seasonal is still not stationary with

an indication that the ACF is still decreasing slowly in seasonal lags, namely lags 48, 96, 144, etc., and in weekly seasonal lags, namely lags 336, 672, etc. For this reason, it is necessary to re-differencing the seasonal pattern ($d = 1, s = 48$) as shown in Figure 3.
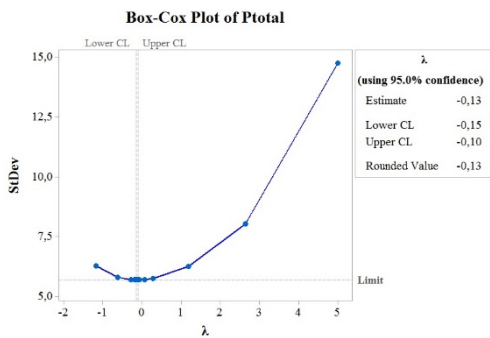

(a)
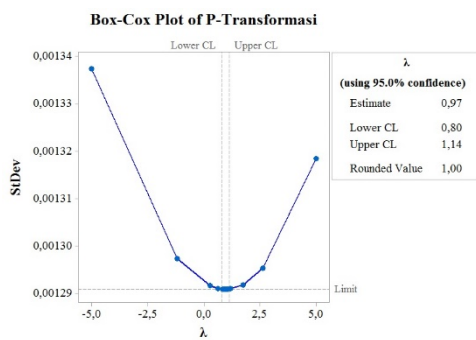
(b)

Fig. 2. Box-Cox transformation diagram; (a) not yet stationary in variance, (b) stationary in variance.

After going through seasonal differencing, there is a strong indication that the data pattern is stationary as the ACF plot is shown in figure 3. The non-seasonal data plot is stationary at lags 1, 2, 3,..., 40. The non-seasonal data pattern tends to die down and will cut. off after lag 7 and lag 8. The ACF plot for the seasonal pattern ($s = 48$) after differencing has also been stationary at lags 48, 96, 144, etc. The data pattern tends to cut off after lag 48. The seasonal pattern ($s = 336$) tends to cut off after lag 336.
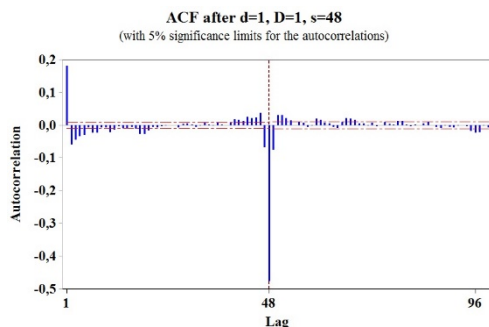
(a)
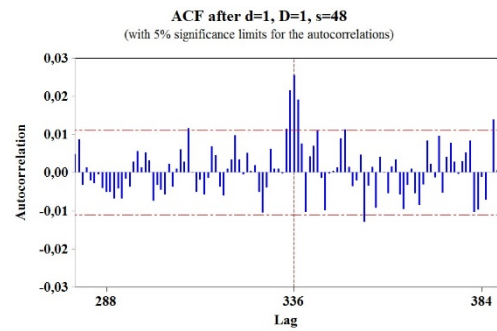
(b)

Fig. 3. Plot of ACF data ($d = 1, s = 48$): (a) daily seasonal plot $s = 48$; (b) weekly seasonal plot $s = 336$.

After all the ACF and PACF data plots have been stationary, the appropriate provisional model guess is the double seasonal ARIMA model $(1,1,1)(0,1,1)^{48}(0,0,1)^{336}$. However, there is a possibility that the white noise has not been met so that it is necessary to add or replace the appropriate order by testing the data. For both seasonal PACF plots die down as shown in Figure 4.
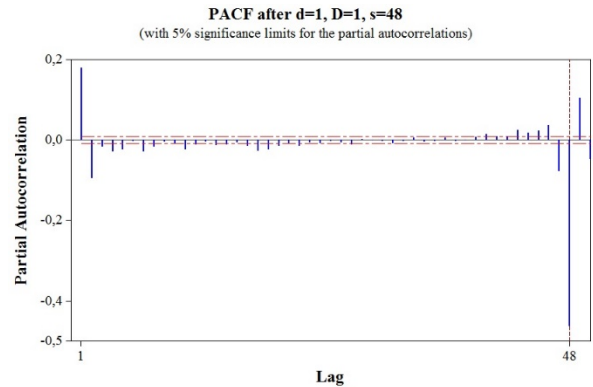

Fig. 4. Plot of PACF after differencing ($d = 1, s = 48$)

B. Parameter Estimation

Estimated time series parameters of the ARIMA model based on the observed values $z_1, z_2, ..., z_n$. Calculations are used to determine the value of p, d, q with the assumption that stationarity has been carried out through the differencing process, the observed parameter d. In this study, parameter estimation uses the least squares method. This method looks for the parameter value that minimizes the number of squares of errors, i.e. the difference between the actual and forecast values [16]. The least square method can be used to estimate AR parameters with a relationship between a value in the current time (z_t) with a value in the previous time (z_{t-k}), plus a random value. MA parameter with the relationship between the value at the present time (z_t) with the residual value at previous times (α_{t-k} with $k = 1, 2, ..$). Through the double seasonal ARIMA model $(1,1,1)(0,1,1)^{48}(0,0,1)^{336}$, the initial estimates of the AR and MA coefficients are shown in Table 1.

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	-0,35184	0,01899	-18,53	< 0,0001	1
MA2,1	0,95734	0,0013007	736,02	< 0,0001	48
MA3,1	-0,04526	0.0045103	-10,03	< 0,0001	336
AR1,1	-0,14578	0,02006	-7,27	< 0,0001	1

Based on Table 1, the data has met the white noise criteria with a p-value greater than the error tolerance value $\alpha = 5\%$, with an alpha significance level of less than 0.0001. In addition, the model at the time of initial estimation has an improvement pattern with 3 MA parameters, namely MA(1,1), MA(2,1) and MA(3,1), so these three parameters must be included in the model estimation. While the residual assumption test which includes the assumption of white noise must meet the independent criteria and have a normal distribution $(0, \sigma^2)$. The Ljung-Box test is used to examine the assumption of independence from the residuals with the following hypothesis[13]:

$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_K = 0$$

H_1 : is at least one ρ_i which is not equal to zero for $i = 1, 2, \ldots, K$.

With an error tolerance of 5%, H_0 is rejected if p-value $< \alpha$, which means that the residual does not meet the white noise assumption. The initial residual test is shown in Table 2 below.

To Lag	ACF Results					
6	-0,002	-0,019	-0,041	-0,017	-0,028	-0,.008
12	-0,033	-0,027	-0,014	-0,009	-0,014	-0,007
18	-0,009	-0,009	-0,008	-0,017	-0,016	-0,023
24	-0,023	-0,018	-0,020	-0,011	-0,013	-0,003
30	-0,009	-0,008	-0,017	-0,008	-0,017	-0,014
36	-0,011	-0,009	-0,002	0,000	-0,010	0,000
42	-0,007	-0,008	0,002	-0,005	-0,004	0,003
48	0,001	0,002	0,006	0,018	0,044	0,060

Based on the estimated parameters of the AR and MA coefficients in Table 2, the residual normal probability plot must meet the white noise assumption with a limit of $< \pm 1.96/\sqrt{n} \approx \pm 0.009$, where n is 50,160 training data. So based on the initial estimation results in Table 2, it is necessary to make an estimate to meet the white noise assumption, namely at lags 2, 3, 4, 5, 7, 8, ,11, 16, 17, 18, 19, 20, 21, 22, 23 , 27,29,30, 31, 46, 47, and 48 based on the ACF results in Table 2. After going through the stage of selecting the initial estimate and then a series of additions and subtractions of parameters that meet the white noise assumption as the results of the residual check are shown in Table 3 below. This estimation result is significant for seasonal lag, especially the estimation result for lag 48.

To Lag	ACF Results					
6	0,000	0,000	-0,002	0,004	0,001	0,.002
12	-0,005	-0,002	0,008	0,000	-0,007	-0,004
18	0,005	0,001	-0,008	0,001	-0,003	-0,003
24	-0,007	-0,000	0,001	0,003	-0,001	0,002
30	-0,005	-0,005	-0,002	0,009	0,001	-0,001
36	-0,005	-0,004	-0,000	-0,001	-0,004	0,008
42	-0,000	-0,005	0,004	0,000	0,000	0,004
48	0,005	0,006	-0,002	-0,002	0,003	-0,000

The best iteration results for AR and MA parameters are shown in Table 4 with the DSARIMA model ([1,2,5,6,7,11, 16,18,35,46], 1, [1,3,13,21,27,46]), $(1,1,1)^{48}(0,0,1)^{336}$.

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0,934	0,01770	52,78	< 0,0001	1
MA1,2	-0,077	0,0072138	-10,64	< 0,0001	3
MA1,3	0,008	0,0038171	2,18	0,0293	13
MA1,4	0,00685	0,0031724	2,16	0,0309	21
MA1,5	0,017	0,0027856	5,92	< 0,0001	27
MA1,6	0,059	0,0067600	8,67	< 0,0001	46
MA2,1	0,98	0,0009744	1003,38	< 0,0001	48
MA3,1	-0.0364	0,0045572	-7,98	< 0,0001	336
AR1,1	1,1464	0,01855	61,81	< 0,0001	1
AR1,2	-0,295	0,0087427	-33,79	< 0,0001	2
AR1,3	-0,0104	0,0052195	-2,00	0,0454	5
AR1,4	0,0189	0,0067496	2,80	0,0051	6
AR1,5	-0,0234	0,0047509	-4,93	< 0,0001	7
AR1,6	-0,004	0,0030582	-1,29	0,1958	11
AR1,7	-0,0083	0,0033299	-2,49	0,0126	16
AR1,8	-0,0125	0,0033252	-3,77	0,0002	18
AR1,9	-0,007	0,0022520	-3,26	0.0011	35
AR1,8	-0,0125	0,0033252	-3,77	0,0002	18
AR1,9	-0,007	0,0022520	-3,26	0.0011	35
AR1,10	0,07	0,0067089	10,62	< 0,0001	46
AR2,1	0,03	0,0050410	5,86	< 0,0001	48

Results of Short Term Load Prediction

Based on the final parameter estimation results in Table 4, the ARIMA coefficient parameters are obtained as follows: AR(1,1) = 1.1464, AR(1,2) = -0.295, AR(1,3) = -0.0104, AR(1,4) = 0.0189, AR(1,5) = -0.0234, AR(1,6) = -0.004, AR(1,7) = -0.0083, AR(1,8) = -0.0125, AR(1,9) = -0.0074, AR(1,10) = 0.07, AR(2,1) = 0.03, MA(1,1) = 0.934, MA(1,2) = -0.077, MA(1,3) = 0.008, MA(1,4) = 0.00685, MA(1,5) = 0.017, MA(1,6) = 0.059, MA(2,1) = 0.98, MA(3,1) = -0.0364.

Based on the prediction model parameters, the DSARIMA model ([1,2,5,6,7,11,16,18,35,46], 1, [1,3,13,21,27,46]), $(1,1,1)^{48}(0,0,1)^{336}$ with the following model equation:

$$(1 - 1.1464B + 0.295B^2 + 0.0104B^5 - 0.0189B^6 + 0.0234B^7 + 0.004B^{11}$$

$$+0.0083B^{16} + 0.0125B^{18} + 0.0074B^{35} - 0.07B^{46})(1 - 0.03B^{48})Z_t^* =$$
$$(1 - 0.934B + 0.077B^3 - 0.008B^{13} - 0.00685B^{21} - 0.017B^{27}$$
$$-0.059B^{46})(1 - 0.98B^{48})(1 + 0.0364B^{336})a_t$$

After going through the reverse transformation of Z_t, a comparison of the predicted electrical load with the actual data (testing) is obtained in Figure 5.
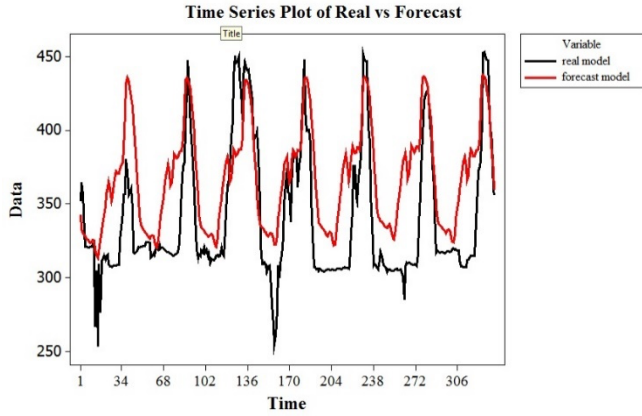

Fig. 5. Comparison of data between actual electric power and forecast results

C. Model Testing and Measuring Accuracy

Basically, to measure the accuracy of the prediction results can be done by various methods. Several statistical methods such as the root-mean-square error (RMSE), mean absolute error (MAE) and the mean absolute percentage error (MAPE). In this study, MAPE was used as a standard for measuring the accuracy of the prediction results. MAPE is defined as follows[10]:

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{Z_i - \hat{Z}_i}{Z_i} \right|}{n} \times 100\% \qquad (5)$$

Where Z_i and $\hat{Z}_i$ are the actual and predicted values, while n is the sum of the predicted values.

Testing the accuracy of the difference between the actual power data and the predicted results obtained by MAPE of 1.56 percent.

V. CONCLUSION

Statistical analysis based on the DSARIMA model is in accordance with the characteristics of electric power with seasonal and fluctuating electrical load patterns. Load changes are always unpredictable at any time depending on the electrical power requirements at the load center. With statistical analysis models, predictions are able to produce the best data. The improvements made in the study were the smoothing of training data and testing data so that there were additional AR parameters, namely at lags 5, 6, and 11. This improvement showed that the residuals from the resulting model had met the white noise assumption with a normal residual probability of $< \pm 0.009$. Based on MAPE, there is an improvement in the measurement of forecasting data with actual data, namely from the percentage error of 2.06 percent to 1.56 percent. Further research that can be developed for

patterns of electrical power demand in large-scale areas such as the Java-Madura-Bali interconnection, Indonesia.

REFERENCES

[1] M. D. Kartikasari and A. R. Prayogi, "Demand forecasting of electricity in Indonesia with limited historical data," 2018.

[2] M. A. Hammad, B. Jereb, B. Rosi, and D. Dragan, "Methods and Models for Electric Load Forecasting: A Comprehensive Review," *Logist. Sustain. Transp.*, 2020.

[3] 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.

[4] E. N. Yaqin *et al.*, "Short Term Load Forecasting for Weekends in Indonesia: Comparison of Three Methods," 2018.

[5] S. A. W. Dinata, M. Azka, M. Faisal, Suhartono, R. Yendra, and M. D. H. Gamal, "Short-Term Load Forecasting Double seasonal ARIMA Methods: An Evaluation Based on Mahakam-East Kalimantan Data," 2020.

[6] Q. W. Luthuli and K. A. Folly, "Short term load forecasting using artificial intelligence," 2016.

[7] K. D. Abeyrathna and C. Jeenanunta, "Hybrid particle swarm optimization with genetic algorithm to train artificial neural networks for short-term load forecasting," *Int. J. Swarm Intell. Res.*, 2019.

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

[9] A. Azadeh, S. F. Ghaderi, M. Sheikhalishahi, and B. P. Nokhandan, "Optimization of Short Load Forecasting in Electricity Market of Iran Using Artificial Neural Networks," *Optim. Eng.*, 2014.

[10] N. Mohamed, M. H. Ahmad, Suhartono, and Z. Ismail, "Improving short term load forecasting using double seasonal arima model," *World Appl. Sci. J.*, 2011.

[11] 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.

[12] X. Gao, X. Li, B. Zhao, W. Ji, X. Jing, and Y. He, "Short-term electricity load forecasting model based on EMD-GRU with feature selection," *Energies*, 2019.

[13] E. R. Ziegel, G. Box, G. Jenkins, and G. Reinsel, "Time Series Analysis, Forecasting, and Control," *Technometrics*, 1995.

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

[15] G. E. P. Box and D. R. Cox, "An Analysis of Transformations," *J. R. Stat. Soc. Ser. B*, 1964.

BIOGRAPHIES OF AUTHORS

ISMIT MADO was born in Banjarmasin City, South Borneo, Indonesia. He received his Bachelor of Electrical Engineering from the Universitas Muhammadiyah Malang, Indonesia, in 1999. His Masters and Doctoral degrees in Departement of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, in 2006 and 2019. Since 2001, he has been a lecturer in the Departement of Electrical Engineering, Universitas Borneo Tarakan, Tarakan City, North Borneo, Indonesia. His research interests include electrical load forecasting, modeling of electric load clusters, and small signal stability and control in power plants.

ANTONIUS RAJAGUKGUK was born in Medan, North Sumatra, Indonesia. and joined Universitas Riau as a lecturer in 1997. He received his B.S. degree in Electronics Engineering from Atma Jaya, Catholic University, Jakarta Indonesia, in 1993, and his M.S. degrees from Institut Teknologi Sepuluh Nopember (ITS) Surabaya, in 2005. He lives in Surabaya and he is a Doctoral student in Electrical Engineering Institut Teknologi Sepuluh Nopember (ITS). His research interest includes applications of power electronics for grid system, power quality and renewable energy.

ARIS TRIWIYATNO, Receive ST., MT., and Dr. degrees in Department of Electrical Engineering, Faculty of Engineering, Sepuluh November Institute of Technology (ITS) Surabaya. Now a permanent lecturer in the Department of Electrical Engineering, Faculty of Engineering, Diponegoro University with research interest: Artificial Intelligent, Software Engineering and Control System.

ARIF FADLLULLAH was born in Tarakan City, North Borneo, Indonesia on May 19, 1991. Graduated with a Bachelor of Informatics Engineering Education at the Department of Computer Engineering, UniversitasNegeri Malang in 2012. Graduated with Masters in Informatics, Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya in 2016. Currently, he works at the Departement of Electrical Engineering, Universitas Borneo Tarakan with expertise in image processing, computer vision, machine learning, and software engineering. Especially for software engineering in the field of electrical engineering, it has developed a Wind Potential Analysis Application as a renewable energy source using Weibull Parameter Estimation Based on the Power Density Method (Case Study: Tarakan City). The research can be seen at https://bit.ly/penelitian_ariffadllullah.