Application of Autoregressive Integrated Moving Average for Modeling and Forecasting of Crude Oil Production

Gbolahan S. Osho^{1*}

 College of Business, Dept of Management and Marketing, Prairie View A&M University, P.O. Box 519, Mail Stop: 2300, Prairie View, Texas USA 77446

* E-mail of the corresponding author: gsosho@pvamu.edu

Abstract

Most of the crude oil production forecasting and modeling studies have focused on the traditional model of decline curve analysis as techniques used for generating forecast values of future oil and gas in place in conventional and unconventional wells. To many production reservoir economists, the importance of forecast estimations is critical because it allows forecast users to have a profound interest in monitoring and improving forecast performance. It also provides clear indications for the directions, strategies, and bottom line of both the National Oil Companies (NOCs) and International Oil Corporations (IOCs). While the traditional DCS model techniques have wellgrounded mathematical underlining. This statistical design does necessarily assure that predicated function regardless of the values of the predictive variables. Moreover, the power traditional oil reservoir production forecasting technique is inefficient. This current research attempts to provide appropriate modeling and forecasting techniques for reservoirs utilizing a time series approach. It reveals how the historical oil production data can be used to project future oil reservoir production and how these projections influence future oil reservoir production decisions. Hence, the main objective of this research study is to practically explore the possibility of the autoregressive integrated-moving average model as a feasible function preference for predicting crude oil production. The historical oil reservoir production time series were used to establish respective autoregressive integrated moving average models through the time series technique by Box-Jenkins and the suitable models were designated with four performance criteria: maximum likelihood, standard error, Schwarz Bayesian criterion, and Akaike criterion for seven elected regions oil reservoir production and the established models conformed to the ARIMA (p, d, q). Once the process is identified, the adequacy of the forecasts will be determined and compared with the traditional decline curve analysis. Thus, as an accurate and effective oil production prediction for stretching a reservoir life cycle and enhancing reservoir productivity and recovery factors. These results provide production economists and reservoir engineers with quick, reliable, consistent, and real-time guidelines in budgeting, planning, and making decisions regarding field development. Finally, ARIMA forecasting models are more precise and deliver operational efficiency for dynamic forecasting of oil reservoir production.

Keywords: Oil and gas production forecasting, decline curve analysis, time series, ARIMA modeling, **DOI**: 10.7176/JETP/13-1-05 **Publication date:** February 28th 2023

1. Introduction

Oil and gas production sustainability is critical to the profitability and bottom line of both the National Oil Companies (NOCs) and International Oil Corporations (IOCs). Therefore, the volatility in crude oil production poses an acute challenge not only to the NOCs and IOCs but also to the global oil and gas market (Uwakonye, Osho & Anucha, 2006). Yikun, *et al* (2022) noted that over the past decades, global crude oil reserves have increased, and production has surged due to hydraulic fracturing technology generally called fracking: a well stimulus practice that deploys liquid compound to generate and reestablish slight fractures in a rock formation to boost oil production and permit increased efficiency for both oil and gas wells (Zhang, He, Wang, & Liang 2022). This makes available for the recovery from a well reservoir that reservoir engineers and production economists once believed was impossible to produce.

In the United States where hydraulic fracturing technology is primarily noticeable, the U.S. Energy Information Administration assesses that the United States oil and gas production would average about 12 million b/d in 2019, a surge of about 10% compared with that of the 2018 daily production output level. By 2020, it suggested that the projected oil and production would increase by one million barrels per day and to a yearly average of about 13 million barrels per day. This plunging daily hydrocarbon output growth level also imitates comparatively fixed hydrocarbon price levels and stalling increases in the general productivity index and oil well performance.

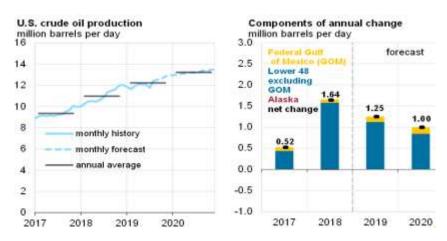


Figure 1. United States Crude Oil Production and Components of Annual Change

Source: U.S. Energy Information Administration: Short-Term Energy Outlook

Therefore, modeling the future crude oil production potential of formations by both conventional and unconventional is one of the reservoir engineers' and production economists' most central responsibilities. As in conventional wells, future crude production for unconventional formations is usually projected by fitting a line over previous production data based on current trends and then implying the curve to forecast future production values, a procedure known as Decline Curve Analysis. It is normally and widely deployed to predict oil and gas production and estimate the amount of oil in place.

The DCA predictions are usually the basis of most oil and gas business decisions and strategies on exploration, field development, facility and property expansions, and loan guarantees. It is also used for economic evaluation to support oil and gas capital budgeting on operational expenditure OPEX and capital expenditures and CAPEX. The traditional Decline Curve Analysis (DCA) forecasting model established by Arps for conventional formations is well grounded on empirical observations of production rate. Nevertheless, it does not have an underline theoretical derivation. The DCA is an oil well production observed method that deduces patterns in the production data from reservoir formations (Osho, Oloyede, Adetosoye, Fernandes, and Samuel 2005). Furthermore, the primary goal of the DCA is twofold: first, to create projected values of future production rates. Second, to provide an estimate of anticipated eventual recoverable productions. Over time, the DCA technique has improved for use in conventional and unconventional formations. The modified DCA method foists an exponential decline with a definite decline rate.

Ayeni and Pilat (1991) noted that the DCA is perhaps only applicable to predicting oil and gas depletion-type of reservoirs with high production capacity. The study suggested that it may not be adequate for reservoirs that are producing at a constant rate or nearly constant with time. Omekara *et al.* (2015) applied the autoregressive moving average to forecast Nigeria's oil production from 2006 - 2015. The importance of their findings suggests that oil production could be used to establish the future oil refinery volume in the Nigerian economy. This production forecast does not consider other relevant information they know in making their predictions. The results indicated that projected rationality is often challenged by determining whether the forecasters' prediction errors are predictable.

Since forecasts are of great importance and generally used in hydrocarbon production estimates. Hydrocarbon production forecast values are typically deployed for estimating residual oil in place, actual reserves, and expected ultimate recoverable reserves. To many production reservoir economists, the importance of forecast estimations is very critical, not only because it allows forecast handlers to check and advance prediction performance but also because it provides clear indications as to the directions, strategies, and bottom line of both the National Oil Companies (NOCs) and International Oil Corporations (IOCs).

While the traditional DCA model techniques have well-grounded mathematical underlining. In addition, they are generally used to optimize production operations, business strategies, and planning. Thus, comprehensive oil and gas production forecast values technically lead to good business strategies and decisions. The distinction of oil and gas production forecast estimations is naturally a keen interest in monitoring and improving oil and gas production performance for both reservoir engineers and production economists. Hence, the principal objective of this research study is to practically investigate the likelihood of the autoregressive integrated-moving average approach as a

feasible process preference for predicting crude oil production. Once the model has been identified, the adequacy of the forecasts will be determined and matched with the traditional decline curve analysis.

In the early 1900s, production engineers noted that oil and gas future production could be projected by fitting an exponential function to historical decline rates data. Towler (2002) noted that DCA is used to investigate the association between oil and gas production flow rate and time for production. Generally, it was observed that the exponential function only suited well for certain oil and gas formations in production at the time. However, it did not sufficiently characterize the performance of some producing formations in depletion drive reservoirs. Furthermore, the traditional decline curve analysis can only extrapolate historical production data as an input-output modeling process and provide history match forecasts by using a well defines hyperbolic differential function in the form:

2. The Theoretical Framework

2.1 Box-Jenkins Models

A historical series sequence is said to be stationary when its mean is finite and autonomous of time. When variables exhibit a constant finite mean, variance, and autocovariance are time-independent finite, constant, and stationary. When the three criteria are met, this time series process then implies weak stationarity. Generally, times series of variables often display nonstationary levels. As noted by Muşetescu, R. C., Grigore, G. E., & Nicolae, S. (2022), and In 19976 Box and Jenkins established a functional procedure for a total class of the autoregressive integrated moving average models. The autoregressive integrated moving average processes are relevant mainly for stationary time series processes, if the mean is normally distributed, the variance and the autocorrelation function stay constant over time. Autoregressive integrated moving average processes are suitable for series with robust pattern characteristics, seasonal and nonseasonal processes, and random walk.

In model selection, the first stage is to identify a rough class of processes and their subclasses. The second stage is for the tentative process to be fitted to the historical series. Then, the third stage is to obtain parameter estimates. The loose estimates attained the at identification stage are utilized as initial values for the estimation of parameters. Lastly, diagnostic checks are deployed to determine the inappropriateness of fit. When no inappropriateness is specified, the estimated process would then be deployed for forecasting. However, when inappropriateness is unveiled, the reiteration processes of identification, estimation, and diagnostic checking were rerun pending when an appropriate process is discovered.

The value of a time series prediction is governed by the performance of the stochastic model that describes it. The basis of parsimony which is widely used by forecasters states that a model must sufficiently represent the data applied to it and as few parameters as possible (Osho, 2019). A major effort is applied to finding an appropriate stochastic model for predicting the future value of the series. The theoretical underpinnings defined in Box and Jenkins (1976) are quite sophisticated, but the nonspecialist can understand the methodology's essence.

$$By_t = y_{t-1} \tag{1}$$

$$(1 - B)y_t = y_t - y_{t-1} = \nabla y_t$$
(2)

A useful process to characterize a non-stationary time series is the autoregressive integrated moving average model ARIMA (p, d, q) model:

$$\phi_p(B)\nabla^d y_t = \theta_a(B)a_t \tag{3}$$

$$\phi_{p}(B)\nabla^{d} y_{t} = C + \theta_{a}(B)a_{t}$$
⁽⁴⁾

The function $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ is the *p*-order AR and the function $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ is the *q*-order MA. Then, for an order of observations y_t and for given $\phi(B)$ and $\theta(B)$, is given as

$$a_t = \frac{\phi(B)}{\theta(B)} (1 - B)^d y_t \tag{4}$$

Therefore, the log-likelihood for generalized $\phi(B)$ and $\theta(B)$ is strictly stated by a linear process of the sum of the squares of the residual:

$$S(\phi, \theta) = \sum_{t=1}^{n} a_t^2 \tag{6}$$

An appropriate conversion is often necessary in some cases to be estimated from data. Assume a process in this order has been fitted. In practice the estimates are usually obtained directly from the error model by differencing the process:

$$\phi(B) = \theta(B)(1-B)^{d} = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_{p+d} B^{p+d}$$
(7)

The error model is generated to articulate the process in terms of the previous term series:

$$y_{t+1} = \phi_1 y_{t+l-1} + \dots + \phi_{p+d} y_{t+1-p-d} - \theta_1 a_{t+l-1} - \dots - \theta_q a_{t+l-q}$$
(8)

The forecasting precision of the processes may be likened numerically using the values of the estimated residual

 σ_a values in the time series of the univariate process. The extent of information is relative to the inverse of the standard deviation. An additional measure for matching the forecasts that are an estimate of the variance and the mean square error. A process with a small mean square error gives more precise forecasts than a process with a large mean square error (Osho, 2018). The mean square error is given as the average square of residuals of the forecasts when the residual is given as follows:

$$MSE = \frac{1}{n} \sum_{i=t_0}^{t_0+n} \left(Y_i - \hat{Y}_i \right)^2$$
(9)

Generally, another criterion that was defined by using the root mean square error of the forecasted value a

$$U = \sqrt{\frac{\frac{1}{n} \sum \left(\hat{\Delta Y}_i - \Delta Y_i \right)^2}{\frac{1}{n} \sum \left(\Delta Y_i \right)^2}}$$
(10)

This statistical benefit of having the point calibrations is equal to zero when the prediction values are perfectly accurate, and it is routinely equal to one when the raw prediction of no change.

3. Results

For the study, oil production monthly data from 2007 - 2020 from seven regions were obtained. The time plots for oil production are presented in Figure 2. To confirm whether the time series data is stationary, plots were shown in Figure 3. It is noted that the time series is non-stationary and with no variation at a static level. First, a visual inspection of each well production data and the time plot indicate the forecast errors have essentially equal variance over time with mean all around zero. A unit root analysis was conducted to establish stationarity for each well production data.

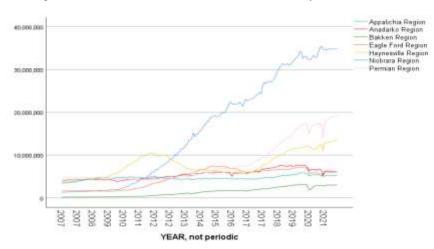


Figure 2. The Time Plots for Oil Production Monthly Data from 2007 – 2020

For this, the augmented Dickey-Fuller (ADF) test indicated that the series is non-stationary. Since the ADF test describes whether the change in $y_{t+1} = \varphi_1 y_{t+l-1} + \ldots + \varphi_{p+d} y_{t+1-p-d} - \theta_1 a_{t+l-1} - \ldots - \theta_q a_{t+l-q}$ is clarified with a lagged value and by a linear trend and estimated for non-stationarity. To predict the oil reservoir production, the time plots for oil production of the time series against periods are presented in Figure 2 for each Well and examined that there is an increasing pattern in the time series and therefore, the observed pattern was eliminated. The results indicated that stationarity was established after the second difference.

www.iiste.org

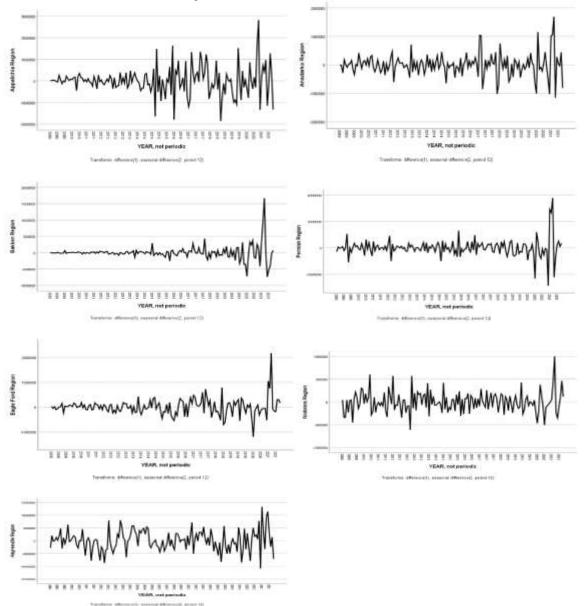


Figure 3. Transforms Difference (2, period 12)

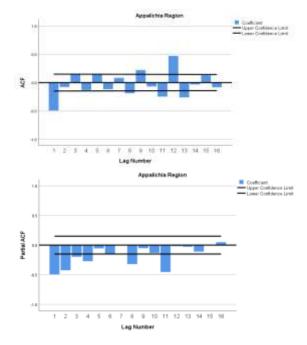
For stationarity, the time series displayed in Figure 3, oscillates about the mean with the ACF and PACF decay to around zero relatively promptly which demonstrates the stationarity. Additionally, to gauge whether the time series is from a stationary process, the unit root Dickey-Fuller test for stationarity was conducted. Figure 4 indicates ACF and Partial ACF with seasonally differenced series for each well. Figure 2 shows the transforms after the difference between the autocorrelation and partial autocorrelation plots and the findings were shown in Table 1. Furthermore, using the ACF and PACF plots of the proposed process for each Well were determined. Accordingly, it is essential to transmute the nonstationary time series to become stationary and generate a stationary, the first differences were obtained as shown in Table 2. The first difference series is displayed as shown in Figure 3. Those ACF and PACF plots of the lagged time signify stationarity.

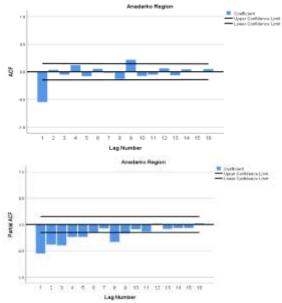
Figure 4. ACF and Partial ACF with Seasonally differenced Series for Each Well

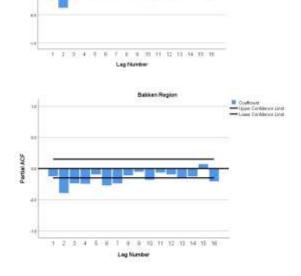
Journal of Energy Technologies and Policy ISSN 2224-3232 (Paper) ISSN 2225-0573 (Online) Vol.13, No.1, 2023

www.iiste.org

Configured Open Englished Land

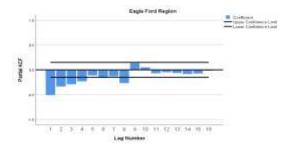






Bakken Region





4

4

۰.

www.iiste.org

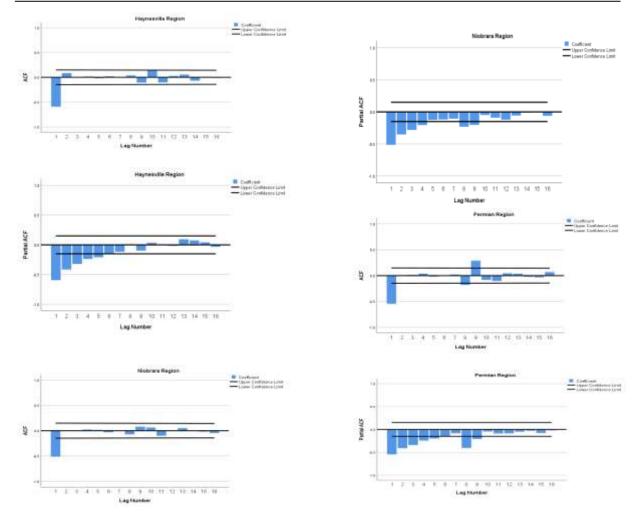


Figure 5 indicates ACF and Partial ACF for ARIMA Model for each Well. To establish whether the errors in the ARIMA forecast are normally distributed, the forecast errors with a time plot were visualized. The coefficient estimates of each model were displayed in Table 2 with the standard errors. It was noted from the results that the coefficient estimates and *p*-values for each AR and MA process are very significant. From these examinations of the times series, it is concluded that the formed ARIMA models are satisfactory.

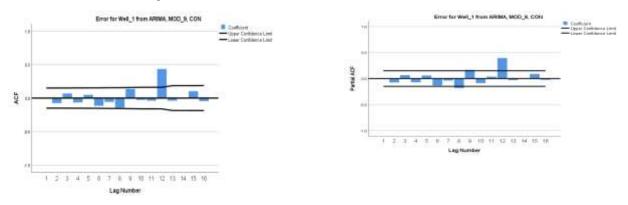
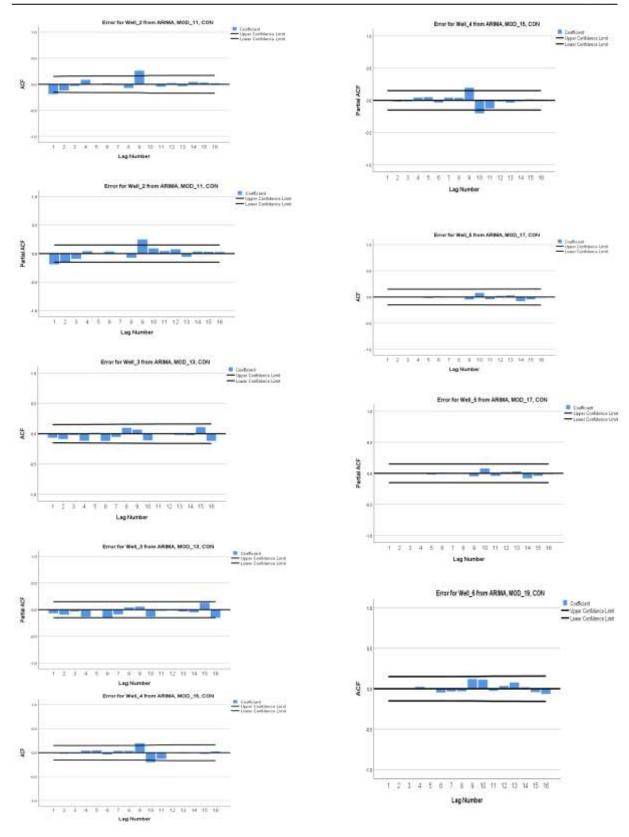


Figure 5. ACF and Partial ACF for ARIMA Model for Each Well

Journal of Energy Technologies and Policy ISSN 2224-3232 (Paper) ISSN 2225-0573 (Online) Vol.13, No.1, 2023 www.iiste.org





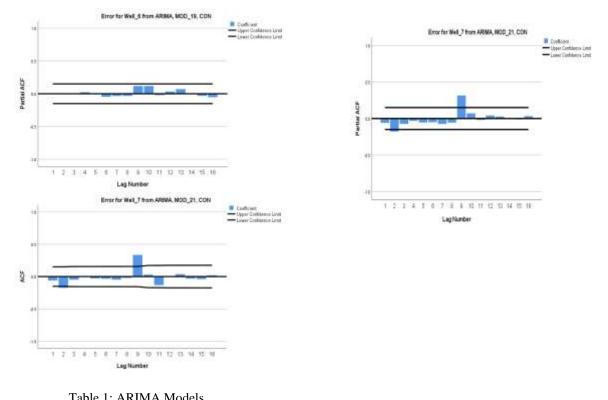


Table 1: ARIMA Models											
Well	AR	d	МА	AIC	BIC	AR	MA				
Region						Sig	sig				
Well 1	1	2	1	5075.2	5084.7	0.042	0.000				
Well 2	1	2	2	4817.0	4829.7	0.000	0.008				
Well 3	2	2	1	4529.1	4541.8	0.000	0.000				
Well 4	1	2	1	4745.3	4754.8	0.064	0.000				
Well 5	2	2	2	4866.9	4882.9	0.006	0.001				
Well 6	2	2	1	4558.5	4571.2	0.008	0.001				

2

1

1

Well 7

Consequently, it was determined that the mean for each Well data is normally constant thru the time series over time and therefore their applicability for time series models was established. The values of error autocorrelation and error partial autocorrelation functions for each well are given in Table 2 and their respective plots are presented in Figure 4. The PACF also has one spike that concluded that an auto-regressive of order one is suitable for the time series data. From the ACF, these suggest that there was one spike in all most for all the time series Well-data and nearly all signaling an AR (1) and MA (1) behavior. In addition, differencing was carried out one time to establish stationarity, this enhances the identification of an autoregressive integrated moving average ARIMA model as a suitable model for oil well production time series data. With the autoregressive and moving average orders identified based on the pattern revealed by ACF and PACF with the first order differencing, the ARIMA model for each Well is specified in Table 1.

5091.3

5101.0

0.009

0.001

Once the models were identified, parameter estimates were obtained using SPSS program precisely intended for ARIMA process construction. Furthermore, each Well time series process was checked for suitability by deploying the diagnostic command in SPSS. For each model, a residual error analysis test was conducted. For each identified, each Well time series process, the standard error test was similarly performed. These results were obtained at a level

of significance (p = 0.05) and the corresponding degrees of freedom are provided in Table 2. Each Well model was found to be satisfactory at p = 0.05.

Well Region	ARIMA (p,d,q)	Parameter estimates	Standard error	Approx. Sig	df	AIC	BIC	log- likelihood
		-0.056	0.080	0.042	174	5075.2	5084.7	-2534.58
Well 1	(1,2,1)	0.946	0.030	0.000				
		-0.945	0.202	0.000	174	4817.0	4829.7	-2404.51
Well 2	(1,2,2)	0.073	0.244	0.008				
		0.922	0.247	0.000				
		0.462	0.075	0.000	174	4529.1	4541.8	-2260.57
Well 3	(2,2,1)	-0.302	0.073	0.000	1/4	7527.1	4541.0	2200.57
wen 5	(2,2,1)	0.957	0.050	0.000				
		-0.038	0.082	0.064	174	4745.3	4754.8	-2369.64
Well 4	(1,2,1)	0.917	0.034	0.000				
		0.466	0.168	0.006	174	4866.9	4882.9	-2428.49
		0.057	0.111	0.610				
Well 5	(2,2,2)	1.624	0.149	0.000				
		-0.725	0.119	0.000				
		-0.045	0.078	0.008	174	4558.5	4571.2	-2275.25
		-0.022	0.078	0.003	1/4	4556.5	4371.2	-2213.23
Well 6	(2,2,1)	0.969	0.078	0.001				
		0.909	0.022	0.001				
XX 11 7	(1.0.1)	-0.196	0.075	0.009	174	5091.3	5101.0	-2542.73
Well 7	(1,2,1)	0.989	0.041	0.001				

Table 2: Parameter Estimates for ARIMA Models

4. Conclusion

This current study discourses fitting of ARIMA model to selected United States oil production monthly data from 2007 - 2020 from seven regions obtained from the United States Energy Information Administration, International Energy Statistics. The primary goal is an attempt to build a forecasting model that may be deployed to generate forecasts values of United States oil production essential and required for oil field economics decision-making and strategic budgeting and planning. Initially for the United States selected oil production wells, the data were non-stationary. After the unit root and differencing were conducted to sure the stationarity. The result of the analysis confirms that the suitable model is a multiplicative autoregressive moving average with a difference order. The historical oil production time series were deployed to establish various ARIMA processes through Box–Jenkins technique and the satisfactory models were designated upon four performance criteria: maximum likelihood, standard error, Schwarz Bayesian criterion, and Akaike criterion for seven elected regions' oil production and the established models conformed to the ARIMA (p, d, q).

The outcomes achieved demonstrate that the forecast models may be to generate forecast values for future oil production. Thus, accurate and effective oil production prediction is crucial for stretching a reservoir life cycle and enhancing reservoir productivity and recovery factors. Finally, ARIMA forecasting models are more precise and deliver operational efficiency for dynamic forecasting of oil reservoir production. The study results confirm that ARIMA modeling can investigate time series dynamics by removing the consequences of hysteresis. The acknowledged oil reservoir production ARIMA models can provide accurate and predictive capabilities. This study further reveals that ARIMA modeling is a robust and efficient technique for oil reservoir production forecasting and

provides consistent decision-making needed for successive oil reservoir production. These results provide production economists and reservoir engineers with quick, reliable, consistent, and real-time guidelines in budgeting, planning, and making decisions regarding field development.

References

Ayeni, Babatunde J., and Richard Pilat. "Crude oil reserve estimation: An application of the autoregressive integrated moving average (ARIMA) model." *Journal of Petroleum Science and Engineering* 8.1 (1992): 13-28.

Box, G. E. P. and D. R. Cox (1964). "An Analysis of transformations" J. Royal Stat. Soc. Series B, 26, 211-252.

Box, G.P. and G. M. Jenkins (1976) Time Series Analysis Forecasting and Control. San Francisco: Holden-Day.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control*, Second Edition. Oakland, CA: Holden-Day.

Box, G. E. P., G.M. Jenkins, and G. C. Rinsel (1994), *Time Series Analysis: Forecasting and Control*, Englewood Cliffs, NJ, Prentice-Hall.

Box, G. E. P., and Jenkins, G. M (1973). "Some Comments on the paper by Chatfield and Prothero and Reviewed by Kendall." *J. R. Statist. Ass.* 70, 70-79.

Dickey, D. A., and Fuller, W.A. (1981). "Likelihood ratio statistics for autoregressive time series with a unit root." *Econometrica* 49 (4) 1057-1072.

Dickey, A. D., Bell, W.R. and Miller, R.B. (1986) "Unit Root in Time Series Models: Tests and Implications," *The American Statistician*, 40, 12-26.

Omekara, C.O., Okereke, O. E. Ire, and K.I. Okamgba, C. O. (2015). "ARIMA Modeling of Nigeria Crude Oil Production" *Journal of Energy Technologies and Policy* Vol.5, No.10, 2015

Osho, G. S., Oloyede B., Adetosoye A., Fernandes., and Samuel R. (2005). DRILLING AN OFFSHORE WELL WITHIN THE ECONOMIC BUDGET CONSTRAINTS. Southwest Review of International Business Research, *16*(1).

Osho, G. S. (2019). A General Framework for Time Series Forecasting Model Using Autoregressive Integrated Moving Average-ARIMA and Transfer Functions. *International Journal of Statistics and Probability*, 8(6).

Mușetescu, R. C., Grigore, G. E., & Nicolae, S. (2022). The Use of GARCH Autoregressive Models in Estimating and Forecasting the Crude Oil Volatility. *European Journal of Interdisciplinary Studies*, *14*(1), 13-38.

Uwakonye, M. N., Osho, G. S., & Anucha, H. (2006). The Impact of Oil and Gas Production on The Nigerian Economy: A Rural Sector Econometric Model. *International Business & Economics Research Journal (IBER)*, 5(2).

Towler, Brian F. "Fundamental principles of reservoir engineering." (2002): 8-9.

Yikun, L. I. U., Fengjiao, W. A. N. G., Yumei, W. A. N. G., Binhui, L. I., Zhang, D., Guang, Y. A. N. G., ... & He, X. U. (2022). The mechanism of hydraulic fracturing assisted oil displacement to enhance oil recovery in low and medium-permeability reservoirs. *Petroleum Exploration and Development*, 49(4), 864-873.

Zhang, Y., He, M., Wang, Y., & Liang, C. (2022). Global economic policy uncertainty aligned: An informative predictor for crude oil market volatility. *International Journal of Forecasting*.

Gbolahan S. Osho, Ph.D.

Dr. Gbolahan S. Osho serves as an Associate Professor of Economics at the College of Business, Dept. of Management and Marketing at Prairie View A&M University. His research has directly influenced and contributed to knowledge in strategic management and public policies, especially regarding the oil & gas industry and natural resources. As a production economist, he is interested in conceptualizing and implementing research programs directed towards corporate governance and profitability, regional and global strategic analysis and economic development, performance measurement in multinationals, oil and gas pricing, derivatives, and policy-related issues facing the natural resource industries, energy sector, and policymakers at the local and global levels. Dr. Osho received his M.S. in Chemical Engineering, M.S. in Economics, and his PhD. in Ag. Economics from Oklahoma State University.