INTRODUCTION

Currently, with the instability of world’s economy, the needs and consumption of crude oil and its supply and demand are progressively becoming high, thus making it important to examine as a whole whether the amount of crude oil supply and demand is enough to establish a continuous, stable economic growth.

In assessing the quantitative analysis of crude oil supply and demand, some complications may emerge in the crude oil demand forecasting. The main objective of this study was to forecast crude oil and condensates demand in Malaysia. All relevant historical data for crude oil and condensates were collected to predict crude oil and condensates demand. The prediction of crude oil depends on the crude oil price where it can adequately guarantee a sudden new supply and demand of crude oil with an increase in the production quality at a lower production cost (Yu et al., 2016).

Crude oil can be defined as a mixture of hydrocarbons that lies in liquid state in natural underground reservoirs and remains liquid at atmospheric pressure after passing through surface separating facilities. A condensate is a very light hydrocarbon with an American Petroleum Institute (API) specific gravity of greater than 50 degrees and less than 80 degrees (Mohammadpoor and Torabi, 2015). Underground, condensates can either exist separately from the crude oil or dissolve in the crude oil. In this paper, we examined the summation of crude oil and condensates data in achieving the forecast.

In the literature, forecasting methods for oil demand and supply and oil price can be categorized into two, which are the traditional statistical method (Chakra et al., 2013; Sheremetov et al., 2013; Chuanping et al., 2010) and a hybridization of the “decomposition and ensemble” method (Li et al., 2012). Traditional statistical models such as ARIMA, Grey Model, Support Vector Machine (SVM), and Artificial Neural Network (ANN) have been extensively implemented in predicting crude oil price and demand. For example, Chakra et al. (2013) employed the higher-order neural network (HONN) model to forecast the production of oil on a sandstone reservoir located in Cambay basin in Gujarat, India. Sheremetov et al. (2013) applied the Time Delay Neural Network (TDNN) and Nonlinear Autoregressive with exogenous input (NARX) methods in predicting monthly Mexican oil and gas production. Aizenberg et al. (2016) also suggested Multilayer Neural Network with Multi-Value Neurons (MLMVN) model to forecast oil production in the Gulf of Mexico. Zang et al. (2010) investigated refined oil demand in China using combined mid-long term forecast in the Grey model. Chuanping et al. (2011) also implemented ARMA method in forecasting China’s oil self-sufficiency. Seyedan et al. (2015) applied Support Vector Machine method in predicting Iran’s oil consumption.

As for the decomposition and ensemble method, most previous literature implemented the hybridization of decomposition and ensemble methods for forecasting oil prices data. For example, Yu et al. (2016) proposed the hybridization of Ensemble Empirical Mode Decomposition (EEMD) and Extended Extreme Learning Machine (EELM) to forecast crude oil spot price data in West Texas Intermediate (WTI). Yu et al. (2008) also suggested forecasting WTI crude oil spot prices and Brent crude oil spot prices using EMD and neural network ensemble learning paradigm. He et al. (2012) analyzed WTI and Brent crude oil prices using a modified hybridization of the wavelet...
decomposed ensemble model. Yu et al. (2015) also used a modified hybridization of EEMD with data characteristic driven reconstruction technique for WTI and Brent oil prices data.

In general, the above models can generate satisfying forecast results if the forecaster follows the whole forecasting process introduced by the inventor of the model. Basically, the forecasting process in both traditional statistical models and hybrid models of decomposition and ensemble method involves many trial-and-error steps, which can be time-consuming. In addition, abundant studies have proved that the forecasting capability might be low if the forecaster does not follow any of the steps suggested by the inventor of the model. Thus, continuous application of traditional statistical models or hybridization of decomposition and ensemble methods in the forecasting procedure might be impractical.

In order to overcome these traditional statistical and hybridization of decomposition and ensemble procedure of a model, especially through trial-and-error process with too many steps to follow in a forecasting method, we are suggesting the Fast Ensemble Empirical Mode Decomposition (FEEMD) model in predicting the time series data. The steps involving in FEEMD model for time series forecasting procedure is very minimal, easy to implement, and most importantly, give very good forecast accuracy.

The rest of the paper is organized as follows. In the next section, we present the ARIMA, ARIMAX, GARCH and FEEMD methods for forecasting in this study. Empirical results after employing the above methods to real data sets of crude oil and condensates are presented in Section 3 where the forecast performances of these methods are compared among each other. Section 4 contains the concluding remarks.

**METHODOLOGY**

In our work, we used traditional statistical methods such as ARIMA, ARIMAX, and GARCH to compare their forecast performances against the FEEMD method on the crude oil and condensates data. Tapis-blend oil prices data was chosen as an exogenous variable for the ARIMAX model to determine the robustness of the proposed FEEMD method in forecasting and to assess the quality of the proposed forecast methods. The forecast performance was measured using the Mean Average Percentage Error (MAPE) method.

**ARIMA model**

The general ARIMA (p, d, q) model consisted of a combination of past values and past errors, or it can be expressed as:

\[
\phi_p(B)(1-B)^dY_t = \theta_q(B)\alpha_t, \tag{1}
\]

with

\[
\phi_p(B) = (1 - \phi_1B - \cdots - \phi_pB^p), \quad \theta_q(B) = (1 - \theta_1B - \cdots - \theta_qB^q)
\]

where B denotes the backshift operator and \( \alpha_t \) denotes a random process in a period \( t \) with zero mean and constant variance. \( \phi_p(B) \) and \( \theta_q(B) \) are regular autoregressive and moving average polynomials of orders p and q. \( Y_t \) is appropriately transformed data series in period \( t \), while \( (1-B)^d \) is the non-seasonal operators. If \( d \) is non-zero, then there is a simple differencing step to remove trend (Box et al., 2008).

**ARIMAX model**

The ARIMAX model is an extension of the ARIMA model, where ARIMAX includes exogenous variables (Pankratz, 1991). The general ARIMAX model can be expressed as:

\[
y_t = \frac{\beta}{\phi(B)}x_t + \frac{\theta(B)}{\phi(B)}Z_t \tag{2}
\]

where \( x_t \) is a covariate at time \( t \) and \( \beta \) is its coefficient. \( \phi_p(B) = (1 - \phi_1B - \cdots - \phi_pB^p) \) and \( \theta_q(B) = (1 - \theta_1B - \cdots - \theta_qB^q) \).

**GARCH model**

The GARCH model can be expressed as follows:

\[
\ln \left( \frac{C_t}{\epsilon_t} \right) = \mu + \epsilon_t \tag{3}
\]

\[
v_t = \omega + \beta v_{t-1} + u_{t-1} \tag{4}
\]

where \( C_t \) is the time-\( t \) of the “crude oil”, \( v_t = \sigma_t^2 \) is the variance, and \( \epsilon_t = \sigma_t z_t \) is the innovation for date \( t \). Here, \( z_t \) is either a standard normal variable or the empirical random variable with a mean of zero and a variance of one. For GARCH(1,1), \( u_t = \alpha \epsilon_t^2 \) (Bollerslev, 1986).

**Fast ensemble empirical mode decomposition (FEEMD) model**

**Empirical mode decomposition (EMD)**

In the theory of EMD, any series \( x(t) \) can be decomposed using the following equation (Huang et al, 1998):

\[
x(t) = \sum_{i=1}^{n} IMF_i(t) + R_n(t) \tag{5}
\]

where \( \{IMF_i(t)\}, (i = 1,2,\ldots,n) \) are the Intrinsic Mode Function (IMF) and \( \{R_n(t)\} \) is the residue. Basically, the computational steps of EMD are as follows:

Step 1: Identify all the local extremes of the raw series \( \{x(t)\} \).
Step 2: Connect all the local maxima and the local minima by two cubic splines to generate the corresponding upper envelop series \( \{x_u(t)\} \) and down envelop series \( \{x_d(t)\} \), respectively.
Step 3: Calculate the mean envelop series \( \{x_{mean}(t)\} \) as:

\[
x_{mean}(t) = \frac{x_u(t)+x_d(t)}{2} \tag{6}
\]

Step 4: Execute the extracting computation as:

\[
Z(t) = x(t) - x_{mean}(t) \tag{7}
\]

Step 5: Examine whether the extracted series \( \{Z(t)\} \) is an IMF component; if yes, set \( IMF(t) = Z(t) \) and replace the series \( \{x(t)\} \) with the residual \( r(t) = x(t) - IMF(t) \); if no, replace the series \( \{x(t)\} \) with the extracted series \( \{Z(t)\} \) and repeat Step 2 to Step 4 until the following terminating threshold is reached:

\[
\zeta_{terminate} = \sum_{j=1}^{m} \frac{|Z_{j+1} - Z_{j}(t)|}{|Z_{j}(t)|} \leq \delta \tag{8}
\]

where \( m \) is the length of the data in the original series \( \{x(t)\} \), \( \delta \) is the terminating threshold, and \( j \) is the number of the iterative computation. Basically, the terminating threshold value falls into the range of \([0.2, 0.3]\).

Step 6: The procedure in Step 1 to Step 5 are repeated until all the IMFs have been obtained.

**Ensemble empirical mode decomposition (EEMD)**

The EEMD algorithm added different realizations of white noise series to the original series \( x(t) \), then adopted the standard EMD procedures to obtain the IMFs (Wu and Huang, 2009). The computational steps of the EEMD are given as follows:

Step 1: Calculate \( x^n(t) = x(t) + \omega^n(t) \), where \( \{\omega^n(t), (k = 1,2,3,\ldots,N)\} \) are the different realizations of white Gaussian noise series, and \( N \) is the number of times to add the noise series inside the original series \( x(t) \).

Step 2: Decompose the series \( \{x^n(t), (k = 1,2,3,\ldots,N)\} \) by executing the standard EMD to obtain their corresponding IMF modes \( \{IMF_k^n(t)\}, (m = 1,2,3,\ldots,M) \), where \( M \) is the number of decomposed IMF modes in the \( k \) round.
Step 3: Compute the average of the corresponding series \( \{IMF_{m}^{p}(t)\} \) as:

\[
\overline{IMF_{m}^{p}(t)} = \frac{1}{N} \sum_{k=1}^{N} IMF_{m}^{k}(t)
\]  

(9)

Step 4: Repeat the above averaging procedure to complete the whole EEMD decomposition. The original signal series \( x(t) \) will be formatted as below:

\[
x(t) = \sum_{m=1}^{P} \overline{IMF_{m}^{p}(t)} + r_{p}(t)
\]  

(10)

where \( \{IMF_{m}^{p}(t)\}, (m = 1, 2, \ldots, p) \) are the EEMD decomposed IMFs, \( r_{p}(t) \) is the corresponding residue, and \( p \) is the number of the final EEMD-based IMFs.

This paper proposed forecasting via the FEEMD method (Liu et al., 2015). The steps to achieve the FEEMD method are as follows:

Step 1: Follow the steps in building the EEMD algorithm. Set the final number of IMF, which is \( p = 1; \) hence during EEMD sifting process, it will give 2 IMFs, which are the first IMF and a residue (IMF2).

Step 2: Compare the MAPE of the residue (IMF2) with the actual data.

Step 3: Forecast the residue (IMF2).

Mean average percentage error (MAPE)

To investigate the performance of the proposed model, MAPE was used as the error performance criterion. The equation for calculating MAPE is as follows:

\[
MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{\hat{x}_{t} - x_{t}}{x_{t}} \right| \times 100
\]  

(11)

where \( x_{t} \) is the actual data, \( \hat{x}_{t} \) is the forecasted data, and \( N \) is the number of samples in the data series.

CRUDE OIL AND CONDENSATES DATA SERIES

Fig. 1 shows the monthly time series plot for crude oil and condensates and the Tapis-blend oil prices from January 2002 to February 2015. The time series plot for crude oil and condensates showed numerous changing turning points, volatility, and fluctuations in the series. It also showed spikes (red oval) during March 2006 and March 2011. The crude oil and condensates demand showed a decreasing pattern after several months and continued decreasing throughout the years. As for the Tapis oil price, it was stable in the beginning and suddenly decreased after a promising increment in July 2008. However, the prices recovered from the decrement after a few months and gradually became stable at the beginning of the year.

FORECASTING RESULTS

For the forecasting process, we implemented ARIMA, ARIMAX, GARCH, and FEEMD procedures in modelling and forecasting crude oil and condensates data series.

ARIMA, ARIMAX, and GARCH modelling

For ARIMA and ARIMAX models, we differenced the crude oil and condensates and Tapis-blend oil prices \((d = 1)\) to make the data stationary and then the ARIMA model was built using Minitab software and ARIMAX model using SAS software. Fig. 2 shows the ACF and PACF for crude oil and condensates and Tapis-blend oil prices after the differencing process while Fig. 3 shows the stationary series.

The model that gave the best MAPE for the crude oil and condensates data series was ARIMA (1,1,0) while the best model chosen for covariates exogenous variable was ARIMAX (1,2,1).

To compute the GARCH model, we used Microsoft Excel 2010 and following GARCH methodology, the values for \( \alpha, \beta, \) and \( \gamma \) were generated directly using Excel 2010 after we maximized the likelihood function. The final value for \( \alpha = 0, \beta = 0.0688, \) and \( \gamma = 0.0004. \)

FEEMD modelling

After we implemented FEEMD algorithm using Matlab, with 0.3 as the standard deviation for the stopping criterion, Fig. 4 depicts the first and second IMFs for crude oil and condensates data series.

The second IMF was compared with the actual data to forecast the crude oil and condensates. For the first IMF, it showed a very high frequency while the second IMF exhibited a smooth trend of the crude oil and condensates data. Various researchers found that the frequency of the first IMF may not be quite useful because its high peak may give low forecast accuracy (Liu et al., 2015). Therefore, in this study, we eliminated the first IMF and only used the second IMF in the forecasting process using FEEMD. We also chose only one iterating process for IMF in order to achieve the simplest forecasting steps and save time for the forecasting process, while still giving highly precise forecasts.

Forecast results

Table 1 shows the MAPE for crude oil and condensates data series using ARIMA, ARIMAX, GARCH, and FEEMD models. The data was divided into two, which are in-sample forecast (152 data points) and out-sample forecast (18 data points).

Table 1 In-sample and out-sample forecasts for crude oil and condensates data series using ARIMA, ARIMAX, GARCH, and FEEMD models.

<table>
<thead>
<tr>
<th>Methods</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In-sample Forecast</td>
</tr>
<tr>
<td>ARIMA</td>
<td>2.8298</td>
</tr>
<tr>
<td>ARIMAX</td>
<td>3.0289</td>
</tr>
<tr>
<td>GARCH</td>
<td>3.0361</td>
</tr>
<tr>
<td>FEEMD</td>
<td>1.4987</td>
</tr>
</tbody>
</table>

Fig. 5 and Fig. 6 demonstrate the time series plot for in-sample and out-sample forecasts of the crude oil and condensates data using ARIMA, ARIMAX, GARCH, and FEEMD models.

From Table 1, Fig. 5, and Fig. 6, FEEMD gave the highest MAPE compared to the other methods. The least precise in-sample forecast of the crude oil and condensates data series was by the GARCH model, followed by ARIMAX model, then ARIMA model. For the out-sample forecast, the least precise model was the ARIMA model, followed by ARIMAX model and then GARCH model. For the in-sample forecast, GARCH model gave the lowest MAPE because the spikes and volatility of the in-sample crude oil and condensates data series were deviated the largest from the actual data, thus affecting the GARCH model’s performance. In comparing between the ARIMA and ARIMAX models for in-sample forecast, ARIMA gave better MAPE than the ARIMAX model. This is because the ARIMAX forecasting model was based on the exogenous variable that can fill in the spikes in the series. Based on the Tapis-blend oil prices data series, it showed a slightly random data pattern with not too many spikes in the Tapis-blend oil prices data series, hence giving low forecast accuracy.

For the FEEMD model, it gave the highest MAPE because the second IMF smoothed the crude oil and condensates data series. However, for out-sample forecast of crude oil condensates using FEEMD (see Fig. 6), the first eleven observations did follow the actual data series but the 12th to 18th observations of the second IMF just became smooth and did not follow the spikes of the data.
Fig. 1 Time series plot for crude oil and condensates (‘000bpd) and Tapis blend oil prices (USD/b) from January 2002 to February 2015.

Fig. 2 ACF and PACF for crude oil and condensates and Tapis-blend oil prices after $d = 1$. 
Fig. 3 Stationary series of crude oil and condensates and Tapis-blend oil prices after $d = 1$

Fig. 4 First and second IMFs for crude oil and condensates data series.
Fig. 5 Crude oil and condensates in-sample forecast with ARIMA, ARIMAX, GARCH, and FEEMD models.

Fig. 6 Crude oil and condensates out-sample forecast with GARCH, and FEEMD models.
CONCLUSIONS

This paper has presented a novel forecasting method of crude oil and condensates based on the FEEMD method. The proposed decomposed and ensemble FEEMD method was compared with traditional statistical methods of ARIMA, ARIMAX and GARCH. Through FEEMD, crude oil and condensates data series were decomposed into two IMFs: the first IMF was the high-frequency crude oil and condensates data and the second IMF was the smooth trend of the crude oil and condensates data series. The second IMF of the crude oil and condensates data series was compared with the actual data. The MAPE of the proposed FEEMD model gave more precise forecast accuracy as compared to the other methods while reducing the trial-and-error process in forecasting that is common in other traditional statistical models, thus saving a lot of time.

ACKNOWLEDGEMENT

This work was financially supported MyBrain15 scholarship and UTM Tier 1 Grant (11H89).

REFERENCES


