Double Seasonal ARIMA for Forecasting Electricity Demand of Kuaro Main Gate in East Kalimantan

M Azka1*, SA Wiradinata1, M Faisal1, and Suhartono2

1 Department of Mathematics, Institut Teknologi Kalimantan, Indonesia.
2 Department of Statistics, Institut Teknologi Sepuluh November, Indonesia.

*Email: muhammad.azka@lecturer.itk.ac.id

Abstract. This study uses intraday electricity load demand data from Kuaro Main Gate data in East Kalimantan as the basis of an empirical comparison of Double Seasonal ARIMA models for prediction up to a day ahead. For the purpose of this study, a one-year hourly Kuaro Main Gate data load demand from 1 January 2018 to December 2018 measured in Megawatt (MW) is used. In multiple times of load demand data, in addition to intraday and intra week cycles, and intra year seasonal cycle is also apparent. We extend the Double Seasonal ARIMA methods in order to accommodate the intra year seasonal cycle. The mean absolute percentage error (MAPE) is used as the measure of forecasting accuracy. A notable feature of the time series is the presence of both an intraweek and an intraday seasonal cycle. We also propose that a Double Seasonal ARIMA model with the one-step-ahead forecast as the most appropriate model for forecasting the two-seasonal cycles Kuaro Main Gate data load demand time series. We use the Statistical Analysis System package to analyze the data. Using the least-squares method to estimate the coefficients in a Double Seasonal ARIMA model, followed by model validation and model selection criteria, we propose the ARIMA \((1,1,1)(0,1,1)_{24}(0,1,1)_{168}\) within-sample MAPE of 0.000992 as the best model for this study. Comparing the forecasting performances by using k-step ahead forecasts and one-step-ahead forecasts, we found that the MAPE for the one-step ahead out-sample forecasts from any horizon ranging from one week lead time to one month one week lead time are all less than 5%. Therefore we propose that a double seasonal ARIMA model with a one-step-ahead forecast must be considered in forecasting time series data with two seasonal cycles.

1. Introduction
East Kalimantan-Indonesia’s geographical status as an archipelago makes domestic electricity coverage and power generation a particularly daunting challenge. The issue of electricity pricing is now dominating public discourse on the sector as the government seeks to reduce the burden of electricity subsidies on the state budget. Electricity demand forecasting is of great potential for the management of power systems. Long-term forecasts of the peak electricity demand are needed for capacity planning and maintenance scheduling [1]. Medium-term demand forecasts are required for power system operation and planning. Short-term load forecasts are required for the control and scheduling of power systems. Short-term forecasts are also required by transmission companies when a self-dispatching market is in operation, power generation, and has exclusive powers over the transmission, distribution, and supply of electricity to the public. There are several such markets in Kalimantan and Indonesia. For example, in Indonesia, one hour-ahead forecasts are a key input to the
balancing market, which operates on a rolling one hour-ahead basis to balance supply and demand after the closure of bi-lateral trading between generators and suppliers [2][3].

More generally, error in predicting electricity load has significant cost implications for companies operating in competitive power markets [4]. It is well recognized that meteorological variables have a very significant influence on electricity demand (see, for example, [5]). However, in short-term forecasting systems, multivariate modeling is usually considered impractical [6]. In such systems, the lead times considered are less than a day ahead, and, due to the expense or unavailability of weather forecasts, univariate methods are sometimes used for longer lead times. In a recent study [7], methods for short-term load forecasting are reviewed, and two intraday load time series are used to compare a variety of univariate methods. One of the aims of this paper is to validate the results of that study. It concluded that a double seasonal ARIMA was the most accurate method, with a new approach based on Box-Jenkins also performing well.

Due to academic interests and industrial needs, short term load forecasting has gained great attention compared to the others. Short term load forecasting plays an important role in a utility company because the accuracy of prediction will affect the power system operations. Forecast errors result in an unbalance between power supply and demand hence increase operations cost. Double seasonal ARIMA modeling has been used by many researchers, in a univariate modeling framework, as a comparison for evaluating alternative models (see [8]). One of the main problems that arise in modeling the use of short-term energy electricity with double seasonal ARIMA is the determination of the optimal order of AR and/or MA models. This is due to the complexity of the variable lag that can be used as the best candidate in the double seasonal ARIMA model. Until now, there is no single time series analysis literature that discusses double seasonal ARIMA models, especially the identification stages for determining the order of models that are in accordance with the data pattern. These orders tend to involve orders ranging from short lags to long lags, especially in the second seasonal order.

This paper will be structured as follows. We start by presenting the Box-Jenkins, double seasonal ARIMA model, specifically formulated to deal with the double seasonality that typically arises in load demand data. This seasonality involves intraday and intraweek seasonal cycles, and describe the details results of a double seasonal ARIMA model. We would hope that the best performing method in our study can serve as benchmarks in future studies with double seasonal ARIMA. Finally, we give our conclusions based on the forecasting evaluation method presented in this study.

The General ARIMA Models More generally, we can write a general ARIMA model as follows:

\[
\prod_{j=1}^{M} \phi_j(B) \prod_{i=1}^{K} (1-B^s)^d Z_t = \theta_0 + \prod_{k=1}^{N} \theta_k(B) a_t
\]

(1)

Thus, the model may contain K differencing factors, M autoregressive factors, and N moving average factors. This extension is useful in describing many nonstandard time series that, for example, may contain a mixture of seasonal phenomena of different periods. Because it is this general form that most time series software use, we now explain this general model in more detail.

Generally, a nonseasonal time series can be modeled as a combination of past values and past error, denoted as ARIMA\((p,d,q)\) can be written as follows:

\[
\phi_p(B)(1-B^s)^d Z_t = \theta_q(B) a_t
\]

(2)

\[
\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p
\]

\[
\phi_q(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_q B^q
\]
Where $Z_t$ is appropriately transformed load demand in period $t$; $(1 - B)^d$ is the nonseasonal differencing operator; $B$ is the backward shift operator; and $a_t$ is the purely random process. As an example let hence the model can be expressed as ARIMA $(0,0,1)$. Consider the model can be written as follows:

$$Z_t = (1 + 0.8395B)a_t = a_t + 0.8395a_{t-1}$$

(3)

2. Methodology
A Modelling Strategy For The Double Seasonal ARIMA Model: In 1972 George E.P Box and Gwilym M. Jenkins developed a method for analyzing stationary univariate time series data. In this section, the importance and general nature of the ARIMA approach to time series analysis are discussed. The novel contributions of this method and limitations are explained. Prerequisites of Box-Jenkins models are defined and explored. Different types of nonstationarity are elaborated. We also discuss tests for detecting these forms of nonstationarity and expound on transformations to stationarity. We then review problems following from the failure to fulfill these prerequisites, as well as common means by which these problems may be resolved. Programming examples with both SAS and MINITAB are included. The modeling procedure of Box-Jenkins ARIMA model involves an iterative five-stage process as follows:

Step 1: Preparation of data including transformation and differencing.
Step 2: Identification of the potential models by looking at the sample autocorrelations and the partial autocorrelations.
Step 3: Estimation of the unknown parameters by some optimization methods.
Step 4: Checking the adequacy of the fitted model by performing a normal probability plot, ACF and PACF on model residuals.
Step 5: Forecast future outcomes based on the known data.

Measuring Forecast Accuracy: The Principal question is how can one compare and contrast several competing models to determine which is the better model. The better model will usually fit the data well. The general model goodness of fit needs to be evaluated. Commonly used measures of goodness/lack of fit include the mean error, the mean percent error, the mean absolute error, and the mean absolute percentage error. Applied to forecasts, these measures are:

Mean absolute error

$$M = \frac{1}{T} \sum_{t=0}^{T} |y_t - \hat{y}_t|$$

Mean absolute percentage error

$$MPE = \frac{100}{T} \sum_{t=0}^{T} \left| \frac{y_t - \hat{y}_t}{y_t} \right|$$

Where $T$ is the total number of time periods (number of observations), is the actual value and is the forecast value at time $t$. These are average measures of percent and absolute error that can be used as indicators of forecast accuracy. Although there is no single absolute level above which the model is unacceptable. Data Set: The data used is half year hourly load demand measured in Megawatt (MW) from August 01, 2017 to December 31, 2017. They are gathered from PLN AP2B SISTEM KALTIM-Balikpapan, East Kalimantan electricity utility company, Balikpapan, Indonesia. PLN (The State-owned electricity company) is one of the biggest and most well-managed power companies in Indonesia in which this utility company has powered for decades through the generation, transmission, and distribution of electricity.
3. Result and Discussion

Plotting the load demand series after three-time differencing which are non-seasonal differencing (d = 1), daily seasonal differencing (D1 = 1, s1 = 24) and weekly seasonal differencing (D2 = 1, s2 = 168) in Figure 1 indicates that load series is stationary series. In order words, this identification step shows that the load data have two seasonal periods, which are daily and weekly seasonality with length 24 and 168 respectively. We then present the ACF and the PACF of the stationary load demand series after three-time differencing in Figure 2.

![Figure 1. Estimated autocorrelation function of hourly electricity demand in Kuaro Main Gate in East Kalimantan](image1)

![Figure 2. Estimated autocorrelation function of hourly electricity demand in Kuaro Main Gate-East Kalimantan](image2)

A Modeling Strategy For The Double Seasonal Arima Model

1. Model 1 : Double Seasonal ARIMA(0,1, 1)(0,1,1)24 (0,1,1)168

All the parameters of this model are significant at alpha 5% significance level with white noise residuals based on Ljung-Box statistic Q* until lags 24. This model gives 5 extreme residual values. In terms of the magnitude of the residuals, there are at Observations. The model’s residuals, however, do not satisfy the Normal Distribution because of the presence of outliers in the data. We considered lag polynomials up to order three. This choice was made arbitrarily, but it is consistent with other load forecasting studies, and it was supported by experimentation with several of the series. We based model selection on the Schwarz Bayesian Criterion, with the requirement that all parameters were significant (at the 5% level). We compared the Schwartz Bayesian information criterion (SBC) for an extensive range of different ARIMA models. The AIC and SBC of this model are 62680.53 and 62800.14 respectively.
2. Model 2 : Double Seasonal ARIMA([2,4,5,7,9,11,12,20,21,22,23,24,48],1,1)(0,1,1)_{24}(0,1,1)^{168}
All the parameters of this model are significant at alpha 5% significance level with white noise residuals based on Ljung-Box statistic Q* until lags 24. This model gives 5 extreme residual values. In terms of the magnitude of the residuals, there are at Observations. The model’s residuals, however, do not satisfy the Normal Distribution because of the presence of outliers in the data. We compared the Schwartz Bayesian information criterion (SBC) for an extensive range of different ARIMA models. The AIC and SBC of this model are 12914.73 and 13013.2 respectively.

3. Model 3 : Double Seasonal ARIMA (1,1,[1,2,4,5,7,9,11,12,20,21,22,23,24,48]) (0,1,1)_{24} (0,1,1)^{168}
All the parameters of this model are significant at alpha 5% significance level with white noise residuals based on Ljung-Box statistic Q* until lags 24. This model gives 5 extreme residual values. In terms of the magnitude of the residuals, there are at Observations. The model’s residuals, however, do not satisfy the Normal Distribution because of the presence of outliers in the data. We compared the Schwartz Bayesian information criterion (SBC) for an extensive range of different ARIMA models. The AIC and SBC of this model are 12752.78 and 12857.4 respectively.

4. Conclusion
We investigated the double seasonal ARIMA model for forecasting the double seasonal (daily and weekly Kuaro Main Gate-East Kalimantan data load demand time series. Comparing the forecasting performances by using k-step ahead forecasts and one-step-ahead forecasts, we found that the MAPE for the one-step ahead out-sample forecasts for any horizon ranging from one week lead time to one month lead time as tabulated in Table 3 are all less than 5%. The MAPE of the k-step ahead forecasts, on the other hand, increases as the time scale increases. It can be concluded that the one-step-ahead forecasts are not greatly influenced by the lead times and are more accurate than k-step ahead forecasts in forecasting Kuaro Main Gate-East Kalimantan load demand.

Acknowledgment
This research was supported by Institut Teknologi Kalimantan under its LPPM with contract number: 1996/IT10.III/PPM.01/2019. The researchers are grateful for this support. The researchers would also like to thank AP2B SITEM KALTIM PLN Balikpapan, East Kalimantan for providing the load data. We also wish to express our appreciation to all friends who help us directly and indirectly.

References
[1] P. E. Mcsharry, S. Bouwman, and G. Bloemhof, “Probabilistic Forecasts of the Magnitude and Timing of Peak Electricity Demand,” vol. 20, no. 2, pp. 1166–1172, 2005.
[2] E. González-romera, M. A. Jaramillo-morán, and D. Carmona-fernández, “Monthly Electric Energy Demand Forecasting Based on Trend Extraction,” vol. 21, no. 4, pp. 1946–1953, 2006.
[3] J. W. Taylor, “Density forecasting for the efficient balancing of the generation and consumption of electricity,” vol. 22, pp. 707–724, 2006.
[4] D. W. Bunn, “Forecasting Loads and Prices in Competitive Power Markets,” vol. 88, no. 2, pp. 163–169, 2000.
[5] J. W. Taylor and R. Buizza, “Neural Network Load Forecasting with Weather Ensemble Predictions,” no. September, p. 2002, 2002.
[6] D. W. Bunn, “Short-Term Forecasting : A Review of Procedures in the Electricity Supply Industry,” vol. 33, pp. 533–545, 1982.
[7] J. W. Taylor, L. M. De Menezes, and P. E. Mcsharry, “A comparison of univariate methods for forecasting electricity demand up to a day ahead,” vol. 22, pp. 1–16, 2006.
[8] G. A. Darbellay and M. Slama, “Forecasting the short-term demand for electricity Do neural networks stand a better chance ?,” vol. 16, pp. 71–83, 2000.