CrossMark

ARTICLE OPEN ACCESS

Oil Demand Forecasting in Malaysia in Transportation Sector Using Artificial Neural Network

Ali Mubarak Al-Qahtani, Jebaraj S.

Department of Mechanical and Manufacturing Engineering Technology, Jubail Industrial College,
Al-Jubail, Kingdom of Saudi Arabia

Email: qahtani_am@jic.edu.sa, sjebaraj74@gmail.com

Abstract

Energy industry in Malaysia is one of critical sector that plays a major role in contributing the nation economic and industrial growth. A forecasting model is required to be developed to provide the oil demand forecast in transportation sector. This research analyses different forecasting models including Artificial Neural Network (ANN) model to predict the future oil demand in transportation sector in Malaysia. In order to select the best forecasting model, the model validation is done using the error analysis techniques. Based on the model validation result, it is found that the Artificial Neural Network (Multivariate) model gives the least error in all the error analysis techniques. The model forecast that the oil demand in transportation sector in Malaysia for the year 2020, 2025 and 2030 would be 559.44, 581.779 and 609.941 kg of oil equivalent respectively.

Keywords: ***Oil consumption; demand forecasting; transportation sector; forecasting models; Aritificial Neural Network (ANN); model validation.***

I. Introduction

Malaysia is the second-largest oil and natural gas producer in the Southeast Asia and the world second largest exporter of liquefied natural gas. The oil reserves in Malaysia are the fourth highest in Asia Pacific after China, India and Vietnam. Although Malaysia are able to produce large number of crude oil and supply sufficient energy for their own demands without relying too much on the imports, these nonrenewable resources are depleting and major plans need to be considered before energy consumption outpace the domestic production. Economic development and population growth for the past few decades have increased the demand for oil consumption in different sectors in Malaysia. One of the major sectors that are affected by this urbanization is transportation sector. In good economic condition, people are able to buy more vehicles for their own convenience due to the high purchasing power and increase in household income. From the year 2006 until year 2017, car registration in Malaysia has increased from 458,294 cars to 28,181,203 cars, which shows an increase of 98.37% of total car registration within the eleven years period. Other reason that leads to the increase in the number of vehicles used in Malaysia is due to limited good and efficient public transport systems in most of the areas. People have to rely on their vehicles to travel within either the short distances or long distances. Most of public transport such as Light Rail Transit (LRT), monorail, commuter and buses are being implemented in urban areas only while those living in rural areas have very limited access to these services. Developing and designing good energy demand in various sectors play an important role in both developed and developing countries for investors and policy makers. Underestimation of future energy demands may cause serious and dreadful consequence to the social life and economic growth of a country while overestimation would lead to inefficient and wasteful energy consumption. The purposes of this study is to develop the various forecasting models for oil consumption in transportation sector, to find the best forecasting model based on the error analysis techniques and to forecast the oil demand in transportation sector in Malaysia from the year 2020 until year 2030.

II. Literature Review

Developing and designing good energy demand in various sectors play an important role in both developed and developing countries for investors and policy makers. Jianliang et al. (2015) had investigated the four types of China's unconventional oil resources comprehensively: heavy and extra-heavy oil, oil sands, broad tight oil and kerogen oil. The results showed that OIP (Oil-in-Place) of these four types of resources amount to 19.64 Gt, 5.97 Gt, 25.74 Gt and 47.64 Gt respectively, while TRRs (technically recoverable resources) amount to 2.24 Gt, 2.26 Gt, 6.95 Gt and 11.98 Gt respectively. A new Monte-Carlo methodology to forecast the crude oil production of Norway and the U.K. based on a two-step process, namely the nonlinear extrapolation of the current/past performances of individual oil fields and a stochastic model of the frequency of future oil field discoveries had been developed by Fiévet et al. (2015). The authors had predicted remaining reserves extractable until 2030 to be 5.7 ± 0.3 billion barrels for Norway and 3.0 ± 0.3 billion barrels for the UK, which are respectively 45% and 66% above the predictions using an extrapolation of aggregate production. A combined forecast of Grey forecasting method and neural network back propagation model, which is called Grey Neural Network and Input-Output Combined Forecasting Model (GNF-IO model), was proposed. The GNF-IO model predicted coal, crude oil, natural gas, renewable and nuclear primary energy consumption volumes by Spain's 36 sub-sectors from 2010 to 2015 according to three different GDP growth scenarios (optimistic, baseline and pessimistic) (Xiuli Liu, 2016). Mass and energy-capital conservation equations had been developed by Fabio Gori (2016) to forecast the oil price evolution with accumulation or depletion of the resources. This work extends the approach of using the mass and energy-capital conservation equations to forecast the price evolution of oil when accumulation or depletion is present. An oil production forecast for China considering economic limits had been carried out by Ke Wang et al. (2016) and developed various scenarios. In the Best Estimate Scenario, China's oil production was expected to reach a maximum of 226.79 million tons in 2020.

Three different mathematical models were proposed to estimate transportation energy demand of Turkey using the artificial bee colony algorithm using gross domestic product, population and total annual vehicle-km as parameters. For transportation energy demand estimations, linear, exponential and quadratic forms of mathematical expressions were used. A 44-year-old historical data from 1970 to 2013 were utilized for the training and testing stages of the models. The authors proposed that the artificial bee colony algorithm revealed the suitability of the optimization method for transportation energy planning and policy developments in Turkey. In addition, the results from various scenarios showed that the energy demand of Turkey will be double that of 2013 by 2034 (Mustafa et al., 2017). Jim Krane (2017) had studied the implications of an increase in Saudi crude oil production capacity and he found that a combination of factors is encouraging Saudi Arabia to consider raising crude oil production capacity beyond the current ceiling of 12.5 million barrels per day. Qingfeng and Xu Sun (2017) had studied the relationship between the change in the price of oil and some of its determinants, using a structural equation model. The authors found that the demand for oil was confirmed to be inelastic to the change in oil price during the sampling period. Myung Suk Kim (2018) have formulated an autoregressive models with exogenous variables reflecting real demand, speculative demand, and supply factors for forecasting monthly global crude oil prices during their periods of decline. To enhance the forecasting ability of China's foreign oil dependence, a study combines the nonlinear metabolic grey model (NMGM) with the linear autoregressive integrated moving average model (ARIMA) had been conducted by Qiang Wang et al., 2018). This technique achieved a mean absolute error of 2.1–2.3%, reflecting its high reliability of the model. The authors used this proposed NMGM-ARIMA model to forecast China's foreign oil dependence for the period of 2017–2030 from two dimensions. Qiang Wang et al.(2018) have hybridized the nonlinear- and linear-forecasting model to a new forecasting technique in two steps to more accurately forecasting U.S. shale oil production,: (i) combining a nonlinear grey model with the mentalism idea to develop nonlinear metabolism grey model (NMGM), (ii) combining the proposed NMGM with Auto Regressive Integrated Moving Average (ARIMA) to develop NMGM-ARIMA technique.

Qiang Wang et al. (2018) have developed a forecasting model to predict the energy demand in China and India by using single-linear, hybrid-linear, and non-linear time series forecast techniques. The forecasting results showed the annual growth rate of India's energy demand from 2017 to 2026 would be 4.49%–5.21% (single-linear), 2.42%–7.04% (hybrid-linear), 0.58%–4.02% (non-linear), respectively. The annual growth rate of China's energy demand from 2017 to 2026 will be 1.36%–1.70% (single-linear), 1.04%–1.49% (hybrid-linear), 1.80%–2.34% (non-linear), respectively. Ali and Maryam (2018) have developed a hybrid model using the exponential smoothing model (ESM), autoregressive integrated moving average model (ARIMA), and nonlinear autoregressive (NAR) neural network to increase the accuracy of forecasting that accounts for problems in accurate diagnosis of linear and nonlinear patterns in economic and financial time series for oil price forecasting. Jingrui Li et al. (2018) did analysis and forecasting of the oil consumption in China based on combination models optimized by artificial intelligence algorithms. 26 combination models using traditional combination method were formulated to increase the prediction accuracy. Yishan Ding (2018) proposed a novel decompose-ensemble methodology with Akaike's information criterion (AIC) - artificial neural network (ANN) approach for crude oil forecasting. Tarek N. Atalla et al. (2018) had conducted a quantitative analysis for gasoline demand, pricing policy, and social welfare in Saudi Arabia. Stavros and George (2018) had proposed the importance of combining high frequency financial information, along with the oil market fundamentals, in order to gain incremental forecasting accuracy for oil prices. The results are both statistically and economically significant, as suggested by several robustness tests. Shaolong et al. (2018) had developed interval decomposition ensemble approach for crude oil price forecasting by integrating bivariate empirical mode decomposition, interval linear programming and interval exponential smoothing method. Niágara et al. (2018) had analyzed the role of Asymmetric Price Response (APR) and Underlying Energy Demand Trend (UEDT) in the Brazilian automotive fuel demand from June 2001 to December 2016. The demand functions of automotive gasoline, ethanol and compressed natural gas (CNG) were estimated by employing the autoregressive distributed lag (ARDL) model and Harvey's Structural Time Series Model (STSM). The importance of considering a more flexible approach incorporating both UEDT and APR was confirmed by the data.

III. Methodology

In this study, different forecasting models were developed based on the past oil consumption data in transportation sector in Malaysia from the year 1980 to 2016. The models are time series regression methods including linear model, exponential model, power model and quadratic model, double moving average method, double exponential smoothing method, triple exponential smoothing method, Auto Regressive Integrated Moving Average (ARIMA) model, Artificial Neural Network (ANN) model (Univariate and Multivariate). The schematic for developing these forecasting models is shown in figure 1. The methodologies of each of the model are discussed as follows.

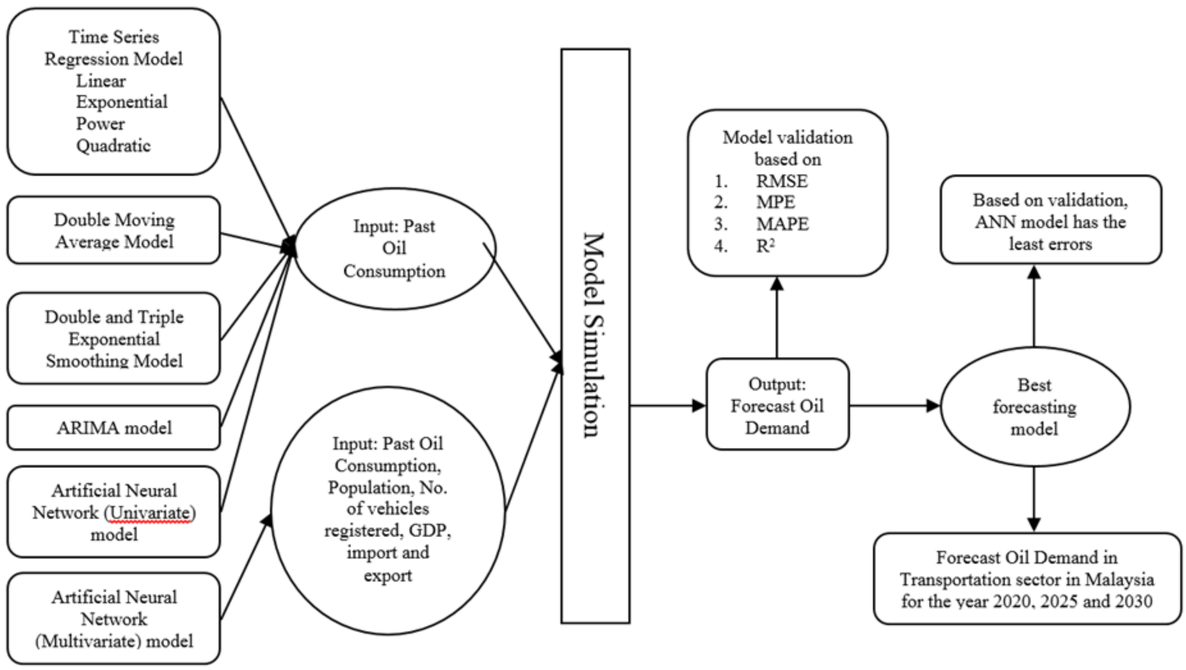

Figure 1: Schematic representation of formulation of oil demand forecasting models

Time Series Regression Methods

Time series regression method is an analysis used to find a best-fit line from the historical data by minimizing the least squared error between the forecast data and past data using the Microsoft excel software. Different types of models were used in this study to forecast the new data and steps are explained in the following section.

1. Linear Model

The linear model finds the best-fitted line using a linear equation by plotting the historical data as a single variable. The equation for linear model is in the form of,

$$y = at + b \tag{1}$$

where y is the forecast data, t is the time period and a and b are the fixed constants.

2. Exponential Model

The exponential model is used to forecast the data by predicting the data to follow a trend of increasing or decreasing at a constant growth rate. The model follows the exponential curve in form of,

$$y = e^{(a+bt)} \tag{2}$$

where y is the forecast data, t is the time period and a and b are the fixed constants

3. Power Model

Power model can be in the form of direct power model, inverse power model or quadratic model. In this study, the model used is direct power model since the data is increasing gradually over time and the equation for the model is as follows,

$$y = at^b \tag{3}$$

where y is the forecast data, t is the time period and a and b are the fixed constants.

4. Quadratic Model

Quadratic model is able to describe the path or trend of the data if the data follows a specific trend without any seasonal variation. The equation of the model is as follows,

$$y = at^2 + bt + c \tag{4}$$

Where y is the forecast data, t is the time period and a, b and c are the fixed constants.

Double Moving Average Method

Moving average method is used for smoothing the time series to find the series trend and to measure the seasonal fluctuations of the data. This model is developed by moving the arithmetic mean values of the data through the time series and applied twice in order to get the smoother time series. In this study, three year data are included in each moving average term.

$$y_{t+2} = \frac{y_t + y_{t+1} + y_{t+2}}{3} \tag{5}$$

The data are then plotted to find the best-fit linear line and the equation is in the form of:

$$y = at + b \tag{6}$$

where y is the new forecast data using double moving average method and t is the time period starting from year 0. a and b are average change in Y and y-intercept based on the best fit line.

Double Exponential Smoothing Method

In the previous method, moving average assigned weight for each previous as equally. In contrast with exponential smoothing, the weight for previous data is decreasing gradually as the time series progress in order to forecast the new data. The model for exponential smoothing is as follow,

$$y_{t+1} = \alpha X_t + (1 - \alpha) y_t \tag{7}$$

where y_t is the forecast data, X_t is the actual data, t is the time period and α is the weighting factor from 0 to 1. The same step is applied twice to get new data. The new data is then plotted to find the best-fit line based on linear model.

Triple Exponential Smoothing Method

Triple exponential smoothing method follows the same procedure as double exponential smoothing method but the step is repeated three times and the forecast equation follows the quadratic model.

Auto Regressive Integrate Moving Average (ARIMA) model

The Auto Regressive Integrate Moving Average (ARIMA) model has two main structures to forecast any set of data. The first one is non-stationary operator, the long term prediction is determined based on the different and the constant between the past data. The second structure is stationary operators, which consist of AR, and MA used to determine the short prediction value. The model predicts future value by combining the linear past value and series of error. The model was developed using SPSS, Statistic software and model validation was done in Microsoft excel.

Artificial Neural Network (ANN) model (Univariate and Multivariate)

The typical structure of an Artificial Neural Network model for univariate or multivariate has three different layers. The first layer is the input layer consist input node depending on the number of variables used in the model. The second layer is the hidden layer and the number of nodes in hidden layer depends on the user. The user can modify the number of nodes in order to find the least error for the model in the training phase. The third layer is the output layer where the values of each node in the hidden layer are summed and multiplied with their associated weights. For each of the input and hidden layer, they have their own bias value to stabilize the weightage of nodes in each layer.

In this study, the training method was done in MATLAB software and follows the feed-forward multilayer neural network with back propagation technique. The structure of the model is shown in figure 2. For univariate model, there is only one input node which is the past oil consumption in transportation sector while for the multivariate model, there are six input nodes which are the past oil consumption in transportation sector, GDP, population, number of vehicles registered, imports and exports.

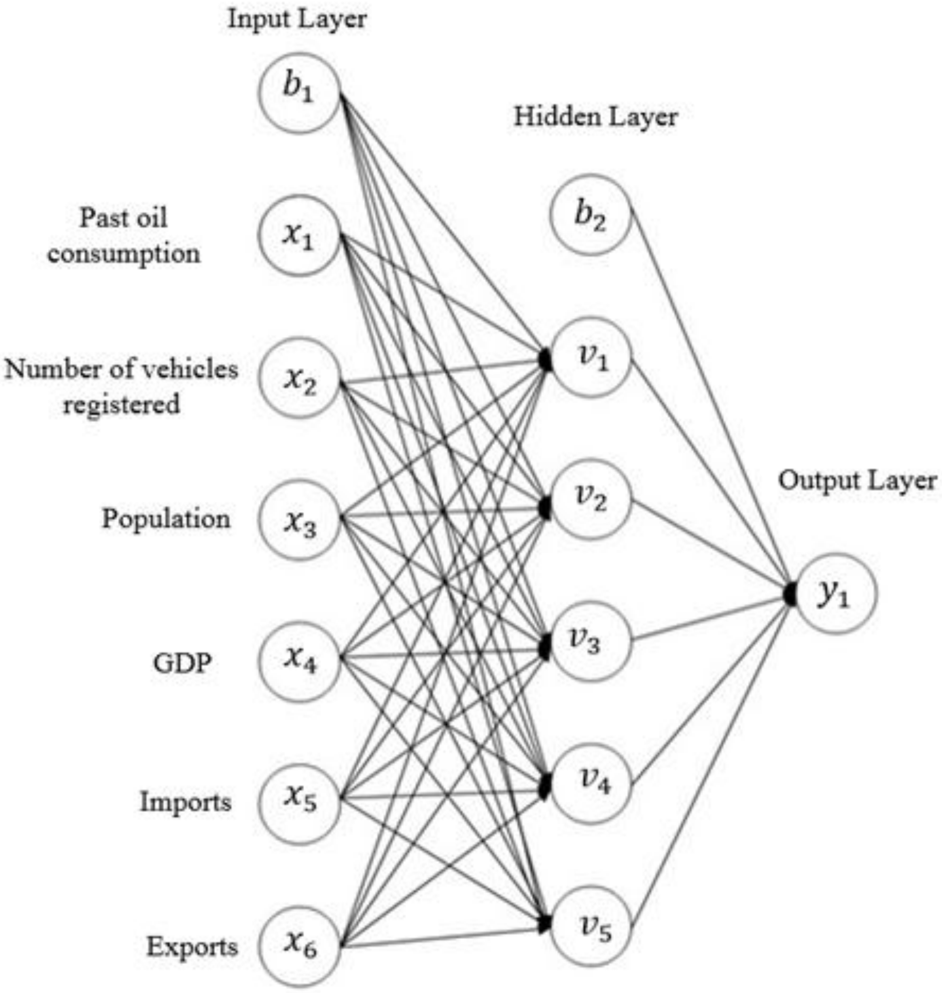

Figure 2: Structure of the multivariate Artificial Neural Network (ANN) model

In error back propagation technique, there are two different passes, which are forward pass, and backward pass. The past data is divided into two parts, for training and validation purpose. During the training process, the weights for each node remains constant in forward pass. The variable data is applied to the input nodes and its effects are propagated through the network from input layer to hidden layer and finally to output layer. The weightage for each node is then adjusted based on the error correction rule in the backward pass to minimize the errors in output layer. The training is done repeatedly to find the least error using the new weightage value for each node.

From the input to hidden layer, the activation transfer function used in the model is log sigmoid function and the model is shown as follows,

$$v_j = \frac{1}{1+\exp(-x_j)}, j = 1,2,3 \ldots n \quad (8)$$

Where v_j value of node is in the hidden network, x_j is total weighted sum of all inputs node plus the bias of neuron j. In the forward pass, the function signal at the output of neuron in each layer is computed as:

$$x_j(n) = \sum_{i=0}^{m} w1_{ij}(n)x_i(n) \quad (9)$$

where m is the total number of inputs excluding the bias inside the input layer and $w1_{ij}(n)$ is the synaptic weight connecting input nodes and hidden node and $x_i(n)$ is the input signal of input node. The index i refers to the input to the ith input terminal to the network. The activation transfer function from hidden layer to output layer is pure linear and the model is shown as below,

$$y_1 = \sum_{i=0}^{k} w2_{ij}(n)v_i(n) \quad (10)$$

where k is the total number of inputs excluding the bias inside the input layer and $w2_{ij}(n)$ is the synaptic weight connecting hidden node and output node and $v_i(n)$ is the hidden signal of hidden node. The index i refers to the input to the i_{th} input terminal to the network. The error between the observed data and network output in back propagation technique is calculated as follows:

The back propagation technique during training process calculates the error between observed data and output of network and the error signal are as follows:

$$E = 0.5(x-y)^2 \quad (11)$$

$$\delta_y = (x-y)y(1-y) \quad (12)$$

$$\delta_v = v_j(1-v_j)\delta_y w_j \quad (13)$$

where d is the observed data and y is the network output. Weights between hidden and output layers and input and hidden layer are adjusted with the following amounts:

$$\Delta w_j^t = \alpha\delta_y v_j + \beta\Delta w_j^{t-1} \quad (14)$$

$$\Delta w_i^t = \alpha\delta_v x_i + \beta\Delta w_i^{t-1} \quad (15)$$

where α is the learning rate, β is the momentum constant, and t is the iteration number of error back propagation.

IV. Model Validation

In order to validate the result of each forecasting model which have been developed in the previous section, model validation is required based on the different error analysis method. Mean Percentage Error (MPE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Correlation coefficient (R^2) between the actual data and observed data were calculated for each model. The objective of this model validation is to find best model which gives the least error for all of the error analysis method. The best model is then selected to forecast the future oil demand in transportation sector in Malaysia.

The forecast data from each model and actual data are plotted and shown in figure 3. The differences between the forecast data and actual data are then calculated to find the Mean Percentage Error (MPE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Correlation Coefficient (R^2). The model validation result is shown in table 1. Based on the results, the Artificial Neural Network (Multivariate) model shows the least error for each error analysis which are 0.440739, -0.1122, 0.153859 and 0.999995 for Mean Percentage Error (MPE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Coefficient Factor Correlation (R^2) respectively.

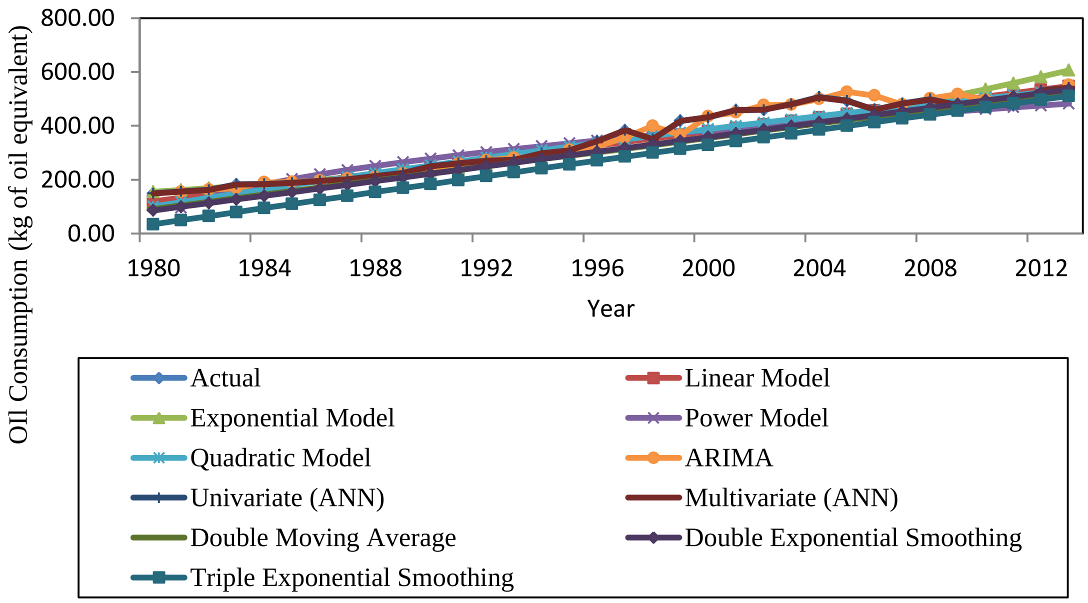

Figure 3: Validation of forecasting models for oil consumption in the transportation sector

V. Oil Demand Forecasting

The input variables used in this study to forecast the future oil demand in transportation sector are past oil consumption, population, GDP, number of vehicles registered, imports and exports. As mentioned in the previous section, different types of forecasting methods were used to predict the future oil demand in transportation sector in Malaysia.

Based on Table 1, Artificial Neural Network (Multivariate) model shows the least error for every error analysis. Thus, the model was chosen to forecast the future oil demand in transportation sector in Malaysia for the years 2020, 2025 and 2030. Figure 4 shows the forecast of oil demand using the Artificial Neural Network (Multivariate) model. The predicted oil demand in transportation sector in Malaysia for the years 2020, 2025 and 2030 are 559.444, 581.779 and 609.41 kg of oil equivalent respectively.

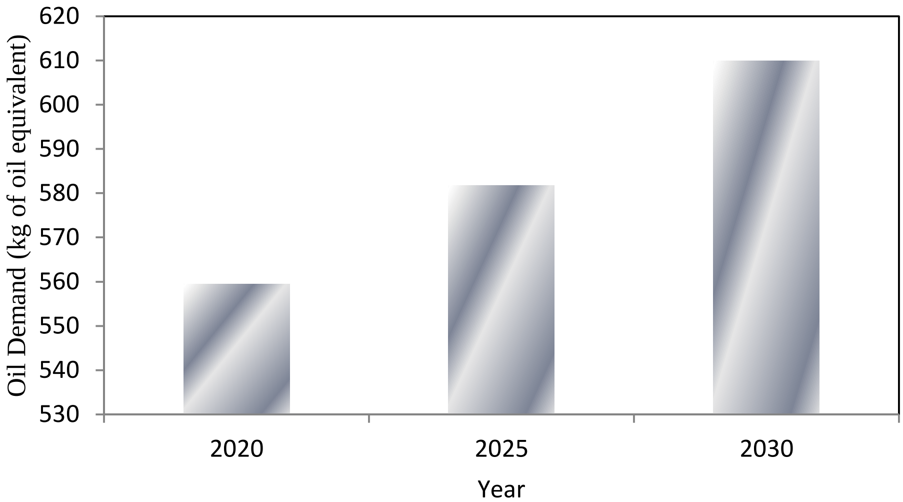

Figure 4: Forecast of oil demand in transportation sector in Malaysia

Table 1: Comparison of forecasting models for oil consumption in transportation sector in Malaysia

No	Forecasting Model	Root Mean Square Error (RMSE)	Mean Percentage Error (MPE) %	Mean Absolute Percentage Error (MAPE)	Correlation Coefficient (R^2)
1	Linear Model ($y = 12.942t + 120.37$)	31.16955	-3.76893	6.587234	0.952479
2	Exponential Model ($y = 155.69e^{0.0412t}$)	45.04872	-3.49365	8.043056	0.896549
3	Power Model ($y = 96.545x^{0.4599}$)	42.41769	-5.48693	11.3304	0.922863
4	Quadratic Model ($y = -0.0936x^2 + 16.218x + 100.71$)	30.2105	-4.43066	7.855017	0.956306
5	ARIMA Model (0,1,0)	21.55421	-2.66135	3.826108	0.974674
6	Artificial Neural Network Model (Univariate)	0.870031	0.156976	0.166166	0.99998
7	Artificial Neural Network Model (Multivariate)	**0.440739**	**-0.112**	**0.153859**	**0.999996**
9	Double Moving Average Method ($y = 13.187x + 92.023$)	46.86917	-12.411	12.77605	0.952479
10	Double Exponential Smoothing Method ($y = 13.609x + 86.041$)	46.49413	-12.658	13.24366	0.952479
11	Triple Exponential Smoothing Method ($y = -0.0213x^2 + 15.119x + 34.661$)	77.25064	-25.5798	25.66361	0.953883

VI. Conclusion

Different forecasting models were developed in this study in order to find the most suitable model to forecast the future oil demand in transportation sector in Malaysia for the years 2020, 2025 and 2030. The input variables used in this study are past oil consumption in transportation sector, number of vehicles registered, population, GDP, imports and exports. Based on the observed result from each model, model validation was done to find the best model which produces the least error for Mean Percentage Error (MPE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The Artificial Neural Network (Multivariate) model is found to show the least error and is used to forecast the future oil demand. The forecast of oil demand in transportation sector in Malaysia using Artificial Neural Network (Multivariate) model for years 2020, 2025 and 2030 are 559.444, 581.779 and 609.41 kg of oil equivalent respectively. I hope that this study will provide more insight for the policy makers and researchers to develop new policy related to energy consumption in transportation sector in Malaysia.

References

[1] Jianliang Wang, Lianyong Feng, Mohr Steve, Xu Tang and Höök Mikael (2015), China's unconventional oil: A review of its resources and outlook for long-term production, Energy, 82, 31-42.

[2] L. Fiévet, Z. Forró, P. Cauwels, D. Sornette (2015), A general improved methodology to forecasting future oil production: Application to the UK and Norway, Energy, 79, 288-297

[3] Xiuli Liu, Blanca Moreno, Ana Salomé García (2016), A grey neural network and input-output combined forecasting model. Primary energy consumption forecasts in Spanish economic sectors, Energy, 115 (1), 1042-1054.

[4] Fabio Gori (2016), Mass and energy-capital conservation equations to forecast the oil price evolution with accumulation or depletion of the resources, Energy, 116 (1), 746-760

[5] Ke Wang, Lianyong Feng, Jianliang Wang, Yi Xiong, Gail E. Tverberg (2016), An oil production forecast for China considering economic limits, Energy, 113, 586-596.

[6] Mustafa Sonmez, Ali Payıdar Akgüngör, Salih Bektaş (2017), Estimating transportation energy demand in Turkey using the artificial bee colony algorithm, Energy, 122, 301-310.

[7] Jim Krane (2017), Beyond 12.5: The implications of an increase in Saudi crude oil production capacity, Energy Policy, 110, 542-547.

[8] Qingfeng Wang, Xu Sun (2017), Crude oil price: Demand, supply, economic activity, economic policy uncertainty and wars – From the perspective of structural equation modelling (SEM), Energy, 133, 483-490.

[9] Myung Suk Kim (2018), Impacts of supply and demand factors on declining oil prices, Energy, 155, 1059-1065.

[10] Qiang Wang, Shuyu Li, Rongrong Li (2018), China's dependency on foreign oil will exceed 80% by 2030: Developing a novel NMGM-ARIMA to forecast China's foreign oil dependence from two dimensions, Energy, 163, 151-167.

[11] Qiang Wang, Xiaoxing Song, Rongrong Li (2018), A novel hybridization of nonlinear grey model and linear ARIMA residual correction for forecasting U.S. shale oil production, Energy, 165, 1320-1331.

[12] Qiang Wang, Shuyu Li, Rongrong Li (2018), Forecasting energy demand in China and India: Using single-linear, hybrid-linear, and non-linear time series forecast techniques, Energy, 161, 821-831.

[13] Ali Safari, Maryam Davallou (2018), Oil price forecasting using a hybrid model, Energy, 148, 49-58.

[14] Jingrui Li, Rui Wang, Jianzhou Wang, Yifan Li (2018), Analysis and forecasting of the oil consumption in China based on combination models optimized by artificial intelligence algorithms, Energy, 144, 243-264.

[15] Yishan Ding (2018), A novel decompose-ensemble methodology with AIC-ANN approach for crude oil forecasting, Energy, 154, 328-336.

[16] Tarek N. Atalla, Anwar A. Gasim, Lester C. Hunt (2018), Gasoline demand, pricing policy, and social welfare in Saudi Arabia: A quantitative analysis, Energy Policy, 114, 123-133.

[17] Stavros Degiannakis, George Filis (2018), Forecasting oil prices: High-frequency financial data are indeed useful, Energy Economics, 76, 388-402.

[18] Shaolong Sun, Yuying Sun, Shouyang Wang, Yunjie Wei (2018), Interval decomposition ensemble approach for crude oil price forecasting, Energy Economics, 76, 274-287.

[19] Niágara Rodrigues, Luciano Losekann, Getulio Silveira Filho (2018), Demand of automotive fuels in Brazil: Underlying energy demand trend and asymmetric price response, Energy Economics, 74, 644-655.