Forecasting Electricity Consumption in Malaysia by Hybrid ARIMA-ANN

Nur Ezzati Mohd Izudin, Rajalingam Sokkalingam, Hanita Daud, Hermiza Mardesci, and Abdullah Husin

Abstract Forecasting electricity consumption is of national interest to any country. Electricity forecast is not only required for short-term and long-term power planning activities but also in the structure of the national economy. Electricity consumption time series data consists of linear and non-linear patterns. Thus, the patterns make the forecasting difficult to be done. Neither autoregressive integrated moving average (ARIMA) nor artificial neural networks (ANN) can be adequate in modeling and forecasting electricity consumption. The ARIMA cannot deal with non-linear relationships while a neural network alone is unable to handle both linear and non-linear pattern equally well. This research is an attempt to develop ARIMA-ANN hybrid model by considering the strength of ARIMA and ANN in linear and non-linear modeling. The Malaysian electricity consumption data is taken to validate the performance of the proposed hybrid model. The results will show that the proposed hybrid model will improve electricity consumption forecasting accuracy by compare with other models.

Keywords Forecasting · Time series · ARIMA · ANN · Hybrid method

N. E. M. Izudin (✉) · R. Sokkalingam · H. Daud
Fundamental and Applied Sciences Department, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak, Malaysia
e-mail: nur_17008265@utp.edu.my

R. Sokkalingam
e-mail: raja.sokkalingam@utp.edu.my

H. Daud
e-mail: hanitadaud@utp.edu.my

H. Mardesci
Fakultas Pertanian, Universitas Islam Indragiri, Riau, Indonesia
e-mail: hermizamardesci@unisi.ac.id

A. Husin
Program Studi Sistem Informasi, Universitas Islam Indragiri, Riau, Indonesia

S. A. Abdul Karim et al. (eds.), *Proceedings of the 6th International Conference on Fundamental and Applied Sciences*, Springer Proceedings in Complexity,
https://doi.org/10.1007/978-981-16-4513-6_66

1 Introduction

In Malaysia, the demand for electricity increases by 4.7 percent per year over the outlook period, to reach 274 TWh in 2030 [1]. The growth in electricity demand is heavily influenced by strong demand from the industrial and residential sectors, which is projected to increase annually at 5.4 and 4.9 percent respectively over the outlook period. The past 15 years (2000–2015) record show that the electricity consumption trend in Malaysia is continuously increasing. Therefore, electricity consumption forecast has fundamental importance in the energy planning of Malaysia. A good forecasting technique is critically important to estimate the level of electricity's demand accurately, thus proper planning could be made by the utility company and the government to meet the country's future. Several techniques are being used for electricity consumption planning to accurately forecast future electricity consumption in Malaysia.

This paper reviews previous research works related to ARIMA, ANN and Hybrid method only which were applied in various applications [2]. Using statistics rule-based approach to forecast peak load electricity demand and they modified the basic regression model using Box-Jenkins autoregressive error. It was produced an adequate model with 2.41% forecasting error. While Haiges et al. forecast using ARIMA method that provides a projection that relies on past historical data to reach a state of statistical equilibrium [3]. Besides, Mohamed Othman et al. using ANN-based forecast the electricity consumption in Malaysia [4]. The authors developed ANN model which involved the creation of numerous feed-forward backpropagation network in MATLAB and select the best ANN model via cross-validation method.

Yan and Zou using Hybrid ARIMA-ANN techniques to forecast water quality [5]. The ARIMA models were first used to do the water quality forecasting and then with the obtained errors ANNs were built taking into the nonlinear patterns. For modeling linear and nonlinear components of a time series. The hybrid ARIMA-ANN combines linear and non-linear models to improve the forecasting performance of the price of Robusta Coffee in India [6]. Mohan and Reddy apply hybrid ARIMA-ANN model to predict the resource usage in server virtualization [7]. The authors using the measurement base approach with time series for prediction and reviewed the effectiveness of the ARIMA model. They compared hybrid ARIMA-ANN to ARIMA and ANN and the result shows that the accuracy has improved 5 times. Mucaj and Sinaj presented three models, ARIMA, NAR, and ARIMA-ANN. The hybrid model proposed to improve the forecasting accuracy and the result shows that the hybrid model has the best results among the three models [8].

Many empirical studies [1, 2, 4, 9–11] has been conducted to forecast electricity consumption using multivariate techniques and time-series analysis such as autoregressive (AR), moving average (MA) algorithm, general exponential smoothing algorithm, ARMA algorithm, and AR integrated MA (ARIMA). The above methods have their own advantages and drawbacks in the variable selection, model selection, segmenting and lack of comparison with a combination of different models to improve the forecasting accuracy.

This paper presents the forecast electricity consumption by using hybrid ARIMA-ANN based on historical data. Therefore, the authors would like to propose a hybrid ARIMA-ANN model for forecasting electricity consumption in Malaysia.

2 Methodology

Figure 1 shows the flowchart model development of hybrid ARIMA-ANN which consists of five stages.

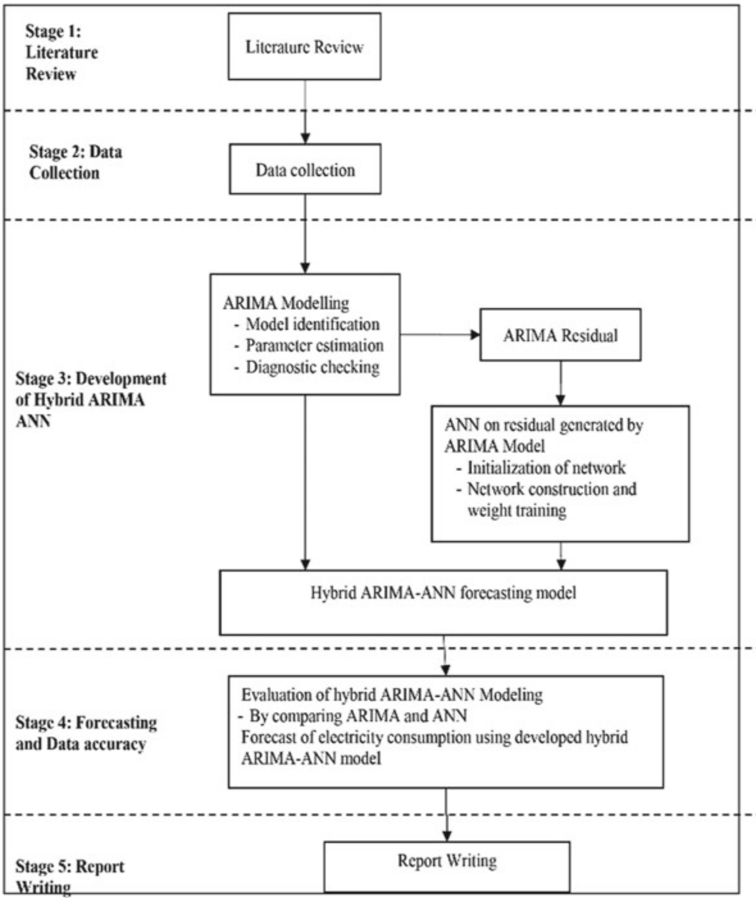


Fig. 1 Model development hybrid ARIMA-ANN

2.1 Data Collection

The dataset was obtained from the Malaysian Energy Commission on a yearly basis for 40 years, between the years 1978 and 2017. Interpolation process was used to obtain more data.

2.2 The ARIMA Model

Box-Jenkins method or ARIMA model was one of the most popular approaches of forecasting and has been introduced this approach in 1976. ARIMA is the combination of the autoregressive and moving average models, the future value of a variable is assumed to be a linear function of random errors and several past observations. The mathematical formula can be expressed as the form of Eq. (1):

$$y_t = \theta_0 + \varphi 1 y_{t-1} + \varphi 1 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_{12} \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \tag{1}$$

where p; number of lags of the considered variable, q; number of lags of the error term, y_t; actual value at time t, y_{t-i}; series in the preceding ith period, φ; ith autoregressive coefficient, ε_t; random error at time t, ε_{t-I}; preceding error term at the ith period, θ_i; ith moving average coefficient. Equation (1) is an important special case of the ARIMA family of models. (1) becomes an AR model of order p when $q = 0$, and the model reduces to an MA model of order q if $p = 0$.

The Box-Jenkins methodology [12] has a fundamental impact on the time series analysis and forecasting application to building ARIMA model. It includes three iterative steps model identification, parameter estimation, and diagnostic checking. To identify the order of the ARIMA model, Box and Jenkins proposed to use the autocorrelation function (ACF) and the partial autocorrelation function (PAFC) of the sample data. The parameter estimated such that an overall measure of errors is minimized by using a nonlinear optimization procedure. Diagnostic checking of model adequacy is the last step, to check if the model assumptions of the errors, ε_t, are satisfied. Plots of residual and diagnostic static can be used to examine the goodness of the model to the past data. If the model is not sufficient, a new model should be identified, which will again be back to the steps of parameter estimation and model verification. To choose the best model the three-step model building is typically repeated several times, and the model for prediction purpose.

2.3 The ANN Model

The model is characterized by a network of three layers of simple processing units connected by acyclic links. The relationship between the output (y_t) and the inputs ($y_{t-1}, \ldots, y_{t-p}$) has the following mathematical representation:

$$y_t = \omega_0 + \sum_{j=1}^{q} \omega_j g\left(\omega_{0j} + \sum_{j=1}^{p} \omega_{i,j} y_{t-i}\right) + e_t \tag{2}$$

where ω_j ($j = 1, 2, .., q$) and $\omega_{i,j}$ ($i = 0, 1, 2, .., p$; $j = 1, 2, .., q$) are the model parameters often called connection weights; p is the number of input nodes and q is the number of hidden nodes. The sigmoid function is often used as the hidden layer transfer function, that is,

$$sig(x) = \frac{1}{(1 + \exp(-x))} \tag{3}$$

Hence, the ANN model, in fact performs a nonlinear functional mapping from the past observations ($y_{t-1}, \ldots, y_{t-p}$) to the future value y_t,

$$y_t = f\left(y_{t-1}, \ldots, y_{t-p,}\omega\right) + e_t \tag{4}$$

where ω is a vector of all parameters and f is a function determined by a network structure and connection weights. Thus the neural network is equivalent to a nonlinear autoregressive model.

2.4 The Hybrid ARIMA-ANN Model

Both ARIMA and ANN models have achieved successes in their own linear o nonlinear domains. However, none of them is a universal model that is suitable for all circumstances. The approximation of ARIMA models to complex nonlinear problems may not be adequate. On the other hand, using ANNs to model linear problems have yield mixed results. Hence, it is not wise to apply ANNs blindly to any type of data. Since it is difficult to completely know the characteristic of the detain a real problem, hybrid model that has both linear and nonlinear modeling capability can be a good choice for forecasting electricity consumption. Different aspect of the underlying patterns may be captured to combining different models. Linear autocorrelation structure and a nonlinear component:

$$Y_t = L_t + N_t \tag{5}$$

where L_t denotes the linear component and N_t denotes nonlinear component. Both of these two parameters have to be estimated from the time series data.

$$e_t = Y_t - YF_t \tag{6}$$

where YF_t is the predicted value of the ARIMA model at time t. With n inputs nodes, the ANN model for the residuals will be:

$$e_t = f(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + u_t \tag{7}$$

where f denotes nonlinear function and u_t denotes random error. NF_t as the forecast from above equation, the combine prediction will be:

$$YF_t = LF_t + NF_t \tag{8}$$

Hybrid ANN and ARIMA system consist of two-step, first ARIMA model is used to analyze the linear part of the problem and the second is neural network model is developed to model the residuals from the ARIMA model.

3 Result and Discussion

3.1 Data Set

The data used in this study were load data of electricity consumption in Malaysia obtained from Malaysia Energy Information hub [13].

3.2 ARIMA Modelling

The historical data of the electricity consumption of Malaysia from 1978 to 2017 was increasing over time as shown in Fig. 2.

Through Fig. 3 of the original series of ACF and PACF, we note that it is non-stationary in the data of the original series as there are some values outside the confidence interval. And to make the series stationary we make differences. The best fit AR parameters and MA parameters should be estimated according to ACF and PACF respectively. Figure 5 show the ACF and PACF. It was obvious that the ACF died off smoothly at a geometric rate after one lag and PACF declined geometrically after one lagged. Therefore, the parameters of AR and MA can be chosen as 1 for the ARIMA model (Fig. 4).

However, in the practical fitting process, any other AR/MA parameters could be selected. For instance, the AR parameters defined as 1 and the MA parameters can

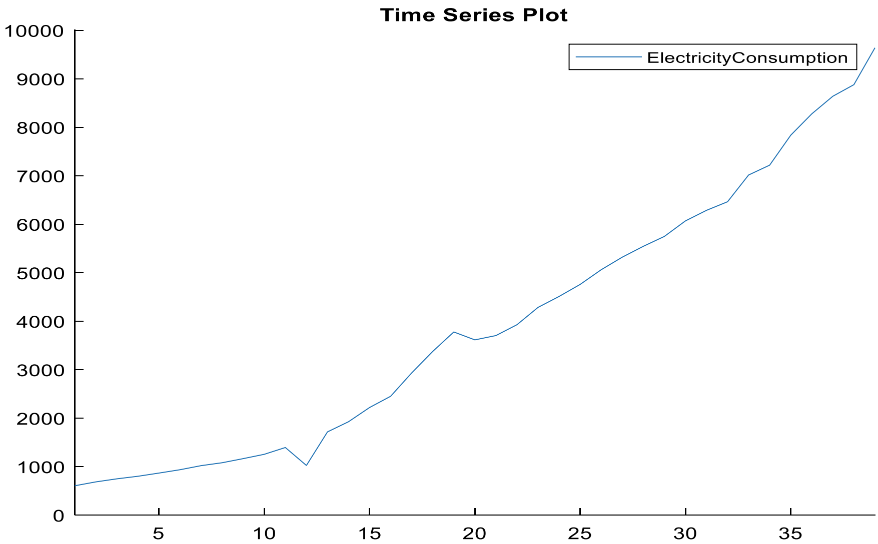


Fig. 2 Electricity consumption of Malaysia from 1978 to 2017 (Kilotonne of oil equivalent)

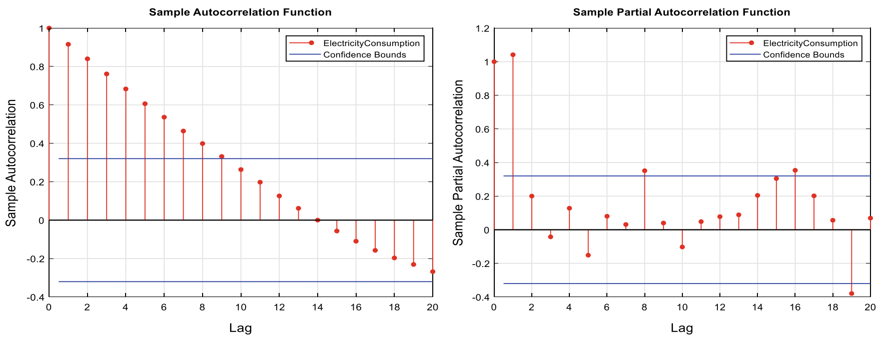


Fig. 3 The ACF and PACF of the series of electricity consumption

be defined as 1 or 3. After fitted, ARIMA(1,2,1) has been found to be the best model among all ARIMA models. Once the ultimately fittest model was identified, the equation form of the model could be obtained:

$$y_t = 2y_{t-1} - 1 - 2y_{t-2} + 10.9595 + e_t + \theta e_t \quad (9)$$

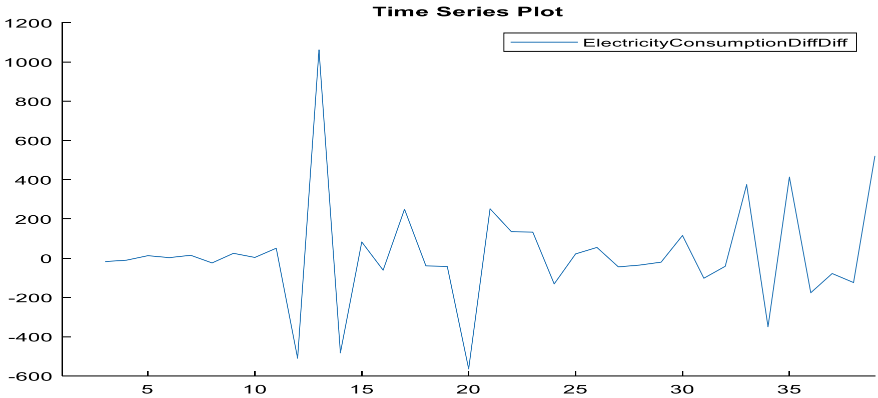


Fig. 4 Graphical representation of the series difference in electricity consumption

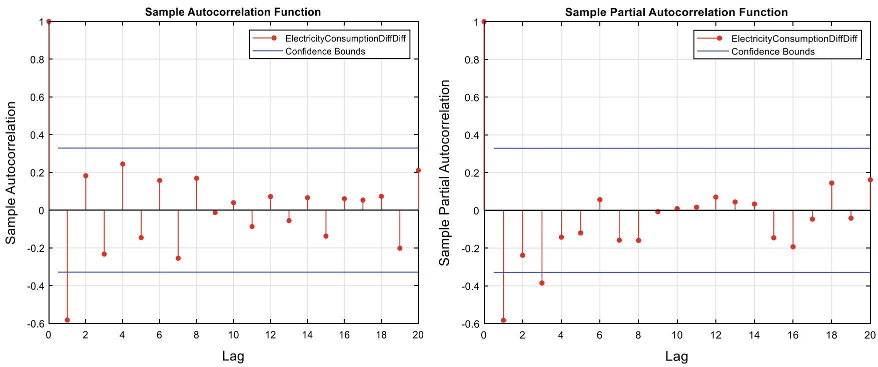


Fig. 5 The ACF and PACF of the series of electricity consumption after difference

3.3 ANN Modelling

A three-layer feedforward neural network model was developed for the forecasting of electricity consumption using an optimized Levenberg–Marquardt training algorithm. The data for the period between 1978 and 2017 were available for the modeling purposes. Electricity consumption time series data were divided into two independent data sets. The first data set of 1978–2013 was used for model training and the other datasets from 2014 to 2017 were used for model verification purposes. In the ANN modeling process, the input and output electricity consumption data sets for each parameter were normalized to the range of [0,1].

The number of neurons in the input and output layers have been set as 5 and 1 respectively. A series of different topologies were used to determine the optimum number of hidden nodes. Compared with the training results, it was found that the training set had the lowest error value when the number of hidden units was 10. 10

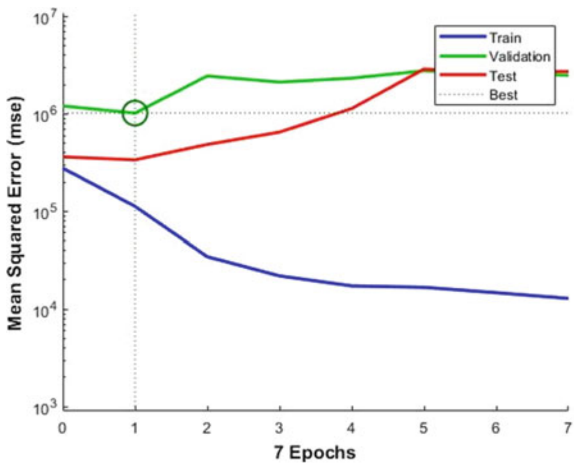


Fig. 6 Training performance of ANN model

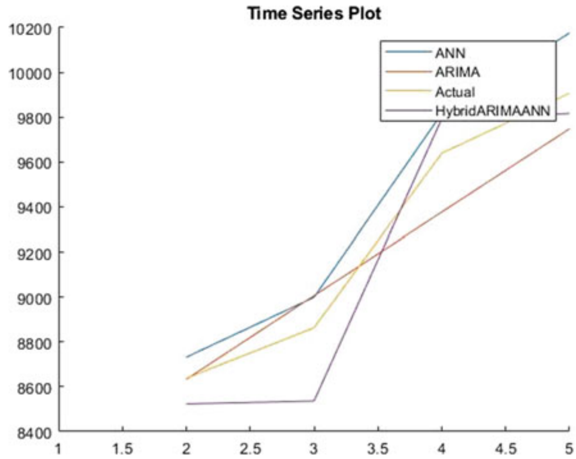
is chosen as the number of hidden nodes. Thus the number of each layer's neurons in the network was 5-10-1 respectively. The parameters of the network were chosen as follows: the transformation function of hidden neuron was "tansig" and "logsig" was the output layer function. The stop criterion of error function was set to 0.001 and the maximum of number of iteration was 1000. Computer program has been performed under MATLAB 2019a. Figure 6 demonstrate the ANN model training performance for electricity consumption parameter.

3.4 Hybrid ARIMA-ANN Modeling

The proposed algorithm of the hybrid system consisted of two steps. In the first step, to analyze the linear part of the problem, an ARIMA model was employed. In the second step, the residuals from the ARIMA model were modeled by using a neural network model. Since the ARIMA model cannot detect the nonlinear structure of the electricity consumption time series data, the residuals of linear model will contain information about nonlinearity. The output from the neural networks can be used as forecasts of the error terms of the ARIMA model.

The hybrid model utilizes the unique feature and strength of ARIMA model as well as ANN model in determining different patterns. Therefore, it may be favorable to model linear and nonlinear patterns separately by using different models and then combine the forecasts to improve the overall modeling and forecasting performance. In the hybrid modeling algorithm, the input and output electricity consumption data sets for each parameter were normalized to the range of [0,1]. In the modeling process, the hybrid model was trained to adjust the model so that the model forecasted electricity consumption parameters match well with observed data. The verification results of 2014–2017 listed in Fig. 7 indicates that the hybrid model forecast results

Fig. 7 Comparing the predicted results with the actual value



reasonable match the observed electricity consumption data. By this model, the forecasting electricity consumption of each year was calculated.

3.5 Comparison of Model Performance

Measure of forecast accuracy should always be evaluated as part of a model validation effort. When more than one forecasting accuracy technique seems reasonable. To evaluate the performance of the forecasting accuracy, the three evaluation statistics which is mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage forecast error (MAPE) are used to evaluate each model.

$$e_t = y_t - y_{t-1} \tag{10}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t| \tag{11}$$

$$RMSE = \frac{1}{n}\sqrt{\sum_{t=1}^{n}(e_t)^2} \tag{12}$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}|\frac{e_t}{y_{t-1}}| \tag{13}$$

Table 1 reports the MAE, RMSE and MAPE for the year 2014–2017 from the ARIMA, ANN and hybrid AIMA-ANN models. It can be seen that the error levels

Table 1 Forecasting performance of different model

	ARIMA	ANN	Hybrid ARIMA-ANN
MAE	2734.9756	239.5	171.3417
RMSE	14,226.0899	275.0873	194.6846
MAPE	0.3930	0.0251	0.0188

in the case of hybrid model are lower than ARIMA and ANN, which leads to the conclusion that the hybrid model show the better consequently and reasonable results.

4 Conclusion

The main objective of this paper is to provide accurate electricity usage prediction models to increase power system reliability. Comparison between the results showed that hybrid ARIMA-ANN model produced better results. The performance of each model is assessed by three statistical measures: RMSE, MAE, and MAPE. The results of the statistical measures will select the best model and will be an effective tool to improve forecasting accuracy. Consequently, forecast of the electricity consumption can be successfully done.

Acknowledgements We would like to thank anonymous referees for many the helpful suggestions which improved the paper considerably. We appreciate the **International Collaborative Research Fund (Universiti Teknologi PETRONAS—Universitas Islam Indragiri) 2021, with cost centre 015-ME0-223** and Fundamental Research Grant Scheme (FRGS) Reference code FRGS/1/2018/STG06/UTP/02/4, Malaysia for sponsoring the research. Lastly, we appreciate Universiti Teknologi PETRONAS where the research was conducted. The views expressed are those of the authors and do not reflect any other authority.

References

1. Aman, S., Ping, H.: Modelling and forecasting electricity consumption of Malaysian large steel mills. Sci. Res. [Internet] **6**(8), 1817–30 (2011). Available from: http://academicjournals.org/sre/PDF/pdf2011/18Apr/Amanetal.pdf
2. Ismail, Z., Yahya, A., Mahpol, K.A.: Forecasting peak load electricity demand using statistics and rule based approach. Am. J. Appl. Sci. **6**(8), 1618–1625 (2009)
3. Haiges, R., Wang, Y.D., Ghoshray, A., Roskilly, A.P.: Forecasting Electricity Generation Capacity in Malaysia: An Auto Regressive Integrated Moving Average Approach. Energy Procedia [Internet] (2017) **105**, 3471–3478 (Oct 2016). Available from: https://doi.org/10.1016/j.egypro.2017.03.795
4. Mohamed Othman, M.S., Johari, D., Musirin, I., Abdul Rahman, T.K., Nik Ismail, N.F.: Artificial neural network-based forecast for electricity consumption in Malaysia. In: 2010 IEEE International Conference on Power and Energy, pp. 24–28 (2010)
5. Yan, H., Zou, Z.: Application of a hybrid ARIMA and neural network model to water quality time series forecasting. J. Converg. Inf. Technol. **8**(4), 59–70 (2013)

6. Naveena, K., Singh, S., Rathod, S., Singh, A.: Hybrid ARIMA-ANN modelling for forecasting the price of robusta coffee in India. Int. J. Curr. Microbiol. Appl. Sci. **6**(7), 1721–1726 (2017)
7. Mohan, B.R., Reddy, G.R.M.: A hybrid ARIMA-ANN model for resource usage prediction. Int. J. Pure. Appl. Math. **119**(12), 12633–12642 (2018)
8. Mucaj, R., Sinaj, V.: Exchange rate forecasting using ARIMA. NAR and ARIMA-ANN Hybrid Model **4**(10), 8581–8586 (2017)
9. Kaytez, F., Taplamacioglu, M.C., Cam, E., Hardalac, F.: Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. Int. J. Electr. Power Energy Syst. [Internet] **67**(May), 431–438 (2015). Available from https://doi.org/10.1016/j.ijepes.2014.12.036
10. Haiges, R., Wang, Y.D., Ghoshray, A., Roskilly, A.P.: Forecasting Electricity Generation Capacity in Malaysia: An Auto Regressive Integrated Moving Average Approach. Energy Proc. [Internet] **105**, 3471–3478 (2017). Available from: https://doi.org/10.1016/j.egypro.2017.03.795
11. Jifri, M.H., Hassan, E.E., Binti Wan Abd Razak, I.A., Miswan, N.H.: Time series performance for electricity load demand in Johor. IEEE Reg. 10 Ann. Int. Conf. Proceedings/TENCON (2017), 998–1003 (2017 Dec)
12. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time Series Analysis : Forecasting and Control [Internet] [cited 2019 Aug 26]. Available from: https://books.google.com.my/books?hl=en&lr=&id=rNt5CgAAQBAJ&oi=fnd&pg=PR7&dq=G.E.P&ots=DJ75AUq1Wz&sig=1K-wN4sDuGNEohthwSvuTfr4d6o&redir_esc=y#v=onepage&q=G.E.P&f=false
13. Statistics—Malaysia Energy Information Hub [Internet] [cited 2018 Apr 8]. Available from: https://meih.st.gov.my/statistics;jsessionid=4BC29F2B42BE4A1C11E4F34DBF2A4A1C?p_auth=Fl82AN7u&p_p_id=Eng_Statistic_WAR_STOASPublicPortlet&p_p_lifecycle=1&p_p_state=maximized&p_p_mode=view&p_p_col_id=column-1&p_p_col_pos=1&p_p_col_count=2&_Eng_Statistic_WA