Implementing Method of Empirical Mode Decomposition based on Artificial Neural Networks and Genetic Algorithms for Crude Oil Price Forecasting

S Herawati¹ and A Djunaidy²

¹Faculty of Engineering, University of Trunojoyo Madura, Bangkalan – Madura, Indonesia
²Dept. Of Information Systems, Institut Teknologi Sepuluh Nopember, Indonesia

Email: sriherawati2019@gmail.com

Abstract. Fluctuations in crude oil prices can affect a country's economic policies. The movement of crude oil prices tends to be nonlinear and non-stationary. One forecasting method that is intended to accommodate these traits is forecasting that integrates empirical mode decomposition (EEMD) ensemble methods based on artificial neural networks and genetic algorithms. In the EEMD method, a white noise signal is added to compensate for the mixture mode that can be formed. Each IMF and residue generated in the decomposition process are used as input to a feedforward neural network (FNN) artificial neural network to obtain forecasting models from each IMF and residue. The genetic algorithm is integrated with the FNN to avoid overfitting, the formation of local optima solutions, and the sensitivity of the selection of FNN parameters. The data in this study uses West Texas Intermediate (WTI) and Brent oil prices. The results of the performance comparison trials for several combination forecasting methods can be concluded that the forecasting results that integrate the EEMD method with JST-GA provide better results compared to the forecasting method that integrates EMD with ANN and EEMD with ANN. The forecasting method developed in this study resulted in forecasting with RMSE / Dstat values of 0.0257 / 61.5936% and 0.0270 / 72.0930% respectively for daily and monthly data from WTI oil types; and the RMSE / Dstat value of 0.0229 / 58.8128% and 0.0300 / 81.5789% respectively for daily and monthly data from the type of Brent oil.

Keywords— crude oil prices; ensemble empirical mode decomposition (EEMD); feedforward neural network (FNN); genetic algorithm;

1. Introduction

The development of crude oil prices is an important concern in the economy [1], industry[2], and government. The effect of crude oil price fluctuations reaches a large number of goods and services that have a direct impact on the economy and society [3][4]. The movement of crude oil prices which tend to be nonlinear [5][6] and non-stationary is influenced by many factors. Demand and supply are factors that influence crude oil price movements[7]. In addition, crude oil price movements are also influenced by many irregular past / present / future events, such as weather, stock levels, Gross Domestic Product, political aspects, and community psychological expectations[8][9].

The characteristics of crude oil price movements that tend to be nonlinear and non-stationary are very important to understand. Most researchers and business practitioners fail to produce good price analysis results consistently, due to the intrinsic complexity of the oil market that is dynamic and
unstable. Empirical Mode Decomposition (EMD) is used to explain new perspectives regarding analytical methods for nonlinear and non-stationary time series. EMD describes a series of time into a number of independent intrinsic modes called Intrinsic Mode Functions (IMFs) based on scale separation. Scale is defined as the distance between two consecutive local extremes, so that this scale can identify the real impact of each mode. For example, intrinsic mode originating from a time series with a three-month scale can be recorded as a seasonal component. However, EMD has the disadvantage of often displaying a mixed mode. The mixed mode defines a single IMF either consisting of widely different scale signals or signals from the same scale in different IMF components.

Ensemble Empirical Mode Decomposition (EEMD) is used to overcome EMD weaknesses [10][11]. Each observation data uses the correct combination of time series and noise. This correct data signal is generated through extraction with the addition of white noise. Thus, the results of the study show that the ensemble average is close to the correct time series if the data is collected by separate observations and different noise levels. This EEMD is used for crude oil price analysis[12]. Crude oil prices with different time frames and frequencies are broken down into several independent intrinsic modes (IMF). Based on an analysis of decomposition of crude oil prices, several forecasting methods can be considered, such as the use of artificial neural networks. One study was conducted by presenting forecasting short-term crude oil prices using the Backpropagation feedforward network [13]. The results showed that feedforward networks were able to predict time series data with the selection of network designs and appropriate training input. However, artificial neural networks often experience overfitting, local optima, and sensitivity to parameter selection[14]. For this reason, artificial neural networks will be optimized through integration with genetic algorithms. Integration of neural networks and genetic algorithms will result in better forecasting than neural networks. Therefore, this study proposes to modify the neural network EMD-based method into EEMD based on artificial neural networks and genetic algorithms for forecasting crude oil prices.

2. Methods
The methods in this research is shown in Figure 1.

![Figure 1. Methods](image-url)
The methods can be explained as follows:

a. Time Series Data

Time series data is data that is displayed based on time and occurs consecutive. Time series data can be used to estimate future, because patterns of data changes in some past periods can be repeated again in the present. This study uses two types of time series data, namely daily and monthly data on West Texas Intermediate (WTI) and Brent.

b. Data Decomposition With EEMD.

The stages of data decomposition with EEMD are done by modifying EMD. Basically, this modification process is related to the decomposition of crude oil prices into several IMF and residues through the addition of white noise. EEMD is used to overcome EMD weaknesses that often display mixed mode. Mixed mode is the result of a signal that is interrupted. EMD is only based on the existence of extrema, the decomposition process will stop if the data does not have the necessary extrema. Noise added to the EEMD decomposition process will be distributed evenly to avoid deterministic deviations in one direction. Every experiment carried out can show very noisy results.

c. Feedforward Neural Network

Before the training and testing process used FNN, the actual data from the EEMD decomposition was carried out by the normalization process. Normalization aims to fulfill the requirements of the binary sigmoid activation function used in this study. Training on the FNN starts with entering the training data pattern into the network to change the weight and bias that connects the neurons. This process is carried out until the network finds the appropriate weights and biases. The process will stop if the number of iterations reaches the maximum value or the specified error value has been reached. Testing is done on input data that has never been trained before. This test uses the weights and biases of the FNN training results. The training and testing process is carried out for each IMF and the residue from the EEMD decomposition. Training and testing on the FNN have need determining data patterns for network architecture. Data patterns are used to determine the number of inputs and network targets. Determination of this data pattern is based on trial and error because there is no theory that can be used as election guidelines.

d. Optimization of weights and biases with genetic algorithms

This step is done by weight optimization and bias with genetic algorithms for FNN output data. The process started from data selection. Selection aims to choose the best parent chromosome for crossovers and mutations. The selection started from generating the initial population randomly. Chromosomes in a population are coded with real numbers. Then the forecasting results are calculated using Adaline. Calculation of forecasting results is done to facilitate the search for fitness values. In this study, the fitness value states the value of the objective function, which minimizes Root Mean Square Error (RMSE).

3. Result

This study uses two types of data, namely daily and monthly. The trial results of comparison forecasting using two prices of WTI crude oil and Brent. Comparisons were made between EEMD based ANN-GA (artificial neural networks - genetic algorithms) with other forecasting methods namely ANN-based EMD and ANN-based EEMD. In this study, the trial was conducted ten times. The results of the trial can be seen in Table 1 and 2 for daily data, while the monthly data can be seen in Table 3 and 4.

| Data Pattern | EMD-JST          | EEMD-JST          | EEMD-JST-GA         |
|--------------|------------------|------------------|---------------------|
|              | RMSE             | Dstat            | RMSE             | Dstat            | RMSE             | Dstat            |
| 15-7-1       | 0.0283           | 56.7592%         | 0.0275            | 59.5345%         | 0.0257           | 61.5936%         |
| 30-3-1       | 0.0308           | 56.2577%         | 0.0285            | 56.6243%         | 0.0275           | 57.7132%         |
| 45-10-1      | 0.0382           | 55.1978%         | 0.0354            | 57.7737%         | 0.0351           | 58.1417%         |
| 60-10-1      | 0.0326           | 56.6449%         | 0.0321            | 57.7425%         | 0.0308           | 58.6754%         |
Table 2. Comparison of forecasting results for Brent daily data

| Data Pattern | EMD-JST | EEMD-JST | EEMD-JST-GA |
|--------------|---------|----------|-------------|
|              | RMSE    | Dstat    | RMSE        | Dstat    | RMSE        | Dstat    |
| 15-10-1      | 0.0306  | 57.4775% | 0.0264      | 59.1892% | 0.0241      | 60.2703% |
| 30-10-1      | 0.0336  | 57.1689% | 0.0236      | 57.3516% | 0.0229      | 58.8128% |
| 45-10-1      | 0.0376  | 56.4815% | 0.0286      | 57.0370% | 0.0265      | 57.5926% |
| 60-10-1      | 0.0293  | 57.6526% | 0.0290      | 60.2817% | 0.0274      | 60.6573% |

Based on Table 1, it can be seen that the results of forecasting using EEMD based on artificial neural networks and genetic algorithms produce the best performance. This forecasting method produces the smallest RMSE in the pattern of 15-7-1 data of 0.0257 and the highest Dstat is 61.5936%. Meanwhile, the results in Table 2 can be seen that forecasting using EEMD based on artificial neural networks and genetic algorithms also produces the best performance. This method produces the smallest RMSE in the 30-10-1 data pattern of 0.0229 and the highest Dstat in the 60-10-1 data pattern of 60.6573%.

Table 3. Comparison of forecasting results for Brent monthly data

| Data Pattern | EMD-JST | EEMD-JST | EEMD-JST-GA |
|--------------|---------|----------|-------------|
|              | RMSE    | Dstat    | RMSE        | Dstat    | RMSE        | Dstat    |
| 6-10-1       | 0.0370  | 60.4651% | 0.0312      | 69.7674% | 0.0270      | 72.0930% |
| 9-10-1       | 0.0398  | 60.0000% | 0.0386      | 62.5000% | 0.0315      | 70.0000% |
| 12-10-1      | 0.0474  | 59.4595% | 0.0393      | 64.8649% | 0.0312      | 70.2703% |

Based on Table 3, it can be seen that the results of forecasting using EEMD based on artificial neural networks and genetic algorithms produce the best performance. This forecasting method produces the smallest RMSE in the 6-10-1 data pattern of 0.0270 and the highest Dstat is 72.0930%. Meanwhile, in Table 4, it can be seen that the best performance forecasting results use EEMD based on artificial neural networks and genetic algorithms. This method produces the smallest RMSE in the 9-10-1 data pattern of 0.0300 and the highest Dstat is 81.5789%.

To see the validation of forecasting results, the results of testing monthly data are used to forecast the prices of WTI crude oil and Brent from January 2013 to June 2013. The forecasting results data are compared with the actual data on crude oil prices as shown in Table 5 for WTI and Brent shown in Table 6. Based on Table 5, it can be seen that the RMSE for WTI data is 6.58 and the variance error is 0.05. While in Table 6 for Brent data produces RMSE of 10.15 and variance error of 0.18. The Variance error generated by the WTI and Brent data is less than 0.3 so that the forecasting results can be concluded that the forecasting is valid.

Table 5. Forecast Results for WTI Monthly Data

| Month      | EEMD - JST - GA | Data Aktual |
|------------|-----------------|-------------|
| January 2013 | 88.99          | 94.76       |
| February 2013 | 89.15         | 95.31       |
| March 2013   | 89.12          | 92.94       |
| April 2013   | 87.87          | 92.02       |
| May 2013     | 86.74          | 94.51       |
| June 2013    | 85.94          | 95.77       |
Table 6. Forecast Results for Brent Monthly Data

| Month       | EEMD – JST - GA | Data Aktual |
|-------------|-----------------|-------------|
| January 2013| 105.77          | 112.96      |
| February 2013| 100.80         | 116.05      |
| March 2013  | 97.51           | 108.47      |
| April 2013  | 96.63           | 102.25      |
| May 2013    | 92.22           | 102.56      |
| June 2013   | 94.27           | 102.92      |

4. Conclusion
From the results of performance comparison trials for several combination forecasting methods it can be concluded that the forecasting results that integrate EEMD with JST-GA (EEMD-JST-GA) provide better results compared to forecasting methods that integrate EMD with ANN (EMD-ANN) and EEMD with ANN (EEMD-ANN). For this, the EEMD-JST-GA forecasting method produces forecasting with RMSE / Dstat values of 0.0257 / 61.5936% and 0.0270 / 72.0930% (respectively for WTI oil daily and monthly data types) and values RMSE / Dstat is 0.0229 / 58.8128% and 0.0300 / 81.5789% (respectively for daily and monthly types of Brent oil). The forecasting performance was better than the performance produced by using the EMD-ANN method (with RMSE / Dstat values of 0.0283 / 56.7592% and 0.0370 / 60.4651% respectively for the types of daily and monthly oil data WTI and 0.0336 / 57.1689% and 0.0503 / 68.4211% respectively for daily and monthly types of Brent oil) and EEMD-ANN methods (with RMSE / Dstat values of 0.0275 / 59.5345 % and 0.0312 / 69.7674% respectively for the types of daily and monthly data of WTI oil and 0.0370 / 72.0930% respectively for the types of daily and monthly data of Brent oil).

5. References
[1] J. Wang, X. Li, T. Hong, and S. Wang, “A semi-heterogeneous approach to combining crude oil price forecasts,” Inf. Sci. (Ny), vol. 460–461, pp. 279–292, 2018.
[2] S. Sun, Y. Sun, S. Wang, and Y. Wei, “Interval decomposition ensemble approach for crude oil price forecasting,” Energy Econ., vol. 76, pp. 274–287, 2018.
[3] S. Kulkarni and I. Haidar, “Forecasting Model for Crude Oil Price Using Artificial Neural Networks and Commodity Futures Prices,” Int. J. Comput. Sci. Inf. Secur., vol. 2, no. 1, 2009.
[4] L. Fan, S. Pan, Z. Li, and H. Li, “An ICA-based support vector regression scheme for forecasting crude oil prices,” Technol. Forecast. Soc. Change, vol. 112, pp. 245–253, 2016.
[5] J. L. Zhang, Y. J. Zhang, and L. Zhang, “A novel hybrid method for crude oil price forecasting,” Energy Econ., vol. 49, pp. 649–659, 2014.
[6] Y. Chen, K. He, and G. K. F. Tso, “Forecasting Crude Oil Prices: A Deep Learning based Model,” Procedia Comput. Sci., vol. 122, pp. 300–307, 2017.
[7] H. Miao, S. Ramchander, T. Wang, and D. Yang, “Influential factors in crude oil price forecasting,” Energy Econ., vol. 68, pp. 77–88, 2017.
[8] L. Yu, S. Wang, and K. K. Lai, “Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm,” Energy Econ., vol. 30, no. 5, pp. 2623–2635, 2008.
[9] A. Safari and M. Davallou, “Oil price forecasting using a hybrid model,” *Energy*, vol. 148, pp. 49–58, 2018.

[10] M. Latif and S. Herawati, “The application of EEMD and neural network based on Polak-Ribiére Conjugate Gradient algorithm for crude oil prices forecasting,” in *MATEC Web of Conferences*, 2016, vol. 58.

[11] L. Li, X. Qu, J. Zhang, H. Li, and B. Ran, “Travel time prediction for highway network based on the ensemble empirical mode decomposition and random vector functional link network,” *Appl. Soft Comput. J.*, vol. 73, pp. 921–932, 2018.

[12] X. Zhang, K. K. Lai, and S. Y. Wang, “A new approach for crude oil price analysis based on Empirical Mode Decomposition,” *Energy Econ.*, vol. 30, no. 3, pp. 905–918, 2008.

[13] I. Haidar, S. Kulkarni, and H. Pan, “Forecasting model for crude oil prices based on artificial neural networks,” *Proc. 2008 Int. Conf. Intell. Sensors, Sens. Networks Inf. Process.*, pp. 103–108, 2008.

[14] M. Nasseri, K. Asghari, and M. J. Abedini, “Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network,” *Expert Syst. Appl.*, vol. 35, no. 3, pp. 1415–1421, 2008.