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 $([1,2,3,4,6,7,9,10,14,21,33],1,8)(0,1,1)^{24}(1,1,0)^{168}$. 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 less MAPE than ARIMA model. Thus, Elman-RNN(22,3,1) is the best method for forecasting hourly electricity load demands in Mengare, Gresik, Indonesia.

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]).

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.

Manuscript received August 18, 2009.

Suhartono is with the Computational Statistics Laboratory, Institut Teknologi Sepuluh Nopember, Surabaya, 60111 Indonesia (phone: 62-81-8376367; e-mail: suhartono@statistika.its.ac.id).

Alfonsus Julanto Endharta is with the Institut Teknologi Sepuluh Nopember, Surabaya, 60111 Indonesia (phone: 62-83-857670077; e-mail: alfonsus_je@yahoo.com).

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-stationery data pattern and d differencing order. The form of ARIMA(p,d,q) is

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t \qquad (1)$$

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

$$\phi_p(B) = (1-\phi_1 B - \phi_2 B^2 - ... - \phi_p B^p),$$

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

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

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)a_t \qquad (2)$$

where s is seasonal period, and

$$\Phi_P(B^s) = (1-\Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_P B^{Ps}),$$

$$\Theta_Q(B^s)=(1-\Theta_1 B^s-\Theta_2 B^{2s}-...-\Theta_Q B^{Qs}).$$

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)(P_1,D_1,Q_1)^{s1}(P_2,D_2,Q_2)^{s2}, has a common form as

$$\phi_p(B)\Phi_{P_1}(B^{s_1})\Phi_{P_2}(B^{s_2})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}Z_t$$

$$=\theta_q(B)\Theta_{Q_1}(B^{s_1})\Theta_{Q_2}(B^{s_2})a_t \qquad (3)$$

where s_1 and s_2 are periods of difference seasonal.

One of method 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

$$f\left(a_t\middle|\sigma_a^2\right)=\left(2\pi\sigma_a^2\right)^{-1/2}\exp\left(-\frac{a_t^2}{2\sigma_a^2}\right) \qquad (4)$$

Because error is independent, the jointly distribution from $a_1,a_2,...,a_n$ is

$$f\left(a_1,a_2,...,a_n\middle|\sigma_a^2\right)=\left(2\pi\sigma_a^2\right)^{-n/2}\exp\left(-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n}a_t^2\right). \qquad (5)$$

Error a_t can be stated in function Z_t, and parameters ϕ,θ,σ_a^2 and also the prior error. Generally a_t form is

$$a_t=Z_t-\phi_1 Z_{t-1}-...-\phi_p Z_{t-p}+$$

$$\theta_1 a_{t-1}+...+\theta_q a_{t-q}. \qquad (6)$$

The likelihood function for model parameters if the observations are known is

$$L\left(\phi,\theta,\sigma_a^2\middle|Z\right)=\left(2\pi\sigma_a^2\right)^{-n/2}\exp\left(-\frac{1}{2\sigma_a^2}S(\phi,\theta)\right) \qquad (7)$$

where

$$S(\phi,\theta)=\sum_{t=1}^{n}\left(Z_t-\phi_1 Z_{t-1}-...-\phi_p Z_{t-p}+\theta_1 a_{t-1}+...+\theta_q a_{t-q}\right) \qquad (8)$$

Then, the log-likelihood function is

$$l\left(\phi,\theta,\sigma_a^2\middle|Z\right)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\left(\sigma_a^2\right)-\frac{1}{2\sigma_a^2}S(\phi,\theta). \qquad (9)$$

The maximum of the log-likelihood function is computed by finding the first-order derivative of Equation (9) to each parameter and equaling with zero.

$$\frac{\partial l\left(\phi,\theta,\sigma_a^2\middle|Z\right)}{\partial\phi}=0;\ \frac{\partial l\left(\phi,\theta,\sigma_a^2\middle|Z\right)}{\partial\theta}=0;\ \frac{\partial l\left(\phi,\theta,\sigma_a^2\middle|Z\right)}{\partial\sigma_a^2}=0.$$

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 calcu-lating the second-order derivative to each parameter, which is notated as l_{ij} where

$$l_{ij}=\frac{\partial^2 l\left(\beta,\sigma_a^2\middle|Z\right)}{\partial\beta_i\partial\beta_j}, \qquad (10)$$

and

$$l(\beta)=-E(l_{ij}). \qquad (11)$$

The parameter variance is notated as $V(\hat{\beta})$ and the parameter standard error is $SE(\hat{\beta})$.

$$V(\hat{\beta})=\left[l(\beta)\right]^{-1} \qquad (12)$$

and

$$SE(\hat{\beta})=\left[V(\hat{\beta})\right]^{1/2}. \qquad (13)$$

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 rese-arch is tangent sigmoid function, and the activation function in output layer is linear function. The form of tangent sig-moid function is

$$f(x)=\frac{1-e^{-2x}}{1+e^{-2x}} \qquad (14)$$

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.

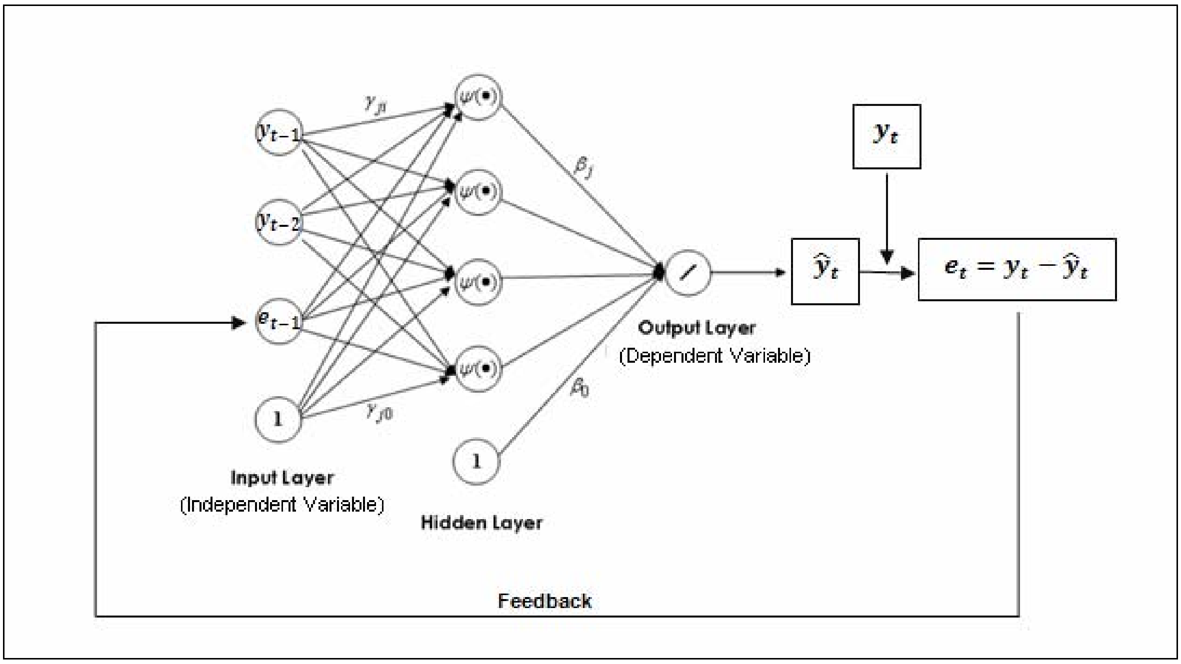


Figure 1. Elman-RNN(2,4,1) or ARMA(2,1)-NN architecture

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.

The weight and the bias in the Elman-RNN model are estimated with backpropagation algorithm. The general RNN with one hidden layer, q input units and p units in the hidden layer is

$$Y = f^o\left[\beta_0 + \sum_{j=1}^{p}\left(\beta_j f^h\left(\gamma_{j0} + \sum_{i=1}^{q}\gamma_{ji}X_i\right)\right)\right] \tag{15}$$

where β_j is the weight of the j^{th} unit in the hidden layer, γ_{ji} is the weight from i^{th} input to j^{th} unit in the hidden layer, $f^h(x)$ is the activation function in the hidden layer, and $f^o(x)$ is the function in the output layer. Chong and Zak (1996) explain that to get the weight and bias we do the estimation by minimize value E in the following equation.

$$E = \frac{1}{2}\sum_{k=1}^{n}\left[Y_{(k)} - \hat{Y}_{(k)}\right]^2 . \tag{16}$$

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

$$w^{(t+1)} = w^{(t)} - \left(m \cdot dw^{(t)} + (1-m)\eta\frac{\partial E}{\partial w}\right) \tag{17}$$

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

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 β_j is

$$\frac{\partial E}{\partial \beta_j} = -\sum_{k=1}^{n}\left[Y_{(k)} - \hat{Y}_{(k)}\right]f^{o'}\left(\beta_0 + \sum_{l=1}^{p}\beta_l V_{l(k)}\right)V_{j(k)} \tag{18}$$

Equation (18) is simplified into

$$\frac{\partial E}{\partial \beta_j} = -\sum_{k=1}^{n}\delta^o{}_{(k)}V_{j(k)} \tag{19}$$

where

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

By the same way, the partial derivatives of E to β_0, γ_{li}, and γ_{l0} are done, so that

$$\frac{\partial E}{\partial \beta_0} = -\sum_{k=1}^{n}\delta^o{}_{(k)} , \tag{20}$$

$$\frac{\partial E}{\partial \gamma_{ji}} = -\sum_{k=1}^{n}\left[Y_{(k)} - \hat{Y}_{(k)}\right]f^{o'}\left(\beta_0 + \sum_{l=1}^{p}\beta_l V_{l(k)}\right)\times \beta_j f^{h'}\left(\gamma_{l0} + \sum_{i=1}^{q}\gamma_{li}X_{i(k)}\right)X_{l(k)} \tag{21}$$

or

$$\frac{\partial E}{\partial \gamma_{ji}} = -\sum_{k=1}^{n}\delta^h{}_{(k)}X_{i(k)} , \tag{22}$$

and

$$\frac{\partial E}{\partial \gamma_{j0}} = -\sum_{k=1}^{n}\delta^h{}_{(k)} , \tag{23}$$

where

$$\delta^h{}_{(k)} = \delta^o{}_{(k)}\beta_j f^{h'}\left(\gamma_{l0} + \sum_{i=1}^{q}\gamma_{li}X_{i(k)}\right). \tag{24}$$

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

$$\beta_j^{(s+1)} = \beta_j^{(s)} - \left(m \cdot dw^{(s)} + (m-1)\eta \sum_{k=1}^{n} \delta^{o}{}_{(k)} V_{j(k)} \right) \quad (25)$$

and

$$\beta_0^{(s+1)} = \beta_0^{(s)} - \left(m \cdot dw^{(s)} + (m-1)\eta \sum_{k=1}^{n} \delta^{o}{}_{(k)} \right). \quad (26)$$

The weight and the bias updating in the hidden layer are

$$\gamma_{ji}^{(s+1)} = \gamma_{ji}^{(s)} - \left(m \cdot dw^{(s)} + (m-1)\eta \sum_{k=1}^{n} \delta^{h}{}_{(k)} X_{i(k)} \right) \quad (27)$$

and

$$\gamma_{j0}^{(s+1)} = \gamma_{j0}^{(s)} - \left(m \cdot dw^{(s)} + (m-1)\eta \sum_{k=1}^{n} \delta^{h}{}_{(k)} \right). \quad (28)$$

dw in Equation (25) to (28) is the change of the related weight or bias, m is the momentum, and η 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 follow:

i. Modeling of double seasonal ARIMA.
ii. 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.
iii. Forecast the out-sample data.
iv. Compare the forecast accuracy between Elman-RNN model and double seasonal ARIMA model.
v. Forecast the electricity consumption for 24-30 September 2007 by using the best model.

IV. EMPIRICAL RESULTS

The result of the hourly electricity consumption descrip-tive 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 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

Day	Observation	Mean	Standard Deviation
Monday	168	2439,0	624,1
Tuesday	168	2469,5	608,2
Wednesday	192	2453,3	584,8
Thursday	192	2447,9	603,9
Friday	192	2427,3	645,1
Saturday	192	2362,7	632,4
Sunday	192	2204,8	660,3

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.

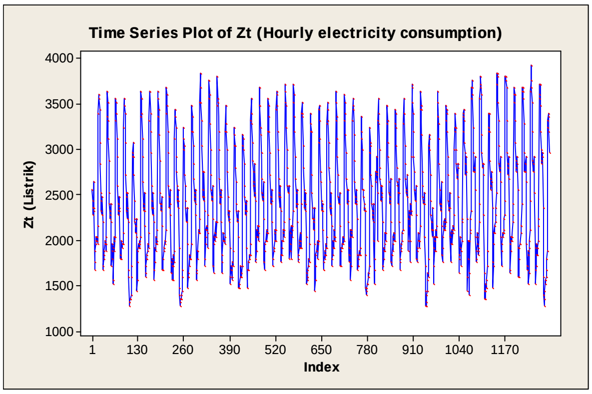


Figure 2. Time series plot of hourly electricity consumption

Figure 3 shows the ACF and PACF plots of the real data. It shows the nonstationerity from the slowly dying down weekly seasonal lags in ACF plot. Hence, daily seasonal differencing (24 lags) should be applied.

After daily seasonal differencing, ACF and PACF plots can be shown in Figure 4. ACF plot shows that regular lags dies down very slowly; hence, it needs regular order differencing.

Daily seasonal and regular order differencing data have ACF and PACF plots in Figure 5. The ACF plot shows that lags 168 and 336 are significant. It is considered that in ACF plot weekly seasonal lags die down very slowly. Therefore, it is necessary to apply weekly seasonal order differencing (168 lags).

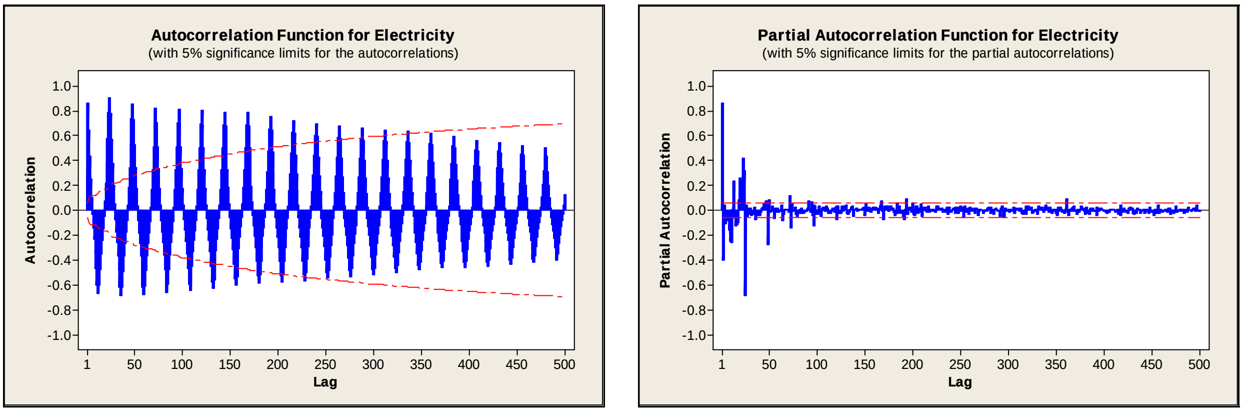


Figure 3. ACF and PACF for electricity data

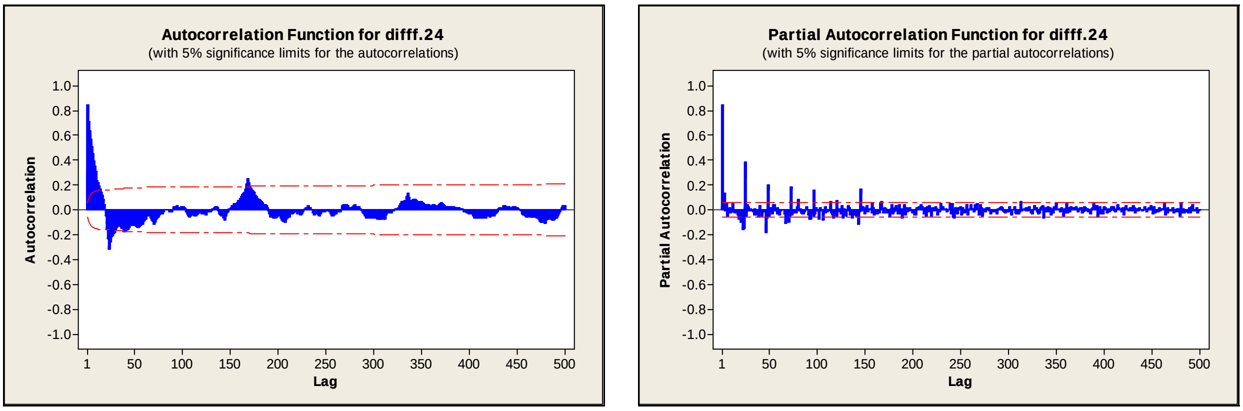


Figure 4. ACF and PACF for weekly seasonal differencing lag 24

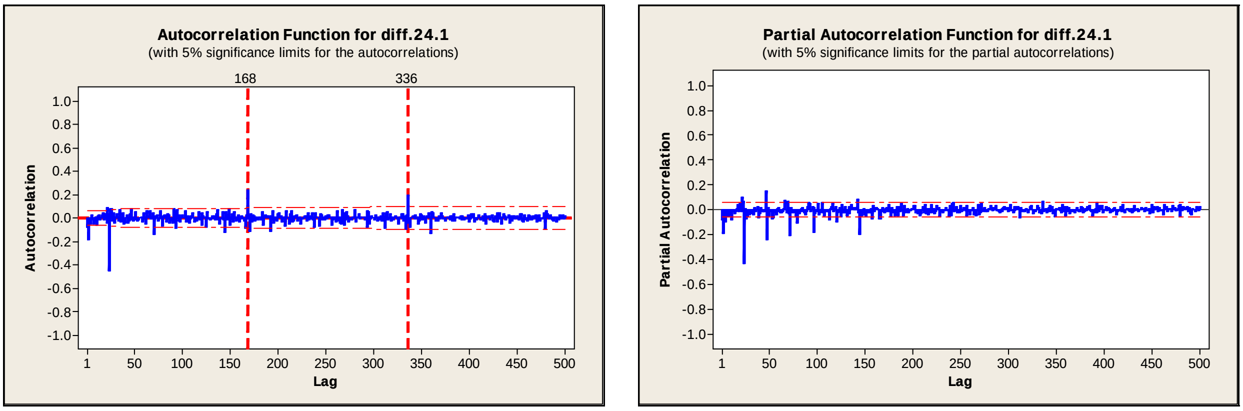


Figure 5. ACF and PACF for weekly seasonal and regular differencing lag 24 and 1

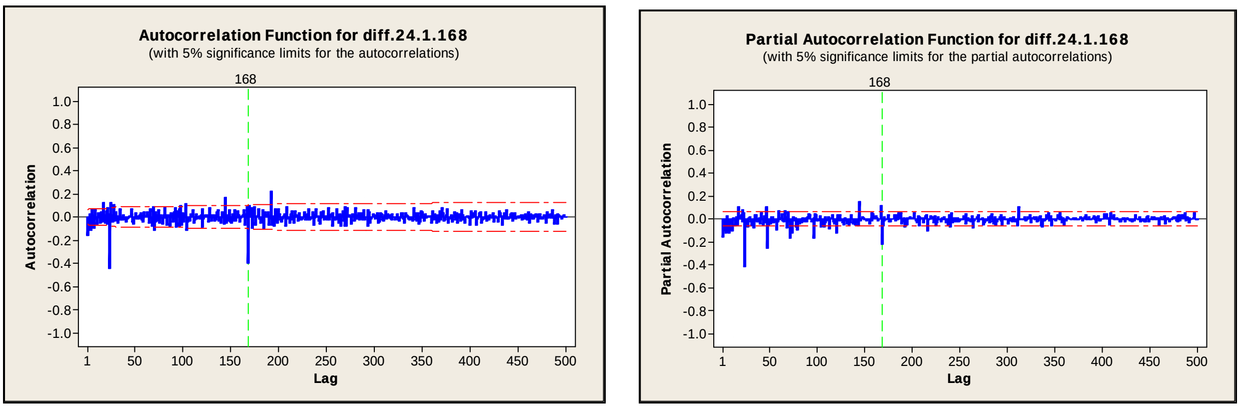


Figure 6. ACF and PACF for data differencing lag 1, 24, and 168

Figure 6 shows the ACF and PACF plot of stationer data, which are the data that has been differenced by lag 1, 24, and 168. Based on ACF and PACF plots of stationer data, predicted double seasonal ARIMA models are two, i.e. ARIMA([1,2,3,4,6,7,9,10,14,21,33],1,[8])$(0,1,1)^{24}(1,1,0)^{168}$ and ([12],1,[1,2,3,4,6,7]) $(0,1,1)^{24}$ $(1,1,0)^{168}$. 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 normal distribution. It is presumed that there are outliers in the data and could be seen completely in [5].

Outlier detection process is only done in model I, because MSE of model I at in-sample data is less than MSE of model II. Outlier detection is done iteratively and we get 14 innovational outliers. Model I has out-sample MAPE about 22.8% and mathematically the model is written as

$$\begin{aligned}&(1+0.164B+0.139B^2+0.155B^3+0.088B^4+\\&0.112B^6+0.152B^7+0.077B^9+0.067B^{10}+\\&0.069B^{14}+0.089B^{21}+0.072B^{22})(1+0.543B^{168})\\&(1-B)(1-B^{24})(1-B^{168})Z_t=(1-0.0674B^8)\\&(1-0.803B^{24})a_t.\end{aligned}$$

Thus, model I with the outliers which is not re-estimated is

$$\begin{aligned}&Z_t=\tfrac{1}{\hat{\pi}(B)}[844I_t^{(830)}-710.886I_t^{(1062)}+621.307I_t^{(906)}+\\&-511.067I_t^{(810)}-485.238I_t^{(1027)}-456.19I_t^{(1038)}+\\&455.09I_t^{(274)}-438.882I_t^{(247)}+376.704I_t^{(1075)}-\\&375.48I_t^{(971)}+362.052I_t^{(594)}-355.701I_t^{(907)}-\\&329.702I_t^{(623)}+308.13I_t^{(931)}+a_t],\end{aligned}$$

where

$$\begin{aligned}&\hat{\pi}(B)=\\&(1+0.164B+0.139B^2+0.155B^3+0.088B^4+\\&0.112B^6+0.152B^7+0.077B^9+0.067B^{10}+\\&0.069B^{14}+0.089B^{21}+0.072B^{22})(1+0.543B^{168})\\&(1-B)(1-B^{24})(1-B^{168})]/[(1-0.0674B^8)\\&(1-0.803B^{24})].\end{aligned}$$

B. *Result of Elman-Recurrent Neural Network*

Elman-RNN method application is done to get the best suitable network for electricity consumption forecast in Mengare Gresik. Defined network elements are the amount of inputs, the amount of hidden units, the amount of outputs, and the activation function in both hidden layer and output layer. The hidden layer used is only one, the activation function in the hidden layer is tangent sigmoid function, and in the output layer is linear function.

The first tried Elman-RNN is a network with inputs same with double seasonal ARIMA model lag. This network use input 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, dan 394. With this input, the network made is Elman-RNN(101,3,1) with MAPE 4.22%.

In the second network, we use ARIMA input and adding 14 detected outliers. These inputs are ARIMA lag input like in the first network and 14 outliers, i.e. in 803^{th}, 1062^{th}, 906^{th}, 810^{th}, 1027^{th}, 1038^{th}, 274^{th}, 247^{th}, 1075^{th}, 971^{th}, 594^{th}, 907^{th}, 623^{th}, and 931^{th} time t. In this scenario, we get Elman-RNN(115,3,1) with MAPE 4.61%. The third network is network with multiplication of 24 lag input. The inputs are lag 24, 48, …, 480. With this network we get Elman-RNN(20,6,1) with MAPE 7.55%. And the last network is lag 1 input and multiplication of 24 lag ± 1. This network input are lag 1, 23, 24, 25, 47, 48, 49, ..., 167, 168, and 169. The best network with this inputs is Elman-RNN(22,3,1) with MAPE 2.78%.

Table 2. Elman-RNN Selection Criteria

Network	In-Sample Criteria			Out-Sample Criteria	
	AIC	SBC	MSE	MAPE	MSE
RNN(101,3,1)	11,061	12,054	9778.1	4.2167	17937.0
RNN(115,3,1)	10,810	12,073	6755.1	4.6108	21308.0
RNN(20,6,1)	11,468	11,413	22955.0	7.5536	44939.0
RNN(22,3,1)	10,228	9,6064	8710.7	2.7833	6943.2

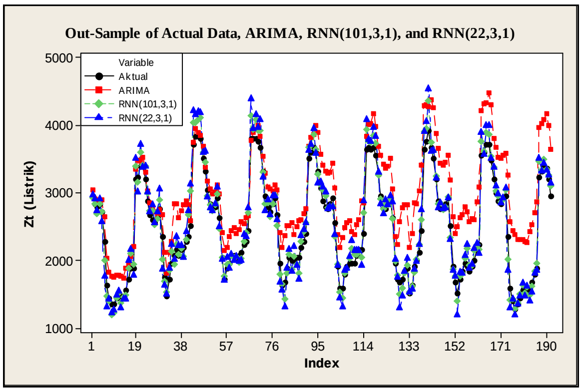
The forecast accuracy comparison between Elman-RNN models can be seen in Table 2. Based on the out-sample MAPE comparison, it can be concluded that Elman-RNN(22,3,1) is the best Elman-RNN for hourly electricity consumption forecasting in Mengare Gresik.

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)$(0,1,1)^{24}(1,1,0)^{168}$ 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.



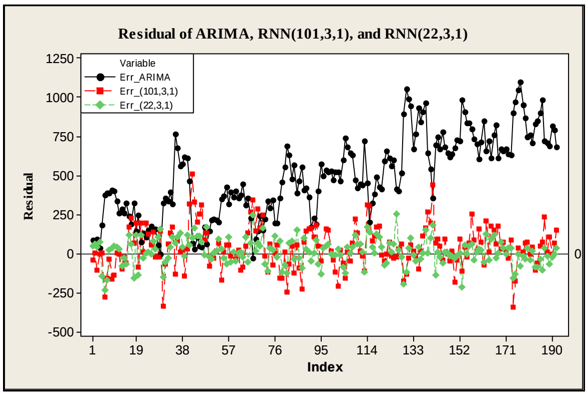


Figure 7. The comparison of out-sample forecast accuracy and forecast residuals between ARIMA model, Elman- RNN(101,3,1), and Elman-RNN(22,3,1)

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.

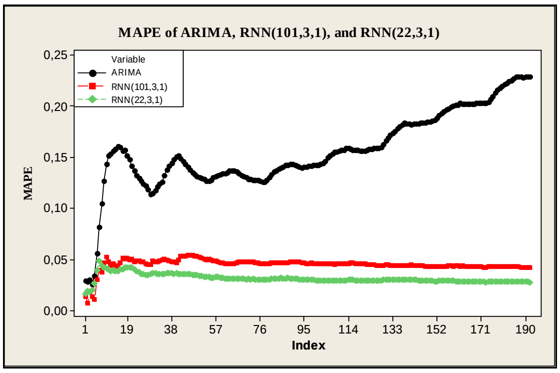


Figure 8. The comparison of out-sample MAPE of ARIMA model, Elman-RNN(101,3,1), and Elman-RNN(22,3,1)

V. CONCLUSION

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

[1] Beale, R. and Finlay, J. 1992. *Neural Networks and Pattern Recognition in Human-Computer Interaction*. England: Ellis Horwood.

[2] Box, G.E.P., Jenkins, G.M., and Reissel. G.C. 1994. *Time Series Analysis Forecasting and Control, 3rd edition*. Prentice Hall.

[3] Chen, J.F., Wang, W.M., and Huang, C.M. 1995. Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting, *Electric Power Systems Research*, **34**,187-196.

[4] Chong, E.K.P. and Zak, S.H. 1996. *An Introduction to Optimization*. John Wiley & Sons, Inc.

[5] Endharta, A.J. 2009. *Forecasting of Short Term Electricity Consumption by using Elman-Recurrent Neural Network*. Bachelor Thesis at Department of Statistics, Institut Teknologi Sepuluh Nopember Surabaya (unpublished).

[6] Husen, W. 2001. *Forecasting of Maximum Short Term Electricity Usage by implementing Neural Network*. Bachelor Thesis at Department of Physics Engineering,

Institut Teknologi Sepuluh Nopember Surabaya (unpublished).

[7] Kalaitzakis, K., Stavrakakis, G.S., and Anagnostakis, E.M. 2002. Short-term load forecasting based on artificial neural networks parallel implementation. *Electric Power Systems Research*, **63**, 185-196.

[8] Kiartzis S.J., Bakirtzis, A.G., and Petridis, V. 1995. Short-term loading forecasting using neural networks. *Electric Power Systems Research,* **33**, 1-6.

[9] Ristiana, Y. 2008. *Autoregressive Neural Network Model* (*ARNN*) *for Forecasting Short Term Electricity Consumption at PT. PLN Gresik*. Bachelor Thesis at Department of Statistics, Institut Teknologi Sepuluh Nopember Surabaya (unpublished).

[10] Suhartono. 2007. *Feedforward Neural Networks for Time Series Forecasting*. Unpublished PhD Dissertation, Department of Mathematics, Gadjah Mada University, Yogyakarta.

[11] Taylor, J.W. 2003. Short-term electricity demand forecasting using double seasonal exponential smoothing, *Journal of the Operational Research Society*, **54**, 799–805.

[12] Tamimi, M. and Egbert, R. 2000. Short term electric load forecasting via fuzzy neural collaboration. *Electric Power Systems Research*, **56**, 243-248.

[13] Taylor, J.W., Menezes, L.M., and McSharry, P.E. 2006. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, **22**, 1-16.

[14] Topalli, A.K. and Erkmen, I. 2003. A hybrid learning for neural networks applied to short term load forecasting, *Neurocomputing*, **51**, 495 – 500.

[15] Trapletti, A. 2000. *On Neural Networks as Statistical Time Series Models*. Unpublished PhD Dissertation, Institute for Statistics, Wien University.

[16] Wei, W.W.S. 2006. *Time Series Analysis: Univariate and Multvariate Methods*. 2nd Edition, Pearson Addison Wesley, Boston.