

## Long-Term Forecasting for World Oil Production

Hosam H. A. Mukhairez  
*Faculty of Information Technology*  
*Islamic University of Gaza*  
*Gaza, Palestine*

### Abstract

*Natural resources are limited and very important in our industrial life and development. Oil is considered the black gold and it is included in hundreds of industrial fields.*

*In this paper, the oil production amounts produced from the world continents and regions (Africa, Asia, Europe, Middle East, North America and S. & Cent. America) was forecasted.*

*The forecasting is conducted using the following forecasting algorithms: Auto-Regressive Integrated Moving Average (ARIMA), Hybrid ARIMA, and Linear Regression with horizon value equals ten (2017-2026). After evaluating tested algorithms using MAPE it is noticed that Hybrid ARIMA was the best forecasting algorithm to be used over Africa (MAPE 5.05%), Asia (MAPE 1.29%) and Europe (MAPE 2.85%) datasets. While ARIMA was the best forecasting algorithm to be used for the Middle East (MAPE 3.06%) and North America (MAPE 18.25%). Finally, Linear regression was the most appropriate forecasting algorithm to be used for South & Cent. America (MAPE 1.62%). The deviation of the forecasted 10 years (2017-2026) of Africa and Asia is decreasing with comparison to the last 10 years (2007-2016), while the forecasted 10 years to the rest of all region is increasing with comparison to last 10 years.*

### 1. Introduction

The term "Oil production" includes crude oil, tight oil, oil sands and NGLs (the liquid content of natural gas where this is recovered separately). The data excludes liquid fuels from other sources such as biomass and derivatives of coal and natural gas.

Oil production presents 32% of the world natural resources production operation, while coal production presents 27%, gas 21% biomass 11% and electricity 9% [1].

The U.S. Energy Information Administration [2] defined the crude Oil by: A mixture of hydrocarbons that exists in liquid phase in natural underground reservoirs and remains liquid at atmospheric pressure after passing through surface separating facilities. Depending upon the characteristics of the crude stream, it may also include (1) Small amounts of hydrocarbons that exist in gaseous phase in natural underground reservoirs but are liquid at atmospheric

pressure after being recovered from oil well (casing head) gas in lease separators and are subsequently comingled with the crude stream without being separately measured. Lease condensate recovered as a liquid from natural gas wells in lease or field separation facilities and later mixed into the crude stream is also included; (2) Small amounts of nonhydrocarbons produced with the oil, such as sulfur and various metals; (3) Drip gases, and liquid hydrocarbons produced from tar sands, oil sands, gilsonite, and oil shale. Liquids produced at natural gas processing plants are excluded. Crude oil is refined to produce a wide array of petroleum products, including heating oils; gasoline, diesel and jet fuels; lubricants; asphalt; ethane, propane, and butane; and many other products used for their energy or chemical content.

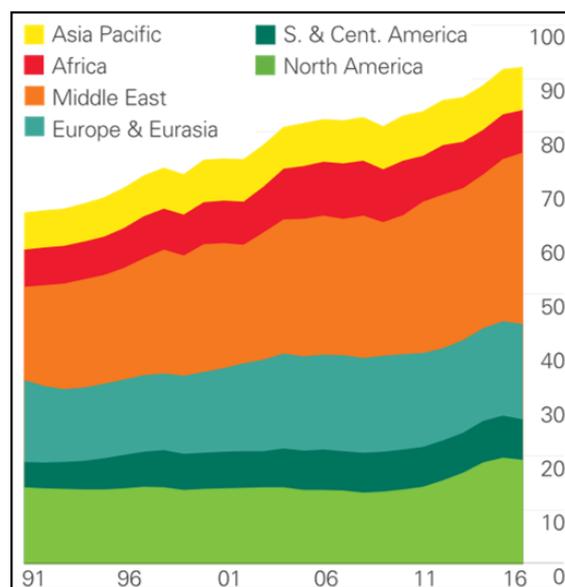


Figure 1. Oil production by regions (million barrels) [3]

Figure 1 presents the regions and continents which is the focus of this research showing that the middle east is the most region in oil production rates. For many years, the top three producers have been (alphabetically) Russia, Saudi Arabia, and the United States. Each of these countries experienced major production declines at different times in the past, but

since 2014 all three have been producing near their peak rates of 9 to 11 million barrels per day [3].

Forecasting task is understanding and recognizing the old behavior of an attributes to predict its future behavior or pattern. It is one of the oldest known predictive analytics techniques. The idea is to use historical data to make forecasts and prediction about future data. Forecasting is one of the most common techniques of time series data analysis. It's used to predict future trends in water demands, retail sales, economic indicators, weather forecasting, stock markets, and many other application scenarios [4].

Depending on the forecasting time horizon term, oil production forecasts can be classified, as suggested by [5] into: (1) Long-Term which forecasts decades, (2) Medium-Term which forecasts years to decade, (3) Short-Term which forecasts up to year. In this research, data was downloaded from BP Global and presented in annual form starting from 1965 to 2016.

In this work some forecasting algorithms were used and chosen the best one depending on the mean absolute percentage error 'MAPE'.

The rest of this paper is organized as follows: Section 2 reviews some related works. Section 3 shows experimental process and results. Section 4 presents discussion. Section 5 the conclusion.

## 2. Related Work

Many authors researched on oil production forecasting and its effects on countries politics actions or decisions, for example Nashawi et al. [6] proposed a model which has multi cycle depending on old data related to oil production and this model updated automatically when new data are available. They observed that the peak in 2014 will be 79 MMSTB/D. They noticed that OPEC has remaining reserve of 909 BSTB, which is about 78% of the world reserves.

Aizenberg et al. [7] used Multilayer Neural Network with Multi-Valued Neurons (MLMVN) to conduct log-term forecasting for oil production. They proposed to use a complex-valued neural network for forecasting and studied some important aspects of the application of ANN models.

Kontorovich et al. [8] studied the key trends of long-term and medium-term scenarios for the world primary energy sources oil, gas and coal.

Abas et al. [9] reviewed the future energy technologies, the noticed that the primary energy sources production rates would continue rising till 2055. Also, they observed that energy demand would increase from 12.5 GTOE 2015 to 25 GTOE by 2035.

Frausto-Solís et al. [10] proposed a new method "SAM-oil". The new method combines evolutionary computing techniques, quality assessment of solutions and statistical analysis with simulated

annealing metaheuristic. They compared the accuracy of "SAM-oil" with ARIMA, neural networks, and some members of the exponential smoothing family. They claim that SAM-oil's is competitive method in oil production forecasting. Efendi and Deris [11] studied the Malaysian oil production and oil consumption. They used fuzzy time series (FTS) model for oil data from 1965 to 2012 and they concluded that FTS is better than regression time series.

## 3. Experimental Results

The chosen forecasting algorithms can be labeled as the most recent and famous algorithms used in time series analysis field. These algorithms are Auto Regressive Integrated Moving Average (ARIMA), ARIMA combined with Neural Networks (Hybrid ARIMA) and Linear Regression forecast. These algorithms were run over datasets to calculate the Mean Absolute Percentage Error (MAPE) to detect which algorithms gives the lowest MPE, which mean the more accuracy algorithm to be used in further.

The data set is real data set collected from BP Global. The data set belongs to the yearly oil produced amount from most regions from the world starting from 1965 to 2016.

The original data consists of 7 columns (year, Africa, Asia, Europe, Middle East, North America, Center and South America). The data sets containing 52 years of indices data, was divided into training set, containing the first 47 years of data 1965 to 2011, and testing set holding the remaining data from 2012 to 2016. Models were trained on the training sets before being applied to the unseen data in the test sets. Table 1 gave a sample of training set for both attributes (consumption and revenue), and Table 2 gave sample of testing set for both attributes (consumption and revenue).

The testing set presents 10% of the overall dataset, and this is the period from 2012 to 2016. To choose the appropriate forecasting algorithms I first evaluated and compared the results from each algorithm with the original data we have. Then three forecasting algorithms were run on the training set and with horizon value 5 to reach 2016 to compare it with the actual data in testing set, at this stage every region was forecasted separately.

### 3.1. Evaluating algorithms over six regions

First step, evaluation process of the selected three algorithms done over the main six areas available in the datasets and compute the mean absolute percentage error MPE for each algorithm.

Table 1. Sample of the training set from 1965 to 2011

Year	Africa	Asia	Europe	Middle East	North America	S. & Cent. America
1965	106.5	44.9	281.9	418.7	489.6	226.3
1966	135.4	50.0	304.7	467.1	521.2	223.9
1967	149.1	53.0	329.0	499.8	557.4	238.3
1968	191.0	62.0	350.7	562.5	581.9	244.7
1969	242.7	71.4	370.3	617.3	596.5	245.1
1970	292.3	98.8	395.0	692.4	627.7	251.8
1971	273.8	118.6	418.6	814.1	625.2	245.9
1972	275.1	139.7	443.5	908.3	639.7	234.3
1973	287.1	166.8	472.4	1055.0	640.8	249.3
1974	264.9	177.8	502.5	1089.0	618.2	227.1

Table 2. Sample of the testing set from 2012 to 2016

Year	Africa	Asia	Europe	Middle East	North America	S. & Cent. America
2012	440.1	400.2	833.6	1344.0	719.6	378.9
2013	408.9	393.9	833.3	1326.1	783.8	379.2
2014	394.2	396.5	834.7	1338.7	869.2	392.9
2015	393.7	400.0	847.3	1411.6	908.3	398.6
2016	374.8	383.0	860.6	1496.9	882.6	384.5

Table 3. Calculated MAPE for three algorithms over six regions

MAPE	Africa	Asia	Europe	Middle East	North America	S. & Cent. America
ARIMA	7.3	4.1	3.1	<u>3.06</u>	<u>18.25</u>	2.1
Hybrid ARIMA	<u>5.05</u>	<u>1.29</u>	<u>2.85</u>	3.52	19.2	2.0
Linear Regression	9.0	6.0	4.0	3.08	19.8	<u>1.62</u>

$$\text{MAPE} = (((\text{Actual} - \text{Predicted}) / \text{Actual}) * 100)$$

This evaluation provides insights in recommending changes to existing models to reduce deviations in future forecasts.

Table 3 presents the values of MAPE for the three selected forecasting algorithms over the six regions. It's observed that Hybrid ARIMA was the best forecasting algorithm to be used over datasets belongs to Africa (MAPE 5.05%), Asia (MAPE 1.29%) and Europe (MAPE 2.85%). While ARIMA was the best forecasting algorithm to be used over datasets belongs the Middle East (MAPE 3.06%) and North America (MAPE 18.25%). Finally, Linear regression was the most appropriate forecasting algorithm to be used for South & Cent. America dataset (MAPE 1.62%).

Now after choosing every appropriate algorithm for each region, I'm going to forecast 10 years in advance for all the six regions.

### 3.2. Forecasting oil production of Africa using Hybrid ARIMA

Figure 2, Figure 4, Figure 5, Figure 6 and Figure 7 show that the trend of oil production in the next 10 years is increasing in Africa, Europe, Middle East, North America and South & Central America, but Figure 3 shows that the trend of oil production is decreasing in Asia.

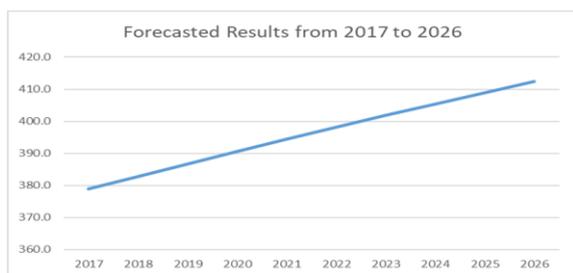


Figure 2. Forecasted results for Africa oil production from 2017 to 2026

### 3.3. Forecasting oil production of Asia using Hybrid ARIMA

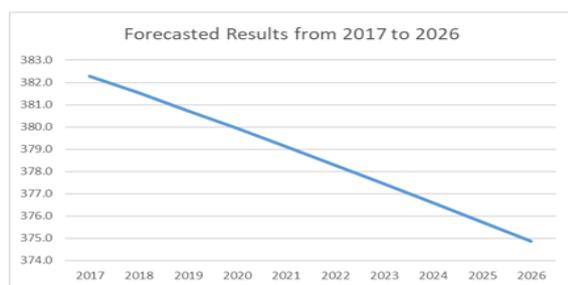


Figure 3. Forecasted results for Asia oil production from 2017 to 2026

### 3.4. Forecasting oil production of Europe using Hybrid ARIMA

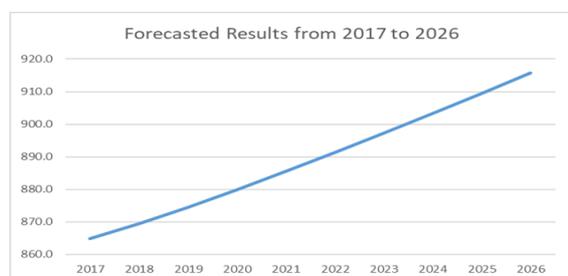


Figure 4. Forecasted results for Europe oil production from 2017 to 2026

### 3.5. Forecasting oil production of Middle East using ARIMA

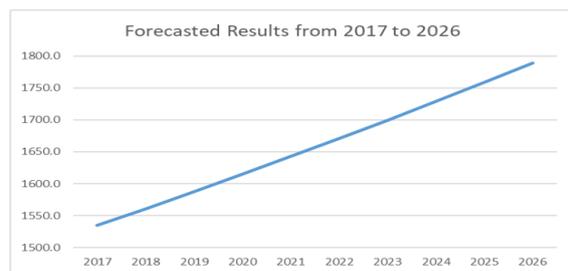


Figure 5. Forecasted results for Middle East oil production from 2017 to 2026

### 3.6. Forecasting oil production of North America using ARIMA

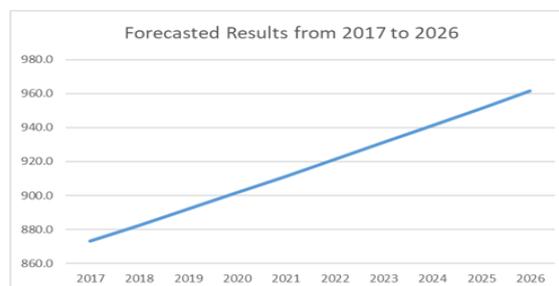


Figure 6. Forecasted results for North America oil production from 2017 to 2026

### 3.7. Forecasting oil production of South and Central America using Linear Regression

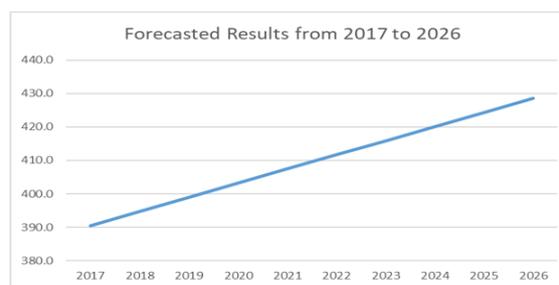


Figure 7. Forecasted results for South & Central America oil production from 2017 to 2026

### 3.8. Results Deviation

Table 4 Shows that the deviation of the sum of forecasted 10 years (2017 - 2026) of oil production with comparison to the last 10 years (2007 - 2016). The deviation of Africa and Asia is decreasing, while the rest of all region is increasing.

Table 4. Deviation of forecasted years compared to last 10 years

Years	Africa	Asia	Europe	Middle East	North America	S. & Cent. America
2007 - 2016	4329.8	3926.1	8494.8	13120.6	7327.9	3825.7
2017 - 2026	3960.3	3786.6	8891.4	16587.7	9166.7	4095.6
Deviation%	-9.3	-3.7	4.5	20.9	20.1	6.6

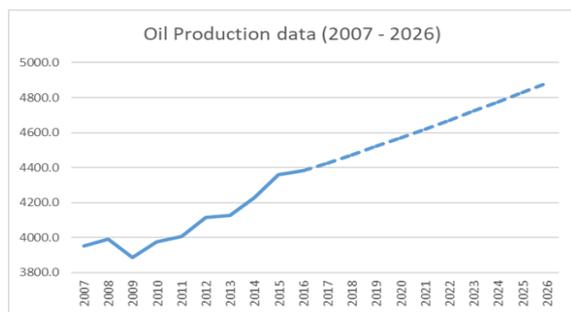


Figure 8. Oil Production data for all region from 2007 to 2026

Figure 1 presents the general chart of oil production amounts for all regions from 2007 to 2026, while the continues line presents the original data and the dotted line presents the forecasted data, and it's clearly noticed that the forecasting trend is growing up for all regions in the next 10 years.

#### 4. Discussion

From Figure 8, it's observed that oil production amounts for the next ten years will increase, and this is logical result, because the population is increasing by the time which means more people will demand oil production types for basic needs and day life use, also the industrial fields are growing rapidly which needs a lot of oil production types to be continues.

In general, all region will have increasing trend of oil production in the next 10 years, except Asia and Africa they have decreasing rate -3.7 and -9.3 respectively

#### 5. Conclusion

In this paper, the oil production produced from all six regions were forecasted in the long-term future. Three main forecasting algorithms were used ARIMA, Hybrid ARIMA and Linear Regression. I applied them on dataset collected from BP Global. It's found that Hybrid ARIMA is the best algorithm to be used for datasets belongs to Africa, Asia and Europe, while ARIMA is the best algorithm to be used for datasets belongs to Middle East and North America, and finally, Linear regression was the most appropriate forecasting algorithm to be used for South & Cent. America dataset.

#### 6. References

[1] Enerdata, Global Energy Statistical Yearbook 2017. 2016.

[2] Administration, U.S.E.I., What is the difference between crude oil, petroleum products, and

petroleum? U.S. Energy Information Administration, 2017: p. 1.

[3] Administration, U.S.E.I., Monthly Energy Review. 2017.

[4] Aggarwal, C.C., Data mining: the textbook. 2015: Springer.

[5] Billings, R.B. and C.V. Jones, Forecasting urban water demand. 2011: American Water Works Association.

[6] Nashawi, I.S., A. Malallah, and M. Al-Bisharah, Forecasting world crude oil production using multicyclic Hubbert model. *Energy & Fuels*, 2010. 24(3): p. 1788-1800.

[7] Aizenberg, I., et al., Multilayer neural network with multi-valued neurons in time series forecasting of oil production. *Neurocomputing*, 2016. 175: p. 980-989.

[8] Kontorovich, A., M. Epov, and L. Eder, Long-term and medium-term scenarios and factors in world energy perspectives for the 21st century. *Russian Geology and Geophysics*, 2014. 55(5-6): p. 534-543.

[9] Abas, N., A. Kalair, and N. Khan, Review of fossil fuels and future energy technologies. *Futures*, 2015. 69: p. 31-49.

[10] Frausto-Solís, J., M. Chi-Chim, and L. Sheremetov, Forecasting oil production time series with a population-based simulated annealing method. *Arabian Journal for Science and Engineering*, 2015. 40(4): p. 1081-1096.

[11] Efendi, R. and M.M. Deris. Forecasting of Malaysian Oil Production and Oil Consumption Using Fuzzy Time Series. in *International Conference on Soft Computing and Data Mining*. 2016. Springer.

[12] Bilgili, F., et al., How did the US economy react to shale gas production revolution? An advanced time series approach. *Energy*, 2016. 116: p. 963-977.