Short Term Load Forecasting Using Double Seasonal ARIMA Model

Article *in* World Applied Sciences Journal · July 2010

4 authors, including:

Maizah Ahmad
Universiti Teknologi Malaysia

Zuhaimy Ismail
Universiti Teknologi Malaysia

Suhartono Suhartono
Institut Teknologi Sepuluh Nopember

Short Term Load Forecasting Using Double Seasonal ARIMA Model

Norizan Mohamed[1]
[1] *Mathematics Department, Faculty of Science And Technology*
Universiti Malaysia Terengganu (UMT), 21030 Kuala Terengganu, Terengganu, MALAYSIA
E-mail: norizan@umt.edu.my,

Maizah Hura Ahmad[2]
Zuhaimy Ismail [3]
[2,3]*Department of Mathematics, Faculty of Science*
Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, MALAYSIA
E-mail: maizah@.utm.my

Suhartono[4]
[4]*Department of Statistics*
Institut Teknologi Sepuluh Nopember, Indonesia.
E-mail: suhartono@statistika.its.ac.id

ABSTRACT

Load demand is a time series data and it is one of the major input factors in economic development especially in a developing country such as Malaysia. Forecasting load demand with high accuracy is hoped to help the country, especially the Malaysian electricity utility company to generate an appropriate load of required power supply which can avoid energy wasting and prevent system failure. A half hourly load demand of Malaysia for one year, from September 01, 2005 to August 31, 2006 measured in Megawatt (MW) is used for this study with the mean absolute percentage error (MAPE) as a forecasting accuracy. We use Statistical Analysis System, SAS package to analyze the data. Using the least squares method to estimate the coefficients in a double SARIMA model, followed by model validation and model selection criteria, we propose the ARIMA(0,1,1)(0,1,1)^{48}(0,1,1)^{336} with in-sample MAPE of 0.9906% as the best model for this study. Comparing the forecasting performances by using k-step ahead forecasts and one-step ahead forecasts, we found that the MAPE for the one-step ahead out-sample forecasts from any horizon ranging from one week lead time to one month one week lead time are all less than 1%. Therefore we propose that a double seasonal ARIMA model with one-step ahead forecast must be considered in forecasting time series data with two seasonal cycles, especially in Malaysia load data.

Keywords: *Load forecasting, double seasonal ARIMA model, k-step ahead forecasts and one-step ahead forecasts*

Introduction

Load demand prediction is important for electric power planning and must be assessed with the greatest precision of any model. The utility power company needs forecasts for different time horizons in order to ensure the uninterrupted energy supply of customer (Tsekouras et al., 2007). Load forecasting can be broadly classified into four main categories which are long term forecasts, intermediate term forecasts, short term forecast and very short term forecast. Long term forecasts are used for system planning, scheduling construction of new generation capacity and purchasing of generating units (Jia et al., 2001). Intermediate term forecasts also called medium term forecast are used for maintenance scheduling, coordination of load dispatching and setting of prices, so that demand can be met with fixed capacity (Jia

ISBN 978-967-363-157-5

et al., 2001). Short-term forecasts are used for optimal generator unit commitment, fuel allocation, maintenance scheduling and buying and selling of power, economic scheduling of generating capacity, scheduling of fuel purchases, security analysis and short-term maintenance scheduling (Jia et al., 2001; Pedregal et al., 2010). Very short term forecast are used for security assessment and economic dispatching, real time control and real time security evaluation (Jia et al., 2001).

The four main categories of time horizons have been studied extensively. Long term forecasts are investigated by Jia et al. (2001), Kermanshashi and Iwamiya (2002), Poa (2007), Al-Saba and El-Amin (1999), Dong and Pedrycz (2008), Al Rashidi and El- Naggar (2010) and Carpinteiro et al. (2007); intermediate term forecasts by Amjady and Keynia (2008), Elkateb et al. (1998), Ghi assi et al. (2006), Tsekouras et al. (2007), Pedregal and Trapero (2010) and Mirasgedis et al. (2006); short-term forecasts by Beccali et al. (2004), Beccali et al. (2007), Cancelo et al. (2008), Gonzalez and Zamarreno (2005), Hobbs et al. (1998), Al-Hamadi and Soliman (2004), Soares and Medeiros (2008), Topalli and Erkmen (2003), Zhang and Dong (2001), Darbellay and Slama (2000), El-Telbany and El-Karmi (2007), Kandil et al. (2006), Xiao et al. (2009), Satish et al. (2004), Srinivasan (1998) and Catalao et al. (2007); and very short term forecasts by Taylor (2008) and Taylor et al. (2006).

Several forecasting methods with varying degrees of success have been implemented for load forecasting including multiple linear regression (Al-Hamadi and Soliman, 2005; Amjady and Keynia, 2008; Ghiassi et al., 2006; Mirasgedis et al, 2006) and nonlinear multivariable regression model (Al Rashidi and El-Naggar, 2010; Tsekouras et al., 2007). Artificial neural network with variety of approaches such as back propagation neural network (Al-Saba and El-Amin, 1999; Ghiassi et al., 2006; Hobbs et al., 1998; Kermanshahi and Iwamiya, 2002; Poa, 2007), particle swarm optimization (El-Telbany and El-Karmi, 2007), dynamic artificial neural network (Ghiassi et al., 2006), Elman artificial neural network (Beccali et al., 2007) and Jordan recurrent neural network (Kermanshahi and Iwamiya, 2002) have been applied. Simple autoregressive (AR) (Al-Saba and El-Amin, 1999), autoregressive moving average (ARMA) (Al-Saba and El-Amin, 1999), and autoregressive integrated moving average (ARIMA) (Al-Saba and El-Amin, 1999; Ghiassi et al., 2006) models have also been proposed.

Due to academic interests and industrial needs short term load forecasting has gained great attention compared to the others. Short term load forecasting plays an important role to utility company because the accuracy of prediction will affect the power system operations. Forecast errors result in unbalance between power supply and demand, hence increase operations cost (Satish et al., 2004; Topalli and Erkmen, 2003). Since generating an appropriate load of required power supply or balancing between power supply and demand are crucial, generating accurate forecasts are extremely important. Accurate forecasts can avoid energy wasting and prevent system failure. Therefore, it is crucial for both researches and the operational planning of utility company to produce short term forecasts with low error in order to relieve the conflict between supply and need (Xiao, 2009).

The current study investigates methods that are appropriate for forecasting Malaysia load demand. Due to the presence of a double seasonal pattern in load demand data which are daily and weekly seasonal, a double seasonal multiplicative ARIMA model is proposed. The multiplicative double seasonal ARIMA model has often been used for univariate forecasting intraday load time series (Cancelo at al., 2008; Darbelly and Slama, 2000; Taylor, 2008; Taylor, 2006; Taylor et al., 2006). The double seasonal ARIMA with polynomial of order one has been studied by Cancelo at al. (2008) and Darbelly and Slama (2000). Taylor et al. (2006) and Taylor (2006) have utilized the double seasonal ARIMA with polynomials of order two and order three. Taylor (2008) has also utilized the double seasonal ARIMA with polynomial of order three and increased the order to five. However for the reason of parsimony he deferred the consideration of higher order models. Basically when we consider the order of polynomial, for example if we consider the polynomial of order k, we are including all lags from lag one to lag k, however by looking at the sample autocorrelations and the partial autocorrelations, there may exist insignificance lags in between lags. It may also indicate the existence of significance lags after lag k

where those lags are not considered in the model earlier. Therefore in this current research we focus on the subset double seasonal ARIMA model in order to include all the significance lags in our tentative model. We hope to show that a double seasonal ARIMA model with one-step ahead forecast improve the forecasting accuracy in short term load demand in Malaysia.

This paper will be organized as follows. We start by presenting the Box-Jenkins double seasonal ARIMA model and discuss the detail results of a double seasonal ARIMA model. Finally we give our conclusions based on forecasting evaluation method presented in this study.

Materials and Methods

A variety of different forecasting approaches is available to forecast time series data and it is important to realize that no single model is universally applicable. An approach presented here is the Box-Jenkins autoregressive integrated moving average (ARIMA) model.

Box-Jenkins ARIMA Model: For more than half a century, the Box-Jenkins ARIMA linear models have dominated many areas of time series forecasting. One of the attractive features of the Box-Jenkins approach for forecasting is that ARIMA processes are a very rich class of possible models and it is usually possible to find a process which provides an adequate description to the data. Generally, a non seasonal time series can be modeled as a combination of past values and past error, denoted as ARIMA(p,d,q) can be written as (Wei, 2006):

$$\phi_p(B)(1-B)^d \dot{Z}_t = \theta_q(B)a_t \tag{1}$$

with

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

$\theta_q(B) = (1-\theta_1 B - \cdots - \theta_q B^q)$

where B is the backward shift operator and a_t is the purely random process.

In practice, many time series contain a seasonal periodic component, which repeats every s observation. To deal with seasonality, the ARIMA model is generalized hence a general multiplicative seasonal ARIMA (SARIMA) model is defined which follows the ARIMA general procedure represented as follows (Wei, 2006):

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D \dot{Z}_t = \theta_q(B)\Theta_Q(B^s)a_t \tag{2}$$

with

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

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

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

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

where B is the backward shift operator and a_t is the purely random process.

The original series Z_t is differenced by appropriate differencing to remove non-stationary terms. $(1-B)^d$ and $(1-B^s)^D$ are the non-seasonal and seasonal differencing operators, respectively (Dong and Pedrycz, 2008; Topalli and Erkmen, 2003). If the integer D is not zero, then seasonal differencing is involved. The above model is called a SARIMA model of order $(p,d,q)(P,D,Q)_s$. If d is non-zero, then there is a simple differencing to remove trend, while seasonal differencing $(1-B^s)^D$ may be used to remove seasonality. Basically, the values of d and D are usually zero or one and rarely two. Currently we present a double seasonal multiplicative ARIMA model due to the presence of a double seasonal patterns in short-term load demand data which are daily seasonal and weekly seasonal. The multiplicative double seasonal ARIMA model (Box et al., 2008):

$$\phi_p(B)\Phi_{P_1}(B^{s_1})\Pi_{P_2}(B^{s_2})(1-B)^d(1-B^{s_1})^{D_1}(1-B^{s_2})^{D_2}\dot{Z}_t = \theta_q(B)\Theta_{Q_1}(B^{s_1})\Psi_{Q_2}(B^{s_2})a_t \tag{3}$$

where $\dot{Z}_t$ is stationary load demand in period *t*; *B* is backward shift operator; $\phi_p(B)$ and $\theta_q(B)$ are regular autoregressive and moving average polynomials of orders p and q ; $\Phi_{P_1}(B^{s_1})$, $\Pi_{P_2}(B^{s_2})$, $\Theta_{Q_1}(B^{s_1})$ and $\Psi_{Q_2}(B^{s_2})$ are autoregressive and moving average polynomials of orders P_1, P_2, Q_1 and Q_2 ; s_1 and s_2 are the seasonal periods; d , D_1 and D_2 are the orders of integration; a_t is a white noise process with zero mean and constant variance. The seasonal cycles, s_1 and s_2 are selected according to the type of load data series. The daily and weekly seasonality are denoted as s_1 and s_1 respectively. Generally, for hourly load, s_1 =24 and s_2 =168 (Taylor et al., 2006), for half-hourly load, s_1 =48 and s_2 =336 (Darbellay and Slama, 2000; Taylor et al., 2006), while for minute by minute load series s_1 =1440 and s_2 =10080 (Taylor, 2008). The multiplicative double seasonal ARIMA model be expressed as $ARIMA(p,d,q)(P_1,D_1,Q_1)_{s_1}(P_2,D_2,Q_2)_{s_2}$. As an example let $d=0$, $D_1=0$, $D_2=0$, $p=0$, $q=1$, $P_1=0$, $P_2=0$, $Q_1=1$, $Q_2=1$, $s_1=48$ and $s_2=336$, hence the model can be expressed as $ARIMA(0,0,1)(0,0,1)^{48}(0,0,1)^{336}$. Consider $\theta_1=0.27$, $\Theta_1=0.77$ and $\Psi_1=0.85$, the model can be written as follows:

$$Z_t = a_t - 0.27a_{t-1} - 0.77a_{t-48} + 0.2079a_{t-49} - 0.85a_{t-336} + 0.2295a_{t-337} + 0.6545a_{t-384} - 0.1725a_{t-385} \tag{4}$$

Box-Jenkins ARIMA Modeling Procedure: The modeling procedure of Box-Jenkins ARIMA Model involves an iterative five-stage process as follows:

- (i) Preparation of data including transformations and differencing
- (ii) Identification of the potential models by looking at the sample autocorrelations and the partial autocorrelations
- (iii) Estimation of the unknown parameters by some optimization methods
- (iv) Checking the adequacy of fitted model by performing normal probability plot, ACF and PACF on model residuals
- (v) Forecast future outcomes based on the known data.

Forecasting evaluation method: Basically, for the purpose of evaluating out of sample forecasting capability different evaluation statistics such as the root mean square error (RMSE), the mean absolute error (MAE) and the mean absolute percentage error (MAPE) are considered. In this study, we consider MAPE as a standard measurement to examine the accuracy of the prediction model. This measure is commonly used in the forecasting literatures. MAPE is defined as (Dong and Pedrycz, 2008):

$$MAPE = \frac{\sum_{i=1}^{n}\left|\frac{Z_i - \hat{Z}_i}{Z_i}\right|}{n} \times 100 \tag{5}$$

where Z_i and $\hat{Z}_i$ are the actual values and the predicted values respectively, while is the number of predicted values.

Data Set: The data used is one year half hourly load demand measured in Megawatt (MW) from September 01, 2005 to August 31, 2005. They are gathered from Malaysian electricity utility company, Tenaga Nasional Berhad (TNB), Malaysia. TNB is one of the biggest and most well-managed power company in Asia in which this utility company has powered for decades through the generation, transmission and distribution of electricity.

Results

A one year Malaysia half hourly load demand, from September 01, 2005 to August 31, 2006 measured in Megawatt (MW) is used in this current study. The data from September 01, 2005 to July 31, 2006 are used for training and the data from August 01, 2006 to August 31, 2006 are used for testing. Malaysia load data is non-stationary data which is clearly shown in Figure 1. The ACF and PACF in Figure 2, shows clearly the presence of seasonal pattern in load data. Plotting the ACF and PACF after non-seasonal differencing (*d=1*) and daily seasonal differencing ($D_1 = 1,\ s_1 = 48$) in Figure 3, indicates the presence of another seasonal pattern which is weekly seasonality with length 336. Figure 4 shows load demand series after non-seasonal differencing (*d=1*), daily seasonal differencing ($D_1 = 1,\ s_1 = 48$) and weekly seasonal differencing ($D_2 = 1,\ s_2 = 336$). It is clear from Figure 4, load demand series is a stationary series after three time differencing. The ACF and PACF of the stationary load demand series are plotted in Figure 5.

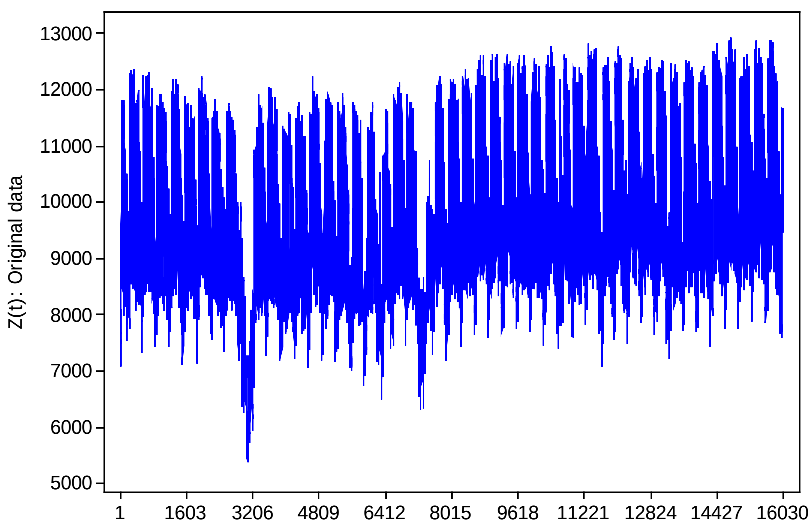

Figure 1: A half hourly load from September 1, 2005 to July 31, 2006.

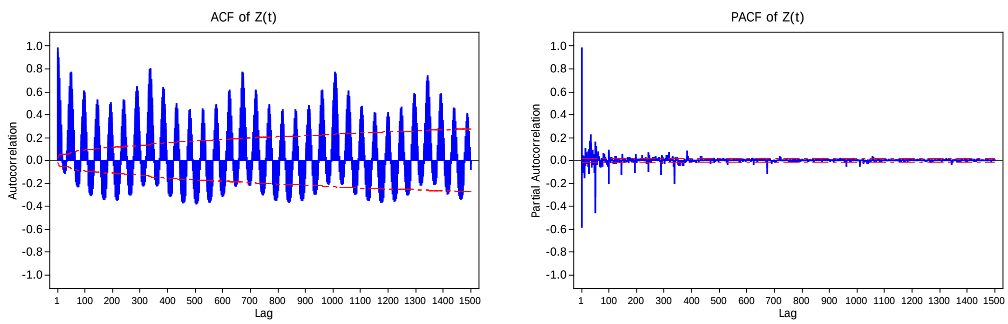

Figure 2: The ACF and PACF of Z_t .

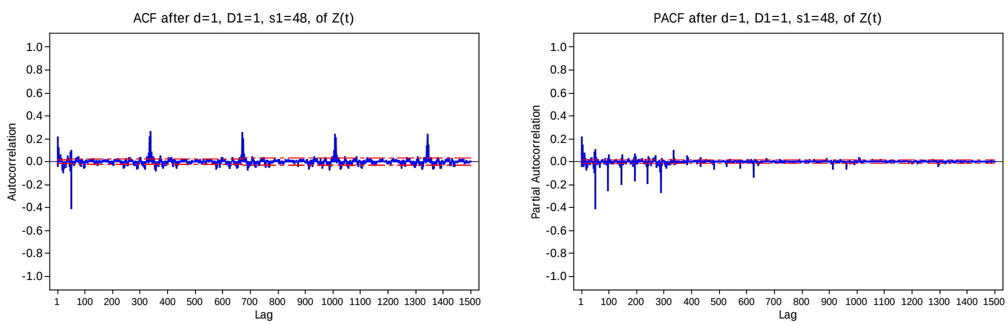

Figure 3: ACF and PACF of Z_t after $d=1$, $D_1 = 1$ and $s_1 = 48$.

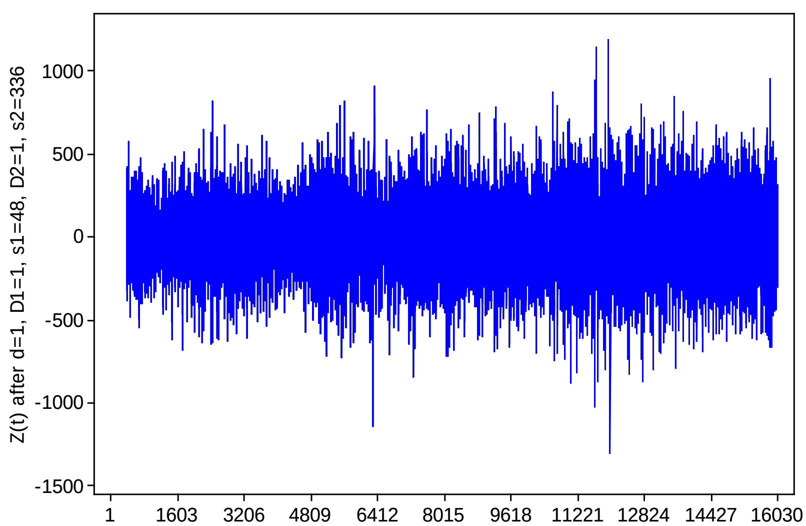

Figure 4: Load demand series after $d=1$, $D_1 = 1$, $s_1 = 48$, $D_2 = 1$ and $s_2 = 336$

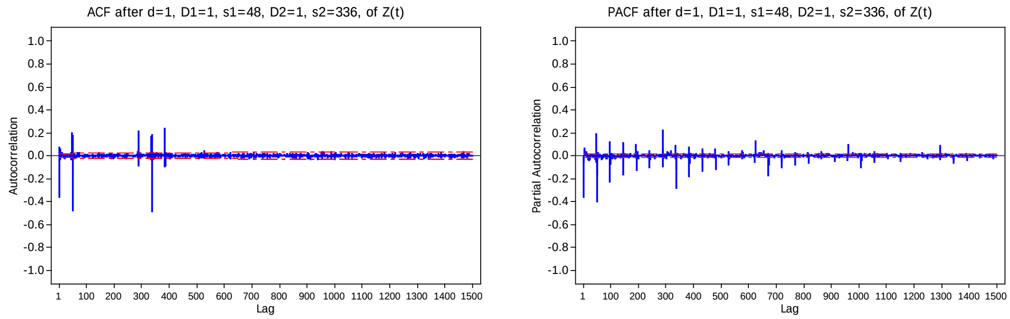

Figure 5: The ACF and PACF of Z_t after *d=1*, $D_1 = 1$, $s_1 = 48$, $D_2 = 1$ and $s_2 = 336$.

For the purposes of analyzing the load data, we have to write a programming for analyzing the double seasonal ARIMA model in Statistical Analysis System, SAS, due to it not being available in statistical packages such as S-plus, MINITAB, MATLAB, Statistical Package for the Social Sciences, SPSS, as well as SAS itself. According to Bowerman and O'Connell (1993) the constant term, δ is included in the tentative model if the absolute value of $\dfrac{\bar{z}}{s_z / \sqrt{n-b+1}}$ is greater than 2 where the δ is statistically different from zero. Since the absolute value less is than 2, 0.0333 therefore the δ is not included in our tentative double seasonal ARIMA model. The integer values *p*, q, P_1, P_2, Q_1 and Q_2 are identified by looking at the sample autocorrelations and the partial autocorrelations of the differenced series. The AR and MA coefficients in a double seasonal ARIMA are estimated by the least squares method. Finally, model validation is made through performing ACF, PACF and normal probability plot of the residuals to determine whether the residuals are white noise and normally distributed. The Akaike's information Criterion (AIC) and the Schwarz Bayesian Criterion (SBC) are used for model selection criteria. Three models are selected from this study.

The first selected model is as follows:

ARIMA([1,2,3,5,11,16,17,18,19,20,23,28,29,34,38,46,47],1,1)$(0,1,1)^{48}(0,1,1)^{336}$

All the parameters are significant at alpha 0.1 significance level with white noise residuals based on Ljung-Box statistic until lags 48. This model gives 10 extreme residual values. In terms of magnitude of the residuals, these are at 11633th, 11632th, 6305th, 7265th, 3041th, 2415th, 10721th, 12659th, 11680th and 11681th observations. The model residual however does not satisfy the Normal Distribution because of the presence of outliers in the data. The AIC and the SBC of this model are 194170.1 and 194330.9 respectively.

The second selected model is as follows:

ARIMA(0,1,[1,2,3,5,11,16,17,18,19,20,21, 22, 24, 28, 29,31,34,36,41,47])$(0,1,1)^{48}(0,1,1)^{336}$

In this model we also found that, all the parameters are significant at alpha 0.05 significance level with white noise residuals based on Ljung-Box Q^* statistic until lags 48. This model gives also 10 extreme residual values. In terms of magnitude of the residuals, these are at 11633th, 11632th, 6305th, 7265th, 3041th, 7456th, 11651th, 2415th, 11681th and 12659th observations. Similar to the first model, the model residual does not satisfy the Normal Distribution. The AIC and the SBC of this model are 194259.2 and 194435.3 respectively.

The third selected model is as follows:

ARIMA(0,1,1)$(0,1,1)^{48}(0,1,1)^{336}$

For this model, all the parameters are significant at alpha 0.05 significance level, however based on Ljung-Box Q^* statistic the residuals are not white noise. This model gives also 10 extreme residual values. In terms of magnitude of the residuals, these are at 11633th, 11632th, 6305th, 7265th, 2945th, 2963th, 2415th, 11652th, 11654th and 11651th observations. Similar to the first and second models, the model residual does not satisfy the Normal Distribution. The AIC and the SBC of the this model are 195054 and 195077 respectively.

Discussion

By looking at the PACF of stationary series, refer Figure 5, we found that our model has a fix pattern with three moving average parameters which are (MA1,1), (MA2,1) and (MA3,1) therefore these three parameters have to be included in our model. It is clear from Table 1, Table 2 and Table 3 that the estimate values of these three parameters of Model 1, Model 2 and Model 3 are greater than 0.2, with highly significant at alpha less than 0.0001 significance level. For the model 1, although all the parameters for autoregressive listed in Table 1 are significant, the estimate values of these parameters are less than ±0.1 except (AR1,1) and (AR1,2). Similar to the model 2, all the parameters for moving average listed in Table 2 are significant. However the estimate values of these parameters are less than ±0.1 except (MA1,2).

The in-sample MAPE and the out-sample MAPE of four time horizons for these three models are summarized in Table 4. One of the conclusions of the M3 Competition, proposed by Makridakis and Hibon (2000) was that the accuracies of various methods depend upon the length of the forecasting horizon involved. Meade (2000) also found that, the forecasting accuracy was less accurate for longer horizons. This current study clearly shows that, by referring to Table 4, the accuracy of out-sample also depends on the length of forecasting horizon which follows the conclusion of the M3 Competition and also the finding by Meade (2000).

Table 1: An Output SAS of model 1.

The ARIMA Procedure

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.65872	0.03451	19.09	<.0001	1
MA2,1	0.79291	0.0052194	151.91	<.0001	48
MA3,1	0.84102	0.0045319	185.58	<.0001	336
AR1,1	0.31643	0.03478	9.10	<.0001	1
AR1,2	0.21626	0.01362	15.87	<.0001	2
AR1,3	0.06673	0.0083353	8.01	<.0001	3
AR1,4	0.02647	0.0083951	3.15	0.0016	5
AR1,5	-0.02746	0.0078218	-3.51	0.0004	7
AR1,6	0.03253	0.0071654	4.54	<.0001	11
AR1,7	-0.02752	0.0080536	-3.42	0.0006	16
AR1,8	-0.03286	0.0083797	-3.92	<.0001	17
AR1,9	-0.05219	0.0087119	-5.99	<.0001	18
AR1,10	-0.03296	0.0092479	-3.56	0.0004	19
AR1,11	-0.02315	0.0093040	-2.49	0.0128	20
AR1,12	0.02819	0.0076927	3.66	0.0002	23
AR1,13	-0.01707	0.0078590	-2.17	0.0299	28
AR1,14	-0.03567	0.0080528	-4.43	<.0001	29
AR1,15	0.02390	0.0073174	3.27	0.0011	34
AR1,16	-0.02226	0.0070429	-3.16	0.0016	38
AR1,17	0.02172	0.0077660	2.80	0.0052	46
AR1,18	0.05786	0.0081684	7.08	<.0001	47

Variance Estimate	14331.19
Std Error Estimate	119.7129
AIC	194170.1
SBC	194330.9
Number of Residuals	15647

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	.	0	.	-0.000	0.000	0.003	-0.005	0.004	-0.008
12	.	0	.	0.001	-0.005	0.007	0.004	0.002	0.013
18	.	0	.	-0.010	-0.005	-0.006	0.001	0.002	0.001
24	10.02	3	0.0184	0.002	0.002	-0.001	-0.002	-0.002	-0.008
30	14.43	9	0.1078	0.008	-0.010	0.008	0.002	0.003	0.006
36	19.63	15	0.1866	0.002	-0.007	-0.014	-0.004	-0.002	0.008
42	28.66	21	0.1224	-0.004	0.002	-0.012	0.002	-0.020	-0.001
48	31.99	27	0.2324	-0.005	0.008	-0.003	-0.006	-0.008	-0.005

The result in Table 4 shows that the third model outperforms the first and second models, even though the third model is the simplest among all. The first conclusion of the M3 Competition proposed by Makridakis and Hibon (2000) was that statistically sophisticated or complex methods do not necessarily provide more accurate forecast than simpler ones. The result from this current study supports that conclusion of the M3 Competition. Koning et al. (2005) mentioned that there is no relationship between complexity and accuracy. Based on this discussion, followed by the performance in Table 4 and also the concept of parsimony where the simplest model is the best model, we suggest that the model 3 is the best model for this current study. We then illustrate the MAPE of in-sample and out-sample forecasts of three models in Figure 7. The model 3 can be expressed as follows:

$$(1-B)(1-B^{48})(1-B^{336})Z_t = (1-0.27184B)(1-0.76592B^{48})(1-0.85019B^{336})a_t \qquad (6)$$

$$\begin{aligned} Z_t &= Z_{t-1} + Z_{t-48} - Z_{t-49} + Z_{t-336} - Z_{t-337} - Z_{t-384} + Z_{t-385} + a_t - 0.27184a_{t-1} - 0.76592a_{t-48} \\ &+ 0.20821a_{t-49} - 0.85019a_{t-336} + 0.23112a_{t-337} + 0.65118a_{t-384} - 0.17702a_{t-385} \end{aligned} \qquad (7)$$

$$Z_t = f(Z_{t-1}, Z_{t-48}, Z_{t-49}, Z_{t-336}, Z_{t-337}, Z_{t-384}, Z_{t-385}, a_t, a_{t-1}, a_{t-48}, a_{t-49}, a_{t-336}, a_{t-337}, a_{t-384}, a_{t-385}) \qquad (8)$$

Table 2: An Output SAS of model 2

The ARIMA Procedure

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.33967	0.0079531	42.71	<.0001	1
MA1,2	-0.10701	0.0083871	-12.76	<.0001	2
MA1,3	-0.03427	0.0079466	-4.31	<.0001	3
MA1,4	-0.04763	0.0075080	-6.34	<.0001	5
MA1,5	-0.03975	0.0075179	-5.29	<.0001	11
MA1,6	0.02036	0.0079699	2.55	0.0107	16
MA1,7	0.03144	0.0084166	3.74	0.0002	17
MA1,8	0.05965	0.0084498	7.06	<.0001	18
MA1,9	0.04398	0.0084533	5.20	<.0001	19
MA1,10	0.04624	0.0084502	5.47	<.0001	20
MA1,11	0.03011	0.0084415	3.57	0.0004	21
MA1,12	0.02646	0.0079903	3.31	0.0009	22
MA1,13	0.02104	0.0075355	2.79	0.0052	24
MA1,14	0.01831	0.0079116	2.31	0.0206	28
MA1,15	0.03519	0.0079182	4.44	<.0001	29
MA1,16	0.02293	0.0075099	3.05	0.0023	31
MA1,17	0.02367	0.0078917	3.00	0.0027	33
MA1,18	-0.02134	0.0079095	-2.70	0.0070	34
MA1,19	-0.02311	0.0075497	-3.06	0.0022	36
MA1,20	0.01561	0.0075243	2.07	0.0380	41
MA1,21	-0.06157	0.0077262	-7.97	<.0001	47
MA2,1	0.78333	0.0052477	149.27	<.0001	48
MA3,1	0.84222	0.0044978	187.25	<.0001	336

Variance Estimate	14411.21
Std Error Estimate	120.0467
AIC	194259.2
SBC	194435.3
Number of Residuals	15647

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	.	0	.	0.001	0.004	0.000	0.004	0.002	-0.001
12	.	0	.	-0.009	0.002	0.008	0.007	-0.003	0.010
18	.	0	.	-0.002	0.003	-0.002	-0.001	-0.002	-0.002
24	7.03	1	0.0080	-0.004	-0.005	-0.003	-0.003	0.003	-0.003
30	12.94	7	0.0736	0.004	-0.016	-0.000	-0.004	-0.006	-0.008
36	18.67	13	0.1336	-0.001	-0.019	-0.004	-0.000	-0.001	0.000
42	21.48	19	0.3109	0.008	0.001	0.000	0.008	0.000	0.007
48	33.36	25	0.1223	-0.004	0.008	0.002	0.024	0.007	0.006

Table 3: An Output SAS of model 3.

The ARIMA Procedure

Conditional Least Squares Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.27184	0.0077170	35.23	<.0001	1
MA2,1	0.76592	0.0052374	146.24	<.0001	48
MA3,1	0.85019	0.0043550	195.22	<.0001	336

Variance Estimate	15181.56
Std Error Estimate	123.2135
AIC	195054
SBC	195077
Number of Residuals	15647

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	374.56	3	<.0001	-0.037	0.112	0.076	0.028	0.057	0.017
12	415.44	9	<.0001	0.000	0.006	0.014	0.013	0.039	0.025
18	597.84	15	<.0001	0.001	-0.003	-0.013	-0.037	-0.054	-0.085
24	887.81	21	<.0001	-0.077	-0.076	-0.055	-0.049	-0.013	-0.035
30	1008.04	27	<.0001	-0.010	-0.031	-0.017	-0.038	-0.061	-0.034
36	1074.44	33	<.0001	-0.038	-0.030	-0.033	0.012	0.003	0.026
42	1085.71	39	<.0001	0.016	0.008	0.002	0.012	-0.010	0.011
48	1230.59	45	<.0001	0.002	0.022	0.017	0.036	0.082	0.022

Table 4: The MAPE of in-sample and out-sample forecasts of three models.

	Model 1	Model 2	Model 3
In-sample forecast	0.9680	0.9711	0.9906
Out-sample one-week forecast	10.1892	9.5818	8.8841
Out-sample two- week forecasts	15.8199	14.9531	13.9414
Out-sample three-weeks forecasts	21.6847	20.5402	19.1838
Out-sample one- month forecast	29.4448	27.8249	25.8641

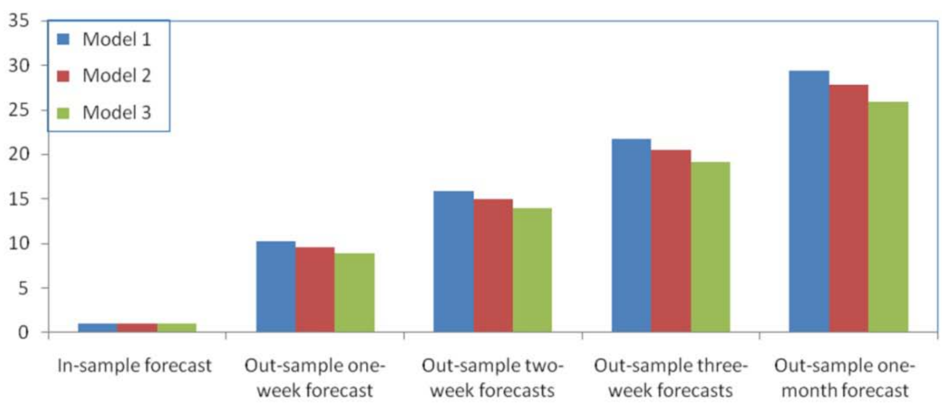

Figure 6: The MAPE of in-sample and out-sample forecasts of three models.

Result of One Step Ahead Out-sample Forecasts

Basically, all statistical packages such as MINITAB, MATLAB, S-plus, SPSS and also SAS provide out-sample forecasts based on k-step ahead. The k-step ahead out-sample forecasts accumulate the error terms resulting in low accuracy in forecasting performances. Therefore the out-sample forecasts based on k-step ahead are highly influenced by lead times as shown in Table 5. However, the one-step ahead out-sample forecasts is not influenced by the forecast lead times. Unfortunately the statistical packages do not provide the one-step ahead out-sample forecasts. We then utilized the Microsoft Office Excel to calculate the one-step ahead out-sample forecasts based on the model gathered from the SAS package. The results of one-step ahead out-sample forecast s of ARIMA$(0,1,1)(0,1,1)^{48}(0,1,1)^{336}$ model are presented in Table 5 which clearly show that the one-step ahead out-sample forecasts is not influenced by the lead times. We illustrate the MAPE of k-step and one-step ahead out-sample forecasts of model 3 in Figure 7; the out-sample forecasts of k-step ahead in Figure 8; and the out-sample forecasts of one-step ahead in Figure 9.

Table 5: The MAPE of k-step and one-step ahead out-sample forecasts of model 3.

	k-step ahead out-sample forecasts	One-step ahead out-sample forecasts
Out-sample one week forecast	8.8841	0.9467
Out-sample two-week forecasts	13.9414	0.9322
Out-sample three-week forecasts	19.1838	0.9046
Out-sample one month forecast	25.8641	0.9778

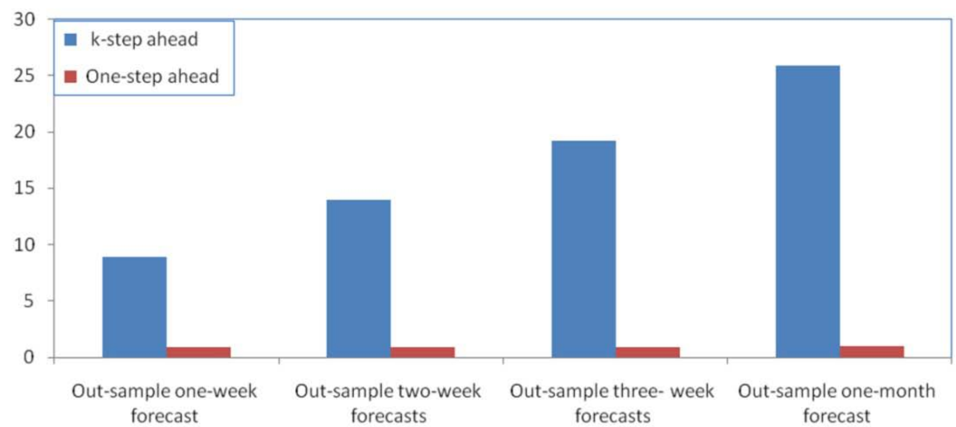

Figure 7: The MAPE of k-step and one-step ahead out-sample forecasts of model 3.

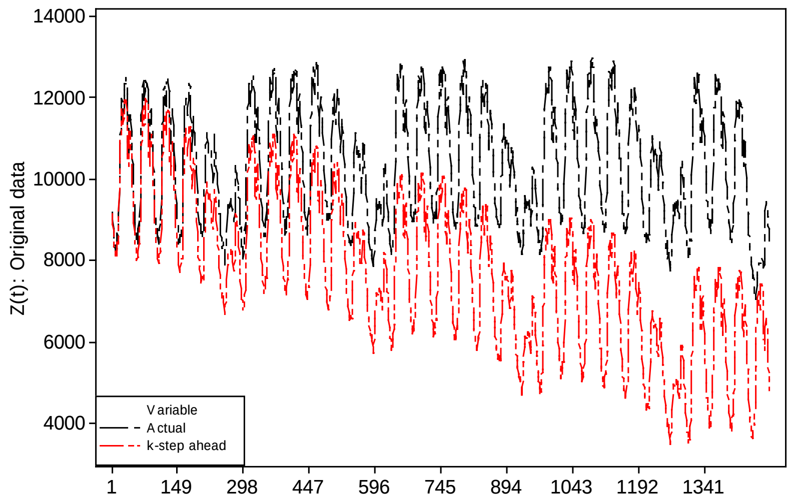

Figure 8: The out-samples of actual data of k-step ahead out-sample forecasts.

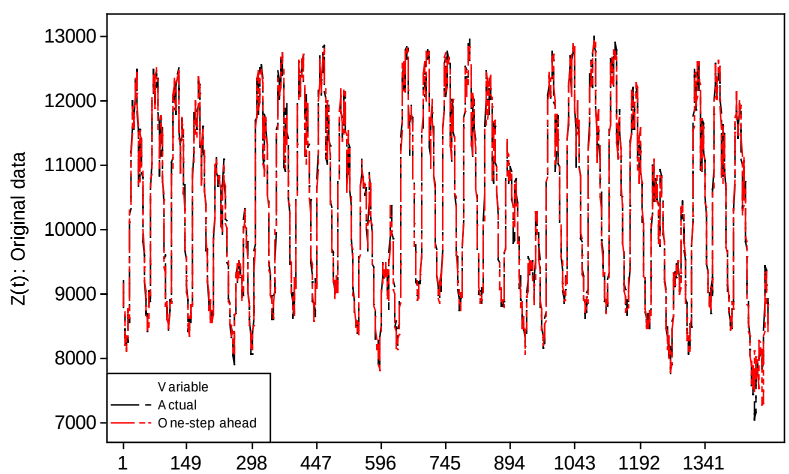

Figure 9: The out-samples of actual data and one-step ahead out-sample forecasts.

We then illustrate out-samples of actual load data, k-step ahead and one-step ahead of model 3 in Figure 10. It is evidenced from the figure that one-step ahead out-sample forecasts follow the actual load data more closely than k-step ahead out-sample forecasts.

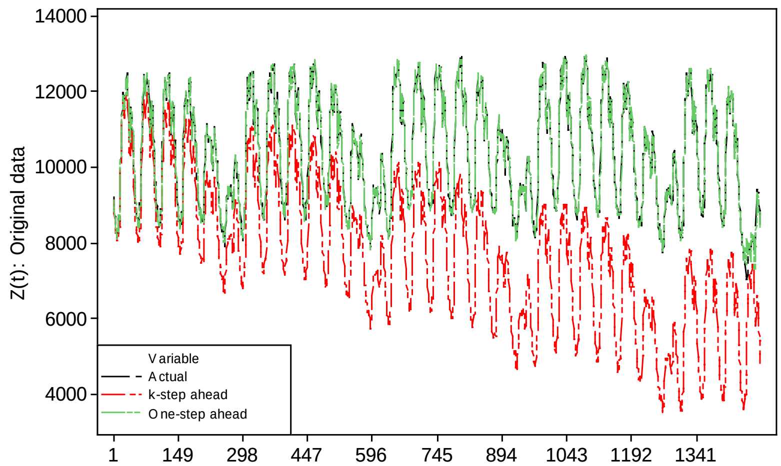

Figure 10: The out-samples of actual data, k-step ahead and one-step ahead out-sample forecasts.

Conclusion

Comparing the forecasting performances by using k-step ahead out-sample forecasts and one-step ahead out-sample forecasts, we found that the MAPE for the one-step ahead out-sample forecasts from any horizon ranging from one week lead time to one month one week lead time as illustrated in Table 5 are all less than 1%. In other words it can be concluded that the one-step ahead out-sample forecasts are not influenced by the lead times. Furthermore, since the MAPE for all lead time horizons for the one-step ahead out-sample forecasts are less the ones obtained using the k-step ahead out-sample forecasts, it can also be concluded that one step head forecasts are more accurate than k-step ahead forecasts. Therefore we propose that a double seasonal ARIMA model with one-step ahead out-sample forecasts must be considered in forecasting time series data with two seasonal cycles, especially in Malaysia load data.

Acknowledgement

This research was supported by University Malaysia Terengganu (UMT). The researchers are grateful acknowledge this support. We wish to thank Tenaga Nasional Berhad (TNB), Malaysia for providing the load data. We also wish to express our appreciation to all friends who help us directly and indirectly.

References

Al-Hamadi, H. M., and Soliman, S. A. (2004). Short-term electric load forecasting based on Kalman filtering algorithm with moving window weather and load model. *Electric Power Systems Research, 68*, 47-59.

Al-Hamadi, H. M., and Soliman, S. A. (2005). Long-term/mid-term electric load forecasting based on short-term correlation and annual growth. *Electric Power System Research, 74,* 353-361.

Al Rashidi, M. R., and El- Naggar, K. M. (2010). Long term electric load forecasting based on particle swarm optimization. *Applied Energy, 87,* 320-326.

Al-Saba, T., and El-Amin, I. (1999). Artificial neural networks as applied to long-term demand forecasting. *Artificial Intelligence in Engineering, 13,* 189-197.

Amjady, N., and Keynia, F. (2008). Mid-term load forecasting of power systems by a new prediction method. *Energy Conversion and Management , 49,* 2678-2687.

Beccali, M., Cellura, M., Lo Brano, V., and Marvuglia, A. (2004). Forecasting daily urban electric load profiles using artificial neural networks. *Energy Conversion and Management, 45,* 2879-2900.

Beccali, M., Cellura, M., Lo Brano, V., and Marvuglia, A. (2007). Short-tem prediction of household electricity consumption: Assessing weather sensitivity in a Mediterranean area. *Renewable and Sustainable Energy Reviews, 12,* 2040-2065.

Bowerman, B. L. & O'Connell, R. T., (1993). *Forecasting and time series: An applied approach, Third Edition.* Duxbury Press, California.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control, Fourth Edition.* John Wiley & Sons, New Jersey.

Cancelo, J. R., Espasa, A., and Grafe, R., (2008). Forecasting the electricity load from one day to one week ahead for the Spanish system operator. *International Journal of Forecasting 24*, 588-602.

Carpinteiro, O. A. S., Leme R. C., de Souza, A. C. Z., Pinheiro, C. A. M., and Moreira, E. M., (2007). Long-term load forecasting via a hierarchical neural model with time integrators. *Electric Power Systems Research, 77*, 371-378.

Catalao, J. P. S., Mariano, S. J. J. S., Mendes, V. M. F., and Ferreira, L. A. F. M. (2007). Short-term electricity prices forecasting in a competitive market: A neural network approach. *Electric Power Systems Research, 77,* 1297-1304.

Dong, R., and Pedrycz, W. (2008). A granular time series approach to long-term forecasting and trend forecasting. *Physica A, 387*, 3253-3270.

Darbellay G. A, Slama, M. (2000). Forecasting the short-term demand for electricity. Do neural networks stand a better chance? *International Journal of Forecasting, 16*, 71-83.

Elkateb, M. M., Solaiman, K., and Al-Turki, Y. (1998). A comparative study of medium-weather-dependent load forecasting using enhanced artificial/fuzzy neural network and statistical techniques. *Neurocomputing, 23*, 3-13.

El-Telbany, M., & El-Karmi, F. (2007) Short-term forecasting of Jordonian electricity demand using particle swarm optimization, *Electric Power Systems Research, 78,* 425-433.

Gonzalez, P. A., and Zamarreno, J. M. (2005). Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Building, 37,* 595-601.

Ghiassi, M., Zimbra, D. K., and Saidane, H. (2006). Medium term system load forecasting with a dynamic artificial neural network model. *Electric Power System Research, 76,* 302-316.

Hobbs, B. F., Helman, U., Jitprapaikulsarn, S., Konda, S., and Maratukulam, D. (1998). Artificial neural networks for short-term energy forecasting: Accuracy and economic value. *Neurocomputing, 23*, 71-84.

Jia, N. X., Yokoyama, R., Zhou, Y.C., and Gau, Z. Y. (2001). A flexible long term load forecasting approach based on new dynamic simulation theory-GSIM. *Electrical power and Energy Systems, 23,* 549- 555.

Kermanshahi, B., and Iwamiya, H. (2002). Up to year 2020 load forecasting using neural nets. *Electric Power and Energy Systems, 24,* 789-797.

Kandil, N., Wamkeue, R., Saad, M., and Georges, S. (2006). An efficient approach for short term load forecasting using artificial neural networks. *Electrical Power and Energy Systems, 28,* 525-530.

Koning, A. J., Frances, P. H., Hibon, M., and Stekler, H. O. (2005). The M3 competition: Statistical tests of the results. *International Journal of Forecasting, 21*, 397-409.

Meade, N. (2000). A note on the Robust Trend and ARARMA methodologies used in the M3 Competition. *International Journal of Forecasting, 16,* 517-519.

Makridakis, S., and Hibon, M. (2000). The M3-Competition: results, conclusion and implications. *International Journal of Forecasting, 16,* 451-476.

Mirasgedis, S., Sarafidis, Y., Georgopoulou, E., Lalas, D. P., Moschovits, M., Karagiannis, F., and Papakonstantinou, D. (2006). Models for mid-term electricity demand forecasting incorporating weather influences. *Energy, 31*, 208-227.

Pao, H-T. (2007). Forecasting electricity market pricing using artificial neural networks. *Energy Conversion and Management, 48,* 907-912.

Pedregal, D. J., and Trapero, J. R. (2010). Mid-term hourly electricity forecasting based on a multi-rate approach. *Energy Conversion and management, 51*, 105-111.

Srinivasan, D. (1998). Evolving artificial neural network for short term load forecasting. *Neurocomputing, 23*, 265-276.

Soares, L. J., and Medeiros, M. C. (2008). Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. *International Journal of Forecasting, 24*, 630-644.

Satish, B., Swarup, K. S., Srivinas, S., and Roa, A. H. (2004). Effect of temperature on short term load forecasting using an integrated ANN. *Electrical Power Systems Research, 72*, 95-101.

Taylor, J. W. (2008). An evaluation of methods for very short-term load forecasting using minute-by-minute British data. *International Journal of Forecasting, 24*, 645-658.

Tsekouras, G. J., Dialynas, E. N., Hatziargyriou, N. D., and Kavatza, S. (2007). A non-linear multivariable regression model for midterm energy of power systems. *Electric Power Systems Research, 77*, 1560-1568.

Topalli, A. K., and Erkmen, I. (2003). A hybrid learning for neural networks applied to short term load forecasting. *Neurocomputing, 51,* 495-500.

Taylor, J. W. (2006). Density forecasting for the efficient balancing of the generation and consumption of electricity. *International Journal of Forecasting, 22*, 707-724.

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

Wei, W. W. S. (2006). *Time Series Analysis, Univariate and multivariate Methods, Second Edition*. Pearson, Addison Wesley, New York.

Xiao, Z., Ye, S-J., Zhang, B., and Cai-Xin Sun, C-X. (2009). BP neural network with rough set for short term load forecasting. *Expert Systems with Applications*, *36*, 273-279.

Zhang, B-L., and Dong, Z-Y. (2001). An adaptive neural-wavelet model for short term load forecasting. *Electric Power Systems Research*, *59*, 121-129.