Australian Journal of Basic and Applied Sciences, 8(15) Special 2014, Pages: 341-351

AENSI Journals

Australian Journal of Basic and Applied Sciences

ISSN:1991-8178

Journal home page: www.ajbasweb.com

A Hybrid of EMD-SVM Based on Extreme Learning Machine for Crude Oil Price Forecasting

Rana Abdullah Ahmed and Ani Bin Shabri

Universiti Teknologi Malaysia, Department of Mathematics, Faculty of Sciences, 81310, Skudai, Johor, Malaysia.

ARTICLE INFO

Article history:
Received 15 September 2014
Accepted 5 October 2014
Available online 25 October 2014

Keywords:
Empirical mode decomposition; forecasting; extreme learning machine; crude oil price and support vector machines.

ABSTRACT

Forecasting crude oil spot prices (COSP) are important as it affects other key sectors of the economy including the stock market. This makes it crucial to develop reliable models that would assist adequately in forecasting the fluctuation of international crude oil price. This is aimed at facilitating the parties involved in taking appropriate action to avoid associated risk. In a current study, a new hybrid method based on empirical mode decomposition (EMD), support vector machine (SVM) and extreme learning machine (ELM) is presented. The crude oil price is adaptively decomposed into a series of smooth intrinsic mode function (IMF) with different scales via EMD. Each extracted IMF was forecasted with different SVM, the final results were obtained by adding together these forecasted results of each IMF. A hybrid method based on an extreme learning machine with adaptive metrics of input is proposed for improving the forecast accuracy of the prediction of all combined IMF. The EMD-SVM-ADD model applies the SVM to predict IMF extracted by EMD and integrates the predicted results, using a simple averaging method. To develop the model, we start by decomposing of the WTI by the EMD extraction process after this for each of the IMF obtained a unique EMD-SVM model is developed which we call IMF-SVM models. The desired EMD-SVM-ADD model is achieved by summing up the predictions from the IMF-SVM. This summing up is where the model derived its name from EMD-SVM because it is simply the additions of the predictions from the smaller IMF-SVM models. To evaluate the efficiency of the model, the study adapts the West Texas International (WTI) crude oil spot price. The results revealed that the new proposed model (EMD-SVM-ELM) performed better when compared with single SVM and EMD-SVM-ADD models judging by their RMSE and MAE. We concluded, based on the results obtained especially with the MAPE is less than 5% that the model is equally suitable for crude oil spot price forecasting.

© 2014 AENSI Publisher All rights reserved.

To Cite This Article: Rana Abdullah Ahmed and Ani Bin Shabri, A Hybrid of EMD-SVM Based on Extreme Learning Machine for Crude Oil Price Forecasting. ***Aust. J. Basic & Appl. Sci.**, 8(15): 341-351, 2014*

INTRODUCTION

Forecasting crude oil spot price (COSP) is among the most significant issues facing energy economists in recent time. Beside that limited the success is realized in arriving at a suitable model to capture the dynamics of crude oil. The events that occur in the oil market over the years have reignited the question if the changes in crude oil price can be predicted, or if it is merely a random walk. This is because of the price increase of crude oil impacts greatly on the price of petrol which has its attendant effect on goods and services produced in the country and by extension the gross domestic product (GDP).

The crude oil price and its forecasting important because it significantly affect the economy and stock markets of the country. A number of researchers worked on crude oil forecasting using Box-Jenkins methodology (Liu, 1991; Chinn *et al*., 2005; Agnolucci, 2009).

In this study, we use the new emerging data decomposition technique proposed by Huang (1998) known as an Empirical Mode Decomposition (EMD) to decompose the popular Brent and WTI crude oil price data into their respective intrinsic mode function (IMF) before combining with appropriate conventional and non-conventional time series models. This is done because of the belief that it will lead to improve the forecast accuracy of the models presented which is the overall goal of any forecast model.

EMD, Fourier and Wavelet transform are tools used for decomposing signals, but EMD is quite different from the other two. With Fourier and Wavelet transforms, there is the need to select basis signal components

Corresponding Author: Rana Abdullah Ahmed Department of Mathematical Sciences, Universiti, Teknologi Malaysia,Skudai, Johor, 81310, Malaysia. U4A -702- college perdana- UTM- Skudai, Johor, 81310, Malaysia
Tel: 01137732792. E-mail: ranooshutm@gmail.com

and also it is needful to also compute the parameters for each of these signals. The aggregates of these signals will compose the original signal. With EMD there is no need to make a priori assumption about the composition of the signal. This EMD does by spline interpolation between the maxima and minima to be able to obtain the intrinsic mode function (IMF) (Labate *et al*., 2013). Since EMD is known to make assumptions about signals, the result is likely to be more meaningful than those of Fourier and wavelet transforms. The IMF, which is products of EMD change over time, and since EMD makes no assumption about the stationary of the signal, this property makes them better suited for nonlinearity and nonstationarity than either of the other two-Fourier and Wavelet.

Lin *et al*. (2012) suggest a hybrid method that combines EMD and LSSVM for predicting foreign exchange rates. Experiment results revealed show that the proposed EMD-LSSVM model performs better than EMD-ARIMA, LSSVM and ARIMA models when the time series is not decomposed.

Li *et al*. (2012) used a hybrid EMD and LSSVM technique to predict country risk of oil exporters. In the paper, EMD was employed in the decomposition of the data into a number of IMF which were then predicted by LSSVM. The authors concluded that the presented model was statistical very robust than other models based on the outcome from data of 10 countries.

Okolobah and Ismail (2013c) presented a combined model for peak load demand forecasting for Nigeria using EMD and ANFIS. In the paper EMD extraction process was used to decompose the series into a number of IMF. Each of the IMF were forecasted using ANFIS and they called this IMF-ANFIS models. The targeted model was achieved by aggregating the outcomes of the IMF-ANFIS models. The results reveal that the EMD-ANFIS model outperforms stand-alone ANFIS model judging by the MAPE of both models.

Liu *et al*. (2013) proposed an innovative model that combines both EMD and RBFNN. GPS PWV time series is decomposing to several signals called intrinsic mode function, the phase-space reconstruction is done, and each of the components is then predicted by RBF. The final prediction value is reconstructed. The result shows that predictive value is close to the actual precipitation, which can better reflect the actual precipitation change.

Xiong *et al*. (2014) suggested the incorporation of end condition models into EMD for decomposing and assembling model frameworks for a- step as well as multi-step ahead forecast model for time series type. Four well developed end condition, mirror, Coughlin's, Slope-based and Ratio's methods in which SVM is utilized for modeling purposes. The outcome of the experiment showed that remarkable improvements are achievable from the proposed technique when end-conditions are incorporated into them.

Extreme learning Machine (ELM) is a neural network training technology, that adopts the least square solution approach. It was introduced by Huang and his team (Huang, 2004). ELM is a single hidden layer feedforward network (SLFNs) architecture. ELM depends on some properties of the network such as; if the weights and biases in the input layer are randomly adjusted and the transfer functions are infinitely differentiable, then for any given training set it is possible to optimally determine its output. ELM has two advantages and they are: it overcomes the drawbacks resulting from gradient descent and it also overcomes the generalization performance of SVM faster with less interference. Below are reviews conducted into previous researches on ELM. The advantages of ELM over traditional algorithms are as follows:

Huang *et al*. (2006) proposed ELM as a learning algorithm based on single layer feedforward networks. In the paper, the network parameters were generated randomly as this makes tuning of the network not to be necessary. The work compared the performance of ELM, SVM and back propagation neural network and concluded that ELM is the best of the three techniques.

Frenay and Verleysen (2010) incorporated ELM algorithm into SVM and called the kernel function ELM kernel. The aim of doing this, the authors say is to eliminate tuning of the network parameters as the parameters used for the configuration of the network were randomly generated just as in ELM.

Huang *et al*. (2012) proposed a unified ELM algorithm for both classification and regression. In the paper, random feature mapping and kernels are used to handle an optimization based ELM. This is after it has presented ELM as an optimization problem having equality constraints. The authors equally said that single optimization can handle regression and classification.

Zhang *et al*. (2013) applied ELM, which depends on rough set theory. The rough set theory undergoes attribute reduction, then applying ELM for training as well as for forecasting the new datasets. The proposed algorithm was found to have better prediction accuracy with an improvement in efficiency.

Yan *et al*. (2013) combined empirical mode decomposition (EMD) and extreme learning machine (ELM) recommends the price of international uranium resources. The IMF is based on the rules of fine to coarse reconstruction, depend on the phase space reconstruction combined into three sub-series. Different ELM models are utilized for modeling and forecast of the three sub-series based on the intrinsic characteristic time scales. It is known that the hybrid prediction method is superior because of the data from the real price of uranium resources show this.

Okolobah and Ismail (2013a) proposed a combine model involving EMD and ARIMA in electricity peak load forecasting using Malaysian load data. Based on MAPE of the combine model and that of ARIMA model concluded that the combine model outperforms the single model and further says that combine model increases the prediction ability of single models. In electricity load forecasting.

ANN using in crude oil forecast, Shambora and Rossiter (2007) and Yu *et al*. (2007a) also used the ANN model to predict crude oil price. Many experiments found that the artificial intelligent (AI), AI-based models often had some advantages over statistical-based models. However, these AI models also have their own shortcomings and disadvantages. For example, ANN often suffers from local minima and over fitting.

Support Vector Machine (SVM) a neural network based algorithm, that has made a remarkable mark in the subject area of forecasting. The SVM has the special property of not allowing for over-fit of models. It is also very stable and effective for application in nonlinear modelling. In addition, SVM is good in optimization of curve fitting problems which often leads to the general explanation in many instances that it does not adhere to explanations that involves exclusions. SVM as compared to other neural networks yield improved performance and as such SVM has been increasingly applied in the field of economic time series (Tay and Cao, 2001; 2002; Kim, 2003; Huang *et al*., 2005). Futhermore, in comparison to traditional neural network models, SVM requires the data set for modelling should be consistent.

Derived from this standard, SVMs will ultimately produce better simplification performances in comparison with other neural networks. Due to such benefits, SVM method has been used in the area of economic time series forecasting (Tay and Cao, 2001; 2002; Kim, 2003; Huang *et al*., 2005).

Qiuge *et al*. (2008) proposed an algorithm called extreme support vector machine (ESVM) classifier. It is a learning algorithm that has random input weights with its feature space; obtain by single hidden layer feedforward network (SLFN). Its obvious advantages are that it has a generalization performance that is better than ELM and is faster than the conventional SVM model.

Zakaria (2012) compared the predictive abilities of multiple linear regression (MLR) and support vector machine (SVM) in food forecasting. In the study two measures were used to evaluate the performance of both models namely MAE and RMSE. The study conducts that SVM performs better than MLR in food forecasting.

Xiao-Lin and Hai-Wei (2012) used three fundamental kernel functions of SVM to develop a prediction model for crude oil price, adopting the particle swarm algorithm for optimizing the model parameters. The result revealed that the prediction models whose parameters have been optimized by a genetic algorithm give a very promising outcome.

The current study proposed new methods for crude oil forecasting based on EMD, SVM and ELM. Crude oil series are decomposed into different components, which display the different frequency characteristics of the prices. Due to the different kernel functions and hyper-parameters, SVM is constructed to forecast each component according to its characteristics independently volatility of crude oil prices take account in forecasting. The final forecasting result can be obtained by adding up all of the forecasting values of each crude oil component. In concrete terms, the original crude oil spot price series, with characteristics of nonlinearity and nonstationarity, were first decomposed into a finite and often small number of intrinsic mode functions (IMF). After these simple IMF components are adaptively extracted by EMD. The rest of this paper is organized as follows: Section 2 Materials and Methods, Section 3 Results. For illustration and verification purposes, West Texas Intermediate (WTI) crude oil spot price is used to test the effectiveness of the proposed methodology, and Discussion in Section 4. Finally, some Conclusion is drawn in Section 5

MATERIALS AND METHODS

Support vector machines:

The Support Vector Machines (SVM) proposed by Vapnik (1995). According to the structured risk minimization (SRM) principle, SVM looks at reducing an upper bound of generalization error rather than an empirical error as in other neural networks. Furthermore, the SVMs models, create the reverse function by concerning a set of high dimensional linear functions. The SVM regression function is formulated as follows:

$$y(x) = w\phi(x) + b \tag{1}$$

where $\phi(x)$ is named the feature, which is nonlinear planed from the input space x. The coefficients w and b are evaluated by minimizing:

$$R(c) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(d_i, y_i) + \frac{1}{2}\|w\|^2, \tag{2}$$

$$L_\varepsilon(d_i, y_i) = \begin{cases} |d-y| - \varepsilon & |d-y| \ge \varepsilon, \\ 0 & others. \end{cases} \tag{3}$$

where both C and ε are prescribed parameters. The first term $L_\varepsilon(d_i, y_i)$ is named the $\varepsilon-$intensive loss function. The d_i is the actual stock price during the ith period. This function shows that errors below ε are not

penalized. Also the term $C\frac{1}{N}\sum_{i=1}^{N} L_{\varepsilon}(d_i, y_i)$ measures the empirical error. The next term, $\frac{1}{2}\|w\|^2$, is the flatness of the function. C assesses the trade-off between the flatness of the model and the empirical risk. ξ and ζ^* were introduced as the positve slack variables, which signify the distance from the actual values to the corresponding boundary values of $\varepsilon-$ tube. Equation (2) is converted to the following constrained formation:
Minimize:

$$R(w,\xi,\xi^*) = \frac{1}{2}ww^T + C^*(\textstyle\sum_{i=1}^{N}(\xi_i + \xi_i^*)) \tag{4}$$

Subjected to:

$$w\phi(x_i) + b_i - d_i \le \varepsilon + \xi_i^*,$$
$$d_i - w\phi(x_i) - b_i \le \varepsilon + \xi_i^*, \tag{5}$$

$$where\ \xi_i, \xi_i^* \ge 0, i = 1,2,...,N. \tag{6}$$

Finally, introducing Lagrange multipliers and maximizing the dual function of Equation (4) we have:

$$R(\alpha_i - \alpha_i^*) = \sum_{i=1}^{N} d_i(\alpha_i - \alpha_i^*) - \varepsilon\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*) \times (\alpha_j - \alpha_j^*)K(x_i, x_j) \tag{7}$$

With the constraints

$$\textstyle\sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \tag{8}$$

$$0 \le \alpha_i \le C, \tag{9}$$

$$0 \le \alpha_i^* \le C\ , \tag{10}$$

$$i = 1,2,...,N.$$

In Equation (7), α_i and α_i^* are called Lagrange multipliers. They satisfy the equalities,

$$\alpha_i * \alpha_i^* = 0,$$
$$f(x,\alpha,\alpha^*)\textstyle\sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i, x_j) + b. \tag{11}$$

Here, $k(x, x_i)$ is named the kernel function. The amount of the kernel is equivalent to the inner product of two vectors x_i and x_j in the feature space $\phi(x_i)$ and $\phi(x_j)$, such that $k(x, x_i) = \phi(x_i) * \phi(x_j)$. Any function that fulfilling Mercer's condition Vapnik (1995) can be applied as the kernel function. The Gaussian kernel function

$$k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 \big/ 2\sigma^2)$$

Is specific in this study. The SVMs were used to evaluate the nonlinear behavior of the predicting data set because Gaussian kernels aim to present good performance under common efficiency assumptions Scholkopf (1997). Parameter C and γ for SVM are estimates using cross-validation technique. Cross-validation is a technique that can be used to estimate the quality of a neural network. When applied to several neural networks with different free parameter values (such as the number of hidden nodes, SVM, and so on), the results of cross-validation can be used to select the best set of parameter values. While there are several types of cross-validation. The best way to get a feel for how k-fold cross-validation can be used with neural networks.

Empirical Mode Decomposition (EMD):

The empirical mode decomposition (EMD) technique, first proposed by Huang (1998), is a form of adaptive time series decomposition technique using the Hilbert-Huang transform (HHT) for nonlinear and nonstationary time series data. The basic principle of EMD is to decompose a time series into a sum of oscillatory functions, namely, intrinsic mode functions (IMFs). In the EMD, the IMFs must satisfy the following two prerequisites:

(1) In the whole data series, the number of extrema (sum of maxima and minima) and the number of zero crossings, must be equal, or differ at most by, one (1).

(2) The mean value of the envelopes defined by local maxima and minima must be zero at all points.
With these two requirements, some meaningful IMFs can be well defined. Otherwise, if one blindly applies the technique to any data series, the EMD may result in a few meaningless harmonics Huang (1999). Usually, an IMF represents a simple oscillatory mode, compared with the simple harmonic function.

Using the definition, any data series $x(t)(t=1,2,...,n)$ can be decomposed, according to the following sifting procedure.

1) Identify all the local extrema, including local maxima and local minima, with $x(t)$,

2) Connect all local extrema by a cubic spline line to generate its upper and lower envelopes $x_{up}(t)$ and $x_{low}(t)$.

3) Compute the point-by-point envelope mean m(t) from upper and lower envelopes, i.e. $m(t)=(x_{up}(t)+x_{low}(t)/2)$.

4) Extract the details, $c(t)=x(t)-m(t)$.

5) Check the properties of $c(t)$:

(i) If $c(t)$ meets the above two requirements, an IMF is derived and $x(t)$ replaced with the residual $r(t)=x(t)-c(t)$;

(ii) If $c(t)$ is not an IMF, replace $x(t)$ with $c(t)$

The EMD extracts the next IMF by applying the above sifting procedure to the residual term $r_1(t)=x(t)-c_1(t)$ where $c_1(t)$ denotes the first IMF. The decomposition process can be repeated until the last residue $r_n(t)$ at has at most one local extremum or becomes a monotonic function, from which no more IMFs can be extracted.

At the end of this sifting procedure, the data series $x(t)$ can be expressed by

$$x(t)=\sum_{j=1}^{n} c_j(t)+r_n(t) \tag{12}$$

where n is the number of IMFs, $r_n(t)$ is the final residue, which is the main trend of, $x(t)$, and $c_j(t)$ $(j=1,2,...,n)$ are the IMFs, which are nearly orthogonal to each other, and all have nearly zero means. Thus, one can achieve decomposition of the data series into n-empirical mode functions and one residue. The IMF components contained in each frequency band are different and they change with variation of time series x while $r_n(t)$ represents the central tendency of data series $x(t)$.

Relative to the traditional Fourier and wavelet decompositions, the EMD technique has several distinct advantages. First of all, it is relatively easy to understand and implement. Second, the fluctuations within a time series are automatically and adaptively selected from the time series, and this process is robust for both nonlinear and nonstationary time series decomposition. Third, it lets the data speak for themselves. EMD can adaptively decompose a time series into several independent IMF components and one residual component. The IMFs and the residual component displaying linear and nonlinear behavior depend only on the nature of the time series being studied. However, in wavelet decomposition, a filter base function must be determined beforehand, but it is difficult for some unknown series to determine the filter base function. Unlike wavelet decomposition, EMD is not required to determine a filter base function before decomposition. In terms of the above merits the EMD can be used as an effective decomposition tool.

Extreme Learning Machine:

The extreme learning machine (ELM) is a flexible computing framework for a broad range of nonlinear problems Huang (2008). A single hidden-layer feed forward network (SLFN) is the most widely used model for forecasting modeling Sun (2008). As shown in Fig.1, the model is characterized by a network of three layers of simple processing units connected by acyclic links. The hidden layers can capture the nonlinear relationship among variables. Each layer consists of multiple neurons that are connected to neurons in adjacent layers. Suppose there are N distinct samples (x_i,t_i) where $x_i=[t_i^1,t_i^2,...,t_i^n]^T\in R^n$ and $t_i=[t_i^1,t_i^2,...,t_i^m]^T\in R^m$ .The SLFN with k hidden neurouns and an activation function vector $g_1(x),g_1(x),....,g_k(x)$ are described as

$$\sum_{i=1}^{k} B_i g_i(w_i.x_j+b_i)=Y_j,\ j=1,2,...N, \tag{13}$$

Where $w_i=[w_{i1},w_{i2},...,w_{in}]^T$ is the weight vector conection the ith hidden neuron and the input neurons, $\beta_i=[\beta_{i1},\beta_{i2},...\beta_{im}]^T$ is the weight vector conecting the ith hidden neuron and the output hidden neurons, and b_i is the threshold of the ith hidden neuron. The operation $w_i.x_j$ is the inner product of w_i and x_j.

Australian Journal of Basic and Applied Sciences, 8(15) Special 2014, Pages: 341-351

The activation function of extreme learning machine is the sigmoid function:

$$g_n(x) = 1/(1+e^{-x}) \tag{14}$$

If the SLFNs can approximate these n samples with a zero error viz. $\sum_{i=1}^{n}|y_i - t_i| = 0$ Thus there also exist parameters w_i, x_i and b_i such that

$$\sum_{i=1}^{k} B_i g_i(w_i . x_j + b_i) = T_j,\ j = 1,2,...N, \tag{15}$$

Thus, above equations can be compactly described as

$$H\beta = T, \tag{16}$$

where H is the hidden-layer output matrix, the ith column of H denotes the *i*th hidden neuron output with respect to inputs $x_1, x_2, ..., x_N$

$$H = \begin{bmatrix} g_1(w_1 .\ x_1 + b_1) & \cdots & g_K(w_k .\ x_1 + b_K) \\ \vdots & \cdots & \vdots \\ g_1(w_1 .\ x_N + b_1) & \cdots & g_K(w_k .\ x_N + b_K) \end{bmatrix} \tag{17}$$

Unlike the traditional function approximation theories which require adjustment of input weights and hidden layer biases, the input weights and hidden biases are randomly generated. Thus, training an SLFN is simply equivelant to finding a least sequares solution $\hat{\beta}$ of the linear function $H\beta = T$:

$$\left\|H\hat{\beta} - T\right\| = \min_\beta \left\|H\beta = T\right\|, \tag{18}$$

The smallest norm least squares solution of the above linear system is

$$\hat{\beta} = H^T T, \tag{19}$$

Where H^T is the Moore-Penrose generalized inverse of matrix *H*. Owing to the Moore-Penrose generalized inverse, the learning speed are dramatically increased for the network

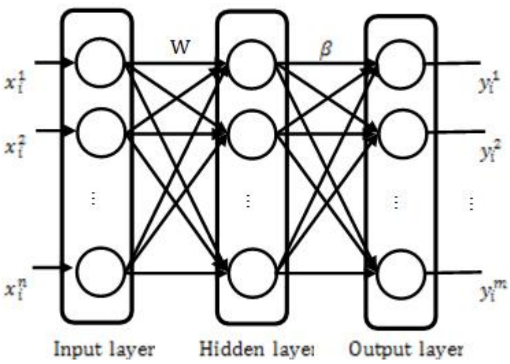

Fig. 1: The Structure of the ELM Model.

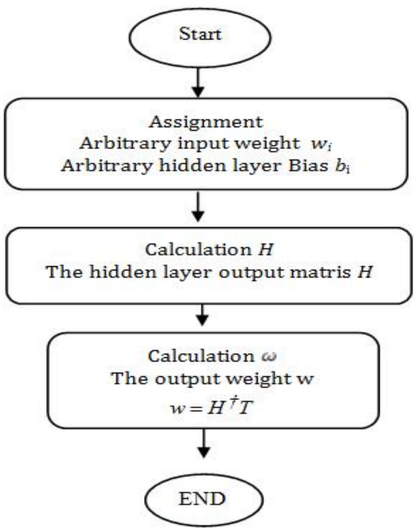

Fig. 2: The learning process of the ELM Model.

The EMD-SVM-ELM Model:

In a time series x_t, *t*=1, 2,…, *N*, one can predict *h*-steps ahead, i.e. x_{t+h}. When *h*=1 means prediction of 1 step ahead and *h*=20 means 20 steps ahead. This depends on the earlier techniques and method presented. We can achieve the formulation of the EMD-SVM-ELM, as presented in Figure 3 and Figure 4.

The following three steps constitute the prediction paradigm of the EMD-SVM-ELM, as demonstrated in Figure 3 and Figure 4:

1. The historical time series x_t, *t*=1, 2,…, N, is decomposed into an IMF component, $c_j(t), j = 1,2,...,n$, and a residue component $r_n(t)$ using an EMD extraction process.
2. The SVM model is used in modelling each of the IMFs and the residue component, and which predictions are made.
3. The residue produced by the SVM, along with the predictions from all the extracted IMF components, is utilized in creating an aggregated output using an ELM model. This is viewed as the outcome of the original time series. To confirm the effectiveness of the presented EMD-SVM-ELM, testing is carried out in the next section on crude oil price series. An additional two variants of this hybrid model are used in the prediction of COSP for the purpose of comparison. These two variants of the hybrid models are the EMD–SVM–ADD, and EMD-SVM-ElM models.

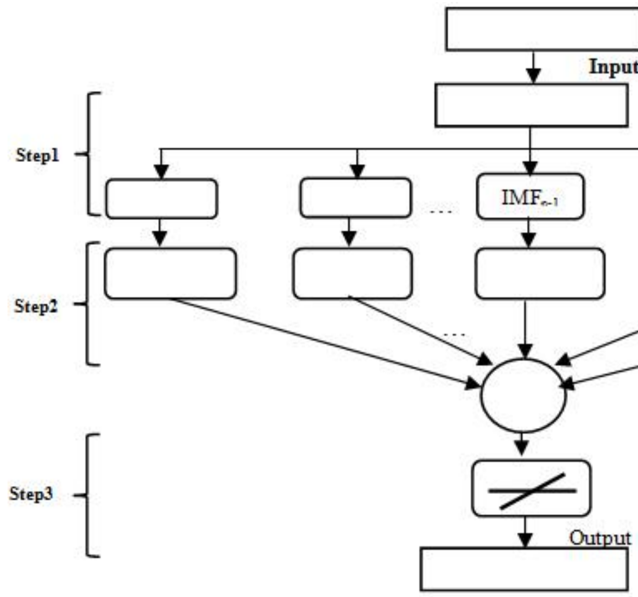

Fig. 3: The flow chart of the proposed EMD- SVM-ADD model.

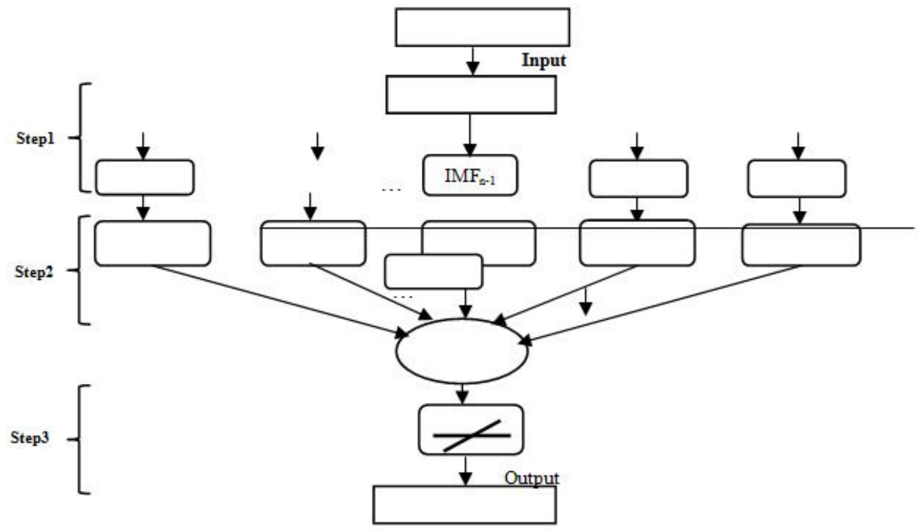

Fig. 4: The flow chart of the proposed EMD- SVM-ELM model.

Experiments:
Data:

Crude oil prices are the approved and agreeable standard price used globally, which makes them an ideal for our purposes. In this paper study, we chose to use the West Texas Intermediate (WTI) crude oil spot price for experimental purposes. The data covers the period January 1, 1986 to September 30, 2006, thereby giving a total of 5237 observations. The data are updated daily and are available from the Energy Information Administration (EIA) website. Figure 5 presents these data.

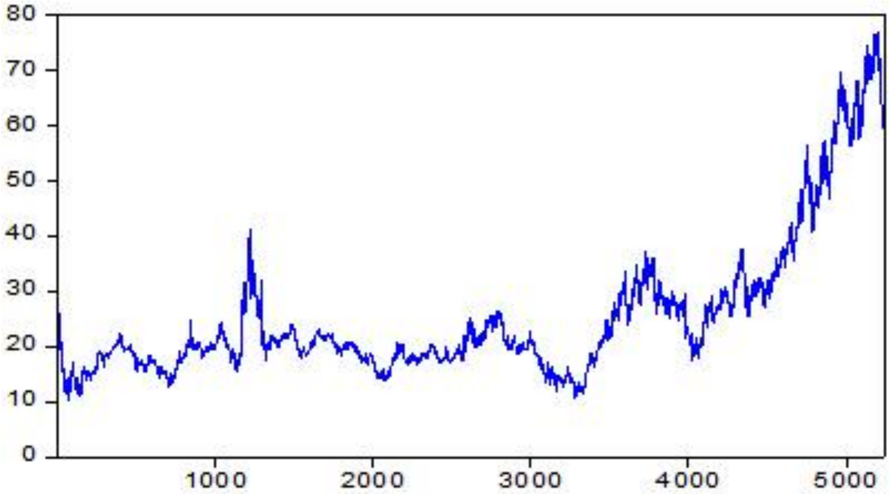

Fig. 5: The time series for WTI daily crude oil prices.

Evaluation of Performance Forecasts:

There are several ways to evaluate the performances of forecasting models. In this research, the Mean Absolute Error (MAE), the Root Mean Squared Error (RMSE) and the mean absolute percentage error (MAPE) will all be used.

Mean Absolute Error (MAE) is given by:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|(\hat{y}_t - y_t)^2\right|$$

The Root Mean Squared Error (RMSE) is given by:

$$RMSE = \left[\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2\right]^{1/2}$$

And the Mean absolute percentage error (MAPE) is obtained as follows:

$$MAPE = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right|$$

The MAPE is a dimensionless quantity that is not too different from the MAE. It Is expressed as a percentage and this makes it very helpful in comparing forecast accuracies from different forecast horizons.

Where

y_t : is the actual values for period t.

$\hat{y}_t$: is the forecast values for period t.

n : is the number of observations.

Results:

The EMD–SVM–ADD model applies the SVM to predict all IMFs, before combining the prediction results using a simple summation method. Thus, prediction of WTI crude oil prices for comparison purposes can be executed by all Three models: SVM model, EMD–SVM–ADD hybrid model, and the proposed EMD–SVM–ELM hybrid model. The prediction experiments are performed as shown in the previous section. Firstly, the two typical crude oil price series can be decomposed into several independent IMFs and one residue, using the EMD technique. Secondly, using EMD, we can create graphical representations of the decomposed results for WTI crude oil prices, as illustrated in Figure 6, that shows the decomposition results in the WTI COSP series.

Clearly, the WTI crude oil spot price series is decomposed into ten IMFs and one residue, which helps us to improve the prediction performance.

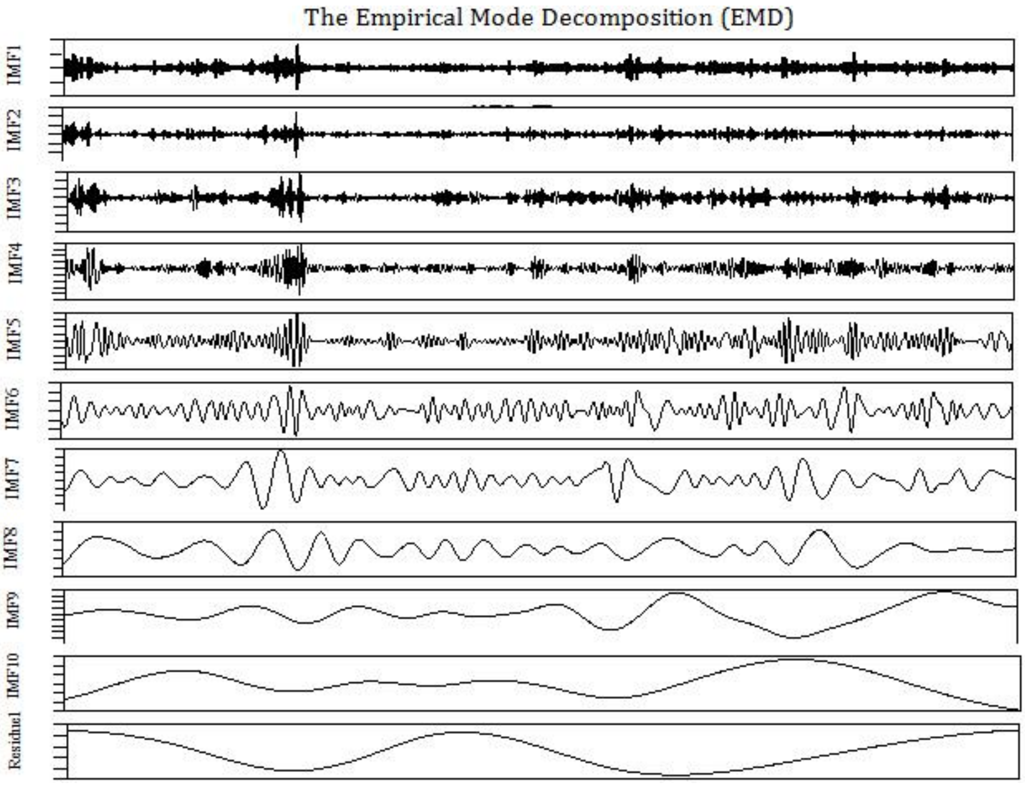

Fig. 6: The decomposition of daily WTI crude oil price.

Table 1 shows the evaluation of prediction results in the WTI crude oil price series via RMSE, MAE and MAPE. From this, we can see that the forecasting results of the proposed EMD-SVM-ELM models, approaches, are very promising for all studied crude oil prices. This is applicable when the measurement of forecasting performance is goodness-of-fit, such as with RMSE and MAE (refer to Table 1), which indicates that the predicted performance of the proposed EMD-SVM-ELM forecasting model is better than that of other models listed in this study. Focusing on the RMSE indicator, the EMD–SVM–ELM method performs the best in all cases, followed by, EMD–SVM–ADD model, and individual SVM model, which could be for two reasons. Firstly, the ELM model is a class of nonlinear forecasting model, which is able to capture nonlinear patterns hidden in the crude oil prices. Secondly, the crude oil market is a high-volatile market and the crude oil prices often show nonlinear and nonstationarity patterns, while the EMD–SVM–ADD model and the single SVM model are nonlinear models, which are not suitable for predicting crude oil price with irregularity. In addition, we also find that the EMD–SVM–ADD model perform much better than the single SVM model, which could be because the EMD decomposition impacts the prediction performance. From Table 1 it can be observed that the forecast accuracy of the models is very good judging by the results posted by the three measures. This is best appreciated by the forecast accuracy of the MAPE for the WTI data that yielded values that are all less than 5% which Lewis (1982) say is highly accurate for forecast models.

Table 1: The RMSE comparisons of different methods.

Methodology	WTI		
	RMSE	MAE	MPAE
Single SVM	0.868	0.630	0.016
EMD-SVM-ADD	0.796	0.605	0.016
EMD-SVM-ELM	0.477	0.363	0.014

Discussion:

The EMD-SVM-ADD model applies the SVM to predict IMF extracted by EMD and integrates the predicted results, using a simple averaging method. To develop the model, we start by decomposing of the WTI data earlier presented in the previous section, by the EMD extraction process. The decomposition obtained is presented in Figure 6. After this for each of the IMF obtained a unique EMD-SVM model is developed which we call IMF-SVM models. The desired EMD-SVM-ADD model is achieved by summing up the predictions from the IMF-SVM. This summing up is where the model derived its name from EMD-SVM because it is simply the additions of the predictions from the smaller IMF-SVM models.

In constructing the EMD-SVM-ELM model we rely upon the IMFs which were earlier extracted following the EMD extraction process presented in the previous section, in which we saw that the WTI data yielded 10 IMFs and a residue. This was employed in developing 11 IMF-SVM models for the WTI data, To achieve the desired model, each of the results obtained from WTI data and this corresponding original values were passed through ELM. That is, the results for all 11 IMF-SVM WTI data as well as the original data for WTI were passed through the ELM process and the result is the desired output for this model.

In Figure 6 we see the decomposition of the WTI data result after going through the EMD extraction process. From the Figure it can be observed that the WTI data yielded 10 IMFs and a residue component. The amplitude of IMF 1 to IMF 8 is very high, while that of IMF 9 and IMF 10 is not as high as the other ones. The long time trend of the data is presented in the residue component. We concluded by presenting the forecast accuracy measure of the model using three different measures: RMSE, MAE and MAPE. It was concluded that the model is very good for forecasting crude oil spot price, especially as the MAPE of the models for WTI of data were less than 5%.

Conclusion:

In this paper a hybrid technique that combines EMD, SVM and ELM is proposed. In the proposed model, the EMD extraction process is first used to decompose the series of stable IMF that are each predicted by separate SVM models. The prediction results from these IMF are combined together to obtain a single SVM model whose forecast accuracy is strengthened by the ELM model. This model is then used for forecasting purpose. This is done because the ELM algorithm improves the forecast accuracy of combined SVM models. The results revealed that the proposed EMD-SVM-ELM outperforms single SVM and EMD-SVM-ADD models in terms of their forecast accuracy. We concluded, based on the results obtained especially with the MAPE is less than 5% that the model is equally suitable for COSP forecasting. Based on the results, we conclude that the ELM algorithm is very good for improving forecast accuracy of SVM models and that the proposed EMD-SVM-ELM method can be adopted for crude oil spot price forecasting across the globe.

REFERENCES

Agnolucci, P., 2009. Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. Energy Economics, 31(2): 316-321.

Chinn, M.D., M. LeBlanc, O. Coibion, 2005. The predictive content of energy futures: an update on petroleum, natural gas, heating oil and gasoline (No. w11033). National Bureau of Economic Research.

Frénay, B., M. Verleysen, 2010. Using SVMs with randomised feature spaces: an extreme learning approach. In ESANN.

Huang, G.B., H. Zhou, X. Ding, R. Zhang, 2012. Extreme learning machine for regression and multiclass classification. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 42(2): 513-529.

Huang, G.B. and L. Chen, 2008. Enhanced random search based incremental extreme learning machine. Neurocomputing, 71(16): 3460-3468.

Huang, G.B., Q.Y. Zhu and C.K. Siew, 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. In Neural Networks. Proceedings. 2004 IEEE International Joint Conference, IEEE.

Huang, G.B., Q.Y. Zhu and C.K. Siew, 2006. Extreme learning machine: theory and applications. Neurocomputing, 70(1): 489-501.

Huang, N.E., Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, H.H. Liu, 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 454(1971): 903-995.

Huang, N.E., Z. Shen and S.R., 1999. Long, A new view of nonlinear water waves: The Hilbert Spectrum 1. Annual review of fluid mechanics, 31(1): 417-457.

Huang, W., Y. Nakamori, S.Y. Wang, 2005. Forecasting stock market movement direction with support vector machine. Computers & Operations Research, 32(10): 2513-2522. Liu, Q., He, Q., & Shi, Z., 2008. Extreme support vector machine classifier. In Advances in Knowledge Discovery and Data Mining, pp: 222-233). Springer Berlin Heidelberg.

Kim, K.J., 2003. Financial time series forecasting using support vector machines. Neurocomputing, 55(1): 307-319.

Labate, D., F.L. Foresta, G. Occhiuto, F.C. Morabito, A. Lay-Ekuakille, P. Vergallo, 2013. Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG: A comparison.Sensors Journal, IEEE, 13(7): 2666-2674.

Lewis, C.D., 1982. Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting. London: Butterworth Scientific.

Li, S., H. Fang, 2014. An EMD-based long-term LSSVR for machine condition prognostics. In Control Conference (CCC), 2014 33rd Chinese, pp: 5133-5138. IEEE.

Lin, C.S., S.H. Chiu and T.Y. Lin, 2012. Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting. Economic Modelling, 29(6): 2583-2590.

Liu, L.M., 1991. Dynamic relationship analysis of US gasoline and crude oil prices. Journal of Forecasting, 10(5): 521-547.

Liu, Y.P., Y. Wang and Z. Wang, 2013. RBF Prediction Model Based on EMD for Forecasting GPS Precipitable Water Vapor and Annual Precipitation. Advanced Materials Research, 765: 2830-2834.

Sun, Z.L., T.M. Choi, K.F. Au, Y. Yu, 2008. Sales forecasting using extreme learning machine with applications in fashion retailing. Decision Support Systems, 46(1): 411-419.

Okolobah, V., Z. Ismail, 2013. FORECASTING PEAK LOAD DEMAND BY ARIMA AND EMD METHOD. Archives Des Sciences, 66(1).

Okolobah, V.A. and Z. Ismail, 2013. New Approach to Peak Load Forecasting based on EMD and ANFIS. Indian Journal of Science and Technology, 6(12): 5600–5606.

Schölkopf, B.,1997. Support vector learning.

Shambora, W.E., R. Rossiter, 2007. Are there exploitable inefficiencies in the futures market for oil?. Energy Economics, 29(1): 18-27.

Tay, E.H. and L.J. Cao., 2001. Application of support vector machines in financial time series forecasting. Omega, 29: 309-317.

Tay, F.E., L.J. Cao, 2002. Modified support vector machines in financial time series forecasting. Neurocomputing, 48(1): 847-861.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. New York: Springer Verlag.

Xiao-lin, Z., W. Hai-wei, 2012. Crude Oil Prices Predictive Model Based on Support Vector Machine and Particle Swarm Optimization. In Software Engineering and Knowledge Engineering: Theory and Practice (pp. 645-650). Springer Berlin Heidelberg.

Xiong, T., Y. Bao and Z. Hu., 2014. Does restraining end effect matter in EMD-based modeling framework for time series prediction? Some experimental evidences. Neurocomputing, 123: 174-184.

Yan, Q., S. Wang and B. Li., 2013. Forecasting Uranium Resource Price Prediction by Extreme Learning Machine with Empirical Mode Decomposition and Phase Space Reconstruction.

Yu, L., K.K. Lai, S. Wang, K. He, 2007. Oil price forecasting with an EMD-based multiscale neural network learning paradigm. In Computational Science–ICCS 2007, pp: 925-932. Springer Berlin Heidelberg.

Zakaria, Z.A., A. Shabri, 2012. Streamflow Forecasting at Ungaged Sites Using Support Vector Machines. Applied Mathematical Sciences, 6(60): 3003-3014.

Zhang, Y., 2013. An Algorithm Research for Prediction of Extreme Learning Machines Based on Rough Sets. Journal of Computers, 8(5).