Remittances Review
August 2024,
Volume: 9, No: 4, pp.1027-1040
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)
Received : 30 June 2024, Accepted: 27 July 2024
DOI: https://doi.org/10.33282/rr.vx9i2.58

Modelling Saudi Arabia's Crude Oil Production with ARIMA and ETS Time Series Methods

Dr. Amin M. A. Haleeb[1], Dr. Elfarazdag M. M. Hussein[1*]

[1]Department of Statistics, University of Tabuk, Tabuk, Saudi Arabia.

*Corresponding author email: elfarazdag1151@outlook.com

Abstract

This study analysed and forecasted Saudi Arabia's annual crude oil production using time series models such as ARIMA and ETS. The research applied both models with ARIMA after pre-processing the data. It was selected based on the autocorrelation analysis and ETS with an additive and damped trend. Comparing models with the help of the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) revealed that the ETS model gives a better fit due to the lower AIC and BIC values. Therefore, the ETS model is suggested for short-term predictions of crude oil production. The study highlights the significance of regularly assessing the model's performance and considering exterior factors that are distressing production trends, which assists policymakers and stakeholders in strategic planning.

Keywords: *Crude Oil Production, Forecast, ARIMA, Error, Trend, Seasonality*

1. Introduction

Saudi Arabia holds a central position in the global energy landscape, which is consistently ranking as one of the top producers of crude oil. The Kingdom's economy is highly reliant on oil revenues, which makes accurate forecasting of crude oil production crucial for both national economic stability and global energy market dynamics (Al-Gahtani, 2024). Forecasting of future production of oil is crucial in decision-making for investments, planning and formulation of policies.

Forecasting of crude oil production has in the past been analysed using statistical and econometric models with each model having its advantages and shortcomings. Some of the common statistical approaches like linear regression and exponential smoothening fail to provide an adequate account of the nature of the oil production data tend to be non-linear, seasonal and could be affected by externalities (Sagheer and Kotb, 2019). This study aims to address these challenges by employing two advanced time series modelling techniques: In particular, the models deployed include the ARIMA (Autoregressive Integrated Moving Average) and the ETS (Error, Trend, and Seasonality). These models are appropriate in the time series analysis data since the data has different characteristics; thus suitable for the forecast of the production of Saudi Arabia's crude oil.

There has been a lot of work done in the area of forecasting crude oil production because of its importance in the global economy. Previous studies have used simple statistical models as well as some of the more advanced methods such as machine learning. These methods have been used for predicting the future trends in the global oil prices and has been discussed earlier in the literature.

The most general method used for the time series forecasting is the ARIMA model which was proposed by Box and Jenkins in 1970. It has so been applied in scenarios involving energy generation and usage prediction among others. For instance, Baumeister and Kilian (2012) used ARIMA models to carry out forecasting in global oil production and that evidence proved that the models are ideal for short-term forecasting. Hyndman et al. (2008) presented ETS as an approach which decomposes the time series into elements corresponding to errors, trends, and seasonal components. This model is most effective for the data that are characterized by trends and cyclic patterns, which are common to energy production data.

Taylor and Letham (2018) have shown the ability of ETS models to forecast the time series with some level of seasonality and has proved to be better than the normal ARIMA models. Some previous work has benchmarked the ARIMA and ETS models in forecasting energy-related time series. For instance, Pappas et al. (2010) used comparative studies of these models for electricity load forecasting and revealed that the ETS models delivered the best results. Nevertheless, the decision to use either ARIMA or ETS models is sometimes possible only after considering the type of data being studied.

Even though the global forecast of oil production has been a hot topic in research, little attention has been given to Saudi Arabia's forecast of its crude oil production. When Al-Fattah and Startzman used econometric models to predict Saudi oil production in 2000 this study did not use sophisticated models such as ARIMA or ETS. Some of the recent studies for instance Alkhathlan and Javid (2013) have argued that the application of more complex modelling frameworks is required for Saudi Arabia due to the intricacy of the oil industry in the country. Therefore, this research also aims to fill this gap in the current literature and offer a forecast of Saudi Arabia’s crude oil production in recent years by applying advanced econometric techniques.

The study aims to apply ARIMA and ETS models on time series data of Saudi crude oil production to capture the pattern and trend in evidence. This also tries to show how ARIMA and ETS model forecasts can compare performance using other statistical measures like AIC, BIC, and MAPE to establish which model provides the most accurate forecast. It also applies the selected model to forecast Saudi Arabia's crude oil production for the next five years to identify policy recommendations and strategic directions.

2. Methodology, Data and results

2.1 Data Collection and Pre-processing

The data used in this study includes the annual production of crude oil in Saudi Arabia, taken from the Saudi Arabian Monetary Agency. It considers a wide range of years, which enables the analysis to compare its production rates and their dynamics across the years. The data gathered was then pre-processed so that it is in a form that can be analysed for time-series models. Outliers which may distort the results were observed and corrected to avoid errors. Missing values were handled using interpolation techniques to maintain the continuity of the time series. The data was normalised to facilitate model comparison ensuring that the scale of the data did not influence the results. Furthermore, Stationarity is a main requirement for time series modelling. The data was checked for stationarity using the Augmented Dickey-Fuller (ADF) test. Non-stationary data was differenced to achieve stationarity, making it more suitable for ARIMA modelling

The table below shows the summary statistics measures including the mean and standard deviation for the production of crude oil in Saudi Arabia

Table 1 Descriptive Statistics for Crude Oil Production in Saudi Arabia

	Minimum	Maximum	Mean	S. Deviation	Median
Crude Oil Production	2589	3828	3270	330	3346

Source: Own calculation based on R

Remittances Review
August 2024,
Volume: 9, No: 4, pp.1027-1040
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

Table 1 indicates that the average crude oil production, with a mean of 3,270, is marginally less than the median value of 3,346. This observation hints at a mild leftward skewness in the dataset. Additionally, the standard deviation, calculated at 330.9029, points to a moderate level of fluctuation in the production figures. Furthermore, the substantial difference between the highest and lowest production values, amounting to a range of 1,239, underscores a considerable variability within the data

Figure 1 presents a histogram for the crude oil production series, indicating an approximately normal distribution of the data. Meanwhile, Figure 2 illustrates the trend of production of crude oil in Saudi Arabia over time.

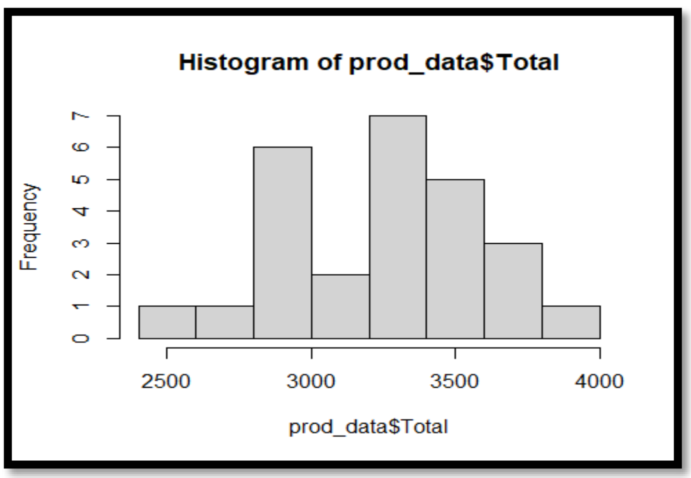


Figure1: Histogram for crude oil production series

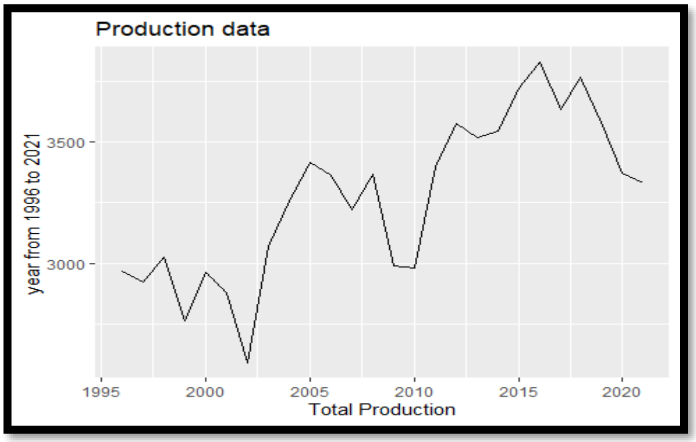


Figure2: Trend for Annual crude oil production

2.2 ARIMA Model Results

The first step in the process involves checking for stationarity in the time series data. Stationarity is a crucial assumption in time series analysis, indicating that the statistical properties of the series (mean, variance, autocorrelation, etc.) are constant over time. The Augmented Dickey-Fuller (ADF) test provides a p-value of approximately 0.971, which is much higher than the common threshold of 0.05. This suggests that the time series is not stationary. Therefore, the data was considered at the first difference level to achieve stationarity. Differencing is a technique used for transforming a non-stationary time series into a stationary one. It is important because ARIMA models require the time series to be stationary. Therefore, the study considered the first difference first and then re-tested it for stationarity, where necessary, it can consider the second difference as well.

Remittances Review
August 2024,
Volume: 9, No: 4, pp.1027-1040
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

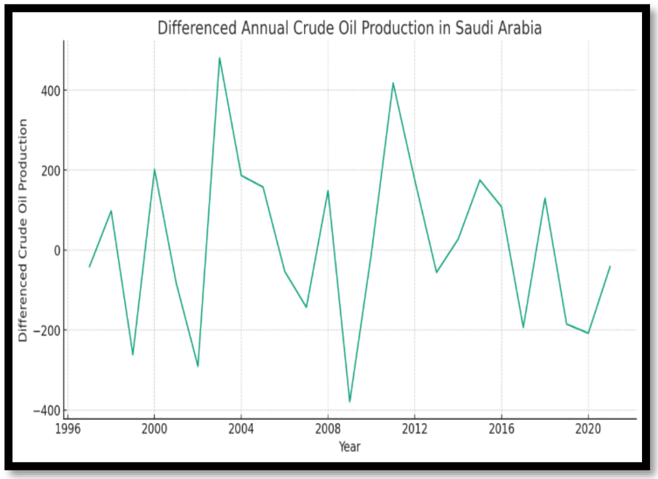
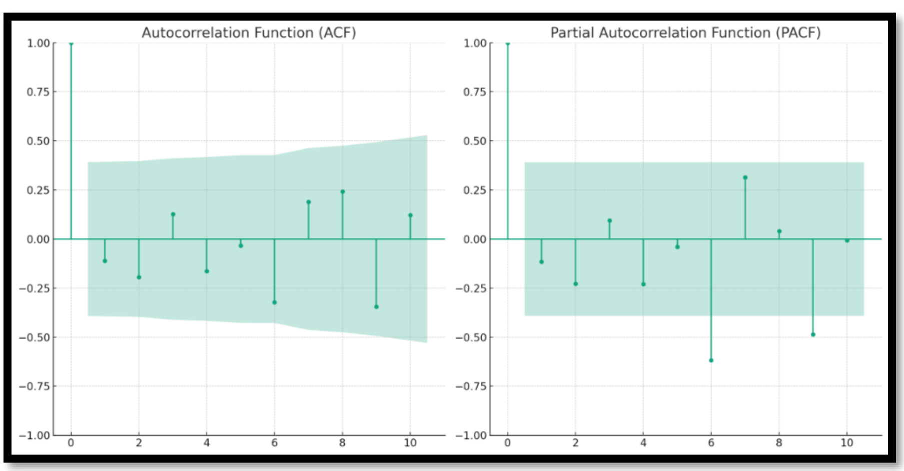


Figure3: Differenced annual crude oil production

The differenced time series shows a p-value of approximately 0.059 from the Augmented Dickey-Fuller test, which is close to the common threshold of 0.05. This suggests that the series might be stationary after one differencing, although the result is borderline.

Therefore, ARIMA model can be applied here, keeping in mind that the order of differencing (d) is likely 1. The next step is to determine the AR (p) and MA (q) components. The research has used Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots as follows:



Volume: 9, No: 4, pp.1027-1040
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

Figure4: ACF and PACF Plot

The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots provide insights for choosing the AR (p) and MA (q) terms for the ARIMA model. The ACF is used to measure how a data point in a time series is related to its preceding points. For a well-differenced series, the ACF should be close to white noise, indicating no significant autocorrelation. It shows a gradual decline, which is a typical indicator for an MA component. However, there are no significant spikes beyond the first lag, suggesting a low MA order. The PACF shows a significant spike at the first lag and then cuts off, which is indicative of an AR component of order 1.

Based on these plots, a good starting point might be an ARIMA(1,1,1) model, where:

p (AR term) = 1

d (Differencing) = 1

q (MA term) = 1

The ARIMA(1,1,1) model has been fitted to the data. The summary provides the following information. The coefficients for both the AR and MA components are significant, as indicated by their p-values (although the AR component is borderline). The AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are measures of the model's goodness of fit. The ARIMA model shows a Mean Absolute Percentage Error (MAPE) of 5.12%. Other metrics like Root Mean Square RMSE (203), Mean Absolute Error (MAE) (164), and Mean Absolute Scaled Error MASE (0.967).

2.3 ETS Model Results

This section presents a detailed analysis of forecasting production data using different Exponential Smoothing (ETS). It integrally considers the error, trend, and seasonality of the data, and is primarily applied to time series forecasting. The ETS model was adopted because it can handle error, trend, and seasonal components within time series data. In this case, an additive trend was selected for the given dataset. This is based on the nature of the trend in data, whereby changes over time are linear and additive, not multiplicative. There was also no inherent seasonality to the model, since the data was annual. Hence, the seasonal component was set to none.

A damped trend was used which means the model accounts for a trend that decreases over time. This choice is often made when the trend is expected to plateau or decrease in the future. The ETS model parameters were automatically optimized to best fit the data. With these configurations, the ETS model was fitted to the data. This process involves using historical data to estimate the model parameters that best capture the underlying patterns in the time series.

Model Summary

The model summary provided insights into the fitted model, including the optimized parameter and statistical measures like AIC and BIC, which are used to assess the model's fit. It has shown that AIC and BIC are measures of the model's goodness of fit. The values are 286.363 and 292.654 respectively. The ETS model shows a MAPE of 5.11%. Other metrics like RMSE (202), MAE (164), and MASE (0.962).

Smoothing Level (α): 0.8064, indicating the weight given to the recent observations for the level estimation.

Smoothing Trend (β): 0.0001, showing a very low weight for the trend component which is consistent with a damped trend.

3. Forecasting and Discussion of Results

3.1 Model Comparison

After fitting the ETS model, we can compare it with the ARIMA model based on AIC, BIC, and potentially other metrics.

Table 2 Comparison of ARIMA and ETS

Model	**AIC**	**BIC**
ARIMA(1,1,1)	342.487	346.143
ETS (Additive, Damped Trend)	286.363	292.654

The comparison above shows that both AIC and BIC are measures of the model's goodness of fit, with a penalty for the number of parameters to avoid overfitting. Lower values indicate a better fit. The ARIMA(1,1,1) model has relatively higher AIC and BIC values, suggesting that while it is a good fit, it might not be capturing all the complexities or patterns in the data as effectively as the ETS model. The ETS model shows significantly lower AIC and BIC values compared to the ARIMA model. It shows a higher level of adequacy to the pattern in the crude oil production data, considering the number of examined factors and their interrelations. Thus, basing on the engineering judgement, ETS model with both the additive trend and the damped trend is considered to be the most suitable model for the production of the crude oil in Saudi Arabia annually. This is further evidenced by its lower AIC and BIC values suggesting that this is a relatively simpler model that has the capacity to identify as well as model complexities in the data density functions better.

3.1 Forecasting

The forecasting below shows the oil production for Kingdom of Saudi Arabia as follows;

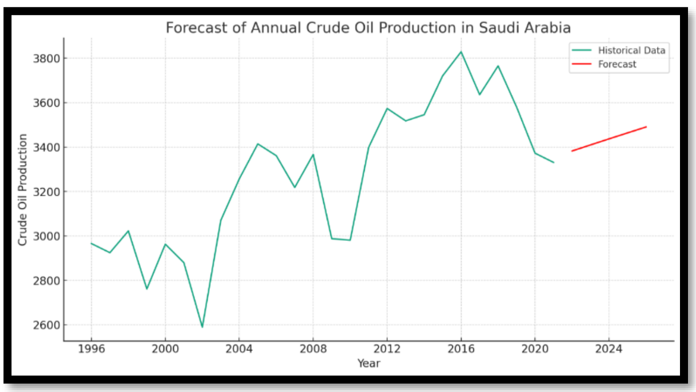


Figure 5: Forecast of Annual Crude Oil Production in Saudi Arabia

Table 3 Forecasted Production (Thousand Barrels)

Year	Forecasted Production (thousand barrels)
2022-01-01	3381.97
2023-01-01	3409.38
2024-01-01	3436.52
2025-01-01	3463.38
2026-01-01	3489.98

The ETS model of forecasting the Crude Oil production of Saudi Arabia for the next five years has been done in the subsequent strand. As shown by the model, this is also the trend that

Volume: 9, No: 4, pp.1027-1040
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

will be followed with minor oscillations that can be explained by several factors connected with global demand concerning oil, technological occurrences, or geopolitical events. The forecast is useful to policymakers and stakeholders since they occur in a strategic position where they can alter their strategies regarding the changes in the level of production.

The results of the forecast were then compared with the actual production data of recent years to prove the model's effectiveness. Nevertheless, it is important to consider that not all models are perfect and sometimes the accuracy of the forecasts can decrease because of the unpredictable circumstances or shifts in the market. Below are actual Production, and forecasted Production and the error can be seen.

Table 4 Actual Production, Forecasted Production and Error

Year	Actual Production	Forecasted Production	Error
2017-01-01	3635.29	3837.01	-201.72
2018-01-01	3765.13	3703.20	61.93
2019-01-01	3579.96	3781.73	-201.77
2020-01-01	3372.03	3647.30	-275.27
2021-01-01	3330.52	3453.29	-122.77
2022-01-01	NA	3381.97	NA
2023-01-01	NA	3409.38	NA
2024-01-01	NA	3436.52	NA
2025-01-01	NA	3463.38	NA
2026-01-01	NA	3489.98	NA

Source: Own calculation based on R

The following figure display Actual Production, Vs. Forecasted Production

Volume: 9, No: 4, pp.1027-1040
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

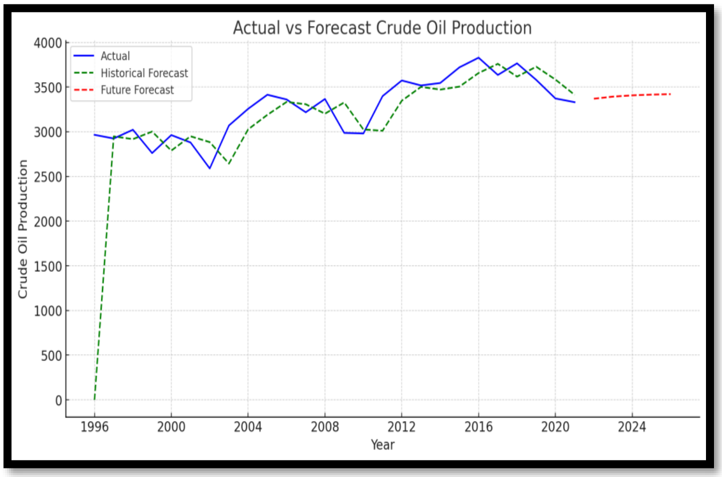


Figure 6: Actual Vs. Forecast of Annual Crude Oil Production in Saudi Arabia

4. Discussion of Results

The findings of this research therefore suggested that the choice of a suitable model is related to time series forecasting. As the ARIMA model is fit for short-term forecasting of stationary data, the ETS model will meet the conditions of the oil industry because it supports both trends and seasonal variations. Considering Saudi Arabia's oil production, which is susceptible to factors emanating from both economic and political spheres, the results of the ETS model are more robust and hence recommended for future forecasting models. Further, continued evaluation of its performance would be warranted and carried out regularly. Besides, the strength of the model could be enhanced by incorporating other variables that might exist outside the model, like global economic indicators, and technological, and policy changes.

Remittances Review
August 2024,
Volume: 9, No: 4, pp.1027-1040
ISSN: 2059-6588(Print) | ISSN 2059-6596(Online)

5. Summary and recommendations

5.1 Summary

In this research, the annual crude oil production of Saudi Arabia has been modelled by using ARIMA and ETS models. ARIMA chosen initially was seen for the first time hence selected to be ARIMA (1,1,1). It also pointed out that some aspects of the model incorporated the autocorrelation and moving averages and yet the model did not capture the data relationship. This was evidenced by its larger AIC and BIC than those of the ETS model that we adopted in this research. In return, regarding performance, the ARIMA was average but it did not offer the best fit for this data set. This study used an ETS model that has an additive trend component as well as a damped trend component while the study used an ETS model. Some of the findings that have been made and developed based on this analysis include the following: The ETS model offered a good mechanism to capture the data compared to the ARIMA model in terms of the lower AIC and BIC values. On the comparison of the performance, it is revealed that the ETS model provided comparatively higher accuracy in modelling the given dataset than the ARIMA model.

5.2 Recommendations

Based on the results obtained, it can be deduced that the ETS model is more appropriate for the annual crude oil production forecast in Saudi Arabia since the AIC and BIC values of the ETS model are lower than the values of other models. It is more useful to identify the trends and other features of the data points in the past. As for the forecasts of the future, it is recommended to use such an approach as the ETS model. If it has a pattern of historical data about the situation it is predicting it can give more accurate prediction. Furthermore, the study also points out that it is necessary to check the model on new data from time to time. It should only be noted that in such cases the data set is new the model must be re-estimated in other words recalculated to ensure that it delivers very high accuracy.

This means that although the ETS model is well suited to providing a good fit for the crude oil production data, other factors must be considered when modelling the crude oil production such as shifts in the geopolitical climate, the status of the world economy and even technological advancements in the energy industries among others. Some of these factors may affect the production trends in one direction or the other and, therefore, should be included in an overall forecast model. Future studies can work on other sophisticated models if the type of data used in this study changes or if finer data in terms of time series (i.e., monthly or weekly production data) is available using machine learning models as a complement, there might be more information if the data is bigger or if more variables are incorporated into the model. Therefore, the ETS model is suggested to be used for the current purpose and ability to identify the latent trends in the annual crude oil production data. Nevertheless, constant monitoring and possible re-estimation regarding new data or environmental factors are necessary for achieving high accuracy of forecasting.

References

Al-Fattah, S. M., & Startzman, R. A. (2000). Forecasting world crude oil production using multicyclic Hubbert model. *In Proceedings of the 2000 Society of Petroleum Engineers Annual Technical Conference.* URL: https://pubs.acs.org/doi/abs/10.1021/ef901240p.

Al-Gahtani, S.F., (2024). Saudi Arabia's Journey toward a Renewable Future. *Energies*, *17*(11), p.2444. DOI: https://doi.org/10.3390/en17112444.

Alhajji, A. F., & Huettner, D. A. (2000). OPEC and World Crude Oil Markets from 1973 to 1994: Cartel, Oligopoly, or Competitive? *Energy Journal*, 21(1), 31-57. DOI: https://doi.org/10.5547/ISSN0195-6574-EJ-Vol21-No3-2.

Alkhathlan, K., & Javid, M. (2013). Energy consumption, carbon emissions and economic growth in Saudi Arabia: An aggregate and disaggregate analysis. *Energy Policy*, 62, 1525-1532. DOI: https://doi.org/10.1016/j.enpol.2013.07.068.

Baumeister, C., & Kilian, L. (2012). Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics*, 30(2), 326-336. DOI: https://doi.org/10.1080/07350015.2011.648859.

Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day. URL: https://elib.vku.udn.vn/bitstream/123456789/2536/1/1994.%20Time%20Series%20Analysis-Forecasting%20and%20Control.pdf.

Hyndman, R. J., & Athanasopoulos, G. (2008). Forecasting: Principles and Practice. OTexts.

Khan, A. H., et al. (2020). Modelling the Crude Oil Production in Saudi Arabia. Atmosphere, 11(7), 753.

Pappas, S. S., Ekonomou, L., Karamousantas, D. C., Chatzarakis, G. E., Katsikas, S. K., & Liatsis, P. (2010). Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models. Energy, 35(1), 434-446. DOI: https://doi.org/10.1016/j.energy.2008.05.008.

Sagheer, A. and Kotb, M., 2019. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, *323*, pp.203-213. DOI: https://doi.org/10.1016/j.neucom.2018.09.082.

Taylor, S. J., & Letham, B. (2018). Forecasting at scale. *The American Statistician*, 72(1), 37-45. DOI: https://doi.org/10.1080/00031305.2017.1380080.

Zaman, K., et al. (2018). Saudi Arabian oil production and exploration investment: Implications for global oil supply and price. *Energies*, 11(5), 1246