Comparative Performance of ARIMA and DES Models in Forecasting Electricity Load Demand in Malaysia

Nor Hamizah Miswan*, Rahaini Mohd Said, Nor Hafizah Hussin, Khairum Hamzah, and Emy Zairah Ahmad
Faculty of Engineering Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal Melaka, Malaysia
*Corresponding email: norhamizah@utem.edu.my

***Abstract*— Malaysia is a developing country which is having a high level of energy demand. Load demand forecasting is essential that is also in line with increasing demand of electricity. The purpose of the current study is to compare the performance of two time series models in forecasting electricity load demand in Malaysia. Two methods are considered, which are Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) and Double Exponential Smoothing (DES). Using Mean Absolute Percentage Error (MAPE) as the forecasting performance measure, the study concludes that ARIMA is more appropriate model.**

***Index Terms*— Load forecasting, time series, ARIMA, DES**

I. INTRODUCTION

Load demand forecasting is a process of predicting the future electricity load demand. Load forecasting can be divided into three categories, which are known as Short-Term Load Forecasting (STLF), Medium Term Load Forecasting (MTLF) and Long Term Load Forecasting (LTLF). STLF focuses on forecasting one hour to one week load demand. It is important for controlling and scheduling power system. MTLF relates to a time frame from one week to one year duration while LTLF is more than one year. MTLF and LTLF are required for maintenance scheduling, hydro and fuel planning and expansion of generation and transmission activities [1].

Accurate load forecasting is important in power system planning and operation, both for regulated and deregulated electricity market [2]. Various methods have been developed to forecast the load profiling data including regression methods, time series approach, Artificial Intelligent and computational intelligent methods [3]. Time series approaches are the commonly used methods for forecasting in general [4]. This is because the behavior of load demand can be analyzed in a time series signal with seasonal periodicities, hourly and daily besides the ability to deal with non-stationary data [6].

Delson [5] used Regression-SARIMA for forecasting the daily peak electricity demand in South Africa. The method was compared to the SARIMA models and as a result, SARIMA models produce more accurate short-term forecast. Intan Azmira [2], on the other hand, used data mining techniques for electricity load forecasting where the forecast was conducted based on five-day types. The load data has been categorized by analyzing the load characteristics among the days in a week. Then, SARIMA methods were applied to forecast the data.

In this paper, forecasting electricity load demand will focus on Malaysia and it is hoped to help electricity utility companies in terms of their management planning and other related parties.

II. METHODOLOGY

A. *ARIMA Modelling*

ARIMA (Auto-Regressive Integrated Moving Average) is the most general part of time series forecasting and can be used when the time series is stationary without missing data. There is no trend for stationary series, where the variation is around the mean and also has constant amplitude. This model was introduced by George. E. P. Box and Gwilym M. Jenkins (Box-Jenkins, 1970), that consists of three parts. The autoregressive part is the lags of the stationarized series in the forecasting equation (AR), the integrated part which is the time series that needs to be differenced to be made stationary (I), and the moving average part is the lags of the forecast error (MA).

A non-seasonal ARIMA model can be represented as ARIMA (p,d,q) where AR(p) is the Autoregressive part of order p, I(d) is the differencing part of order d, and MA(q) represents the Moving Average part of order q.

Phase 1 : Model Identification

In order to identify the appropriate ARIMA model for Y, first we need to determine the order of differencing (d) to stationarize the series. Let y be the d^{th} difference of order Y, which means:

$$\text{If } d=0: \; y_t = Y_t \quad (1)$$
$$\text{If } d=1: \; y_t = Y_t - Y_{t-1} \quad (2)$$
$$\text{If } d=2: \; y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} \quad (3)$$

Note:
(1) $d=0$ means that there are no difference on Y, so the data is stationary. (2) $d=1$ means that it is a first difference of Y to made the data stationary. (3) $d=2$ means that if the first difference is not achieved, then the second order differencing is perfomed.

When the time series has been stationarized by differencing, the second step is to use the ACF and PACF plot to identify the potential model. Based on the autocorrelation function (ACF) and partial auto correlation function (PACF) plots of the difference series, we can determine the number of terms for AR(p) and MA(q) that are needed.

Using the combinations of AR and MA term, then we can detect the potential model for our forecasting. Note that the MA(q) model is represented by the ACF plot and the AR(p) model is represented by PACF plot.

Phase 2 : Estimation and Validation

From the term obtained from AR(p), I(d), and MA(q), we can try a different combination between the terms. Hence from the combination, we can check the AIC (Akaike's Information Criterion) criterion which is given by:

$$AIC = -2 \ln (\text{maximum likelihood}) + 2p \qquad (4)$$

where

p is the number of parameters involved

This value will be different for every combination. The best model for forecasting will the one that has the lowest value of AIC.

Phase 3 : Model Application

The most adequate model for ARIMA forecasting can be chosen using the forecast accuracy criteria such as Mean Square Error (MSE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), which are given by the respective equations.

$$MSE = \frac{\sum_{n}^{t} (y_t - \hat{y}_t)^2}{n}$$

$$RMSE = \sqrt{\frac{\sum_{n}^{t} (y_t - \hat{y}_t)^2}{n}}$$

$$MAPE = \sum_{t=1}^{n} \frac{|(y_t - \hat{y}_t)/y_t * 100|}{n}$$

Where y_t and $\hat{y}_t$ are the actual observe value and the predicted values, resectively. While n is the number of predicted value. Figure 1 shows the detailed process for ARIMA modelling.

B. Double Exponential Smoothing (DES)

One of the basic ideas of Exponential Smoothing is to construct the future values as weighted averages of past observations with the more recent observations carrying more weight in determining forecasts than observations in the more distant past. DES method, also known as Brown Exponential Smoothing, is used when the time series have linear trend and irregular fluctuations. The general representation of DES method is:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \qquad (5)$$

where

y_t is the forecasting model
β_0 is the time-varying mean term
β_1 is the time-varying slope term
ε_t is the random error

The most adequate model for DES forecasting can be chosen using the forecasting accuracy criteria such as Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD) and Mean Squared Deviation (MSD), which are given by the respective equations:

$$MAPE = \sum_{t=1}^{n} \frac{|(y_t - \hat{y}_t)/y_t * 100|}{n}$$

$$MAD = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|^2}{n}$$

$$MSD = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{n}$$

The lowest MAPE, MAD and MSD values will provide a better fit for DES models.

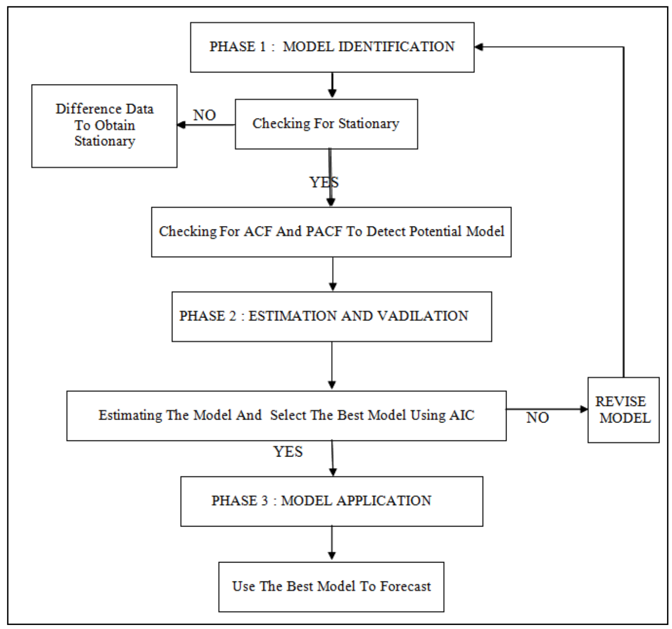

Fig. 1. Process in the ARIMA model

III. RESULT AND DISCUSSION

A. ARIMA Modelling

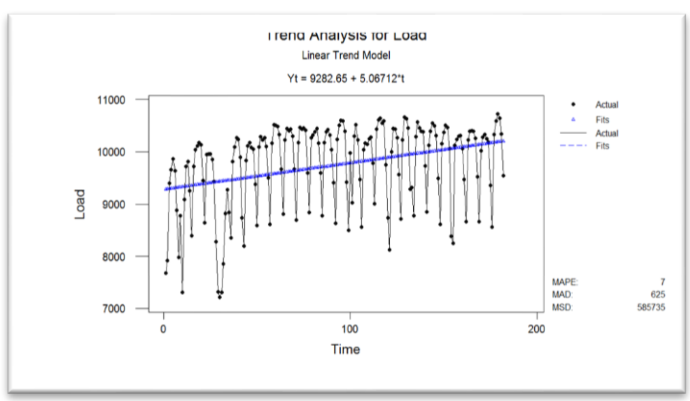

Fig. 2. Linear trend analysis of the original data

Figure 2 shows the linear trend analysis for the original electricity load demand data. The plot suggests that the original data set is not stationary. Therefore, we analyze the first difference of the series, $\Delta y = y_t - \hat{y}_t$ and we can conclude that the data is stationary.

The graph of the plot suggests an appropriate ARIMA model of the data. AR (*p*) model is represented by the PACF and the MA (*q*) model is represented by the ACF. Based on the plot and the significant spike, the following nine models have been estimated and identified using S-Plus software. The potential models are shown in Table I below.

TABLE I
THE LIST OF THE POTENTIAL ARIMA MODELS

MODEL	AIC	RMSE	MAPE
ARIMA (2,1,2)	15.81060	637.9573	5.4739
ARIMA (2,1,3)	15.76337	619.5779	5.437263
ARIMA (2,1,4)	15.75109	612.3372	5.202139
ARIMA (3,1,2)	15.77827	624.0926	5.268022
ARIMA (3,1,3)	15.74956	611.7324	5.220027
ARIMA (3,1,4)	15.68860	590.0231	5.124535
ARIMA (4,1,2)	15.73980	608.6154	5.154816
ARIMA (4,1,3)	15.74896	607.9474	5.127413
ARIMA (4,1,4)	15.68240	584.7161	5.004309

The estimated ARIMA model for forecasting the electrical load with their corresponding AIC values is given in Table 1. It is shown that the ARIMA (4,1,4) has the minimum AIC, MAPE and RMSE values. It is shown that ARIMA (4,1,4) is best model modelling and forecasting among the other ARIMA models.

B. *Double Exponential Smoothing*

By using equation 5 for DES models, we need to estimate the values of β_0 for level and β_1 for trend. The accurate combination of these two parameters will be selected as the best DES models for load demand data. The accurate combination will give the lowest values of MAPE, MAD and MSD. Table 2 shows the potential combination of β_0 and β_1 for DES methods.

From Table 2, the lowest MAPE, MAD and MSD values are from Model 1, where the values of β_0 and β_1 are 0.1. Hence, Model 1 is the best forecasting model for DES methods. The general equation for Model 1 is given by the following equation,

$$y_t = 0.1 + 0.1t + \varepsilon_t$$

TABLE II
THE LIST OF THE POTENTIAL DES MODELS

MODEL	MODEL PARAMETERS Level, β_0	Trend, β_1	MAPE	MAD	MSD
Model 1	0.1	0.1	7	648	624846
Model 2	0.1	0.2	7	665	674849
Model 3	0.1	0.5	8	743	863762
Model 4	0.2	0.1	7	671	642511
Model 5	0.2	0.2	7	695	697529
Model 6	0.2	0.5	8	745	832174
Model 7	0.5	0.1	7	681	661554
Model 8	0.5	0.2	8	715	726666
Model 9	0.5	0.5	9	831	955451
Model 10	1	1	8	767	1206869

C. *Comparative Performance for ARIMA and DES Models*

MAPE will be used as a forecast accuracy criterion in order to measure the performance of the best models from ARIMA and DES. The MAPE values are tabulated in Table III.

TABLE III
COMPARATIVE PERFORMANCE OF THE BEST FORECASTING MODELS FOR ARIMA AND DES MODELS

BEST FORECASTING MODELS	MAPE
ARIMA(4,1,4)	5.004309
DES (Model 1)	7

From Table III, the lowest MAPE values are from ARIMA (4,1,4) model. Hence, ARIMA models are the best models for modelling and forecasting electricity load demand data in Malaysia as compared to DES models.

IV. CONCLUSION

The forecasting of electricity load demand has become one of the major fields of research in recent years. This paper presents an attempt to forecast the load demand by using two different time series models, namely ARIMA and DES, and finding the appropriate model. ARIMA has been considered as the best model as compared to DES model due to the lowest MAPE value. This model can be used in forecasting the electricity load demand in Malaysia for the future.

ACKNOWLEDGEMENTS

The main author would like to acknowledge the support of the Faculty of Engineering Technology (FTK), Universiti Teknikal Malaysia Melaka (UTeM).

REFERENCES

[1] Fadhilah Abd. Razak, Mahendran Shitan, Amir H. Hashim and Izham Z. Abidin, 2009, "Load Forecasting Using Time Series Models", Jurnal Kejuruteraan 21, 53-62

[2] Intan Azmira Wan Abdul Razak, Shah Majid, Mohd Shahrieel Mohd Aras and Arfah Ahmad, 2012, "Electricity Load Forecasting using Data Mining Technique, Advances in Data Mining Knowledge Discovery and Applications", 235-254

[3] Heiko Hanh, Silja Meyer-Nieberg and Stefan Pickl, 2009, "Electric Load Forecast Methods: Tools for Decision Making", European Journal of Operational Research 199, 902-907

[4] Eisa Almeshaiei and Hassan Soltan, 2011, "A Methodology for Electric Power Load Forecasting", Alexandria Engineering Journal 50, 137-144

[5] Delson Chikobvu and Caston Sigauke, 2012, "Regression-SARIMA Modelling of Daily Peak Electricity Demand in South Africa", Journal of Energy in Southern Africa 23, 23-30

[6] J. C. Hwang and C. S. Chen, "Customer Short Term Load Forecasting by using ARIMA Transfer Function Model", Electrical Engineering, pp. 317-322

Nor Hamizah Miswan was born in Johor, Malaysia. She is a Statistics Lecturer of Mechanical Engineering Technology department. Her first Degree was Mathematical Science majoring in Statistics from Universiti Teknologi Malaysia in 2011, then completed her MSc in Mathematical Science in 2013 from Universiti Teknologi Malaysia. Her main interests of research are Applied Statistics, Time Series, Data Science, TRIZ, Stocastic Process and Reliability Analysis. She is one of the reviewers for the Applied Mathematical Sciences Journal.

Rahaini Mohd Said was born in Johor, Malaysia. She is a Statistics Lecturer of Electronic & Computer Technology department. Her first Degree was Statistics from Universiti Teknologi MARA in 2009, and then completed her MSc in Mathematical Science in 2011 from Universiti Teknologi Malaysia. Her main interests of research are Applied Statistics, Big data, Design of Experiment and TRIZ.

Khairum Hamzah was born in Kelantan, Malaysia. He is a mathematics lecturer of Manufacturing Engineering Technology Department. His first degree was Bachelor of Science with Honours Mathematics from Universiti Teknologi Mara, Shah Alam Malaysia in 2008, and then completed his MSc in Mathematics in 2011 from Universiti Sains Malaysia. His main interests of research are Applied Mathematics, Mathematics Education, Fuzzy Assessment, CAGD and Ergonomic. He is one of the reviewers for the International Journal of Education and Practice and Online Journal of Social Sciences Research.

Nor Hafizah Hussin was born in Johor Bahru. She is a Statistics Lecturer of Electrical Engineering Technology department. Her first degree was Mathematical Science majoring in Statistics from University of Malaya in 2011 and then completed her MSc Statistics in 2013 also from University of Malaya. Her main interests of research are Applied Staistics, Time Series Analysis and Markov Chain Monte Carlo (MCMC) methods.

Emy Zairah Ahmad was born in Taiping. She is an Electrical Engineering Lecturer of Electrical Engineering Technology department. Her first degree was Electrical Engineering from University of Sheffield, UK in 2010 and then completed her Msc in Global Production Engineering majoring in Solar Technology in 2013 from TU Berlin, Germany. Her main interests of research are Photovoltaic, Solar Thermal Systems, Electrical Power Systems and Energy Efficiency.

IJENS