Short Term Electricity Load Demand Forecasting in Indonesia by Using Double Seasonal Recurrent Neural Networks

Suhartono and Alfonsus Julanto Endharta

Abstract—Neural networks have apparently enjoyed considerable success in practice for predicting short-term hourly electricity demands in many countries. Forecasting of short-term hourly electricity in some countries usually is done by employing classical time series methods such as Winter’s method and Double Seasonal ARIMA model. Recently, Feed-Forward Neural Networks (FFNN) is also applied for electricity demand forecasting, including in Indonesia. The application of Double Seasonal ARIMA for forecasting short-term electricity load demands in most cities in Indonesia shows that the model contains both order of autoregressive and moving average. Moving average order can not be represented by FFNN. In this paper, we use an architecture of Neural Network that able to represent moving average order, i.e. Elman-Recurrent Neural Network (RNN). As a case study, we use data of hourly electricity load demand in Mengare, Gresik, Indonesia. The results show that the best ARIMA model for forecasting these data is ARIMA ((1,2,3,4,6,7,9,10,14,21,33),0,1) **(1,1,0)**. There are 14 innovational outliers detected from this ARIMA model. We use 4 different architectures of RNN particularly for the inputs, i.e. the input units are similar to ARIMA model predictors, similar to ARIMA predictors plus 14 dummy outliers, the 24 multiplied lagged of the data, and the combination of 1 lagged and the 24 multiplied lagged plus minus 1. The results show that the best network is the last one, i.e., Elman-RNN((22,3,1)). The comparison of forecast accuracy shows that Elman-RNN yields that the best network is the last one, i.e., Elman-RNN((22,3,1)). The lagged and the 24 multiplied lagged plus minus 1. The results show that the best ARIMA model for forecasting short-term electricity load demands from time to time. PLN works by predicting electricity power use prior to electricity power which is consumed by customers hourly. The prediction made is based on prior electricity power use. The amount of electricity power use prediction is held to optimize electricity power used by customers, so that there will not be any electricity extravagancy or extinction. The prediction can be done by using some forecasting methods, such as double seasonal ARIMA model and Neural Network (NN) method. Some researches that are related to short-term electricity power forecasting can be seen in [3], [4], [6], [7], [8], [9], [11], [12], [13] and [14]. Neural network methods used in those researches are Feedforward Neural Network, which is known as AR-NN model. This model can not get and represent moving average order effect in time series. Some prior researches, in other countries or in Indonesia, show that ARIMA model for the electricity consumption data tends to include MA order (see [9] and [13]).

The aim of this research is to study further about other NN type, i.e. Elman-Recurrent Neural Network (RNN) which can explain AR and MA order effects simultaneously for double seasonal time series data forecast, and compare the forecast accuracy with double seasonal ARIMA model.

II. FORECASTING METHODS

There are many quantitative forecasting methods based on time series approach. In this section, some forecast methods used in this research, such as ARIMA model and Neural Network, will be explained concisely.

A. ARIMA Model

One of time series models which is popular and mostly used is ARIMA model. Based on [16], autoregressive (AR) model shows that there is a relation between a value in the present (Z_t) and values in the past (Z_{t-k}), added by random value. Moving average (MA) model shows that there is a relation between a value in the present (Z_t) and residuals in the past (a_{t-k} with k = 1,2,…). ARIMA(p,d,q) model is a mixture of AR(p) and MA(q), with a non-stationary data pattern and d differencing order. The form of ARIMA(p,d,q) is

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

where p is AR model order, q is MA model order, d is differencing order, and

\[ \phi_p(B) = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_p B^p) \]

\[ \theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q) \]

Generalization of ARIMA model for a seasonal pattern data, which is written as ARIMAR(p,d,q)(P,D,Q)s, is [16]

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

where s is seasonal period, and

\[ \Phi_p(B^s) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_p B^{ps}) \]

Keywords—Double Seasonal, ARIMA, Recurrent Neural Network, short-term electricity load demand.

I. INTRODUCTION

PT PLN (Perusahaan Listrik Negara) is Government Corporation that supplies electricity needs in Indonesia. This electricity needs depend on the electronic tool used by public society, so that PLN must fit public electricity demands from time to time. PLN works by predicting electricity power which is consumed by customers hourly. The prediction made is based on prior electricity power use. The amount of electricity power use prediction is held to optimize electricity power used by customers, so that there will not be any electricity extravagancy or extinction. The prediction can be done by using some forecasting methods, such as double seasonal ARIMA model and Neural Network (NN) method. Some researches that are related to short-term electricity power forecasting can be seen in [3], [4], [6], [7], [8], [9], [11], [12], [13] and [14]. Neural network methods used in those researches are Feedforward Neural Network, which is known as AR-NN model. This model can not get and represent moving average order effect in time series. Some prior researches, in other countries or in Indonesia, show that ARIMA model for the electricity consumption data tends to include MA order (see [9] and [13]).
Short-term electricity consumption data has a double seasonal pattern, including daily seasonal and weekly seasonal. ARIMA model with multiplicative double seasonal pattern as a generalization of seasonal ARIMA model, written as ARIMA\((p, d, q)\)(1, 1, \(\Theta\))\((D, \Theta, Q)\), has a common form as
\[
\phi_p(B)\Phi_p(B^{12})\phi_q(B)\Phi_q(B^{12})\left(1-B\right)^d\left(1-B^{12}\right)^D Z_t = \theta_q(B)\Theta_q(B^{12})\theta_p(B)\Theta_p(B^{12})a_t
\]
where \(s_1\) and \(s_2\) are periods of difference seasonal.

One of methods that can be used to estimate ARIMA model parameter is Maximum Likelihood Estimation (MLE) method. The assumption needed in MLE method is that error \(a_t\) distributes normally [2]. Therefore, the cumulative distribution function is
\[
f\left(a_i\Big|\sigma_a^2\right) = \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{a_i^2}{2\sigma_a^2}\right)
\]
Because error is independent, the jointly distribution from \(a_1, a_2, \ldots, a_n\) is
\[
f\left(a_1, a_2, \ldots, a_n \Big|\sigma_a^2\right) = \left(\frac{1}{\sqrt{2\pi\sigma_a^2}}\right)^n \exp\left(-\frac{1}{2\sigma_a^2} \sum_{i=1}^{n} a_i^2\right)
\]
Error \(a_t\) can be stated in function \(Z_t\) and parameters \(\phi, \theta, \sigma_a^2\) and also the prior error. Generally \(a_t\) form is
\[
a_t = Z_t - \phi_1 Z_{t-1} - \ldots - \phi_p Z_{t-p} + \theta_q a_{t-q} + \ldots + \theta_q a_{t-q}.
\]
The likelihood function for model parameters if the observations are known is
\[
L(\phi, \theta, \sigma_a^2 |Z) = \left(\frac{1}{\sqrt{2\pi\sigma_a^2}}\right)^n \exp\left(-\frac{1}{2\sigma_a^2} S(\phi, \theta)\right)
\]
where
\[
S(\phi, \theta) = \sum_{t=1}^{n} \left(Z_t - \phi_1 Z_{t-1} - \ldots - \phi_p Z_{t-p} + \theta_q a_{t-q} + \ldots + \theta_q a_{t-q}\right)
\]
Then, the log-likelihood function is
\[
l(\phi, \theta, \sigma_a^2 |Z) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_a^2) - \frac{1}{2\sigma_a^2} S(\phi, \theta).
\]
The maximum of the log-likelihood function is computed by finding the first-order derivative of Equation (9) to each parameter and equalling with zero.
\[
\frac{\partial l(\phi, \theta, \sigma_a^2 |Z)}{\partial \phi} = 0; \quad \frac{\partial l(\phi, \theta, \sigma_a^2 |Z)}{\partial \theta} = 0; \quad \frac{\partial l(\phi, \theta, \sigma_a^2 |Z)}{\partial \sigma_a^2} = 0.
\]
Information matrix which is notated as \(l(\phi, \theta)\) is used to get the standard error of parameter estimated by MLE method [2]. This matrix is found by calculating the second-order derivative to each parameter, which is notated as \(l_{ij}\) where
\[
l_{ij} = \frac{\partial^2 l(\beta, \sigma_a^2 \mid Z)}{\partial \beta_i \partial \beta_j},
\]
and
\[
l(\beta) = -E(l_{ij}).
\]
The parameter variance is noted as \(V(\hat{\beta})\) and the parameter standard error is \(SE(\hat{\beta})\).
\[
V(\hat{\beta}) = [l(\beta)]^{-1}
\]
and
\[
SE(\hat{\beta}) = \sqrt{V(\hat{\beta})}.
\]

B. Neural Network

Generally Neural Network (NN) has some components, i.e. neuron, layer, activation function, and weight. NN modeling is seen from the network form which is including the amount of neuron in the input layer, the amount of neuron in the hidden layer, and the amount of neuron in the output layer, and also the activation function used. Feed-Forward Neural Network (FFNN) is the mostly used NN model for time series data forecasting [10]. FFNN model in statistics modeling for time series fore-casting can be seen as a non-linear autoregressive model. This form has a limitation, which can only sense and represent autoregressive (AR) effects in time series data.

One of NN form which is more flexible than FFNN is Recurrent Neural Network (RNN). RNN model is said to be flexible because the network output is set to be the input to get the next output [1]. RNN model is also called Autoregressive Moving Average-Neural Network (ARMA-NN), because besides some response or target lag as the inputs, it also includes lags of the difference between the target prediction and the actual value, which is known as the error lags [15]. Generally the RNN model architecture is same with ARMA\((p, q)\) model. The difference is that the time series function is non-linear in RNN model and linear in ARMA\((p, q)\) model. So that RNN model is said to be the non-linear autoregressive moving average.

The activation function used in hidden layer in this research is tangent sigmoid function, and the activation function in output layer is linear function. The form of tangent sigmoid function is
\[
f(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}.
\]
And linear function is \(f(x) = x\). The architecture of Elman-RNN, for example ARMA(2,1)-NN and 4 neuron units in hidden layer is shown in Figure1.
Elman-RNN(2,4,1) or ARMA(2,1)-NN is a nonlinear model. This network has 3 inputs, such as $Y_{t-1}$, $Y_{t-2}$, and residual $e_{t-1}$. Four hidden units in the hidden layer with activation function $\psi(\bullet)$ and one output in the output layer with linear function. The difference among all NN types is that in Elman-RNN, there is a feedback process, a process representing the output as the next input. Therefore, the advantage of using Elman-RNN is the fits are more accurate, especially for data having moving average order.

To solve the equation, we do the partial derivative of $E$ to each weight and bias $w$ with chain rules. The partial derivative of $E$ to the weight $j$ is

$$ \frac{\partial E}{\partial \beta_j} = -\sum_{k=1}^{n} [Y_{(k)} - \hat{Y}_{(k)}] f^{o'}(\beta_0 + \sum_{l=1}^{p} \beta_l V_{l(k)}) V_{j(k)} $$

Equation (18) is simplified into

$$ \frac{\partial E}{\partial \beta_j} = -\sum_{k=1}^{n} \delta^{o(k)} V_{j(k)} $$

where

$$ \delta^{o(k)} = [Y_{(k)} - \hat{Y}_{(k)}] f^{o'}(\beta_0 + \sum_{l=1}^{p} \beta_l V_{l(k)}) $$

By the same way, the partial derivatives of $E$ to $\beta_0$, $\gamma_j$, and $\gamma_{10}$ are done, so that

$$ \frac{\partial E}{\partial \beta_0} = -\sum_{k=1}^{n} \delta^{o(k)} $$

$$ \frac{\partial E}{\partial \gamma_j} = -\sum_{k=1}^{n} [Y_{(k)} - \hat{Y}_{(k)}] f^{o'}(\beta_0 + \sum_{l=1}^{p} \beta_l V_{l(k)}) \times \beta_j f^{k'}(\gamma_{10} + \sum_{i=1}^{q} \gamma_i X_{i(k)}) X_{k(k)} $$

$$ \frac{\partial E}{\partial \gamma_{10}} = -\sum_{k=1}^{n} \delta^{h(k)} X_{i(k)} $$

Minimization of Equation (16) is done with Gradient Descent method with momentum. Gradient Descent method with momentum $m$, $0 < m < 1$, is formulated as

$$ w^{(i+1)} = w^{(i)} - \eta \left( m \cdot dw^{(i)} + (1-m) \frac{\partial E}{\partial w} \right) $$

where $dw$ is the change of the weight or bias, $\eta$ is the learning rate which is defined, $0 < \eta < 1$.  

Figure 1. Elman-RNN(2,4,1) or ARMA(2,1)-NN architecture
Based on the derivatives the weight and the bias can be estimated with Gradient Descent method with momentum. The weight and the bias updating in the output layer are

\[ \beta_j^{(s+1)} = \beta_j^{(s)} - \left( m \cdot dw_j^{(s)} + (m - 1) \eta \sum_{k=1}^{\infty} \delta_k^{(s)} V_{j(k)} \right) \]  \hspace{1cm} (25)

and

\[ \beta_0^{(s+1)} = \beta_0^{(s)} - \left( m \cdot dw_0^{(s)} + (m - 1) \eta \sum_{k=1}^{\infty} \delta_k^{(s)} \right). \]  \hspace{1cm} (26)

The weight and the bias updating in the hidden layer are

\[ \gamma_{ji}^{(s+1)} = \gamma_{ji}^{(s)} - \left( m \cdot dw_{ji}^{(s)} + (m - 1) \eta \sum_{k=1}^{\infty} \delta_k^{(s)} X_{i(k)} \right) \]  \hspace{1cm} (27)

and

\[ \gamma_{ji}^{(s+1)} = \gamma_{ji}^{(s)} - \left( m \cdot dw_{ji}^{(s)} + (m - 1) \eta \sum_{k=1}^{\infty} \delta_k^{(s)} \right). \]  \hspace{1cm} (28)

dw in Equation (25) to (28) is the change of the related weight or bias, m is the momentum, and \( \eta \) is the learning rate.

### III. METHODOLOGY

The data used in this research is electricity consumption data, which is a secondary data from PLN Gresik. The data taken as the case study is hourly electricity consumption data in Mengare Gresik, which is recorded from 1 August to 23 September 2007. In-sample data is taken from 1 August to 15 September 2007, and the out-sample data is 16-23 September 2007. The variable in this research is hourly electricity consumption. The steps of the analysis are as follows:

1. Modeling of double seasonal ARIMA.
2. Modeling of Elman-RNN with 4 kinds of input, i.e.:
   a. Input based on double seasonal ARIMA model.
   b. Input based on double seasonal ARIMA model and outlier dummies.
   c. Input multiplication of 24 lag up to lag 480.
   d. Input lag 1 and multiplication of 24 lag ± 1.
3. Forecast the out-sample data.
4. Compare the forecast accuracy between Elman-RNN model and double seasonal ARIMA model.
5. Forecast the electricity consumption for 24-30 September 2007 by using the best model.

### IV. EMPIRICAL RESULTS

The result of the hourly electricity consumption descriptive in Mengare Gresik from 1 August to 23 September 2007 shows that highest electricity consumption is at 19.00 about 3537 kW, and the least one is at 07.00 about 1665.2 kW. It is presumed that at 07.00 customers turn off the lamps off, get ready for work, and leave for the office. Customer work hours begin at 09.00 and end at 17.00, so that household electricity consumption is less or beyond the overall electricity consumption average. At 18.00 customers turn the night lamps on and at 19.00 customers has been back from work, and do many kinds of activities at house, that use a large amount of electricity such as electronics use.

Descriptive of the daily electricity consumption can be seen in Table 1. Based on the result in Table 1 we know that on Tuesday the electricity consumption is the largest, about 2469.6 kW, and the least electricity consumption is on Sunday, about 2204.8 kW. The electricity consumption averages on Saturday and Monday are beyond the overall average because those days are week-end days, so that customers tend to spend their week-end days with their family outside the house.

### Table 1. Descriptive Statistics of the Daily Electricity Consumption

<table>
<thead>
<tr>
<th>Day</th>
<th>Observation</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>168</td>
<td>2439</td>
<td>624.1</td>
</tr>
<tr>
<td>Tuesday</td>
<td>168</td>
<td>2469</td>
<td>608.2</td>
</tr>
<tr>
<td>Wednesday</td>
<td>192</td>
<td>2453</td>
<td>584.8</td>
</tr>
<tr>
<td>Thursday</td>
<td>192</td>
<td>2447</td>
<td>603.9</td>
</tr>
<tr>
<td>Friday</td>
<td>192</td>
<td>2427</td>
<td>645.1</td>
</tr>
<tr>
<td>Saturday</td>
<td>192</td>
<td>2362</td>
<td>632.4</td>
</tr>
<tr>
<td>Sunday</td>
<td>192</td>
<td>2204</td>
<td>660.3</td>
</tr>
</tbody>
</table>

#### A. Result of Double Seasonal ARIMA Model

ARIMA model building process is based on Box-Jenkins procedure [2], starting with model order identification from the stationer data. Figure 2 shows a non-stationer hourly electricity consumption data pattern, especially in the daily and weekly periods. Data stationery is found by differencing lag 1, 24, and 168.

![Time series plot of hourly electricity consumption](image)

Figure 2. Time series plot of hourly electricity consumption
Figure 3. ACF and PACF for electricity data

Figure 4. ACF and PACF for weekly seasonal differencing lag 24

Figure 5. ACF and PACF for weekly seasonal and regular differencing lag 24 and 1

Figure 6. ACF and PACF for data differencing lag 1, 24, and 168
Figure 6 shows the ACF and PACF plot of stationery data, which are the data that has been differenced by lag 1, 24, and 168. Based on ACF and PACF plots of stationery data, predicted double seasonal ARIMA models are two, i.e. ARIMA([1,2,3,4,6,7,9,10,14,21,33],1,[8])(0,1,1)(1,1,0)\textsuperscript{168} and ([12,1],[1,2,3,4,6,7])(0,1,1)(1,1,0)\textsuperscript{68} Parameters significance test and diagnostic check for both model with Ljung-Box test show that the residuals are white noise. Normality test of the residual with Kolmogorov-Smirnov test shows that the residuals for both models do not satisfy Normality test with Kolmogorov-Smirnov test. Outlier detection is done iteratively and we get 14 detected outliers. These inputs are ARIMA lag input like ARIMA lag input like lag 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 15, 21, 22, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 38, 39, 45, 46, 57, 58, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 182, 183, 189, 190, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 206, 207, 213, 214, 225, 226, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 350, 351, 357, 358, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 374, 375, 381, 382, 393, 394. With this input, the network made is Elman-RNN(101,3,1) with MAPE 4.22%.

Table 2. Elman-RNN Selection Criteria

<table>
<thead>
<tr>
<th>Network</th>
<th>AIC</th>
<th>SBC</th>
<th>MSE</th>
<th>MAPE</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNN(101,3,1)</td>
<td>11,061</td>
<td>12,054</td>
<td>9778.1</td>
<td>4.2167</td>
<td>17937.0</td>
</tr>
<tr>
<td>RNN(115,3,1)</td>
<td>10,810</td>
<td>12,073</td>
<td>6755.1</td>
<td>4.6108</td>
<td>21308.0</td>
</tr>
<tr>
<td>RNN(20,6,1)</td>
<td>11,468</td>
<td>11,413</td>
<td>22955.0</td>
<td>7.5536</td>
<td>44939.0</td>
</tr>
<tr>
<td>RNN(22,3,1)</td>
<td>10,228</td>
<td>9,6064</td>
<td>8710.7</td>
<td>2.7833</td>
<td>6943.2</td>
</tr>
</tbody>
</table>

The forecast accuracy comparison between Elman-RNN models can be seen in Table 2. Based on the out-sample MAPE comparison, it can be concluded that Elman-RNN(22,3,1) is the best Elman-RNN for hourly electricity consumption forecasting in Mengare Gresik.

C. Comparison between Double Seasonal ARIMA and Elman-RNN

ARIMA model compared is ARIMA model without outliers because the software (SAS package) can not model the outliers, which are many in the double seasonal ARIMA model. The best ARIMA model for hourly electricity consumption data forecasting in Mengare is ARIMA([1,2,3,4,6,7,9,10,14,21,33],1,[8])(0,1,1)(1,1,0)\textsuperscript{168} and the best NN is Elman-RNN(22,3,1). The comparison is also done with Elman-RNN(101,3,1), which is the network that has the same input as the double seasonal ARIMA model.

The comparison of forecast and forecast residual graphically for the out-sample data can be seen in Figure 7. Based on these results, we can conclude that the residual of Elman-
RNN is near with zero, compared with the residual of ARIMA model. Besides that, the forecast done with Elman-RNN is more accurate than the forecast done with ARIMA model.

The comparison process is also done for iterative out-sample MAPE. The comparison is resulted in Figure 8. From this figure, we can see that Elman-RNN(22,3,1) gives less forecast error than double seasonal ARIMA model and another Elman-RNN. Overall, the forecast accuracy comparison shows that Elman-RNN is a better model than double seasonal ARIMA model, for forecasting the electricity consumption in Mengare Gresik.

Figure 8. The comparison of out-sample MAPE of ARIMA model, Elman-RNN(101,3,1), and Elman-RNN(22,3,1)

Based on the results of data analysis in the previous section, we conclude that:

a. The appropriate ARIMA model for hourly short-term electricity consumption forecasting in Mengare Gresik is ARIMA([1-4,6,7,9,10,14,21,33],1,8)(0,1,1)24(1,1,0)168 with in-sample MSE 11417.426. The MAPE at out-sample data is 22.8%.

b. The best Elman-RNN to forecast hourly short-term electricity consumption in Mengare Gresik is Elman-RNN(22,3,1) with inputs lag 1, 23, 24, 25, 47, 48, 49, 71, 72, 73, 95, 96, 97, 119, 120, 121, 143, 144, 145, 167, 168, and 169. The activation function used in the hidden layer is tangent sigmoid function and in the output layer is linear function. This network gives MAPE 3% at out-sample data.

c. The comparison of model forecast accuracy shows that Elman-RNN method, i.e. Elman-RNN(22,3,1), is the best model to forecast hourly electricity consumption in Mengare Gresik.

The result of this research also shows that there is a restriction of SAS package in estimating double seasonal ARIMA model parameter with adding outlier effect from the outlier detection process. This condition gives opportunity to do a further research related to statistic package improvement, especially for double seasonal ARIMA model involving long lags and the outlier detection.

REFERENCES


