IJSA (2025), Vol. 9, No. 2, 230-239
DOI: https://doi.org/10.29244/ijsa.v9i2p230-239
Indonesian Journal of Statistics and Its Applications
Available at https://journal-stats.ipb.ac.id/index.php/ijsa
e-ISSN: 2599-0802

OPEC Crude Oil Price Forecasting Using ARIMA with Ensemble Empirical Mode Decomposition

Tiara Lutfiah Adisti[1‡], Agus Mohamad Soleh[1], Aam Alamudi[1], Septian Rahardiantoro[1], and Akbar Rizki[1]

[1]School of Data Science, Mathematics, and Informatics, IPB University, Indonesia
‡corresponding author: agusms@apps.ipb.ac.id

*

Abstract

World crude oil prices fluctuate every day. One source of crude oil traded is oil from crude oil exporting countries that are members of the Organization of the Petroleum Exporting Countries (OPEC). In the total of 40% of world crude oil is produced by OPEC. This makes forecasting the price of crude oil OPEC's policy very necessary in order to maintain world oil market stability. Fluctuating oil price data is made simpler and easier to interpret by applying the Ensemble Empirical Mode Decomposition (EEMD) method. The EEMD method decomposes the data into a number of Intrinsic Mode Functions (IMF) and residual of the IMF. In this study, the ARIMA forecasting model is compared using the original data and the decomposition results in the form of IMF components and IMF residuals. The comparison of the two methods is seen based on the overall and average MAPE value of the forecasting results in five time ranges. The EEMD-ARIMA method has an average MAPE value of 9.09% and standard deviation MAPE value of 7.39%. OPEC crude oil price forecast in January-August 2021 ranges from $42.22 to $60.6 per barrel. The final result of the analysis in this study shows that the ARIMA method with decomposition data (EEMD-ARIMA) is better than the ARIMA method using original data.

Keywords: ARIMA, EEMD, EMD, forecast, OPEC crude oil price.

* Received: Nov 2025; Reviewed: Apr 2022; Published: Dec 2025

1. Introduction

Crude oil is a non-renewable natural resource. High demand for oil causes fluctuations in global crude oil prices. Crude oil prices are influenced by several factors, one of which is oil production by member countries of the Organization of the Petroleum Exporting Countries (OPEC). According to Leoresta and Sulasmiyati (2017), OPEC countries' oil production to supply oil is quite large, accounting for about 40% of total global crude oil production. Similar to other types of oil, OPEC crude oil prices also experience instability caused by various factors, one of which is uncertain oil production restrictions.

High crude oil prices will cause an increase in the price of fuel products (Cologni and Manera 2008). The negative effects of rising oil prices also affect investment and stock prices. Therefore, forecasting OPEC crude oil prices is needed to prevent deficits in various countries, help companies maintain production, and make it easier for investors to predict stock price movements.

One method that can be used to analyze crude oil prices is the Autoregressive Integrated Moving Average (ARIMA). The ARIMA method assumes a linear function based on past data from variables used for forecasting. Nonlinearity and nonstationarity in the data can be overcome by simplifying the data using the Empirical Mode Decomposition (EMD) method, which will produce a number of Intrinsic Mode Functions (IMF) and IMF residuals (Huang et al. 1998). Wu and Huang (2005) developed the Ensemble Empirical Mode Decomposition (EEMD) method, which can overcome mode mixing, a weakness of EMD. The EEMD method will be applied to crude oil price data to convert non-linear and non-stationary data into IMF components and IMF residuals that tend to be stationary. In previous research, Fitri et al. (2020) successfully applied EEMD to forecast beef prices and produced a Mean Absolute Percentage Error (MAPE) value of less than 5%, which means that the forecasting method is very good. Therefore, this study compares the forecasting models produced by the ARIMA method without EEMD integration and the ARIMA method previously integrated with the EEMD method to forecast OPEC crude oil prices.

The objectives of this study are (1) to apply the EEMD method to monthly OPEC crude oil prices, (2) to build ARIMA and EEMD-ARIMA models on OPEC crude oil price data, (3) to evaluate the ARIMA and EEMD-ARIMA models in forecasting OPEC crude oil prices, and (4) to forecast monthly OPEC crude oil prices from January to August 2021 using the best method.

2. Methodology

2.1 Data

The data used in this study is secondary data obtained from www.opec.org, the official website of the Organization of the Petroleum Exporting Countries (OPEC). The data obtained is daily world crude oil price data in US dollars ($) per barrel as determined by OPEC. This study uses monthly data from January 2008 to August 2021, which is the result of the average daily price of crude oil for each month.

2.2 Data Analysis Procedures

The data analysis procedure to be performed is as follows:

1. Explore the data by plotting the monthly price movements of OPEC crude oil from January 2008 to December 2020.
2. Determine the IMF and IMF residuals according to the EEMD algorithm as follows:
   a. Add white noise to the time series data. *White noise*$\sim\mathrm{N}(0, \sigma^2)$.
   b. Decomposing the data that has been given white noise into several IMFs and IMF residuals using the EMD algorithm.
   c. Repeating steps a and b iteratively *N* times. Each iteration is performed with different white noise.
   d. Calculating the IMF from the ensemble mean $c_i(t) = \frac{1}{N}\sum_{j=1}^{N} c_{ij}(t)$ and the decomposition residual formed from the iteration, $r = \frac{1}{N}\sum_{j=1}^{N} r_j$, where j is the number of iterations performed $(j = 1, 2, \ldots, N)$ and i is the IMF (Su *et al.* 2016).

   Based on the research by Zhang *et al.* (2008) ensemble iterations in the EEMD algorithm can be performed 100 times, and the standard deviation of white noise that can be used is 0.1 or 0.2.
3. Determine the contribution of each IMF to the original data using the period mean, Pearson correlation, and percentage of variance ratio.
4. Creating a forecasting model using the ARIMA method from ensemble decomposition data and original data. The training data is divided into 90% and the test data into 10%. The division of training data and test data can be seen in Table 1.

   Tabel 1: Training and test data distribution

No.	Training data	Test Data
1	January 2008 – December 2015	January – December 2016
2	January 2009 – December 2016	January – December 2017
3	January 2010 – December 2017	January – December 2018
4	January 2011 – December 2018	January – December 2019
5	January 2012 – December 2019	January – December 2020
6	January 2008 – December 2018	January 2019 – December 2020

   ARIMA modeling for original data and decomposed data is performed using the following steps:
   a. Check stationarity in the mean and variance. Stationarity in the mean is checked using the Augmented Dicky Fuller (ADF) test, and differencing is performed if the data is not stationary in the mean. Stationarity in the variance is checked using the Levene test. Data that is non-stationary in variance is transformed.
   b. Identify the ARIMA model based on the ACF and PACF plots on data that is stationary in variance and mean.
   c. Estimate the parameters and test the significance of the parameters of the tentative model that is formed.
   d. Perform diagnostic tests on the model with significant parameter estimates.
   e. Perform overfitting by adding orders to the obtained model and recheck the significance of the parameters and model diagnostics.
   f. Determine the best model among the candidate models that meet all assumptions and have the smallest Akaike's Information Criterion (AIC) value (Montgomery et al. 2008)
5. Perform forecasting for each test data period based on Table 1, and evaluate each model with Mean Absolute Percentage Error (MAPE). Calculate the MAPE value as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \times 100\% \quad (1)$$

6. Compare the overall, average, and standard deviation values of the MAPE for the ARIMA and EEMD-ARIMA methods.
7. Perform monthly crude oil price forecasting for the OPEC period from January to August 2021 using the model based on the best method.

3. Results and Discussion

3.1 Data Exploration

The monthly price of OPEC crude oil from 2008 to 2020 fluctuated as shown in Figure 1. The highest price occurred in July 2008 with an oil price of \$131.22 per barrel. According to Mawikere (2016), this price increase was caused by economic and population growth, the dominance of the US dollar, and geopolitical issues in the Middle East, which forced OPEC to regulate oil production from member countries. In 2008, oil prices also experienced a significant decline, as in mid-2014. The decline in crude oil prices was caused by an oversupply that did not balance the market as shale gas supplies from the United States continued to increase. In 2015, oil prices stabilized again as oil demand increased by around 1.4% (Ilahi 2016).

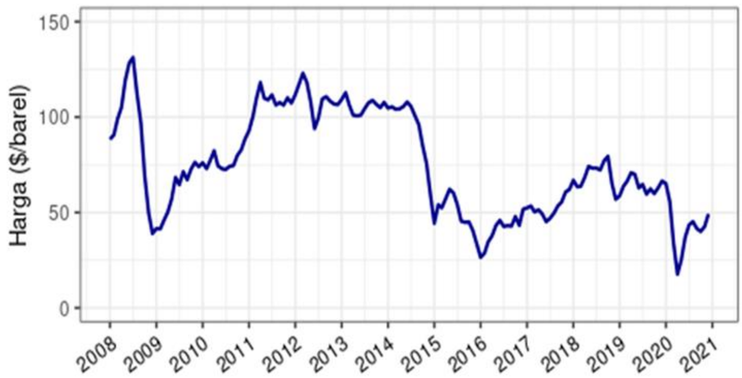


Figure 1: Monthly crude oil prices of OPEC (2008-2020)

The lowest price occurred recently in April 2020 at \$17.66 per barrel. This price was the lowest point of the decline in oil prices that occurred during the Covid-19 pandemic. Oil prices rose again in May until the end of 2020 because OPEC countries agreed to continue cutting production by 7.2 million barrels per day. This decision was accompanied by a return to normal global economic activity, causing oil prices to rise again.

3.2 Data Decomposition

Monthly crude oil price data for the period 2008-2020 was analyzed using the Ensemble Empirical Mode Decomposition (EEMD) method in stages, starting from high frequency decomposition to low frequency. This EEMD process produced five IMFs and residual IMFs. The IMF components produced are independent of one another. The decomposition of OPEC crude oil monthly prices produced five IMFs and an IMF residual (Figure 2).

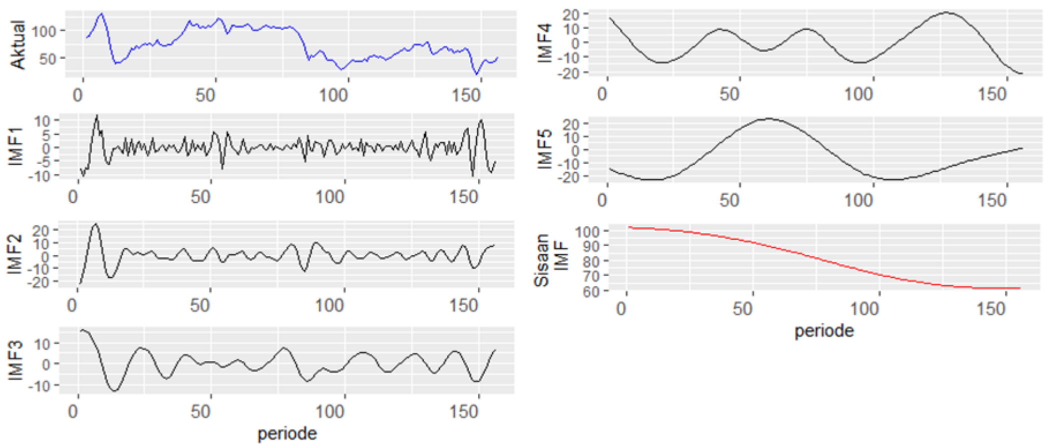


Figure 2: Results of ensemble EMD decomposition of oil prices

The contribution of the IMF and the IMF residual can be seen from the number of peaks and troughs, the period mean, the correlation with the original data (input), the variance, and the percentage of the variance ratio shown in Table 2. The decreasing frequency can be seen from the decreasing number of peaks and troughs, which makes the period mean larger. The contribution of the IMF can also be seen from the closeness of the relationship between the decomposition results and the actual data based on their correlation. The greater the correlation coefficient, the more relevant the IMF is to the input.

Table 2: Description of IMF components and IMF error

	Number of Peaks	**Number of Valleys**	**Period Average (months)**	**Period Average (years)**	**Pearson Correlation**	**Variety**	**Variety Ratio (%)**
Actual						731.02	
IMF1	48	49	3.22	0.27	0.18	11.84	1.62
IMF2	18	18	8.67	0.72	0.30	38.40	5.25
IMF3	10	10	15.60	1.30	0.43	28.80	3.94
IMF4	3	3	52.00	4.33	0.41	112.95	15.45
IMF5	1	2	104.00	8.67	0.64	219.36	30.01
Error					0.56	221.19	30.26
Total							86.53

Table 2 shows the contribution of the IMF to the actual data based on the variance value and variance ratio percentage. IMF5 and the remaining IMF have a fairly large percentage of variance, which is around 30%. The greater the percentage of variance in the decomposition results, the more actual data information is explained by the IMF. The decomposition results show a good contribution in the four- and eight-year periods. Overall, the total variance that can be explained by all IMF and the remaining IMF is only 86.53%. The decomposition component that is unable to explain all of the actual data variance may be due to a combination of rounding errors, nonlinearity of the input, and variance based on the cubic spline interpolation process (Peel et al. 2010).

3.3 Oil Price Forecasting with EEMD-ARIMA

The results of the ensemble decomposition of five IMF components and IMF residuals for the period January 2008 – December 2018 were used to create a forecasting model. Each IMF and IMF residual was modeled using the ARIMA method. These components were tested for stationarity in mean and variance before modeling. IMF and IMF residuals that were not stationary were differentiated and transformed until they became stationary. The selection of the best model was also based on the smallest Akaike's Information Criterion (AIC) value for each component series. The model forecast results from each IMF and IMF residual were summed to produce a complete forecast of OPEC crude oil prices. Furthermore, the forecast results are compared with the test data, namely the IMF components and IMF residuals for the period January 2019 to December 2020. The best ARIMA model for each IMF and IMF residual can be seen in Table 3.

Table 3: The best model for each IMF and residual IMF monthly crude oil prices for the period January 2008 – December 2018

Component	Model	AIC
IMF1	ARIMA(3;0;1)	634.07
IMF2	ARIMA(2;0;5)	251.86
IMF3	ARIMA(4;0;2)	-335.69
IMF4	ARIMA(3;2;0)	-1385.54
IMF5	ARIMA(0;3;2)	-1668.22
IMF error	ARIMA(1;5;0)	-288.81

Based on Figure 3, the plot between the test data and the forecast has a large difference in several months with a MAPE value of 23.1%. The plot shows that the forecast results follow a downward pattern similar to the actual data, although they are less accurate in the early to mid-2020 period. The unexpected Covid-19 pandemic caused the oil price movement plot to decline more sharply before starting to increase towards the end of 2020.

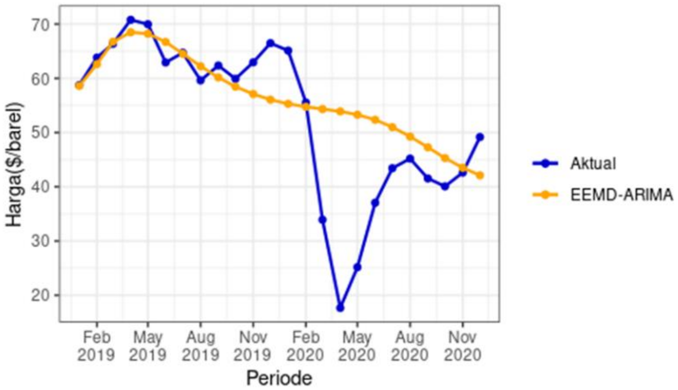
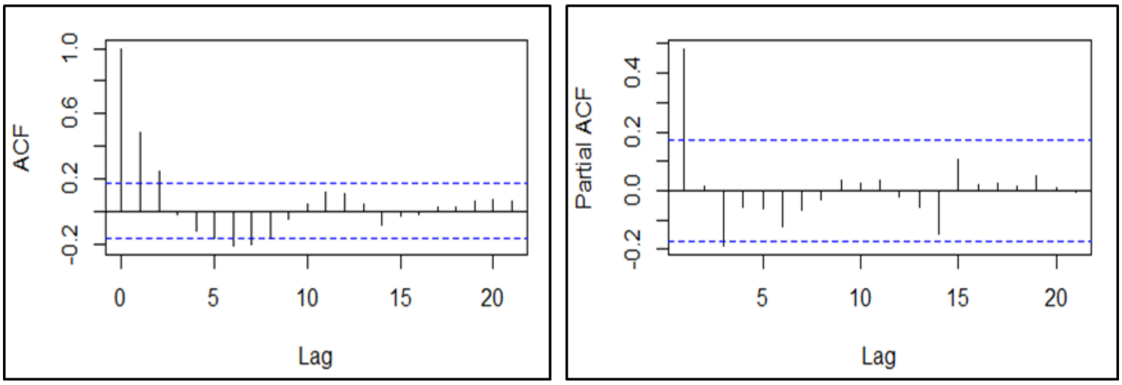


Figure 3: Actual data and oil price forecast results with EEMD-ARIMA (testing)

3.4 Oil Price Forecasting with ARIMA

The ARIMA forecasting method for monthly crude oil price data was identified based on ACF and PACF plots. Modeling for data from January 2008 to December 2018 was identified after the data was found to be stationary in mean and variance. The data was tested for stationarity in the mean by looking at the ACF and PACF plots and the Augmented Dickey Fuller (ADF) test. Based on the ADF test, a p-value of 0.511 was obtained, which is greater than the 5% significance level, meaning that the data is not stationary in the mean. A one-time differentiation was performed on the data. Based on Figure 4, the ACF and PACF plots did not form a specific pattern, and the ADF test produced a p-value of 0.01, meaning that the data was stationary in mean. Next, a test for stationarity in variance was conducted. Based on the Levene test, a test statistic of



2.529 and a p-value of 0.0831 were obtained, meaning that the data was stationary in variance.

Figure 4: Plot of stationarity tests that are stationary in the mean. (a) Plot of ACF and (b) plot of PACF after differencing

Data that has been stationary in variance and mean is then identified based on the ACF and PACF plots resulting from differencing (Figure 4). The initial tentative models that were the best model candidates based on the ACF and PACF plot patterns were ARIMA(1;1;0), ARIMA(1;1;1), ARIMA(1;1;2), ARIMA(3;1;0), ARIMA(3;1;1), and ARIMA(3;1;2). Parameter estimation was performed for the five model candidates with the results shown in Table 4.

Table 4: Estimated values of tentative ARIMA model parameters

Model	Parameter	Koefisien parameter	*P-value*	AIC
ARIMA(1,1,0)	AR(1)*	0.488	<0.0001	833.88
ARIMA(1,1,1)	AR(1)*	0.509	<0.0001	835.83
	MA(1)	-0.028	0.8337	
ARIMA(1,1,2)	AR(1)	0.155	0.5125	833.43
	MA(1)	0.325	0.1511	
	MA(2)*	0.293	0.0139	
ARIMA(3;1;0)	AR(1)*	0.482	<0.0001	832.33
	AR(2)	0.124	0.1987	
	AR(3)*	-0.207	0.0181	
ARIMA(3,1,1)	AR(1)*	1.096	<0.0001	832.37
	AR(2)	-0.179	0.3008	
	AR(3)*	-0.202	0.0343	
	MA(1)*	-0.651	0.0027	
ARIMA(3,1,2)	AR(1)*	0.889	0.0116	834.05
	AR(2)	0.120	0.7945	
	AR(3)	-0.310	0.0755	

MA(1)	-0.438	0.2298
MA(2)	-0.202	0.5341

* significant parameters at the actual level 5%

Based on the results in Table 4, only the ARIMA(1;1;0) model has all significant parameter estimates. The model was then tested for model diagnostics using the Ljung-Box test and overfitting. The results of the overfitting model parameter estimation were not significant, so the ARIMA(1;1;0) model is the best model to use for oil price forecasting.

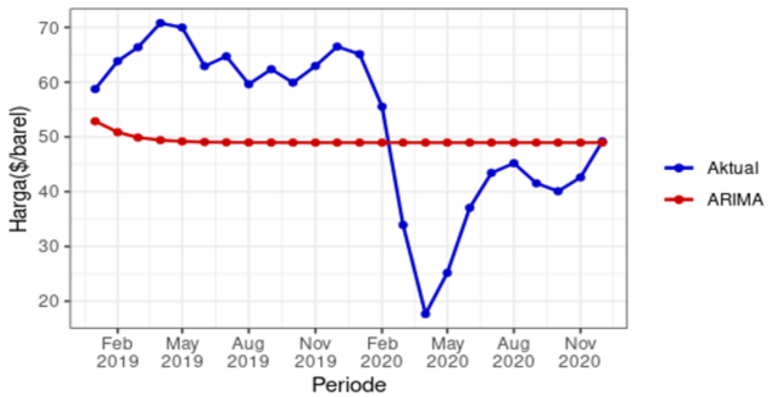


Figure 5: Plot actual data and oil price forecast results with ARIMA (testing)

The ARIMA forecasting results based on Figure 5, the actual data plot, and the forecasting results for the test data are very different, with a MAPE value of 30.36%. Exploratively, this shows that the forecasting results are not good for predicting crude oil prices.

3.5 Comparison of ARIMA and EEMD-ARIMA Methods

The best method for forecasting OPEC crude oil prices can be obtained by comparing the MAPE values of the forecast results from the candidate methods. Repeated forecasting analysis with several different data ranges can be performed to ensure that the method is the best method that can be used for creating a crude oil price forecasting model. A comparison of the MAPE values of the forecast results from five data sets with different time ranges can be seen in Table 5.

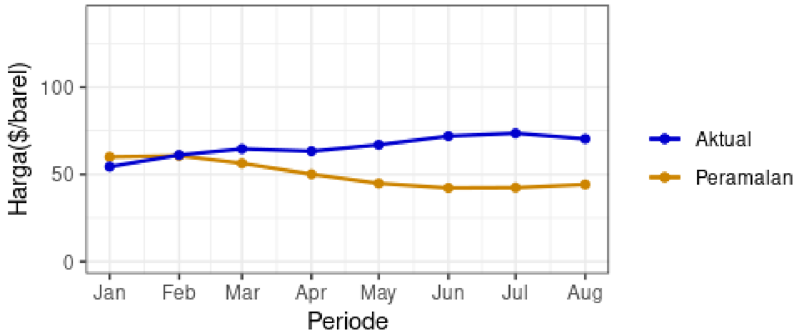
Table 5: Comparison of MAPE values between ARIMA and EEMD-ARIMA methods

Data	MAPE test data (%)	
	ARIMA	EEMD-ARIMA
2008-2016	35.82	11.56
2009-2017	9.52	3.03
2010-2018	10.70	6.96
2011-2019	17.61	3.16
2012-2020	83.42	20.74
Average	31.41	9.09
Standard Devation	30.92	7.39

Based on Table 5, the EEMD-ARIMA method has a smaller MAPE value than the ARIMA method in all data ranges. The average MAPE value is 9.56% and the standard deviation is 7.39%, indicating that the EEMD-ARIMA method is better for forecasting OPEC crude oil prices.

3.6 Oil Price Forecast for January-August 2021

Based on a comparison of the forecasting results between the ARIMA and EEMD-



ARIMA methods, the forecasting of OPEC crude oil prices for the January-August 2021 period was carried out using the best method, namely EEMD-ARIMA. The forecasting was carried out using all data from January 2008 to December 2020. Based on Figure 6, it can be seen that the actual data and the forecast results for OPEC crude oil prices for the period January-August 2021 are not much different.

Figure 6: Crude oil price forecast for January-August 2021

4. Conclusions and Recommendations

4.1 Conclusions

The ensemble-decomposed OPEC crude oil price data is easier to interpret and provides more detailed information to explain movements during a specific period. Based on modeling five data ranges, the EEMD-ARIMA method has an average MAPE value of 9.09%, while the ARIMA method has an average MAPE value of 31.41%. In addition, the standard deviation of the MAPE value of the EEMD-ARIMA method is 7.93% and that of the ARIMA method is 30.92%. Therefore, forecasting crude oil prices using IMF components and IMF residuals is better than using the original data.

4.2 Recommendations

The results of oil price data forecasting using the EEMD-ARIMA method still produce a relatively large MAPE. This is due to the influence of unexpected events called interventions. A suggestion for further research is to use an intervention model integrated with an EMD ensemble to improve the accuracy of the forecasts produced.

Reference

Cologni, A., Manera, M. (2008). Oil prices, inflation and interest rates in a structural cointegrated VAR model for the G-7 countries. *Energy Economics*, (30), 856–888.

Fitri, H. (2020). *Analisis harga daging sapi menggunakan ensemble empirical mode decomposition* [tesis]. Bogor: Institut Pertanian Bogor.

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. (1998). The Empirical Mode Decomposition and the Hilbert

Spectrum for Nonlinear and Nonstationary Time Series Analysis. *Proc. Roy. Soc. Lond A*, (454), 903-995.

Ilahi, R. (2018). Dampak kebijakan pemangkasan produksi minyak dunia oleh Organization Of The Patroleum Exporting Countries (OPEC) terhadap Indonesia. *Jurnal Online Mahasiswa FISIP Universitas Riau*, 5(1),1-5.

Leoresta, M.P.A., Sulasmiyati, S. (2017). Pengaruh produksi minyak OPEC, GDP manufacture output, konsumsi minyak, dan net ekspor manufaktur terhadap fluktuasi harga minyak OPEC (Studi pada 5 negara manufaktur terbesar dan perbandingannya dengan Indonesia periode 1980-2015). *Jurnal Administrasi Bisnis*, 50(5), 152-151.

Mawikere, J.C. (2008). Implikasi Kuota Produksi Minyak. Organization of the Petroleum Exporting Countries. (OPEC) dengan Kebijakan Keanggotaan dan Harga. *Jurnal Analisis Hubungan Internasional*, 5(3), 126-137.

Montgomery, D.C., Jennings, C.L., Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. New Jersey(US): Wiley.

Peel, M.C., Amirthanathan, G.E., Pegram, G.G.S., McMahon, T.A., Chiew, F.H.S. (2005). Issues with the application of empirical mode decomposition analysis. *International Congress on Modelling and Simulation,* 1681 – 1687.

Su, H., Li, H., Chen, Z., Wen, Z. (2016). An approach using ensemble empirical mode decomposition to remove noise from prototypical observations on dam safety. *Springer Plus*, 5, 650. DOI 10.1186/s40064-016-2304-4.

Wu, Z., Huang, N.E. (2005). Ensemble empirical mode decomposition: a noise assisted data analysis method. *Advances in Adaptive Data Analysis,* 1(1), 1-41.

Zhang, X., Lai, K.K., Wang, S.Y. (2008). A New Approach for Crude Oil Price Analysis Base on Empirical Mode Decomposition. *Energy Economics*, 30,905-918.