Comparison of Short-Term Load Forecasting Based on Kalimantan Data*

Syalam Ali Wira Dinata¹‡, Muhammad Azka¹, Primadina Hasanah¹, Suhartono², and Moh Danil Hendry Gamal³

¹Department of Mathematics, Institut Teknologi Kalimantan, Balikpapan, Indonesia
²Department of Statistics, Institut Teknologi Sepuluh November, Surabaya, Indonesia
³Department of Mathematics, University of Riau, Pekanbaru, Indonesia
*Corresponding Email: asyalam.ali.wira_dinata@lecturer.itk.ac.id

Abstract

This paper investigates a case study on short term forecasting for East Kalimantan, with emphasis on special days, such as public holidays. A time series of load demand electricity recorded at hourly intervals contains more than one seasonal pattern. There is a great attraction in using a modelling time series method that is able to capture triple seasonalties. The Triple SARIMA model has been adapted for this purpose and competitive for modelling load. Using the least squares method to estimate the coefficients in a triple SARIMA model, followed by model building, model assumptions and comparing model criteria, we propose and demonstration the triple Seasonal Autoregressive Integrated Moving Average model with AIC 290631.9 and SBC 290674.2 as the best model for this study. The Triple seasonal ARIMA is one of the alternative strategy to propose accurate forecasts of electricity load Kalimantan data for planning, operation maintenance and market related activities.

Keywords: forecasting, time series, triple SARIMA.

Copyright © 2021 Syalam Ali Wira Dinata, Muhammad Azka, Primadina Hasanah, Suhartono, and Moh Danil Hendry Gamal. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
1. Introduction

Load forecasting is an important problem especially in obtaining the high accuracy level. Short-term load forecasting has been a fundamental of major interest for the electricity industry. In this decade, the short time load forecasting is frequently developed by researchers. Traditionally, hourly forecasts with a lead time of between one hour and seven days are required for the scheduling and control of power systems. Accurate short-term forecasts of electricity demand (load) are crucial for making informed decisions regarding unit commitment, energy transfer scheduling, and load frequency control of the power systems.

Currently anomalous load conditions models (see, Rahman & Bhatnagar, (1988), Hyde & Hodnett, (1993), Hyde & Hodnett, (1997)) can capture special days. For handling not normal working days pattern, the model may need another anomalous method. anomalous load conditions is a one of contemporary approaches that is commonly used to model time series with special days (see, Kim et al., 2000), Song et al., 2005, Fidalgo & Lopes, 2005). Some researchers have suggested anomalous approaches with load conditions in modelling electricity data can be found in Dordonnat et al., (2008), Soares & Medeiros, (2008), and Srinivasan et al., (1995). Arora and Taylor Arora & Taylor, (2013) implemented anomalous for Great Britain with nine years of half-hourly load data. In this paper, triple seasonal ARIMA performance for forecasting accuracy with consider and select polynomial function the orders for AR and MA. The study also assesses the advantage of integrating a residual autocorrelation and partial autocorrelation.

Electricity demand is often modelled in terms of weather variable. However, univariate methods are frequently considered to be sufficient for short lead times because the weather variables tend to change in a smooth fashion over short time frames, and this will be captured in the demand series itself. A variety of univariate methods have been used for short-term load forecasting.

Many different methods and models have been proposed by researchers using Triple seasonal ARIMA time series for load forecasting including anomalous load conditions (Kim et al., 2000), (Arora & Taylor, 2013), and (Fidalgo & Lopes, 2005), Neural Network for anomalous load periods (Lamedica et al., 1996). Fuzzy linear regression method for load forecasting with variety of approaches such as the holidays (Song et al., 2005), Fuzzy Neural computation (Srinivasan et al., 1995), State space model (Dordonnat et al., 2008) and Triple seasonal methods have also been applied for load forecasting.

In this paper, our primary contribution is the extention of the triple seasonal methods to include the yearly seasonal cycle. However, for the reason of parsimony and to find one that there is true AR or MA model for the consideration of higher order selection models. We also consider the polynomial of order triple ARIMA including all lags by looking at the sample ACF and PACF and autocorrelation check for white noise.

2. Methods

This is a special case of the general rule based triple seasonal ARIMA model as follows:
\begin{align*}
(1 - \cdots - \phi_p B^p)(1 - \cdots - \Phi_P B^{P_S})(1 - \cdots - \Omega_P B^{P_{S_2}})(1 - \cdots - \Gamma_P B^{P_{S_3}}) \nabla^d \nabla_{S_1}^D \nabla_{S_2}^D \nabla_{S_3}^D \\
(I_{N_t} \zeta(B^{S_1}) + (1 - I_{N_t}) \zeta(B^{S_1}))(Z_t - c) = (1 - \cdots - \theta_q B^q)(1 - \cdots - \Theta_Q B^{Q_S})(1 - \cdots - \Psi_Q B^{Q_{S_2}}) \\
(1 - \cdots - \Lambda_{Q_3} B^{Q_{S_3}})(I_{N_t} \lambda(B^{S_1}) + (1 - I_{N_t}) \kappa(B^{S_1}))((I_{N_t} a_t^{(N)} + (1 - I_{N_t}) a_t^{(S)}))
\end{align*}

where

\begin{align*}
Z_t &= \text{the load observed at period } t, \\
c &= \text{constant parameter,} \\
B &= \text{the backward shift operator or lag operator,} \\
\phi_p, \Phi_P, \Omega_P, \Gamma_P &\text{ are AR polynomial functions of order } p, P, P_1, \text{ and } P_3, \\
\theta_q, \Theta_Q, \Psi_Q, \Lambda_{Q_3} &= \text{MA polynomial functions of order } q, Q, Q_1, \text{ and } Q_3, \\
a_t^{(N)} &\sim \text{NID}(0, \sigma_N^2) \quad \text{the model errors for normal,} \\
a_t^{(S)} &\sim \text{NID}(0, \sigma_S^2) \quad \text{the model errors for special days,}
\end{align*}

variances \( \sigma_N^2 \) and \( \sigma_S^2 \) while NID refers to a Normal and Independently distributed process.

The function \( \zeta(B^{S_1}) \) and \( \lambda(B^{S_1}) \) accommodate the yearly seasonal effect for normal days, the function \( \zeta(B^{S_1}) \) and \( \kappa(B^{S_1}) \) accommodate the yearly seasonal effect for special days.

For example the multiplicative rule based triple seasonal ARIMA model can be expressed as \( \phi_1 = 0, \Phi_1 = 0, \Omega_1 = 0, \Gamma_1 = 1 \) and differencing is a technique that can also be used to remove seasonal components and trends with \( d = 1, D_1 = 1, D_2 = 1, D_3 = 1, S_1 = 24, S_2 = 168, S_3 = 8760 \) and \( \theta_1 = 0, \Theta_1 = 0, \Psi_1 = 0, \Lambda_1 = 0 \) hence the model can be expressed as model rule based triple seasonal ARIMA \((0,1,0)(0,1,0)24(0,1,0)168(0,1,1)8760\).

\begin{align*}
&1 - \Gamma_1 B^{8760} (1 - B)(1 - B^{24})(1 - B^{168})(1 - B^{8760})(I_{N_t} \zeta(B^{S_1}) + (1 - I_{N_t}) \zeta(B^{S_1}))(Z_t - c) = \\
&(1 - I_{N_t} \lambda(B^{S_1}) + (1 - I_{N_t}) \kappa(B^{S_1}))((I_{N_t} a_t^{(N)} + (1 - I_{N_t}) a_t^{(S)}))
\end{align*}

where

\begin{align*}
\zeta(B^{S_1}) &= 1 + \tau_1 B^{8760} + \tau_2 B^{8760} + \tau_3 B^{8760} + S_3(t - 8760) + S_3(t - 8760) + S_3(t - 8760) \\
\zeta(B^{S_1}) &= 1 + \omega_1 B^{8760} + \omega_2 B^{8760} + \omega_3 B^{8760} + S_3(t - 8760) + S_3(t - 8760) + S_3(t - 8760) \\
\lambda(B^{S_1}) &= 1 + \mu_1 B^{8760} + \mu_2 B^{8760} + \mu_3 B^{8760} + S_3(t - 8760) + S_3(t - 8760) + S_3(t - 8760) \\
\kappa(B^{S_1}) &= 1 + v_1 B^{8760} + v_2 B^{8760} + v_3 B^{8760} + S_3(t - 8760) + S_3(t - 8760) + S_3(t - 8760)
\end{align*}

3. Results and Discussion

The data used is year hourly load measured in Megawatt (MW) from January 01, 2015 to December 31, 2018. They are gathered from PLN AP2B SISTEM KALTIM-Balikpapan, KM. 15 Karang Joang Nort Balikpapan, Mahakam East Kalimantan electricity utility company, Balikpapan Indonesia.

The data used in this study were obtained from the East Kalimantan electricity utility company, PLN, main gate Mahakam Balikpapan. The data were divided into
sets: Initialization set and test set. Figure 2 plots the initialization set data, it is clear from Figure 1 that Mahakam-East Kalimantan load demand data is non-stationary. The ACF and PACF in figure 2 and figure 3 respectively, shows the capture of seasonal pattern, plotting the figure 4 is plot load series with the first differencing \([ (1 - B), d = 1 ] \), figure 5 is the plot data series after the daily seasonal differencing with length 24\([ (1 - B)(1 - B^{24}), d = 1, D_1 = 1 ] \), figure 6 shows load the output plot series after the weekly seasonal differencing with length 168 \([ (1 - B)(1 - B^{24})(1 - B^{168}), d = 1, D_1 = 1, D_2 = 1 ] \) and it is clear from figure 7, load electricity is stationary pattern after fourth times differencing with non-seasonal, daily, weekly and yearly periods and figure 7 is load series for \([ (1 - B)(1 - B^{24})(1 - B^{168})(1 - B^{8760}), d = 1, D_1 = 1, D_2 = 1, D_3 = 1 ] \). The ACF and PACF in figure 8 for differencing for \([ (1 - B), d = 1 ] \) load, figure 9 for seasonal differencing for \([ (1 - B)(1 - B^{24}), d = 1, D_1 = 1 ] \) load, figure 10 daily and weekly seasonal differencing for \([ (1 - B)(1 - B^{24})(1 - B^{168}), d = 1, D_1 = 1, D_2 = 1 ] \) load and daily, weekly and yearly in figure 11. differencing for \([ (1 - B)(1 - B^{24})(1 - B^{168})(1 - B^{8760}), d = 1, D_1 = 1, D_2 = 1, D_3 = 1 ] \) load shows clearly the presence of seasonal.

Figure 1: Electricity load in Mahakam - East Kalimantan
Figure 2: Autocorrelation Function load before non-seasonal and seasonal differencing

Figure 3: Partial Autocorrelation function load before non-seasonal and seasonal differencing
Figure 4: An output plot series \([ (1 - B), d = 1 ] \)

Figure 5: An output plot series \([ (1 - B)(1 - B^{24}), d = 1, D_1 = 1 ] \)
Figure 6: An output plot series $[(1 - B)(1 - B^{24})(1 - B^{168}), d = 1, D_1 = 1, D_2 = 1)]$

Figure 7: Load series for $[(1 - B)(1 - B^{24})(1 - B^{168})(1 - B^{8760}), d = 1, D_1 = 1, D_2 = 1, D_3 = 1)]$
Figure 8: Differencing for \([(1 - B), d = 1]\) load

Figure 9: Differencing for \([(1 - B)(1 - B^{24}), d = 1, D_i = 1]\) load
Figure 10: Differencing for \([(1-B)(1-B^{24})(1-B^{168}), d = 1, D_1 = 1, D_2 = 1)] \text{load}

Figure 11: Differencing for \([(1-B)(1-B^{24})(1-B^{168})(1-B^{8760}), d = 1, D_1 = 1, D_2 = 1, D_3 = 1)] \text{load}
Table 1: Descriptive Statistics after Non-seasonal and seasonal Differencing

### Non-seasonal differencing with length 1 [(1 − B), d = 1]

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>43816</td>
<td>0.0041</td>
<td>0.0915</td>
<td>-0.525</td>
<td>19.155</td>
<td>366.924</td>
</tr>
</tbody>
</table>

### CoefVar

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>469921.07</td>
<td>178.500</td>
<td>1.60672E+07</td>
<td>-394.190</td>
<td>-10.650</td>
<td>-1.575</td>
</tr>
</tbody>
</table>

### Q3

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.370</td>
<td>412.240</td>
<td>806.430</td>
<td>20.020</td>
<td>-0.29</td>
<td>284.610</td>
</tr>
</tbody>
</table>

### Daily seasonal differencing with length 24 [(1 − B)(1 − B^{24}), d = 1, D_1 = 1]

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>43816</td>
<td>0.0002</td>
<td>0.0798</td>
<td>-0.0165</td>
<td>16.693</td>
<td>278.646</td>
</tr>
</tbody>
</table>

### CoefVar

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>9421570.01</td>
<td>7.750</td>
<td>1.21883E+07</td>
<td>-407.400</td>
<td>-6.090</td>
<td>0.0100</td>
</tr>
</tbody>
</table>

### Q3

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.980</td>
<td>419.680</td>
<td>827.080</td>
<td>12.070</td>
<td>0.11</td>
<td>339.818</td>
</tr>
</tbody>
</table>

### Weekly seasonal differencing with length 168 [(1 − B)(1 − B^{24})(1 − B^{168}), d = 1, D_1 = 1, D_2 = 1]

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>43816</td>
<td>0.000</td>
<td>0.112</td>
<td>0.034</td>
<td>23.297</td>
<td>542.734</td>
</tr>
</tbody>
</table>

### CoefVar

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5034691.92</td>
<td>20.140</td>
<td>2.36220E+07</td>
<td>-424.650</td>
<td>-9.140</td>
<td>0.070</td>
</tr>
</tbody>
</table>

### Q3

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.055</td>
<td>416.230</td>
<td>840.880</td>
<td>18.195</td>
<td>-0.02</td>
<td>670.109</td>
</tr>
</tbody>
</table>

### Yearly seasonal differencing with length 8760

[(1 − B)(1 − B^{24})(1 − B^{168})(1 − B^{8760}), d = 1, D_1 = 1, D_2 = 1, D_3 = 1]

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>43816</td>
<td>0.001</td>
<td>0.153</td>
<td>-0.013</td>
<td>28.533</td>
<td>814.125</td>
</tr>
</tbody>
</table>

### CoefVar

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2305709.03</td>
<td>42.900</td>
<td>2.82225E+07</td>
<td>-449.660</td>
<td>-8.930</td>
<td>0.000</td>
</tr>
</tbody>
</table>

### Q3

<table>
<thead>
<tr>
<th>Count</th>
<th>Mean</th>
<th>SE Mean</th>
<th>TrMean</th>
<th>StDev</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.680</td>
<td>436.710</td>
<td>886.370</td>
<td>17.610</td>
<td>-0.02</td>
<td>1017.924</td>
</tr>
</tbody>
</table>
The load demand series after three time differencing which are non-seasonal differencing, daily seasonal differencing and weekly seasonal differencing indicates that load series is stationary series are shown in figure 3. There are several non-seasonal lags (lag 1, lag 2, ..., lag 8) and the ACF tends to be cut off after lag 1 whereas PACF diminishes dies down. On the other hand, ACF dan PACF at seasonal lags (lag 12, lag 24, ...) tend to cut off after lag 12, lag 24, lag 168 and lag 8760.

**Model 1**

The first selected Triple Seasonal ARIMA $(0,1,1) (0,1,1)^{24}(0,1,1)^{168}(0,1,1)^{8760}$

\[
\beta(B)(1 - B)(1 - B^{24})(1 - B^{168})(1 - B^{8760})Z_t = \delta(B)\psi(B)\varsigma(B)a_t
\]  

(3)

where

\[
\beta(B) = (1 + 0.27997B) \\
\delta(B) = (1 - 0.2895B)(1 - 0.82912B^{24}) \\
\psi(B) = (1 - 0.87918B^{168}) \\
\varsigma(B) = (1 - 0.45053B^{8760})
\]

**Model 2**

The second selected Triple Seasonal ARIMA $(0,1,2) (0,1,2)^{24}(0,1,2)^{168}(0,1,2)^{8760}$

\[
\beta(B)(1 - B)(1 - B^{24})(1 - B^{168})(1 - B^{8760})Z_t = \delta(B)\psi(B)\varsigma(B)a_t
\]  

(4)
where

\[ \beta(B) = (1 - 0.03302B) \]
\[ \delta(B) = (1 - 0.03302B^2)(1 - 0.83622B^{24}) \]
\[ \psi(B) = (1 - 0.88137B^{168}) \]
\[ \zeta(B) = (1 - 0.45277B^{8760}) \]

**Model 3**
The third selected Triple Seasonal ARIMA \((1,1,0)(1,1,0)^{24}(1,1,0)^{168}(1,1,0)^{8760}\)

\[ (1 + 0.27997B)(1 - 0.1391B^2)(1 - 0.82863B^{24})(1 - 0.87918B^{168})(1 - 0.45047B^{8760})Z_t = a_t \quad (5) \]

**Model 4**
The fourth selected Triple Seasonal ARIMA \((1,1,2)(1,1,0)^{24}(1,1,0)^{168}(1,1,0)^{8760}\)

\[ (1 + 0.24893B)(1 + 0.47536B^{24})(1 + 0.4926B^{168})(1 + 0.42524B^{8760})Z_t = a_t \quad (6) \]

**Model 5**
The fifth selected Triple Seasonal ARIMA \((2,1,1)(2,1,1)^{24}(2,1,1)^{168}(2,1,1)^{8760}\)

\[ (1 + 0.03621B^2)(1 - 0.473B^{24})(1 - 0.491B^{168})(1 - 0.425B^{8760})Z_t = a_t \quad (7) \]

We have seen that a simple and widely applicable stochastic model for the analysis of nonstationary time series, which contains seasonal component is triple seasonal multiplicative Model 1.

Table 2: the autoregressive and moving average coefficients in Triple Seasonal ARIMA

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Approx Pr &gt;</th>
<th>Lag</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA1,1</td>
<td>0.13910</td>
<td>0.005543</td>
<td>25.09</td>
<td>&lt;.0001</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>MA2,1</td>
<td>0.82863</td>
<td>0.0030545</td>
<td>271.28</td>
<td>&lt;.0001</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>MA3,1</td>
<td>0.87918</td>
<td>0.0026383</td>
<td>333.24</td>
<td>&lt;.0001</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8760</td>
</tr>
<tr>
<td>MA4,1</td>
<td>0.45047</td>
<td>0.005338</td>
<td>84.39</td>
<td>&lt;.0001</td>
<td>8760</td>
</tr>
<tr>
<td>AR1,1</td>
<td>-0.27997</td>
<td>0.005375</td>
<td>-52.08</td>
<td>&lt;.0001</td>
<td>1</td>
</tr>
</tbody>
</table>

Variance Estimate 265.1986
Std Error Estimate 16.28492
AIC 290631.9
SBC 290674.2
Number of Residuals 34523
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. Table 1 shows three of our diagnostic tools in one display a sequence plot of the standardized residuals, the sample ACF of the residuals, and p-values for the Ljung-Box test statistic for a whole range of values of k from 6 to 48. The horizontal dashed line at 5% helps judge the size of the p-values. It is seen that the series has significant autocorrelation at lags 1 and 24.

The standard error on the lag 1 = 0.0055, lag 24= 0.0030, lag 168 = 0.0026 and lag 8760 = 0.005, models seem to be catch on the dependence structure of the color property time series quite well. The horizontal dashed line at 5% helps judge the size of the p-values. It is seen that the series has significant autocorrelation at lags 1 and 24. The estimated MA1,1 with 13%, MA2,1 with 82%, MA3,1 with 87%, MA4,1 with 45%, and AR1,1 with -27% models seem to be catch on the dependence structure of the color property time series quite well. Specifically, the models are first-order moving average, or MA1,1, MA2,1, MA3,1, MA4,1, AR1,1 models with standard error 0.5%, 0.3%, 0.2%, 0.5% and 0.5% respectively.

Table 3: The SAS output of the first model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MA1,1</th>
<th>MA2,1</th>
<th>MA2,2</th>
<th>MA2,3</th>
<th>MA3,1</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA1,1</td>
<td>1.000</td>
<td>-0.026</td>
<td>-0.007</td>
<td>0.005</td>
<td>-0.274</td>
</tr>
<tr>
<td>MA2,1</td>
<td>-0.026</td>
<td>1.000</td>
<td>-0.223</td>
<td>-0.052</td>
<td>0.035</td>
</tr>
<tr>
<td>MA3,1</td>
<td>-0.007</td>
<td>-0.223</td>
<td>1.000</td>
<td>-0.047</td>
<td>0.011</td>
</tr>
<tr>
<td>MA4,1</td>
<td>0.005</td>
<td>-0.052</td>
<td>-0.047</td>
<td>1.000</td>
<td>0.004</td>
</tr>
<tr>
<td>AR1,1</td>
<td>-0.274</td>
<td>0.035</td>
<td>0.011</td>
<td>0.004</td>
<td>1.000</td>
</tr>
</tbody>
</table>

These two criteria penalize the sum of squared residuals for including additional parameters in the model. Models that have small values of the AIC or SBC are considered good models. Good models are obtained by minimizing either the AIC or BIC. Our preference is to use the SBC. It generally results in smaller, and hence simpler model and so its use is consistent with the time-honored model-building principle of parsimony (all other thing being equal, simple models are preferred to complex ones. In this section, we present the AIC and SBC results for hourly lead times up to one day ahead calculated for the one month post sample period of each of the two series of load data. The MAPEs of one-step and l-step ahead out-sample forecasts using Model 1 presented in Table 2. Because the model in (4) is adequate, we can use it to forecast the future employment figures. For reasons of parsimony, we deferred from considering higher order models. We investigated differencing, but this led to poorer forecast accuracy. As discussed, for a given forecast origin, say t = 9098, ..., 43480, forecasts can be calculated directly from the difference equation.

Thus, the l-step ahead forecast from the time origin t = 9098, ..., 43480 is given by
The simulated l-step-ahead forecast errors are typically used for forecast monitoring. The reason for this is that changes in the underlying time series will also typically be reflected in the forecast errors.

\[ Z_{t+l} = Z_{t+l-1} + Z_{t+l-24} - Z_{t+l-25} + Z_{t+l-336} - Z_{t+l-337} - Z_{t+l-384} + Z_{t+l-385} + \]
\[ a_{t+l} - 0.05a_{t+l-1} + 0.76a_{t+l-24} - 0.04a_{t+l-25} + 0.85a_{t+l-336} - 0.23a_{t+l-337} - 0.65a_{t+l-384} \] 

Figure 13: The simulated l-step ahead in-sample forecasts the Model 1 using triple seasonal ARIMA.

There is no guarantee that allowing for the daily, weekly and yearly cycle in a time series model will lead to improved accuracy for forecasting to just one day ahead. If a time series consists of uncorrelated observations in this time series are normally distributed, the time series is Gaussian white noise. Ideally forecast errors are Gaussian white noise. The normal probability plot indicate the residuals are approximately normally distributed and Figure 15 shows that, for the time series plot of the residual from forecasting Model 1.
Residuals are useful in checking whether a model has adequately captured the information in the data. A good forecasting method will yield residuals with the residuals have zero mean. If the residuals have a mean other than zero, then the forecasts are biased. The underlying rationale for this is that if the model were correct and the parameter values known exactly, then the residuals would be simply the true errors from which the data are assumed to have been generated. Therefore, if the form of the model is correct and the parameters are estimated, the residuals calculated from the fitted model should (for large n, at least, as the parameter
estimates generally are consistent) behave nearly as though they were white noise; and diagnostic checks and tests of fit based on this similarity of white noise. Plot of the residuals from this model indicate that in addition to an underlying outlier, there is additional structure.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>289528.9</td>
<td>302484.8</td>
<td>291126</td>
<td>290631.9</td>
<td>293053.1</td>
</tr>
<tr>
<td>SBC</td>
<td>289571.2</td>
<td>302527.1</td>
<td>291159.8</td>
<td>290674.2</td>
<td>293086.9</td>
</tr>
<tr>
<td>Variance Estimate</td>
<td>256.8598</td>
<td>374.3094</td>
<td>269.0294</td>
<td>265.1986</td>
<td>284.4739</td>
</tr>
</tbody>
</table>

Once parameters have been estimated, we check on the adequacy of the model for the load data series. The estimate values of these regular, seasonal and nonseasonal parameters of Model 1 until Model 5 are greater than 2%, with highly significant at alpha less than 0.0001 significance level.

4. Conclusion

In this paper, presented a case study on load forecasting for East Kalimantan, with emphasis on forecasting load on special days using a triple Seasonal ARIMA method for Mahakam load data. In comparison with that study, modelling load for East Kalimantan is more challenging, due to the relatively large number of different types of special days in Mahakam.

Acknowledgments. This research was supported by RISTEK DIKTI under its LPPM No. 2837/IT10.II/PPM.01/2020 The researchers are grateful for this support.

References


