

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Electric Load Forecasting an Application of Cluster Models Based on Double Seasonal Pattern Time Series Analysis

Ismi Mado

Abstract

Electricity consumption always changes according to need. This pattern deserves serious attention. Where the electric power generation must be balanced with the demand for electric power on the load side. It is necessary to predict and classify loads to maintain reliable power generation stability. This research proposes a method of forecasting electric loads with double seasonal patterns and classifies electric loads as a cluster group. Double seasonal pattern forecasting fits perfectly with fluctuating loads. Meanwhile, the load cluster pattern is intended to classify seasonal trends in a certain period. The first objective of this research is to propose DSARIMA to predict electric load. Furthermore, the results of the load prediction are used as electrical load clustering data through a descriptive analytical approach. The best model DSARIMA forecasting is $([1, 2, 5, 6, 7, 11, 16, 18, 35, 46], 1, [1, 3, 13, 21, 27, 46]) (1, 1, 1)^{48} (0, 0, 1)^{336}$ with a MAPE of 1.56 percent. The cluster pattern consists of four groups with a range of intervals between the minimum and maximum data values divided by the quartile. The presentation of this research data is based on data on the consumption of electricity loads every half hour at the Generating Unit, the National Electricity Company in Gresik City, Indonesia.

Keywords: electric loads, DSARIMA model, descriptive analytic, clustering, forecasting, time series

1. Introduction

Fluctuations in electrical power greatly affect the performance of power generation systems. Changes in electrical power due to variations in demand for electrical power momentarily result in an imbalance of electricity generated by the electric power absorbed. If the power supplied is greater than there will be energy waste. And if the power supplied is smaller then there will be overload which will result in a blackout. This means that the amount of electric power generated must be balanced or not too far from the nominal value of the electrical power requirements at the load center. In fact, the use of electrical energy tends to change at any time.

For this reason, it is necessary to predict the use of electric power that is able to maintain a balance between supply and consumption of electric power in the power generation system. Research of electricity load forecasting is very important in the power plant system operation plan [1]. Load forecasting studies are classified into three categories: long-term, medium-term and short-term predictions. Long-term predictions are needed for planning the peak load capacity and system maintenance schedule [2], medium-term predictions are needed for the planning and operation of the power plant system [3], and short-term predictions are needed to control and schedule the generating system [4]. So that load forecasting studies play a role in ensuring the economic value of financing, system reliability, stability and quality of electricity system services.

Fluctuations in electrical power at the load center contain a set of time-based information. The characteristics of the load from the period of use both by household, commercial, industrial and public costs, are needed so that fluctuations can be analyzed. The load characteristics, besides being able to be analyzed also contain a series of load patterns tendencies due to usage. This conduct of using electric loads contains seasonal patterns. Daily use tends to recur on certain days, as well as weekly load patterns. This trend is then analyzed through the load cluster approach to achieve load usage patterns based on seasonal patterns.

The Box-Jenkins time series study approach conducted in this research was able to increase the estimated usage and application of seasonal patterns based on electricity load clusters. The time series prediction model is an accurate choice and continues to grow to this day [5–7]. Researchers have carried out load forecasting study activities with 2.06 percent MAPE [8]. In research, the parameter estimation pattern was developed again with the least squares method which is better. And then the load cluster modeling is developed to classify the trend based on seasonal patterns.

2. Electrical load characteristics

The main purpose of an electric power distribution system is to distribute electric power from substations or sources to a number of customers or loads. The most important main factor in the distribution system planning is the characteristics of various electrical loads.

The electrical load characteristics are needed so that the system voltage, the thermal effect of loading and the loading pattern can be analyzed properly. The analysis is included in determining the initial projections in the next planning.

The characteristics of the electrical load are very dependent on the type of load it serves. This will be clearly seen from the results of recording the load curve in a time interval. The following are several factors that determine the load characteristics according to the needs of this study [9].

2.1 Load factor

Load factor is the ratio between average load and peak load measured in a certain period. Average load and peak load can be expressed in KiloWatt (KW), KiloVolt-Ampere (KVA) and so on, but the units of both must be the same. Load factor can be calculated for a certain period usually used in units of daily, monthly or yearly.

The peak load referred to in this study is a momentary peak load or average peak load in a certain interval (maximum demand), generally a maximum demand of

15 minutes or 30 minutes is used. In this study, the load data used is 30-minute interval load data.

The definition of the load factor can be written in the following equation:

when you are citing sources, the citations should be set in numbered format. All the references given in the list of references should be cited in the body of the text. Please set citations in square brackets keeping the below points in mind.

$$\text{Load factor} = \frac{\text{average load in a certain period}}{\text{peak load in a certain period}} \quad (1)$$

The load factor can be known from the load curve. As for the estimation of the magnitude of the burden factor in the future, it can be approached with existing statistical data as was done in this study.

When applied to the power plant, it is formulated into

$$\text{Load factor} = \frac{P_{\text{average}}}{P_{\text{peak}}} \times \frac{T}{T} \quad (2)$$

If T is in a year, an annual expense factor is obtained. If in 1 month the monthly load factor is obtained, as well as the daily load factor.

2.2 Daily load

Daily load factors vary according to the characteristics of the load area, whether it is a dense residential area, industrial area, trade or a combination of various types of customers.

This daily load factor will also affect the weather conditions and certain days such as holidays and so on.

2.3 Load curve

Load curves illustrate the variation of loading on a substation measured by KW or KVA as a function of time. Measurement time intervals are usually determined based on the use of measurement results, for example intervals of 30 minutes, 60 minutes, 1 day or 1 week.

The load curve shows the demand or load requirements at different time intervals. With the help of this load curve, we can determine the magnitude of the largest load and then the generating capacity can also be determined.

2.4 Peak load

Peak load or maximum demand is defined as the biggest load of needs that occurs during a certain period. Certain periods can be in the form of daily, monthly or annual periods. Furthermore, the peak load must be interpreted as the average load during a certain interval, where the possibility of such load. For example, the daily load of a distribution transformer where the peak load during an interval of 1 hour, ie between 19:00 (point A) and 20:00 (point B). The average value of the A - B curve is its peak requirement.

Keep in mind here that peak needs are not instantaneous needs, but on average during a certain time interval, usually a certain time interval is 15 minutes, 30 minutes or 1 hour.

The characteristics of the burden between holidays are different from ordinary days so that they have different load variants. Load characteristics can also be

distinguished by the factor of loading outside the time of the peak load, or who are at the time of the peak load. So we need load forecasting with the aim of preparing operating generating units. When electricity demand increases, it will be balanced with adequate electricity supply to prevent power outages, otherwise if electricity consumption decreases, electricity supply will be reduced so as not to over supply.

3. Electrical load analysis based on time series model

Box and Jenkins popularized the use of ARIMA models and the Box-Jenkins methodology became highly popular in the 1970s among academics [10]. The ARIMA model is also called the Box-Jenkins time series. A time series is a series of observations taken sequentially based on time [11]. The observation process is carried out at the same interval, for example in hour, daily, weekly, monthly, yearly or other intervals. The purpose of time series analysis is twofold, namely to model the stochastic mechanism found in observations based on time and to predict the value of observations in the future. The value of a variable can be predicted if the nature of the variable is known in the present and in the past.

3.1 ARIMA model classification

The ARIMA model is divided into several groups, namely: autoregressive (AR), moving average (MA), and ARMA. The ARIMA model is a nonstationary ARMA model that has gone through a differencing process so that it becomes a stationary model. The ARIMA model also contains seasonal patterns. Defined as a pattern that repeats in a fixed time interval. The application of this seasonal pattern has been developed into a double seasonal pattern [12–14]. Double seasonal ARIMA model is written with notation, as follows.

$$ARIMA(p, d, q)(P_1, D_1, Q_1)^{S_1}(P_2, D_2, Q_2)^{S_2} \quad (3)$$

This model consists of two components, namely the first level which is usually developed from a linear forecasting model to explain seasonal trends from data or known as potential load. And at the second level developed from the ARIMA model to capture autoregressive patterns from data or called irregular loads. For stationary data, the seasonal factor can be determined by identifying the coefficient of autocorrelation at two or three time intervals that are very different from zero. So that this seasonal pattern can be identified whether it contains a tendency to have a seasonal pattern or multiple seasonal patterns and has the following general form [15]:

$$\phi_p(B)\Phi_{P_1}(B^{S_1})\Phi_{P_2}(B^{S_2})(1-B)^d(1-B^{S_1})^{D_1}(1-B^{S_2})^{D_2}Z_t = \theta_q(B)\Theta_{Q_1}(B^{S_1})\Theta_{Q_2}(B^{S_2})a_t \quad (4)$$

With

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Phi_{P_1}(B^{S_1}) &= 1 - \Phi_{11} B^{S_1} - \Phi_{21} B^{2S_1} - \dots - \Phi_{P_1} B^{P_1 S_1} \\ \Phi_{P_2}(B^{S_2}) &= 1 - \Pi_{12} B^{S_2} - \Pi_{22} B^{2S_2} - \dots - \Pi_{P_2} B^{P_2 S_2} \\ \theta_q(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \\ \Theta_{Q_1}(B^{S_1}) &= 1 - \Theta_{11} B^{S_1} - \Theta_{21} B^{2S_1} - \dots - \Theta_{Q_1} B^{Q_1 S_1} \\ \Theta_{Q_2}(B^{S_2}) &= 1 - \Psi_{12} B^{S_2} - \Psi_{22} B^{2S_2} - \dots - \Psi_{Q_2} B^{Q_2 S_2}. \end{aligned}$$

3.2 ARIMA Box-Jenkins procedure

The prediction procedure of ARIMA Box-Jenkins model through five stages of iteration, as follows:

- i. Preparation of data, including checking of data stationary
- ii. Identification of ARIMA model through autocorrelation function and partial autocorrelation function
- iii. Estimation of ARIMA model parameters: p , d , and q
- iv. Determination of ARIMA model equations
- v. Forecasting.

3.3 Identification

Identification requires calculation and general review of the results of the autocorrelation function (ACF) and the partial autocorrelation function (PACF). The results of these calculations are needed to determine the appropriate ARIMA model, whether ARIMA ($p, 0, 0$) or AR (p), ARIMA ($0, 0, q$) or MA (q), ARIMA ($p, 0, q$) or ARMA (p, q), ARIMA (p, d, q). Meanwhile, to determine the presence or absence of the d model value, it is determined by the data itself. If the data form is stationary, d is 0, while the data form is not stationary, the value of d is not equal to 0 ($d > 0$). Likewise, the dual seasonal ARIMA model also refers to the autocorrelation function (ACF) and partial autocorrelation function (PACF) as well as knowledge of the system or process being studied.

Identification can be done after fixed time series data. The application of the model after ACF and PACF data has a tendency according to the reference to **Table 1** and for the seasonal data patterns determined by referring to **Table 2** [11].

3.4 Parameter approximation

There are two basic ways to get this parameter:

- a. By trial and error, test several different values and choose one of these values (or a set of values, if more than one parameter is estimated) that minimizes the sum of squared residuals.
- b. Iterative approach, choosing an initial estimate and then letting the computer correct the iterative approximation.

ACF patterns	PACF patterns	ARIMA parameters
Heading to zero after lag q	Decreasing gradually/bumpy	ARIMA ($0, d, q$)
Decreasing gradually/bumpy	Heading to zero after lag q	ARIMA ($p, d, 0$)
Decreasing gradually/bumpy (until lag q is still different from zero)	Decreasing gradually/bumpy (until lag q is still different from zero)	ARIMA (p, d, q)

Table 1.
PACF and ACF patterns.

Model	ACF	PACF
AR (p)	Dies down (decreases exponentially) in seasonal lags	Cut off after lag p^s
MA (q)	Cut off after lag q^s	Dies down (decreases exponentially) in seasonal lags
ARMA (p, q)	Dies down (decreases exponentially) in seasonal lags	Dies down (decreases exponentially) in seasonal lags

Table 2.
PACF and ACF seasonal patterns.

3.5 Parameter testing

Parameter testing phase is to test whether the selection of parameters p, d, q is true and correct. The model is said to be good if the error value is random, meaning that it no longer has a certain pattern. In other words, the model obtained can capture well the existing data patterns. To see the error value of the test carried out testing the value of the autocorrelation coefficient of the error, using one of the following two statistics:

1. Q Box dan Pierce Test

$$Q = n' \sum_{k=1}^m r_k^2 \tag{5}$$

2. Ljung-Box Test

$$Q = n'(n' + 2) \sum_{k=1}^m \frac{r_k^2}{(n' - k)} \tag{6}$$

Spread by chi squared (χ^2) with free degrees (db) = ($m - p - q - P - Q$)
Where

$$n' = n - (d + SD) \tag{7}$$

3.6 Testing criteria

If $Q \leq \chi^2(\alpha, db)$, meaning: error value is random (model is accepted)
If $Q > \chi^2(\alpha, db)$, meaning: error value is not random (model cannot be accepted)

3.7 Parameter estimation

This study uses the least squares method in estimating parameters [15]. The ARIMA model parameters are based on the time series observed with $Z_1, Z_2, \dots, Z_1$. The quadratic method assumes that the best curve is the curve that has the least square error of the data set. The parameter values of the ARIMA models p, d , and q are determined through the stationary ACF and PACF chart plots.

3.8 Measuring accuracy level of forecasting result

Basically, to measure the accuracy of forecasting result can be done by various methods. Some statistical methods such as as Root Mean Square Error

(RMSE), Mean Absolute Error (MEA) and Mean Absolute Percentage Error (MAPE). In this research, MAPE is used as a standard measurement of the accuracy offorecasting result. MAPE is defined as follows [13]

$$\text{MAPE} = \frac{\sum_{i=1}^n \left| \frac{Z - \hat{Z}_i}{Z_i} \right|}{n} \times 100\% \quad (8)$$

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

3.9 Electric load cluster modeling

Cluster analysis performed in this study refers to the statistical description of the analysis technique. Descriptive statistics are methods relating to the collection and presentation of a group of data so as to provide useful information [16]. This description analysis includes several things, namely: frequency distribution, measurement of central tendency, and measurement of variability [17].

The data that has been obtained from a study which is still in the form of random data that can be made into grouped data is data that has been arranged into certain classes. Lists containing grouped data are called frequency distributions or frequency tables. Frequency distribution is the arrangement of data according to certain interval classes or according to certain categories in a list. Frequency distribution can be presented in groups, distribution based on rank order or ranking of distribution classes, distribution in groups, and distribution charts.

Measuring central tendency is a statistical analysis that specifically describes a representative score. The central tendency shows the location of the largest part of the value in the distribution including a general description of data frequencies such as mode, media, and mean or mean count.

While the measurement of variability to describe the degree of dispersion of quantitative data. This measure consists of interquartile range, quartile deviation, mean deviation, standard deviation and coefficient of variation, and variance. Measurement of variability serves to determine the homogeneity or heterogeneity of data. A data may have the same central tendency value but have different variance values.

4. DSARIMA-based load forecasting

The data used in this study is the consumption of electric power every 30 minutes during January 2, 2009 to November 19, 2011 in the Generating Unit service, the National Electricity Company in Gresik City, Indonesia.

The data is distributed on: 1. Data for training during January 2, 2009 to November 12, 2011, 2. Data for testing with the assumption of real data compared to training data from forecasting results during November 13–19, 2011.

Statistical Analysis System (SAS) is used as a simulation of electricity load forecasting and Minitab programming is used to analyze the electricity load cluster model.

4.1 Parameter identification

To identify data, the first step that must be taken is to plot the time series of the data. The time series plot is displayed to see the data patterns and stationarity of the

data which aims to determine the ARIMA model. The pattern of data as shown in **Figure 1** is very volatile. This condition is likely influenced by the integrated power distribution system in the Java-Madura-Bali Indonesia interconnection system.

When referring to **Figure 1(a)**, it can be seen that the data are not stationary in variance or mean. For more details, it will be seen in the autocorrelation function as shown in **Figure 2**. And if it refers to time series patterns there is a tendency for the data to contain seasonal patterns as shown in **Figure 1(b)**.

The data is not stationary in the variance, so it is necessary to transform the data as follows. Testing stationarity in variance if the p -value or $\lambda = 1$. Based on the results of the transformation, the data is not stationary in the variance marked with the value $\lambda = -0.13$ as shown in **Figure 3a**. After going through the process of transformation the data becomes significant with the value $\lambda = 1$ as shown in **Figure 3b**.

After the data is transformed it will be transformed back to get the active data value, as follows

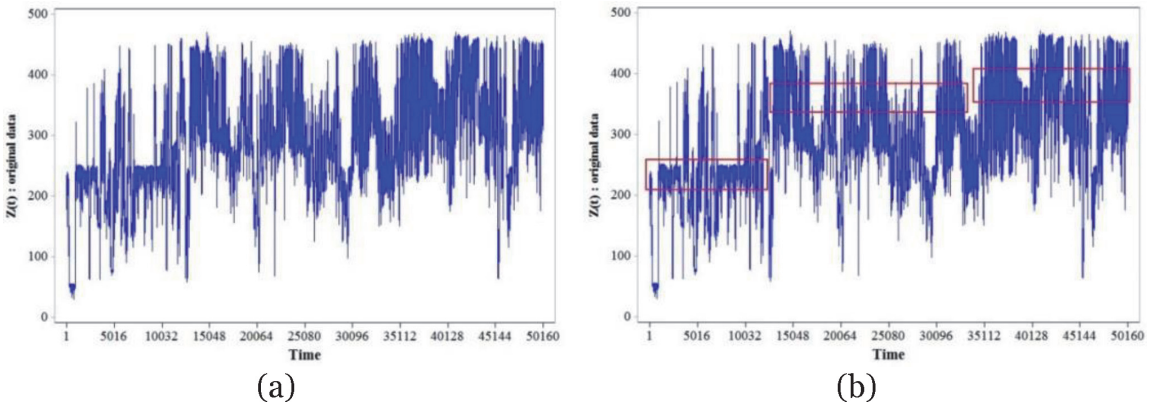


Figure 1.
(a) Data plot of electricity usage every 30 minutes during January 2, 2009 to November 12, 2011; (b) plot of electrical load data with seasonal patterns (red box).

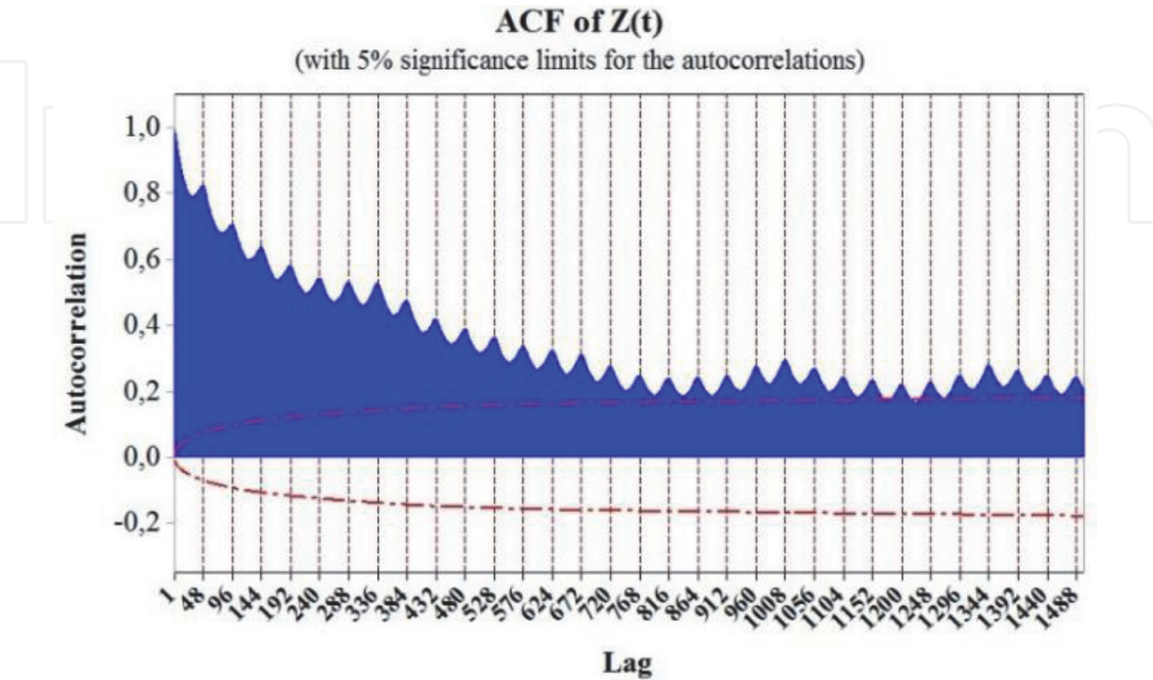


Figure 2.
ACF plot data.

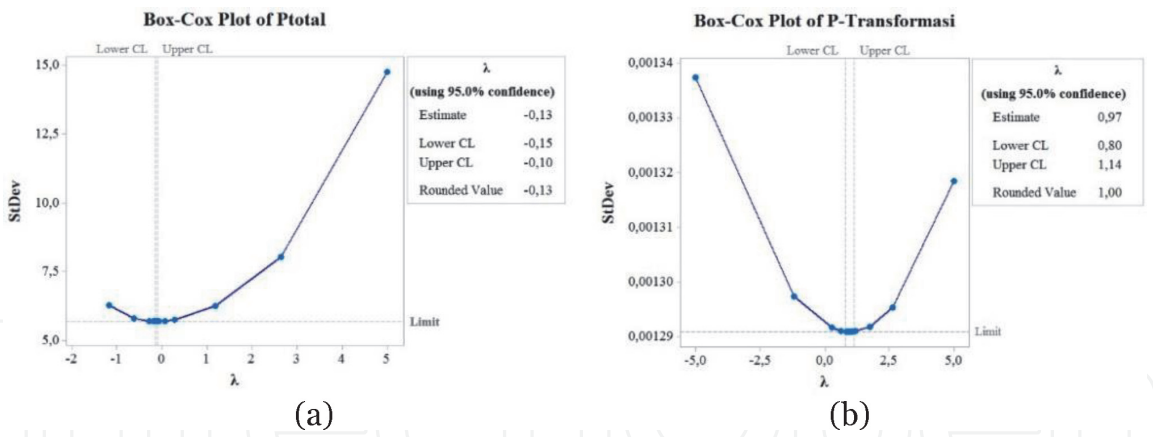


Figure 3.
(a) Box-Cox transformation; (b) after transformation.

$$Z_t^* = Z_t^{-0,13} \tag{9}$$

Then

$$Z_t = (Z_t^*)^{-\frac{100}{13}} \tag{10}$$

The data is stationary in variance, but the transformation results in **Figure 2b** are not stationary in the mean. Data has not shown a constant value in the middle. The stationarity of the data can also be seen through the plot of the autocorrelation function (ACF). From **Figure 2**, it can be seen that the coefficient of autocorrelation is significantly different from zero and slowly decreases. The pattern shows that the data is not stationary in particular not stationary in the mean, while the ARIMA method requires data that is stationary.

The ACF plot also shows that there are strong indications of having a seasonal pattern in both daily and weekly seasonal averages as shown in **Figure 4**, below.

In **Figure 4a**, it can be seen that the electricity load data has a seasonal pattern that is the daily seasonal as seen in lags 48, 96, 144, etc. And in **Figure 4b**, the data also contains weekly seasonal as seen in lag 336, 672, 1008, 1344, etc.

Because the data is not stationary in the mean, it is necessary to do differencing ($d = 1$). The ACF plot of differencing data results is shown in **Figure 5** below.

Based on the ACF plot in **Figure 5**, it appears that the nonseasonal data has been stationary. However, seasonal plots are still not stationary with an indication that ACF is still falling slowly in daily seasonal lags, ie lags 48, 96, 144, etc., and weekly seasonal lags, ie lags 336, 672, etc.

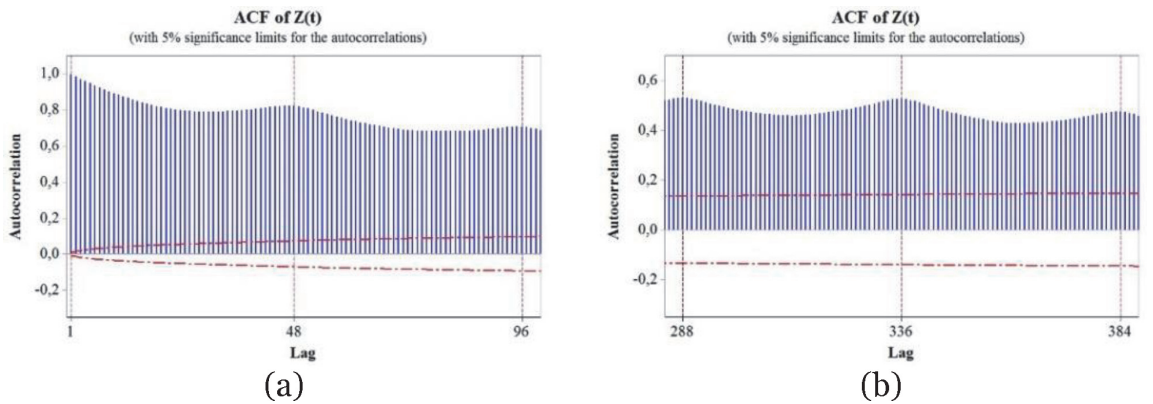


Figure 4.
ACF plots with seasonal patterns: (a) daily seasonal; (b) weekly seasonal.

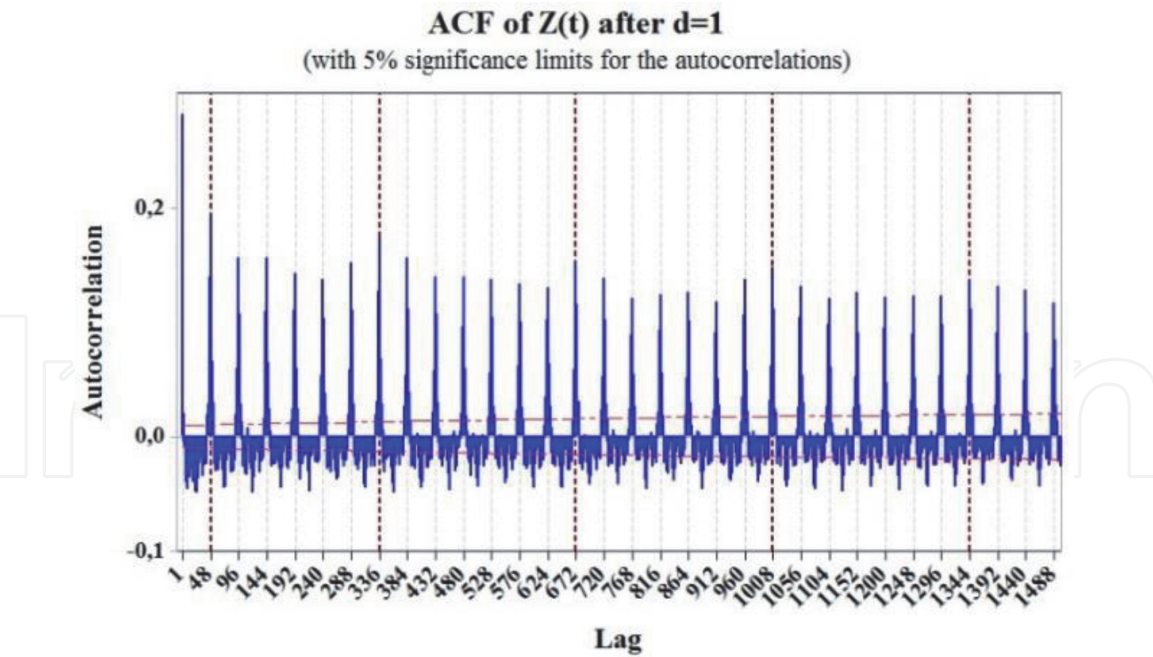


Figure 5.
ACF plot after differencing ($d = 1$.)

It is necessary to do differencing data once more in the seasonal pattern ($d = 1, D = 1, s = 48$). After going through seasonal differencing there are strong indications that the data patterns have been stationary.

Based on the ACF plot for differencing ($d = 1, D = 1, s = 48$) it is clear that the data as a whole has been stationary in the mean. The nonseasonal data plot has been stationary in lags 1, 2, 3, ..., 40. The data pattern tends to dies down and will be cuts off after lag 7 and lag 8 in **Figure 6a**.

The ACF plot for seasonal patterns $s = 48$ after differencing has also been stationary at lags 48, 96, 144, etc. The data pattern tends to be cuts off after lag 48 in **Figure 6b**. The seasonal pattern $s = 336$ tends to be cuts off after lag 336 in **Figure 6c**.

For PACF plots both seasonal ($s = 48$) and ($s = 336$) dies down as shown in **Figure 6d**. Based on the provisions in **Tables 1** and **2**, the parameter identification results can be rewritten in the following **Table 3**.

The ACF and PACF data plots are stationary, the alleged nonseasonal ARIMA models are in accordance with the stationary topology in **Table 1** and the seasonal ARIMA in **Table 2**. The temporary model of ARIMA provisional model is double seasonal based on **Table 3** is DSARIMA $(1, 1, 1)(0, 1, 1)^{48}(0, 0, 1)^{336}$. However, there is a possibility that white noise has not been fulfilled, so it is necessary to add or change the order in accordance with the test.

4.2 Parameter estimation

AR and MA coefficients in the DSARIMA model are estimated using the least squares method. The initial estimate that has been obtained is used as the initial value of the estimation method iteratively. Obtained initial estimates of AR and MA coefficients from the interim model DSARIMA $(1, 1, 1)(0, 1, 1)^{48}(0, 0, 1)^{336}$ as shown in **Table 4** in the following.

Based on **Table 4**, AR and MA parameters have met the criteria for white noise with a p-value greater than the error tolerance value $\alpha = 5\%$, with an alpha significance level of less than 0.0001. However, it is necessary to re-test the residual

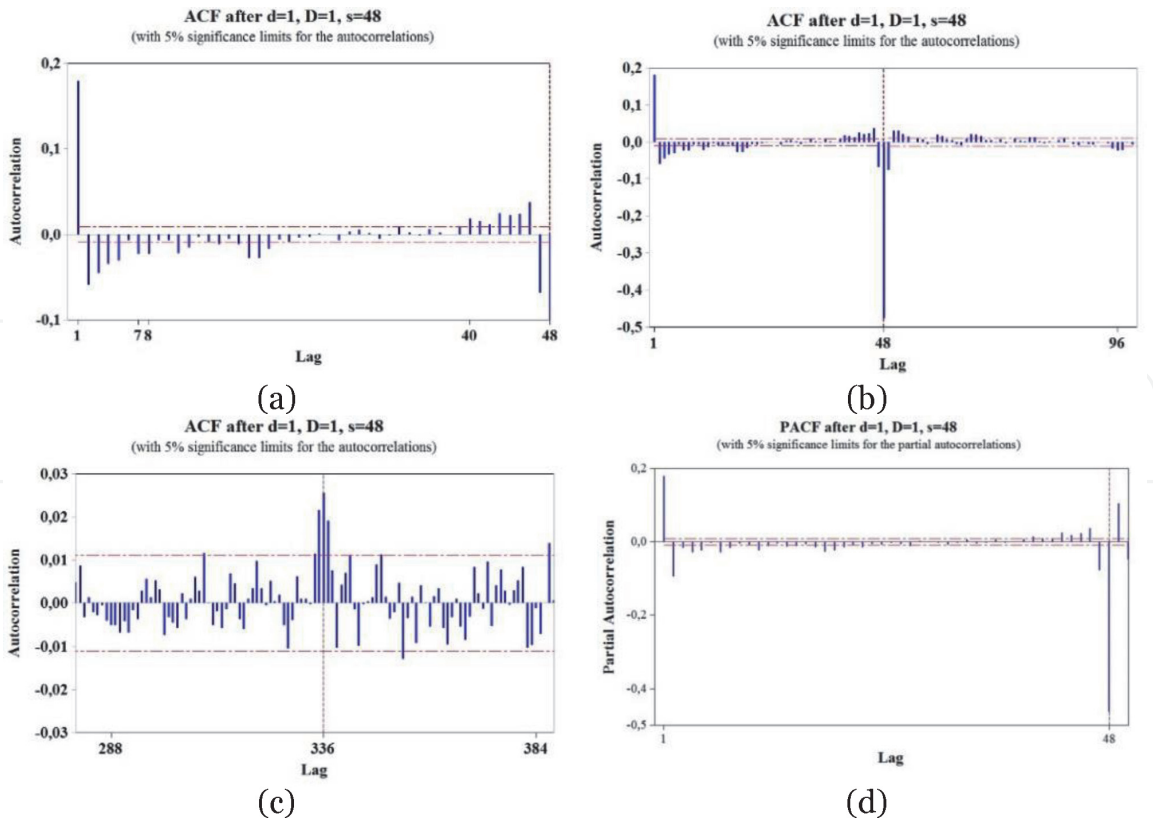


Figure 6.
ACF and PACF plot after differencing ($d = 1, D = 1, s = 48$)

Models	ACF	PACF	Estimated parameters
Nonseasonal	Dies down	Dies down	ARMA (1, 1)
Seasonal ($s = 48$)	Cuts off	Dies down	MA (1) ⁴⁸
Seasonal ($s = 336$)	Dies down	Dies down	MA (1) ³³⁶

Table 3.
Identification plots for ACF and PACF.

Parameter	Estimate	Standard error	t value	Approx Pr $> t $	Lag
MA 1.1	-0.35184	0.01899	-18.53	<0.0001	1
MA 2.1	0>95734	0.0013007	736.02	<0.0001	48
MA 3.1	-0.04526	0.0045103	-10.03	<0.0001	336
AR 1.1	-0.14,578	0.02006	-7.27	<0.0001	1

Table 4.
An output SAS of model with CLS iterative.

assumptions which include the white noise assumption and meet the independent criteria and are normally distributed $(0, \sigma^2)$.
Ljung-Box Test is used to check the assumption of independence from residuals with the following hypotheses:

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

$$H_1: \text{there is at least one } \rho_i \text{ that is not equal to zero for } i = 1, 2, \dots, K$$

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

Based on the estimated AR and MA coefficient parameters in **Table 5**, the residual normal probability plot must meet the assumption of white noise with a limit of $< \pm \frac{1.96}{\sqrt{n}} \approx \pm 0.009$, where n as many as 50,160 training data. Then based on the initial estimation results in **Table 5**, it is necessary to estimate to meet the white noise assumption, namely by including an estimate on the lag 2, 3, 4, 5, 7, 8, 9, 11, 16, 17, 18, 19, 20, 21, 22, 23, 27, 29, 30, 31, 46, 47, and 48. The results of the residual check are shown in **Table 6** below. The estimation results are significant for seasonal lag, which is lag 48.

Based on residual checking, namely by adding and subtracting AR and MA parameters, it can be seen that all lags have met the assumption of white noise with a limit of $< \pm \frac{1.96}{\sqrt{n}} \approx \pm 0,009$ (see ACF Results). The best iteration results of the AR and MA parameters are shown in **Table 7** below.

Based on **Table 7**, the DSARIMA model is obtained with the coefficients $([1, 2, 5, 6, 7, 11, 16, 18, 35, 46], 1, [1, 3, 13, 21, 27, 46]) (1, 1, 1)^{48}(0, 0, 1)^{336}$, which have met the assumption of white noise.

To Lag	ChiSq	DF	Pr > ChiSq	ACF results					
6	153.39	2	<0.0001	−0.002	−0.019	−0.041	−0.017	−0.028	−0.008
12	274.15	8	<0.0001	−0.033	−0.027	−0.0114	−0.009	−0.014	−0.007
18	342.13	14	<0.0001	−0.009	−0.009	−0.008	−0.017	−0.016	−0.023
24	422>74	20	<0.0001	−0.023	−0.018	−0.020	−0.011	−0.013	−0.003
30	43.05	26	<0.0001	−0.009	−0.008	−0.017	−0.008	−0.017	−0.014
36	489.03	32	<0.0001	−0.011	−0.009	−0.002	0.000	−0.010	0.000
42	497.60	38	<0.0001	−0.007	−0.008	0.002	−0.005	−0.004	0.003
48	804.03	44	<0.0001	0.001	0.002	0.006	0.018	0.044	0.060

Table 5.
An output SAS of model with ACF check of residuals.

To Lag	ChiSq	DF	Pr > ChiSq	ACF results					
6	—	0	—	0.000	0.000	−0.002	0.004	0.001	0.002
12	—	0	—	−0.005	−0.002	0.008	0.000	−0.007	−0.004
18	—	0	—	0.005	0.001	−0.008	0.001	−0.003	−0.003
24	18>10	5	0.0028	−0.007	−0.000	0.001	0.003	−0.001	0.002
30	24.78	11	0.0098	−0.005	−0.005	−0.002	0.009	0.0011	−0.001
36	31.03	17	0.0198	−0.005	−0.004	−0.000	−0.001	−0.004	0.008
42	33.77	23	0.0686	−0.000	−0.005	0.004	0.000	0.000	0.004
48	37.61	29	0.1314	0.005	0.006	−0.002	−0.002	0.003	−0.000

Table 6.
An output SAS of model with ACF check of residuals.

Parameter	Estimate	Standard error	t value	Approx Pr > t	Lag
MA 1.1	0.934	0.01770	52.78	<0.0001	1
MA 1.2	−0.077	0.0072138	−10.64	<0.0001	3
MA 1.3	0.008	0.0038171	2.18	0.0293	13
MA 1.4	0.00685	0.0031724	2.16	0.0309	21
MA 1.5	0.017	0.0027856	5.92	<0.0001	27
MA 1.6	0.059	0.0067600	8.67	<0.0001	46
MA 2.1	0.98	0.0009744	1003.38	<0.0001	48
MA 3.1	−0.0364	0.0045572	−7.98	<0.0001	336
AR 1.1	1.1464	0.01855	61.81	<0.0001	1
AR 1.2	−0.295	0.0087427	−33.79	<0.0001	2
AR 1.3	−0.0104	0.0052195	−2.00	0.0454	5
AR 1.4	0.0189	0.0067496	2.80	0.0051	6
AR 1.5	−0.0234	0.0047509	−4.93	<0.0001	7
AR 1.6	−0.004	0.030582	−1.29	0.1958	11
AR 1.7	−0.0083	0.0033299	−2.49	0.0126	16
AR 1.8	−0.0125	0.0033252	−3.77	0.0002	18
AR 1.9	−0.007	0.0022520	−3.26	0.0011	35
AR 1.10	0.07	0.0067089	10.62	<0.0001	46
AR 2.1	0.03	0.0050410	5.86	<0.0001	48

Table 7.
An output SAS of model with CLS iterative.

4.3 Electrical load forecasting results

Based on the final results of the estimated parameters 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 obtained DSARIMA models $([1, 2, 5, 6, 7, 11, 16, 18, 35, 46], 1, [1, 3, 13, 21, 27, 46])(1, 1, 1)^{48} (0, 0, 1)^{336}$ with the model equation as follows:

$$\begin{aligned} & (1 - 1.1464B + 0.295B^2 + 0.0104B^5 - 0.0189B^6 + 0.0234B^7 + 0.004B^{11} \\ & + 0.0083B^{16} + 0.0125B^{18} + 0.0074B^{35} - 0.07B^{46})(1 - 0.03B^{48})Z_t^* = \\ & (1 - 0.934B + 0.077B^3 - 0.008B^{13} - 0.00685B^{21} - 0.017B^{27} \\ & - 0.059B^{46})(1 - 0.98B^{48})(1 + 0.0364B^{336})a_t \end{aligned}$$

After going through a reverse transformation Z_t electrical load for the comparison of predicted results with actual data (testing) in **Figure 7** below.

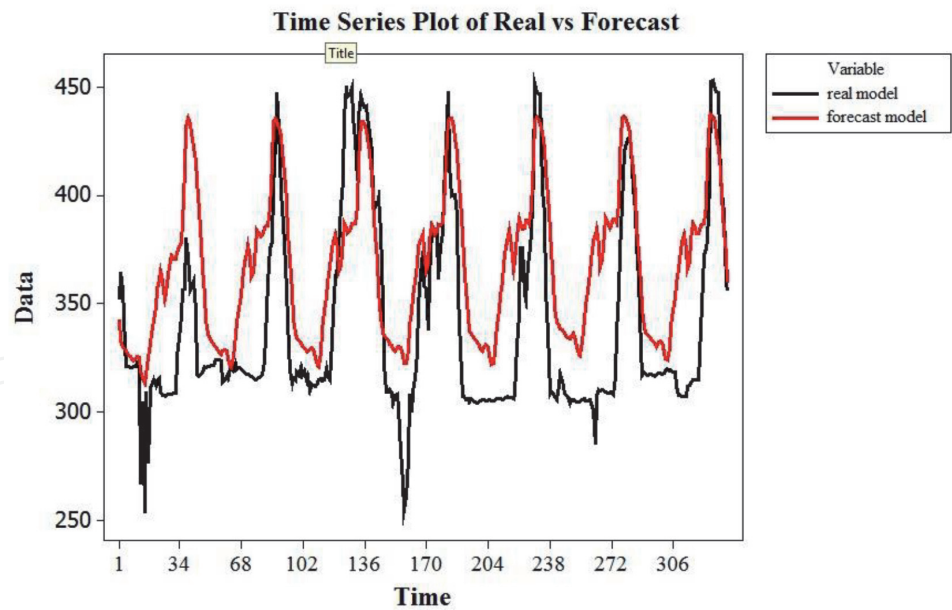


Figure 7.
Comparison of actual power with forecast power.

4.4 Model testing and measuring forecasting accuracy

Accuracy testing between actual power data and prediction results. Test using the MAPE procedure and obtained at 1.56 percent.

5. Electric load modeling

The application of descriptive analytic methods in this book is presented to obtain significant information in managing optimal electrical energy as the author did [18]. Through frequency distribution, data can be arranged based on certain criteria. Data categories are presented based on rank orders that contain ranking data from the top or highest load to the lowest data value.

5.1 Data distribution forecasting results

This electricity load forecasting data is a usage data for a week at intervals every half hour measurement at the power generation. This electricity load forecasting data sample is 336 (N = 336) with mean of 370.56 MWh, meaning that the value is centered at 370.526 MWh. Standard deviation of 36.2582 or the value of this deviation is not too large, this shows the diversity of data is not too large, which means the data is homogeneous.

Descriptive Statistics						
N	N*	Mean	StDev	Median	Minimum	Maximum
336	0	370,526	36,2584	370,551	312,912	445,746

Furthermore, forecasting the data shown in the time measurements every half-hour of electric power consumption in the load center in **Figure 8** below.

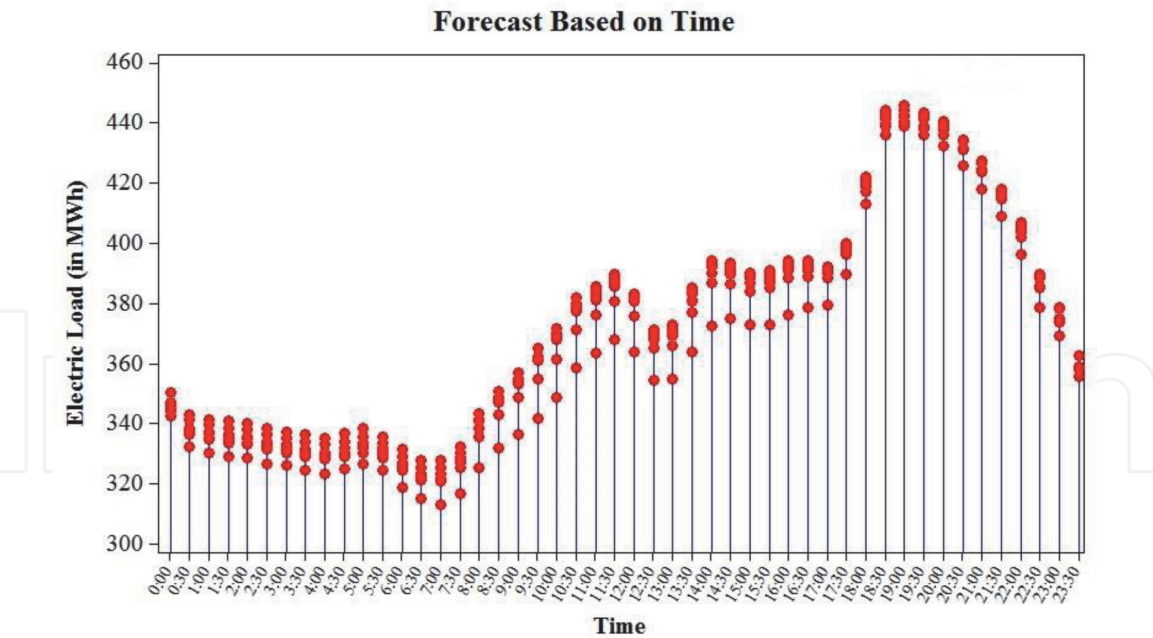


Figure 8.
Plot data forecasting.

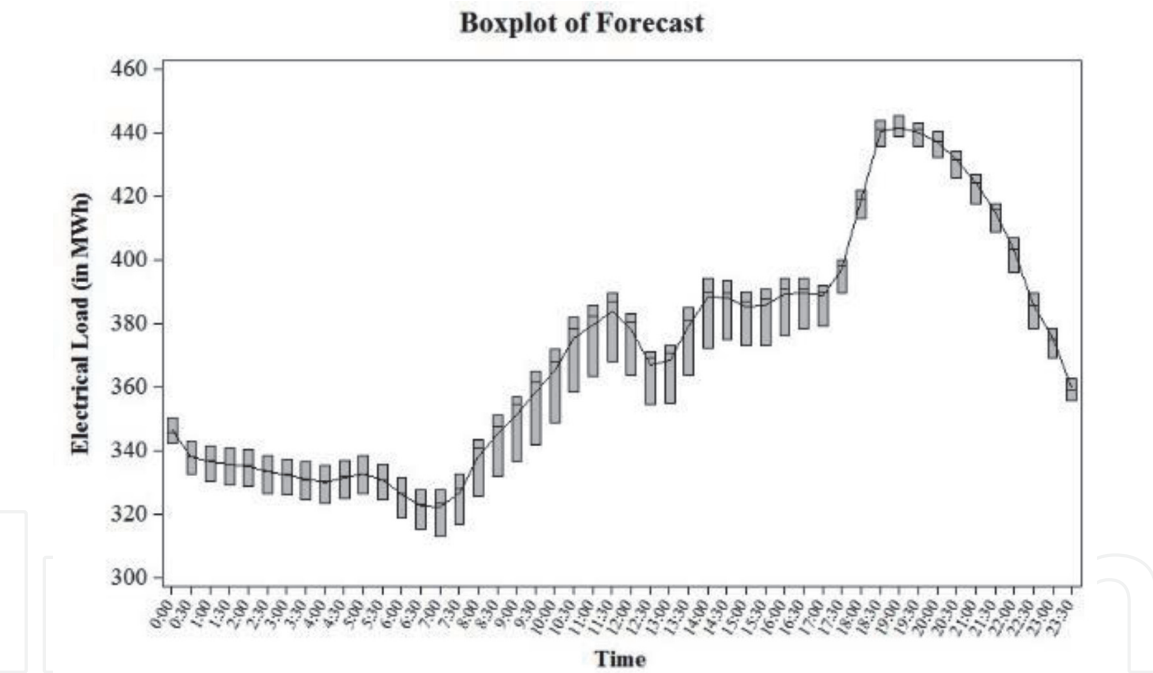


Figure 9.
Graph forecast boxplot.

Visualizations in other forms can be displayed in the form of boxplot graphics. **Figure 9** shows of range (in a box) every hour of measurement and the average value line of every half hour of measurement.

Figure 9 shows that data tend to be at the minimum level, first quartile and the median value. Electricity load increases at third quartile intervals and the maximum load. This condition occurs between 18:30 until 21:30 at night.

Each measurement of electric power absorption at the load center has a peak load. Based on the measurement data, it can be seen that the peak power load absorption occurs at 19:00 and generally the peak load tendency occurs at that hour.

No	Days	Mean	StDev	Median	Minimum	Peak Load	Time of peak load
1	Friday	375.143	35.4253	375.832	327.509	444.234	19:00
2	Saturday	373.635	36.2699	375.208	325.378	445.746	19:00
3	Sunday	361.193	36.8101	357.005	312.912	438.985	19:00
4	Monday	368.672	36.5413	368.417	320.639	440.478	19:00
5	Tuesday	370.616	36.2821	371.793	321.685	439.731	19:30
6	Wednesday	371.619	36.3137	372.718	322.735	441.976	19:00
7	Thursday	372.806	36.6616	374.899	323.262	442.727	19:30

Table 8.
Daily data samples.

Henceforth processing this distribution data through seasonal data that can be presented in the form of daily data, as follows.

The sample data used is Friday data and then the data will be presented in **Table 8** below.

Friday’s electricity load data—samples of electric load data are 48 (N = 48) with mean of 375.143 MWh, meaning that the value is centered at 375.143 MWh. Standard deviation of 35.4253 or the value of this deviation is not too large, this shows the diversity of data is not too large, which means the data is homogeneous.

Descriptive Statistics						
N	N*	Mean	StDev	Median	Minimum	Maximum
48	0	375,143	35,4253	375,832	327,509	444,234

On Friday shown in **Figure 10**, the peak load occurred at 19:00 amounting to 444.234 MWh with a minimum electric absorption range of 327.509 MWh. On Friday, the data has mean of 375.143 MWh.

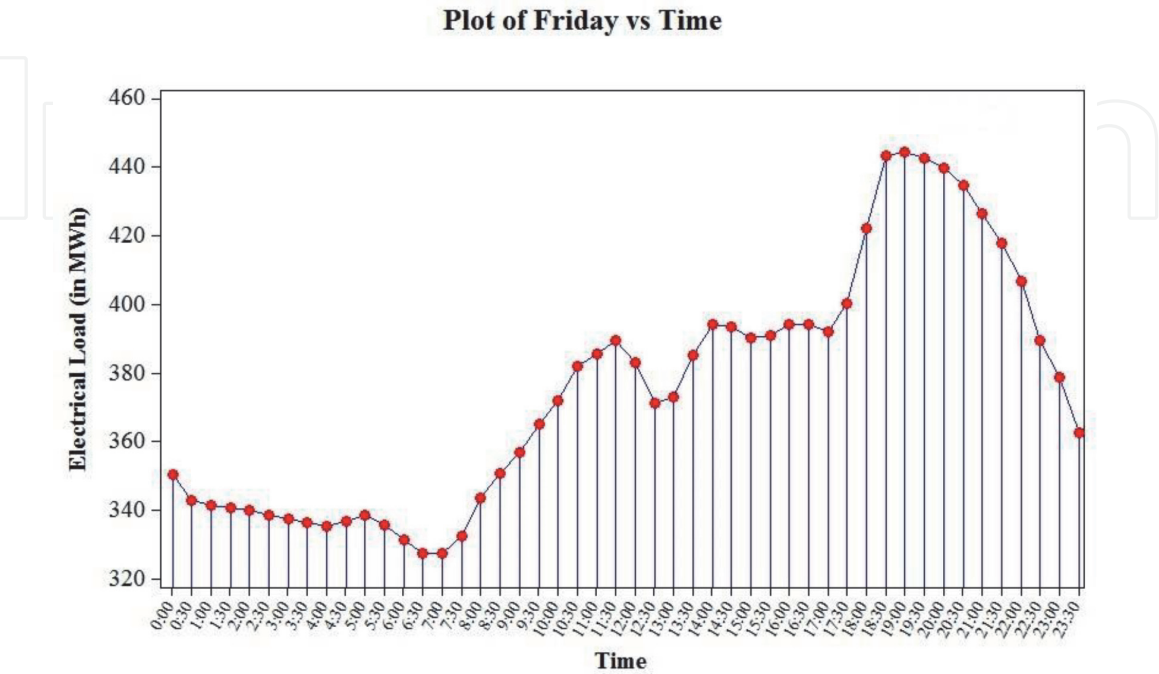


Figure 10.
Data plot on Friday.

Clusters	Interval range	Frequency
1	Min–Q1	97
2	Q1–Median	83
3	Median–Q3	86
4	Q3–Max	70
		N = 336

Table 9.
Range of clusters in the data variant.

Furthermore, seasonal electricity load data on a daily scale can be restated in the form of **Table 8** below.

5.2 Predicted cluster data

In descriptive analysis, frequency distribution, measurement of central tendencies and measurement of variability can be presented in the frequency distribution graph. The purpose of the presentation and information provided in addition to being able to describe the tendency of the data to form certain patterns, this analysis can also be used as a reference for changes in electric power in the power generation system.

The degree of data dispersion can be determined based on the range of interquartile intervals that indicate the homogeneity of the data. In this study, the electrical load cluster is defined as the range of quartile intervals to median value or is shown in the electrical load data below.

Descriptive Statistics: Forecast									
Variable	N	Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum
Forecast	336	370,53	36,26	1314,67	312,91	335,64	370,55	390,83	445,75
N for									

It can be seen that the data sample with N = 336 has an average of 370.53 MWh which means that the centralized data distribution is rated median. Standard deviation of 36.26 or the value of this deviation is not too large, this shows the diversity of data is not too large, which means the data is homogeneous.

Quartile intervals that divide data over median values form a cluster pattern, with the distribution of data presented in **Table 9** below

An important aspect of this data sample analysis is the presentation of data with seasonal variants. Data development by taking into account the seasonal variants of the hours and daily helped to optimize the management and operational decisions of the generating system both in scheduling and controlling.

6. Conclusion

One of the research trends in electrical engineering is time series analysis. This research includes forecasting studies and modeling of electrical load clusters. The time series analysis method is very suitable with the characteristics of the electrical load that is always fluctuating. This method is also able to produce different data or not included in the training data process.

For the purposes of this electrical load research, forecasting study using the DSARIMA method is an appropriate choice. This method accurately considers the seasonal parameters of the electricity load with MAPE of 1.56 percent when compared with the actual data.

Whereas the modeling of electrical load clusters based on descriptive analytic methods, obtained knowledge of the dynamics of electrical loads. The electrical load pattern has seasonal characteristics at daily and weekly intervals. This pattern forms a unique load characteristic at all times.

So, forecasting studies and modeling of electricity load clusters are able to answer the challenges of electricity energy utilization policies and the operation of generating systems that are able to maintain the balance of supply and demand.

Nomenclature

T	period of time (hours)
$P_{average}$	average load in period T (watts)
P_{peak}	peak load in the T (watts)
p, d, q	nonseasonal parts of the model
P, D, Q	seasonal parts of the model
S_1, S_2	1st and 2nd period seasonal
D_1, D_2, d	order of differences
S	number of period per season
m	maximum lag time
r_k	autocorrelation or time-lag , 2, 3, ... , k
Z_t	time series process in period T
Z_t^*	forecasting process in transformation in period T
Q_1	quartile 1
Q_3	quartile 3

Greek symbols

λ	Box-Cox transformation number
α_t	white noise
$\theta_q(B)$	regular MA polynomials of order q
$\Theta_{Q_1}(B^{S_1}), \Theta_{Q_2}(B^{S_2})$	MA polynomials of orders
$\varphi_p(B)$	regular AR polynomials of orders p
$\Phi_{P_1}(B^{S_1}), \Phi_{P_2}(B^{S_2})$	AR polynomials of orders

Abbreviations

MAPE	mean absolute percentage error
MWh	mega watt hours

IntechOpen

IntechOpen

Author details

Ismit Mado
University Borneo Tarakan, Tarakan City, Indonesia

*Address all correspondence to: ismitmado@borneo.ac.id

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Tsekouras GJ, Dialynas EN, Hatziargyriou ND, Kavatza S. A on-linier multivariable regression model for midterm energy forecasting of power systems. *Electric Power Systems Research*. 2007;**77**(12):1560-1568
- [2] McSharry PE, Bouwman S, Bloemhof G. Probabilistic forecasts of the magnitude and timing of peak electricity demand. *IEEE Transactions on Power Systems*. 2005;**20**(2):1166-1172
- [3] Gonzalez-Romera E, Jaramillo-Moran MA, Carmona-Fernandez D. Monthly electric energy demand forecasting based on trend extraction. *IEEE Transactions on Power Systems*. 2006;**21**(4):1946-1953
- [4] Taylor JW, Mc Sharry PE. Short-term load forecasting methods: An evaluation based on european data. *IEEE Transactions on Power Systems*. 2007;**22**(4):2213-2219
- [5] Hahn H, Meyer-Nieberg S, Pickl S. Electric load forecasting methods: Tool for decision making. *European Journal of Operational Research*. 2009;**199**(3): 902-907
- [6] Nia MM, Din J, Lam HY, Panagopoulos AD. Stochastic approach to a rain attenuation time series synthesizer for heavy rain regions. *International Journal of Electrical and Computer Engineering*. 2016;**6**(5):2379
- [7] Kamley S, Jaloree S, Thakur RS. Performance forecasting of share market using machine learning techniques: A review. *International Journal of Electrical and Computer Engineering*. 2016;**6**(6):2088-8708
- [8] Mado I, Soeprijanto A, 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*. 2018;**8**(6):4892-4901. ISSN: 2088-8708. DOI: 10.11591/ijece.v8i6
- [9] Suswanto D. Karakteristik beban tenaga listrik. In: Suswanto D, editor. *Sistem Distribusi Tenaga Listrik*. Indonesia: Departement of Electrical Engineering, Universitas Negeri Padang; 2009
- [10] Makridakis S, Hibon M. ARMA models and the Box-Jenkins methodology. *Journal of Forecasting*. 1997;**16**(3):147-163
- [11] Wei WW. *Time Series Analysis: Univariate and Multivariate Methods*. Vol. 2. Boston, MA: Pearson Addison Wesley; 2006;1:1:108-109
- [12] Soares LJ, Medeiros MC. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. *International Journal of Forecasting*. 2008;**24**(4):630-644
- [13] Mohamed N, Ahmad MH, Ismail Z, Suhartono S. Short term load forecasting using double seasonal ARIMA model. In: *Proceedings of the Regional Conference on Statistical Sciences*. Vol. 10. 2010. pp. 57-73
- [14] Kim SY, Jung HW, Park JD, Baek SM, Kim WS, Chon KH, et al. Weekly maximum electric load forecasting for 104 weeks by seasonal ARIMA model. *Journal of the Korean Institute of Illuminating and Electrical Installation Engineers*. 2014;**28**(1):50-56
- [15] Cryer JD, Chan KS. *Time Series Analysis: With Application*. R. Springer Science & Business Media; 2008;7:154-158
- [16] Walpole RE. *Pengantar Statistika*. Jakarta: PT Gramedia Pustaka Utama; 1993. ISBN 979-403-313-8

[17] Wiyono BB. Statistik Pendidikan,
Buku Ajar Mata Kuliah Statistik. 2001

[18] Mado I, Soeprijanto A, Suhartono S.
Electrical load adat clustering in PJB UP
Gresik based on time series analysis
approach. In: Proceedings of the
International Conference on Vocational
Education and Electrical Engineering.
Surabaya, Indonesia; 2015. pp. 261-266.
Available from: <http://digilib.unimed.ac.id/23841/1/Fulltext.pdf>