A Review of Crude Oil Prices Forecasting using Hybrid Method

Nurull Qurraisya Nadiyya Md-Khaira[1], Ruhaidah Samsudina[2], Ani Shabrib[3]

[1,2]Department of Software Engineering, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[3]Department of Science Mathematic, Faculty of Science, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

Article Info	ABSTRACT
	Crude oil is considered as a crucial energy source in modern days. Consequently, the fluctuation of crude oil prices can cause a significant impact on economic activities. Researchers have proposed many hybrid forecasting models on top of single forecasting methods which are utilized to predict crude oil prices movement more accurately. Nevertheless, many limitations still existed in hybrid forecasting models and models that can predict crude oil prices as accurate as possible is required. The motivations of this review paper are to identify and assess the mostly used crude oil prices forecasting methods and to analyse their current limitations. 12 studies that used "decomposition-and-ensemble" framework was selected for review. Wavelet transform is identified as the mostly used data decomposition method while some limitations have been recognized. Future researches should include more studies to further elucidate the limitations in existing forecasting method so that subsequent forecasting methods can be improved.

Corresponding Author:

Nurull Qurraisya Nadiyya Md-Khaira,
Department of Software Engineering, Faculty of Computing,
Universiti Teknologi Malaysia, 81310, Malaysia.
Email: shashalala92@gmail.com

1. INTRODUCTION

Crude oil is one of the important energy source and plays a crucial rule in the world economy. Diesel, gasoline, heating oil, lubricant and other forms of petrochemicals a some of the end products produced using crude oil and these products are indispensable for daily uses. Economic progress, social balance and national security can be affected significantly by crude oil prices changes [1]. For instance, rise in crude oil prices will surely increase the gasoline prices which in return affect the fundamental goods and services needed by the citizen. Just like other commodities, supply and demand basically determines the fluctuation of crude oil prices. Other factors like natural disaster, political event and military conflict also affect the changes in crude oil prices [2]. For example, the incident of 2005 hurricane season has led to the closure of oil and natural gas production as well as refineries. Consequently, prices for petroleum-based products rose substantially as market supplies deteriorate. All the issues mentioned earlier illustrate the importance of crude oil prices fluctuation to us. For that reason, awareness of crude oil prices fluctuation is very crucial and one way of doing so is to utilize time series forecasting methods that are proposed and proven by many studies.

Time series forecasting can assist practitioners in predicting forthcoming movement thus providing advantage for future planning. An example of time series forecasting method is Autoregressive Integrated Moving Average (ARIMA) method. This single forecasting method used by Yusof et al [3] in their study to predict the crude oil production in Malaysia. Hybrid forecasting methods have also been proposed by researchers to improve forecasting accuracy. Hybrid forecasting models are proven to be more powerful

compared to single models because the combination focuses on the strengths of these single models while eliminating their original limitations [4]. Recently, a novel "decomposition-and-ensemble" framework for hybrid forecasting has proven its efficacy in time series forecasting [5], which breaks down an intricate time series into simpler components, forecasts them individually and ultimately ensembles them into final result. One of the examples is the crude oil prices forecasting method proposed by Shabri et al [6] which is based on hybridizing wavelet and artificial neural network (ANN) model. Their study concluded that the hybrid method produced a huge forecasting improvement as compared to the ANN model alone. Furthermore, Wang et al [7] also conducted a study for crude oil prices forecasting using ARIMA and Back Propagation Neural Network Combinatorial Algorithm. Their result showed that highest forecasting accuracy is achieved by the combinatorial algorithm rather than to directly utilize both methods individually.

Even though many hybrid forecasting methods have shown great improvement in crude oil prices forecasting, many limitations can still be observed and a method that can predict the COSP as accurate as possible is very necessary. A categorization between the existing hybrid forecasting methods for crude oil prices can give us an indication of how a better method can be proposed by observing which performs better than the other one. Apart from that, the limitations of each categorization must also be analyzed so that future researches can comprehend the limitations thus improving new hybrid forecasting methods. Therefore, the motivations of this review paper are to identify and categorize the existing methods in crude oil prices hybrid forecasting model and to analyse the limitations in each category. Keep in mind that this review only focuses on the hybrid models that make use of the "decomposition-and-ensemble" framework because it has gained a lot of attention lately and has been implemented in many recent studies. The structure of the paper is as follow. Section 2 includes a brief explanation about the forecasting methods that are incorporated into hybrid models of our selected studies. The review for the existing hybrid forecasting models which includes the categorization details and the limitations identified for each category is presented in Section 3. Lastly, Section 4 concludes the review with some recommended future investigation.

2. HYBRID FORECASTING METHODS

As stated earlier, the initial step in "decomposition-and-ensemble" framework is getting a lot of attention in hybrid forecasting model. The initial step in this framework is usually called the data decomposition. Wavelet transform is one of the popular data decomposition methods that are incorporated in a hybrid forecasting model. In general, wavelets mean small waves [8]. They have a distinguished amount of fluctuations thus they can be exploited to resemble variables in time or space. According to Schlüter et al. [9], wavelet transform acts as a data decomposition method that breaks an original data series down into a linear combination of distinct frequencies. Therefore, it can localize and pinpoint the diversity of different frequencies in a time series [10]. Figure 1 and Figure 2 which are extracted from Md-Khair et al, [11] shows the graph of monthly crude oil prices series from WTI and the decomposed components series using wavelet transform respectively. Notice that in Figure 2, the smooth series represents the original prices series but more stable in term of volatility because the original fluctuations are captured by the detail series. Another most commonly used data decomposition technique is empirical mode decomposition (EMD), firstly proposed by Huang et al, [12]. Yu et al, [13] explained that the fundamental concept of EMD is to decompose a time series into several oscillatory functions which is called intrinsic mode functions (IMFs). In addition, its decomposition is based on the local characteristic time scale of the data series thus it is very efficient. Nevertheless, the main flaw in this method is the mode mixing problem which can cause the IMFs to be weak in the physical meaning [14], [15]. This leads to another data decomposition method which is called ensemble EMD (EEMD), introduced by Wu et al. [16] to overcome the weakness. Unlike wavelet transform, EEMD does not requires basis function for decomposition and only expect two parameters which are the number of ensemble and the standard deviation of Gaussian white noise [5].
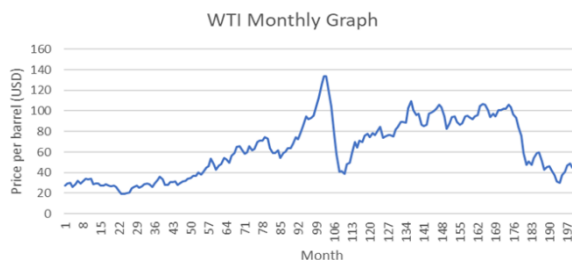

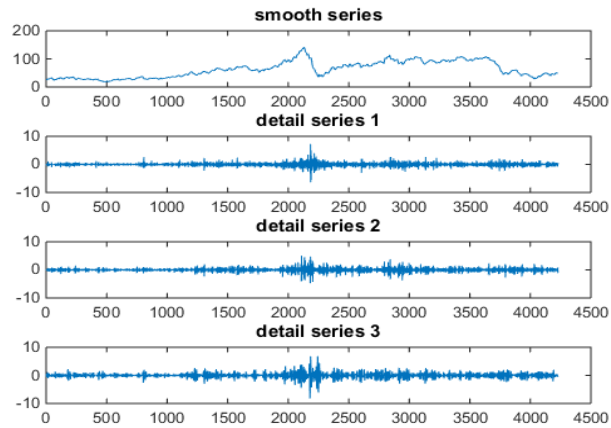Figure 1. Monthly dataset for WTI COSP

Figure 2. Decomposed wavelets for WTI dataset

After data decomposition step is done, each individual component is forecasted using suitable forecasting method. There exist two main types of forecasting methods which are used by researchers which are statistical method and artificial intelligence (AI) method. Statistical methods are broadly used because the comprehension and execution is pretty simple with the ability to be analyzed in great detail [17]. One example of statistical method is autoregressive integrated moving average (ARIMA). It is formed based on several single models which are autoregressive (AR), moving average (MA) and the consolidation of AR and MA named ARMA model [18]. ARMA model is utilized in forecasting stationary data series. In cases where the data series is non-stationary, differentiation needs to be done to the data series therefore forming ARIMA introducing I in the model where differentiation is done. A popular ARIMA modelling introduced by Box and Jenkins consists of model identification, parameter estimation and model validation [19]. Figure 3, taken from Md-Khair et al. [11] presents a typical flowchart of Box-Jenkins ARIMA model.
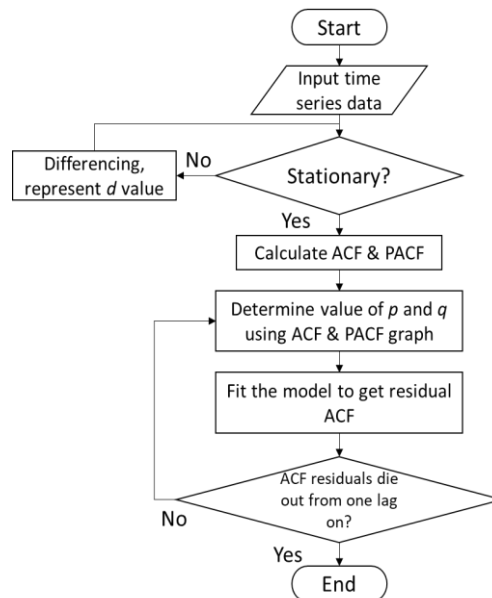

Figure 3. Flowchart of Box-Jenkins ARIMA model

Unfortunately, statistical forecasting methods relies on the assumption that the pattern which exists in the past will hold true for the future and linear in nature. This means that they can only provide good forecasting results for linear and near linear data series and are not suitable for non-stationary of non-linear data series such as the crude oil prices [6]. Therefore, AI forecasting methods are introduced. AI methods

imitate the human mind in problem solving thus they can be exploited to find approximate answers for real-world problems which contain inaccuracies and uncertainties. An example of AI method is artificial neural network (ANN). ANN. According to Zhang et al. [17], ANN model was introduced in time series forecasting from 1964 and has been widely used in that particular field since then. Basically, ANN is a mathematical design which has a highly connected structure [6]. The structure consists of input layer where data are presented to the network, hidden layer where the processing is done and the output layer where the results are produced. Nonetheless, ANN has these flaws where it suffers from local minima and overfitting. Furthermore, it is challenging to determine the network structure because it can only be established using trial-and-error approach [20].

In addition to ANN, least square support vector machine (LSSVM) is another example of commonly used AI forecasting methods. LSSVM is based on support vector machine (SVM) that used in classification and nonlinear function estimation. LSSVM is proposed by Suykens et al. [21] to solve the major drawback of SVM higher computational burden for the constrained optimization programming [22]. LSSVM only needs to solve a set of linear equations, rather than quadratic programming which is much easier and computationally simpler. Furthermore, LSSVM method utilizes equality constraints instead of inequality constraints and adopts the least square linear system as its loss function, making it computationally attractive, good in convergence and high in precision. Figure 4, taken from Md-Khair et al. [11] shows a typical design for LSSVM model development.
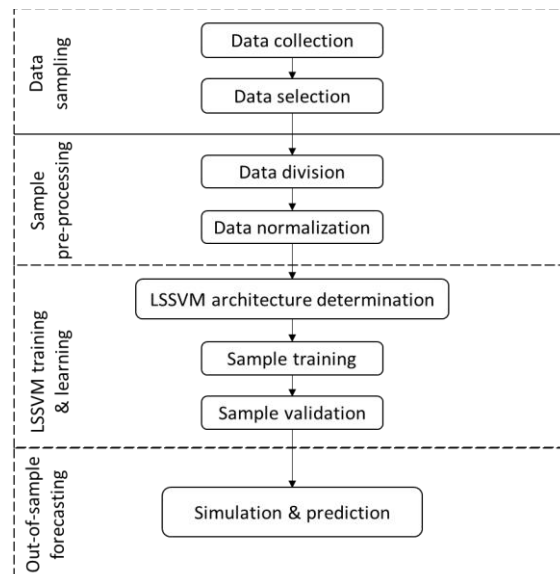

Figure 4. LSSVM model development

3. REVIEW ON HYBRID FORECASTING METHODS

A total of 12 studies that proposed hybrid methods in crude oil prices forecasting have been included in this review. In addition, only methods that follow the "decomposition-and-ensemble" framework was selected because this framework is proven to increase forecasting accuracy significantly. Moreover, this framework has gained a lot of attention when it has been implemented in many recent hybrid forecasting studies which clearly proved its efficacy. To summarize the selected study, they are tabulated in Table 1 which shows the associated study column, approach column that consist of data decomposition and forecasting method and lastly the grade of crude oil prices used column.

As can be observed from Table 1, some approaches contain more than one method in the forecasting method column. Even though all of them follow the "decomposition-and-ensemble" framework, we identified that all the approaches can be clustered into a more specific categorization. From our investigation, two categories were introduced which are (1) hybrid model with single forecasting method, (2) hybrid model with multiple forecasting method and (3) hybrid model with single or multiple forecasting method(s) and optimization method. Each category is explained in more details in the following sub section.

Table 1. Description of Selected Studies

Associated Study	Approach Data Decomposition	Forecasting Method	Crude Oil Prices Grade
[23]	EEMD	LSSVM – PSO (forecast nonlinear component) GARCH (forecast time-varying component)	WTI (daily, weekly, monthly)
[24]	Wavelet Transform	RBF neural network PCA (reduce number of input variables)	Brent (weekly)
[25]	Wavelet Transform	PSO (determine optimal parameters of MLR) MLR (model oil prices)	WTI (daily)
[6]	Wavelet Transform	ANN	WTI & Brent (daily)
[13]	EMD	FNN (model each extracted components) ALNN (aggregate all components)	WTI & Brent (daily)
[26]	Wavelet Transform	LSSVM	WTI & Brent (monthly)
[27]	Wavelet Transform	Simple Averaging Ensemble	WTI & Brent (daily)
[5]	EEMD	APSO (Optimize parameters in RVM) RVM (forecast each component)	WTI (daily)
[28]	EMD	SBM (restrain the end effect occurs during sifting process of EMD FNN (model each extracted components)	WTI (weekly)
[29]	Wavelet Transform	SVM	WTI (daily)
[30]	Wavelet Transform	LSSVM	WTI & Brent (monthly)
[11]	Wavelet Transform	ARIMA (forecast smooth series) LSSVM (forecast detail series)	WTI & Brent (monthly)

3.1. Hybrid Model with Single Forecasting Method

In this category, an approach consists of a data decomposition method and only one forecasting method. An example is the study conducted by Shabri et al. [6] which utilized wavelet transform as data decomposition and ANN as the forecasting method. In this approach, wavelet transform decomposes the crude oil prices into sub components which are then become as the inputs to the ANN model to forecast the prices series. The result obtained in their study shown that their proposed approach performed better than the single method ANN, ARIMA and GARCH. Other studies also proven that this category of hybrid model is more effective than implementing the single method itself without data decomposition [26], [29], [30].

3.2. Hybrid Model with Multiple Forecasting Methods

This category does not introduce much different from the first category. The only thing that is different is that in this category, more than one forecasting methods are incorporated into a hybrid model. According to Wang et al. [31], the advantages of this category over the previous one is that weakness in individual forecasting methods can be discarded to improve forecasting accuracy. The study from Md-Khair et al. [11] is an example that proposed such model. In this study, they proposed a hybrid method that utilized wavelet transform as the data decomposition method while using a combination of ARIMA to forecast the smooth series and LSSVM to forecast the detail series. Their experiment proved that the proposed model is superior than the single models and hybrid models with one forecasting method in term of prediction accuracy. From the results of these studies, we can conclude that hybrid model with multiple forecasting methods performs better than hybrid model with single forecasting method.

3.3. Hybrid Model with Single or Multiple Forecasting Method(s) and Optimization Method

For this category, an approach incorporates a data decomposition method with one or more forecasting method with added optimization method. The advantage of methods in this category over the previous one is that optimization method can improve the required parameters in certain forecasting methods thus increasing forecasting accuracy. For instance, the approach proposed by Zhang et al. [23] which make use of LSSVM with particle swarm optimization (LSSVM-PSO) method to forecast the nonlinear component and generalized autoregressive conditional heteroskedasticity (GARCH) to forecast the time-varying component. The PSO in their study is the optimization method which is used to obtain the optimal parameters for the LSSVM method. The comparison made with EEMD plus GARCH method, EEMD plus PSO-LSSVM method and PSO-LSSVM method shown that their proposed method outperforms the others with more accurate forecasting. Another study in this category is the one conducted by Li et al. [5]. The proposed approach in this study utilized EEMD for data decomposition and relevance vector machine (RVM) for forecasting. Adaptive PSO (APSO) act as the optimization method which is used to simultaneously optimize the weights and parameters of RVM kernels. The result also showed that the proposed approach out performs all the single models and the hybrid models included in their experiment. From the results of these studies,

we can conclude that hybrid model with single or multiple forecasting method(s) and optimization method performs better than hybrid model with single forecasting method and hybrid model with multiple forecasting method.

3.4. Drawback in Hybrid Forecasting Model

From the review done, some noticeable drawbacks can be elucidated. The first one is in the hybrid model with single forecasting method category. In this category, if there is any weakness in the forecasting method, it cannot be eliminated from the model. For example, if ARIMA method is utilized, only dataset that is linear and near linear will give good forecasting results [6]. The problem is, in real world forecasting, data series will not be entirely linear or non-linear therefore a single forecasting model will not be sufficient because no single model can successfully recognize all the patterns contained in a data series. Therefore, this category cannot provide optimal forecasting solution for crude oil prices because it contains volatility, nonlinearity, and irregularity which will decrease the forecasting accuracy.

Another drawback is in the hybrid model with single or multiple forecasting method(s) and optimization method category. Even though this category provides the best forecasting accuracy between these three categories, the optimization process is usually time-consuming. This drawback is proven when Li et al. [5] stated that a lot of time is required to find the optimal parameters and to compute the combined kernel. Furthermore, according to them, it is quite difficult to replicate the experiment with the same exact result because the optimization method PSO uses many random values in the evolutionary process. Another drawback is the complexity of models in this category is quite high. According to Zhang et al. [23], even though this model can grab to complex volatility of crude oil prices very well, the calculation process is more intricate than with most previous methods in their experiment.

4. CONCLUSION AND FUTURE WORKS

The complexity of crude oil prices fluctuations is a major concern. Therefore, a lot of forecasting methods have been proposed by researchers to comprehend the movements thus giving advantages for future planning. Throughout the period, many forecasting methods have been studied and optimized to further improve the forecasting accuracy. In this review, a total of 12 studies that proposed hybrid methods in crude oil prices forecasting based on the "decomposition-and-ensemble" framework have been included to be analysed. Several categorizations for clustering the selected studies were introduced to give us an indication of how a better method can be proposed by observing which performs better than the other one. Furthermore, the limitations of each categorization were analyzed so that future researches can comprehend the limitations thus improving new hybrid forecasting methods. For future researches, it is recommended that more related studies to be included so that a more thorough and solid categorization can be established, and more drawbacks can be identified and highlighted.

REFERENCES

[1] Wu G, Zhang Y-J. Does China factor matter? An econometric analysis of international crude oil prices. *Energy Policy*. 2014; 72: 78-86.
[2] Cheong C W. Modeling and forecasting crude oil markets using ARCH-type models. *Energy policy*. 2009; 37(6): 2346-2355.
[3] Yusof N M, Rashid R S A, Mohamed Z. *Malaysia crude oil production estimation: an application of ARIMA model*. International Conference on Science and Social Research (CSSR). 2010; 1: 1255-1259.
[4] Chiroma H, Abdulkareem S, Abubakar A, Usman M J. Computational intelligence techniques with application to crude oil price projection: a literature survey from 2001-2012. *Neural Network World*. 2013; 23(6): 523.
[5] Li T, Zhou M, Guo C, Luo M, Wu J, Pan F, Tao Q, He T. Forecasting Crude Oil Price Using EEMD and RVM with Adaptive PSO-Based Kernels. *Energies*. 2016; 9(12): 1014.
[6] Shabri A, Samsudin R. Daily crude oil price forecasting using hybridizing wavelet and artificial neural network model. *Mathematical Problems in Engineering*. 2014; 2014.
[7] Wang G, Wu J. *Crude oil price forecasting based on the ARIMA and BP neural network combinatorial algorithm*. International Conference of Logistics Engineering and Management (ICLEM). Chengdu, China. 2012482-487.
[8] Crowley P M. A guide to wavelets for economists. *Journal of Economic Surveys*. 2007; 21(2): 207-267.

[9]   Schlüter S, Deuschle C, *Using wavelets for time series forecasting: Does it pay off?* 2010, IWQW discussion paper series.

[10] Kriechbaumer T, Angus A, Parsons D, Casado M R. An improved wavelet–ARIMA approach for forecasting metal prices. *Resources Policy*. 2014; 39: 32-41.

[11] Md-Khair N Q N, Samsudin R, Shabri A. Forecasting Crude Oil Prices using Discrete Wavelet Transform with Autoregressive Integrated Moving Average and Least Square Support Vector Machine Combination Approach. *International Journal on Advanced Science, Engineering and Information Technology*. 2017; 7(4-2): 1553-1561.

[12] Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N-C, Tung C C, Liu H H. *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*. Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences. 1998; 454: 903-995.

[13] Yu L, Wang S, Lai K K. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*. 2008; 30(5): 2623-2635.

[14] Zhu B, Wei Y. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega*. 2013; 41(3): 517-524.

[15] Zhang Y, Zhang T, Teng J, Su H. Improved Ensemble Empirical Mode Decomposition for rolling bearing fault diagnosis. *Indonesian Journal of Electrical Engineering and Computer Science*. 2014; 12(1): 539-549.

[16] Wu Z, Huang N E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*. 2009; 1(01): 1-41.

[17] Zhang G, Patuwo B E, Hu M Y. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*. 1998; 14(1): 35-62.

[18] Ediger V Ş, Akar S, Uğurlu B. Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model. *Energy Policy*. 2006; 34(18): 3836-3846.

[19] Azevedo V G, Campos L M. Combination of forecasts for the price of crude oil on the spot market. *International Journal of Production Research*. 2016; 54(17): 5219-5235.

[20] Kişi Ö. Neural network and wavelet conjunction model for modelling monthly level fluctuations in Turkey. *Hydrological processes*. 2009; 23(14): 2081-2092.

[21] Suykens J A, Vandewalle J. Least squares support vector machine classifiers. *Neural processing letters*. 1999; 9(3): 293-300.

[22] Razak I A W A, Abidin I Z, Siah Y K, Abidin A A Z, Rahman T K A, Baharin N, Jali M H B. An Hour Ahead Electricity Price Forecasting with Least Square Support Vector Machine and Bacterial Foraging Optimization Algorithm. *Indonesian Journal of Electrical Engineering and Computer Science*. 2018; 10(2).

[23] Zhang J-L, Zhang Y-J, Zhang L. A novel hybrid method for crude oil price forecasting. *Energy Economics*. 2015; 49: 649-659.

[24] Qunli W, Ge H, Xiaodong C. *Crude oil price forecasting with an improved model based on wavelet transform and RBF neural network*. International Forum on Information Technology and Applications (IFITA'09). 2009; 1: 231-234.

[25] Shabri A, Samsudin R. Crude oil price forecasting based on hybridizing wavelet multiple linear regression model, particle swarm optimization techniques, and principal component analysis. *The Scientific World Journal*. 2014; 2014.

[26] Bao Y, Zhang X, Yu L, Lai K K, Wang S. An integrated model using wavelet decomposition and least squares support vector machines for monthly crude oil prices forecasting. *New Mathematics and Natural Computation*. 2011; 7(02): 299-311.

[27] He K, Yu L, Lai K K. Crude oil price analysis and forecasting using wavelet decomposed ensemble model. *Energy*. 2012; 46(1): 564-574.

[28] Xiong T, Bao Y, Hu Z. Beyond one-step-ahead forecasting: evaluation of alternative multi-step-ahead forecasting models for crude oil prices. *Energy Economics*. 2013; 40: 405-415.

[29] Samsudin R, Shabri A. Crude oil price forecasting with an improved model based on wavelet transform and support vector machines. *EJ. Artif. Intell. Comput. Sci*. 2013; 1: 9-15.

[30] Bao Y, Zhang X, Yu L, Lai K K, Wang S. *Hybridizing wavelet and least squares support vector machines for crude oil price forecasting*. Proceedings of the 2nd international workshop on intelligent finance. 2007; 1: 1-15.

[31] Wang J, Hu J. A robust combination approach for short-term wind speed forecasting and analysis–Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*. 2015; 93: 41-56.