Forecasting electricity loads on national holidays in East Kalimantan.

P Hasanah\(^*\), SA Wiradinata\(^2\), M Azka\(^3\), CA Sari\(^4\)
\(^{1,2,3,4}\)Department of Mathematics, Institut Teknologi Kalimantan, Balikpapan, Indonesia

Email*: primadina@lecturer.itk.ac.id

Abstract. In the last decades, electricity becomes essential problems in East Kalimantan. Some previous research has conducted to assess the power plant systems and to examine the future electricity demand during the period. Moreover, there are some particular events such as national holidays that bring different impacts on the electricity demand. Forecasting electricity loads on national holidays will give significant information to the stakeholders in order to prevent the power outage of electricity. Two samples of hourly electricity loads in national holidays; New Year and Independence Day during 2015-2018 is analyzed in this research. Furthermore, the SARIMA model is used to estimate the suitable model to predict the future electricity loads. The result shows that SARIMA (2,2,0)(0,1,0)\(^{24}\) and SARIMA (1,1,0)(1,1,0)\(^{24}\) are the appropriate model for New Year and Independence Day, respectively. In addition, it is predicted that in 2020 the electricity demand on New Year will rise to 400 MegaWatt based on 95% of confidence intervals. On Independence Day, the electricity consumption will reach 300-450 MegaWatt in 2019 and 2020 with significant growth up to 10% each year.

1. Introduction
Electricity has been a popular issue in East Kalimantan for several years. Nowadays, electrification in East Kalimantan reaches 68.56% and the ratio of electrified villages is estimated up to 92.46% [1]. Currently, Electricity loads in East Kalimantan is provided by Mahakam System, an interconnected system through 150 kV transmission network. To overcome the lack of generating capacity for Mahakam System, Indonesian’s State Electricity Company, namely PLN, has improved their productivity gradually [2]. However, the experience of power outage remains as frequent problems among people of East Kalimantan.

To response the electricity problems in East Kalimantan, some previous research was conducted to solve the problems. Some research focused on the infrastructure and power plant systems, such as the research by Windarta [3] and Wahyuda [4]. The others proposed a solution by providing the future electricity forecasting for the Mahakam System. For instance, in 2018 Kresnawan [5] conducted a long term forecasting of electricity generation sectors. They used Long-range Energy Alternative Planning (LEAP) to examine the electricity demand and supply rate. The other research was the short term prediction of electricity demand conducted by Azka,et.al [6] using double seasonal arima.
All previous research focused on global electricity demand in East Kalimantan. Nonetheless, the electricity usage was influenced by some particular events such as the number of working days, holiday events and national celebrations. Electricity loads have been assumed to have different behaviour between normal and special days, which the electricity utilization during normal days are higher than the special days. Even though it has a different pattern, forecasting electricity loads on some special days will give essential information to PLN or the government to prevent the power outage especially in the special days. Therefore, a study of electricity forecasting in the special days should be discovered.

2. Methodology
This research was conducted with several steps. Firstly, the behaviour of the time series should be analyzed to determine the appropriate forecasting methods. The analytical method consists of determining trends, cycle, and seasonal series on the data. From the first step, we knew that the data set involved a seasonal pattern during years. Therefore, we tried to develop forecasting methods by Seasonal Arima (SARIMA) approach. Secondly, the preprocessing forecast was carried out to make stationary data. Then, it followed by determining the appropriate order of SARIMA to estimate electricity loads. Finally, the last step known as the diagnostic check, was measuring forecasting accuracy to choose the best SARIMA model.

2.1 Descriptive Statistics
This research examined the dataset of hourly electricity loads in the Mahakam System on New Year and Independence Day from 2015 to 2018. This data was obtained from PT. PLN (Persero) AP2B East Kalimantan System. In details, time series data for each year consists of load electricity data in each hour from 01.00 AM to 24.00 PM. The figure out of the data can be seen as below.

![Figure 1. Time series plot of new year electricity loads](image1)

![Figure 2. Time series plot of independence day electricity loads](image2)
2.2 Seasonal Autoregressive Integrated Moving Average (SARIMA)

ARIMA is the most popular method in time series forecasting. This model has been used widely in many time series problems. The ARIMA model aims to explain the existing autocorrelation in the data. ARIMA consists of Autoregressive (AR) and Moving Average (MA) models, whereas both the models integrated with differencing conditions in the data. There are lots of references regarding ARIMA models, however, the fundamental models can be found at [7], [8]. The Autoregressive models (AR) define as

\[ \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \cdots + \phi_p \tilde{Z}_{t-p} + a_t \]  

(1)

Where \( \tilde{Z}_t \) is time series data at time \( t \), \( p \) is the order of AR and \( a_t \) is white noise. This is like a multiple regression with lagged values \( \tilde{Z}_t \) as predictors and the model refers to this as an AR(P) model.

Otherwise, the Moving Average (MA) models explain as

\[ \tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q} \]  

(2)

Which \( q \) denotes the MA order.

The integrated model, ARMA will combine the model as follows

\[ \tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \cdots + \phi_p \tilde{Z}_{t-p} + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} \]  

(3)

Then the model is denoted as ARMA (p,q), where \( p \) and \( q \) are the order of AR and MA, respectively. If there were non stationary data, then there should be a differencing process. The degree of differencing, denoted as \( d \) used in integrated model ARIMA (p,d,q). The mathematics equation given below.

\[ \phi_p (B)(1 - B) d \tilde{Z}_t = \theta_q (B) a_t, \]  

(4)

where

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

(5)

and

\[ \theta_p (B) = (1 - \theta_1 B^1 - \cdots - \theta_q B^q) \]  

(6)

The identification of ARIMA order, (p,d,q), was commonly known from the behaviour of ACF and PACF plots. The significant lags of ACF and PACF determine the value of \( p \) and \( q \). The summary of the identification process given in table 1.

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

In many cases, there is a seasonal series in the dataset. Therefore, ARIMA models should be developed to accommodate the seasonal terms. This model is well known as Seasonal ARIMA or famously called as SARIMA model. A seasonal ARIMA model is formed by including additional seasonal terms in the ARIMA models we have seen so far. It is written as SARIMA \((p, d, q)(P, D, Q)_m\).
which the first order, \((p, d, q)\) refers to the non seasonal part of the models, otherwise \((P, D, Q)\) is the seasonal part of the model, and \(m\) determines the seasonal period of the data. The model describes as:

\[
\phi_p(B) \Phi_p(B^s)(1 - B)^d(1 - B^s)^D Z_t = \delta + \Theta_q(B^s)\theta_q(B) a_t .
\]  

(7)

In details,
\(\delta\) = intercept,
\(Z_t\) = time series data at time \(t\)
\((1 - B)^d\) = integrated,
\(\phi_p(B)\) = Autoregressive parameters at order-\(p\),
\(\theta_q(B)\) = Moving Average at order-\(q\),
\((1 - B^s)^D\) = seasonal integrated,
\(\Phi_p(B^s)\) = \(1 - \Phi_1 B - \cdots - \Phi_p B^{ps}\),
\(\Theta_q(B^s)\) = \(1 - \Theta_1 B - \cdots - \Theta_q B^{qs}\),
\(a_t\) = error at-\(t\),
\(d\) = regular differencing,
\(D\) = seasonal differencing.

2.3 Stationary Time Series

This step aimed to determine stationary time series for each dataset. Firstly, trend and seasonality must be removed from the dataset. It can be seen in Figure 1 and 2 that each dataset involved seasonality with the period of 24 hour. In addition, for Independence Day (Figure 2) dataset, it seems to have trend in the last two years. It was clear that those datasets were not only non stationary, but it involved a seasonality period.

To overcome the problems, based on [9] we should take a seasonal difference for each dataset. The first difference for seasonality components of New Year can be seen on Figure 3, while for Independence Day given on Figure 4.

![Figure 3. Seasonal difference for new year](image)

![Figure 4. Seasonal difference for independence day](image)
Regarding figure 3, it can be inferred on the seasonal lag (24) and its multiplication, there are non-significant spikes in New Year data. It means AR(0) and MA(0) can be used to estimate the seasonal order of SARIMA. In contrast, for Independence day, it can be seen that there are significant spikes around the seasonal lag (24) in PACF. Therefore, AR(1) can be used to determine seasonal order of SARIMA. Moreover, it is clear from the figures that the dataset is non-stationary, so it requires the next differences from non-seasonal components. Owing to Augmented Dickey Fuller (ADF) test for stationarity, the stationary condition for New Year reached after second differences. The non seasonal differences can be seen at Figure 5 and Figure 6.

![Figure 5. Stationary plot for new year](image)

From the stationary plot given in figure 5, we should determine the seasonal arima order based on ACF and PACF. The ACF plot informed the exponential decrease under the higher lag, that indicated MA(0), while the PACF plot spikes on the first and second lag. Thus, the AR(1) and AR(2) could be identified as alternative models. To sum up, the alternative SARIMA model for the New Year data set were SARIMA (1,2,0)(0,1,0)$_{24}$, and SARIMA (2,2,0)(0,1,0)$_{24}$.

Meanwhile, for The Independence Day’s dataset, after the deseasonality process, the dataset has been identified to be non stationary. Then the first difference on the non seasonality component successfully made it stationary based on the ADF test. The ACF and PACF plot can be seen below

![Figure 6. Stationary plot for independence day](image)

Furthermore, form ACF plot in figure 6 illustrated that the data followed MA(0). On the other hand, there were significant spikes on first and third lag in PACF that lead to AR (1) and AR(3) models. In conclusion, the alternative SARIMA model for Independence Day dataset were SARIMA (1,1,0)(1,1,0)$_{24}$ and SARIMA (3,1,0)(1,1,0)$_{24}$.

3. Result and Discussion

3.1 SARIMA model for New Year.

From the stationary test, there were two approximation models to estimate SARIMA model for New Year; SARIMA (1,2,0)(0,1,0)$_{24}$, and SARIMA (2,2,0)(0,1,0)$_{24}$. Despite those models, it was important
to consider the overfitting model by approaching the higher order of SARIMA. The overfitting model conducted on SARIMA \((3,2,0)(0,1,0)_{24}\). The estimation result given in the table below.

**Table 2.** SARIMA models for new year

| Models                  | Order | p-value  | Sig. |
|-------------------------|-------|----------|------|
| \((1,2,0)(0,1,0)_{24}\) | AR(1) | 2.275e-05| Yes  |
| \((2,2,0)(0,1,0)_{24}\) | AR(1) | 1.558e-07| Yes  |
|                        | AR(2) | 0.008094 | Yes  |
| \((3,2,0)(0,1,0)_{24}\) | AR(1) | 1.277e-08| Yes  |
|                        | AR(2) | 0.001167 | Yes  |
|                        | AR(3) | 0.078034 | No   |

It is obvious from Table 2 that the overfitting model for AR(3) did not significantly under the 5% of the significant level. It described that the AR(1) or AR(2) are adequate order to approach the SARIMA model. To continue the process, the diagnostic checks will assess for residuals in each model. It consists of checking the white noise condition, the normality, and the accuracy.

**Table 3.** Diagnostic checking for SARIMA models of new year

| Models                  | Sig. | White Noise | Normality | AIC     |
|-------------------------|------|-------------|-----------|---------|
| \((1,2,0)(0,1,0)_{24}\) | Yes  | Yes         | Yes       | 397.09  |
| \((2,2,0)(0,1,0)_{24}\) | Yes  | Yes         | Yes       | 392.64  |
| \((3,2,0)(0,1,0)_{24}\) | No   | Yes         | Yes       | 391.66  |

Each residual model was already fulfilling the white noise condition based on Ljung-Box statistic and the normality residuals based on Kolmogorov Smirnov test. Moreover the accuracy, represented by AIC, used to determine the best SARIMA model. The lowest AIC goes to SARIMA \((2,2,0)(0,1,0)_{24}\) with AIC value of 392.64. Therefore, SARIMA \((2,2,0)(0,1,0)_{24}\) has been chosen as the appropriate model to estimate the New Year dataset.

![Plot of Actual Vs Predicted Value for New Year](image)

**Figure 7.** Actual and predicted plot value for new year
Figure 8. Forecast for new year electricity loads

Prediction plots in Figure 8 illustrate the electricity loads in the next New Year holiday, that are the forecasting for New Year holiday in 2019 and 2020. In general the figure showed the decreasing trend during times. Furthermore, the behaviour in every 24 hours is relatively similar between the historical data. Based on the figure, the highest electricity consumption was approximately around 7-9 pm. While the pattern between forecast in 2019 and 2020 is similar, the electricity demand on New Year of 2020 is projected around 100-200 MegaWatt. However, based on 95% of confidence intervals, the electricity demand on New Year holiday can reach up to 400 MegaWatt.

3.2 SARIMA model for Independence Day

Using the same steps with the previous dataset, first of all is estimating the SARIMA models for alternative order then it was followed by testing with overfitting model. There were two models given from stationary test, SARIMA \((1,1,0)(1,1,0)_{24}\) and SARIMA \((3,1,0)(1,1,0)_{24}\) which has AR(3) as the maximum order. Then, SARIMA \((2,1,0)(1,1,0)_{24}\) was chosen as the overfitting model. The results of the coefficient test were given in the following table.

| Models                      | Orde | p-value     | Sig. |
|-----------------------------|------|-------------|------|
| \((1,1,0)(1,1,0)_{24}\)     | AR(1)| 0.01727     | Yes  |
|                             | SAR(1)| 2.948e-06  | Yes  |
| \((2,1,0)(1,1,0)_{24}\)     | AR(1)| 0.0146     | Yes  |
|                             | AR(2)| 0.58784    | No   |
|                             | SAR(1)| 3.385e-06  | Yes  |
| \((3,1,0)(1,1,0)_{24}\)     | AR(1)| 0.01464    | Yes  |
|                             | AR(2)| 0.58606    | No   |
|                             | AR(3)| 0.93067    | No   |
|                             | SAR(1)| 3.433e-06  | Yes  |
It is clear from Table 4 that only SARIMA (1,1,0)(1,1,0)_{24} which has all significant value for its coefficient. Therefore, the maximum of significant order was AR(1) and SAR(1). The process is followed by a diagnostic test to determine the best model. The result is provided in table 4.

Table 5. Diagnostic checking for SARIMA models of independence day

| Models                | Sig. | White Noise | Normality | AIC   |
|-----------------------|------|-------------|-----------|-------|
| (1,1,0)(1,1,0)_{24}   | Yes  | Yes         | Yes       | 625.15 |
| (2,1,0)(1,1,0)_{24}   | No   | Yes         | Yes       | 628.85 |
| (3,1,0)(1,1,0)_{24}   | No   | Yes         | Yes       | 626.86 |

Based on the diagnostic check, Sarima (1,1,0)(1,1,0)_{24} was chosen as the best model to estimate electricity loads on Independence Day. The model has the lowest AIC value at 625.15 and all the coefficients were significant under the significant level of 5%. Moreover, the fitted model can be shown at figure 9 and the future prediction is described at figure 10.

Figure 9. Forecast for independence day electricity loads

Figure 10. Forecast for independence day electricity loads

Electricity loads on Independence day are forecasted to be growth gradually in 2019 and 2020. From the figure it is clear that based on a 95% confidence interval, the electricity demand was projected around 200-500 MegaWatt. Otherwise, the electricity forecast between 2019 and 2020 had a similar pattern, where the demand in 2020 is predicted to be slightly higher than 2019. The maximum loads in 2019 and 2020 was approximately between 7-9 PM around 400 and 450, respectively.
4. Conclusion

In general, the electricity demand during the holidays has a similar pattern with the historical data which is the pick points span around 7-9 PM. It is literally obvious because at that time, households or other sectors were using electricity devices. However, this research was impactful to provide the range of electricity loads that is required by people in East Kalimantan. On the one hand, with the historical data from 2015-2018, the projection of electricity demand in the New Year holiday seems to decrease in 2019 and 2020. However, based on a 95% confidence interval, the electricity in New Year can reach to 400 MegaWatt. On the other hand, the electricity loads on Independence day is projected to rise around 10% from 2019 to 2020 with the range value around 300-450 MegaWatt.

In conclusion, the Indonesian’s State Electricity Company should provide the appropriate amount of electricity, especially in the national holidays. Based on this research, PLN should distribute around 300-450 MegaWatt Electricity during the national holidays with the growth requirement approximately 10% in every year.

Acknowledgement

This research was supported by Kemenristek/BRIN through LPPM ITK with contract number:2846/IT10.II/PPM.01/2020. The researchers are grateful for this support. The researchers would also like to thank AP2B SISTEM 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] Muslimin Tambunan W and Wahyuda 2019 Cooperation between power plant in East Kalimantan by integrating renewable energy power plant IOP Conf. Ser. Mater. Sci. Eng. 528
[2] Muslimin and Utomo D S 2018 Power plant allocation in East Kalimantan considering total cost and emissions IOP Conf. Ser. Mater. Sci. Eng. 337
[3] Windarta J et al. 2020 Planning for the utilization of hydro power in the Belayan river, East Kalimantan J. Phys. Conf. Ser. 1524
[4] Wahyuda Wahyuda Muslimin and Widodo S U 2019 Development of Transportation Models for Scheduling Power Plants and Electricity Distribution Icoemics 17 pp. 78–86
[5] Kresnawan M R Safitri I A and Darmawan I 2018 Long term projection of electricity generation sector in east kalimantan province: LEAP model application Proc. - 12th SEATUC Symp. p. 1–5.
[6] Azka M Wiradinata S A Faisal M and Suhartono 2020 Double Seasonal ARIMA for Forecasting Electricity Demand of Kuaro Main Gate in East Kalimantan IOP Conf. Ser. Mater. Sci. Eng. 846 pp. 1–6.
[7] Makridakis S, Wheelwright, S.S., McGee V E 1983 Forecasting Methods and Applications 2nd ed. Wiley.
[8] Montgomery, Jennings K 1983 Time Series Analysis and Forecasting Methods. IEE Colloq. 91.
[9] Hyndman R J and Athanasopoulos G 2018 Forecasting: Principles and Practice Princ. Optim. Des. pp. 421–455.