



A MATHEMATICAL MODEL FOR PREDICTING OUTPUT IN AN OILFIELD IN THE NIGER DELTA AREA OF NIGERIA

M. H. Oladeinde^{1,*}, A. O. Ohwo² and C. A. Oladeinde³

^{1,2,3} PRODUCTION ENGINEERING DEPARTMENT, UNIVERSITY OF BENIN, BENIN CITY, EDO STATE, NIGERIA

*Email addresses:*¹ moladeinde@uniben.edu,² austinhwo@yahoo.com,³ kallyama@yahoo.com

ABSTRACT

In this paper, a multiple linear regression model was developed for forecasting crude oil production volume in an oilfield in the Niger Delta area of Nigeria. Data of crude oil production volume and six significant parameters affecting crude oil volume were collected for a two year window. A six variable linear model was developed. The significance of each regression coefficient at $\alpha = 0.1$ in the six variable model was subjected to further analysis. The regression coefficient screening process resulted in a two variable model for predicting oil field output. The resultant model was found to have greater utility in predicting oil field output as it produced less residual. The present model can be of invaluable use to the oil field manager.

Keywords: linear regression, model, oilfield, crude [oil], production volume forecasting.

1. INTRODUCTION

The exploration of crude oil deposits has become a major foreign exchange earner for a number of countries. For instance, the Nigerian economy is heavily dependent on the oil sector which, according to the World Bank, accounts for over 95 percent of export earnings and about 85 percent of government revenues [1]. Consequently, correctly predicting the oilfield output is important for oilfield development [2] and is key to planning for the Nigerian government. The increase in crude oil production has been reported to cater for the rising energy needs of our present world. Since replacing coal and becoming the world's most important primary source of energy, oil has played an increasingly central role in the development of human society and exhibited an ever-closer relationship to the economy [3].

At the oilfield level, planning responsibility and decision making by the oilfield manager is facilitated by reliable forecasts of oil production volume. By far, there are many methods to predict the output of oilfield such as Multiple Linear Regression, Artificial Neural Network, Grey Prediction method and Logistic Curve Method which have different applicable environments and limits [4]. The forecast of oil field output has received attention from a number of researchers. The paper [5] presented a study on

forecast of oil reserves and production in an oilfield in China. Scaling laws from percolation theory was used to predict oilfield performance by [6], developed Multicyclic Hubbert modeling as a method for forecasting future Petroleum production was the theme of [7]. A Grey Forecasting model based on BP Neural Network for Crude Oil Production and Consumption in China was carried out by [8]. Artificial Neural Networks are useful for predicting oilfield output because they require few a priori assumptions about the model [9]. The optimization of oilfield production using neural network was carried out by [10]. Multi linear regression has been used for oilfield output prediction in some oilfields outside Nigeria. The multi linear regression has been used by [11]. Least square fitting method has been used by [12] to predict oilfield production. An improved multi linear regression method has been used for oilfield output forecasting by [13]. Software for prediction of oilfield output based on life cycle model was developed and applied by [14]. Hyperbolic -Exponential decline models was used by [15] for cumulative production forecast of an oil well. The use of expert systems for prediction of oilfields was undertaken by [16] with the aid of different fuzzy functions

In most studies oilfield production volume is predicted on an annual basis. It is expected that

disaggregation of annual production forecasts into monthly forecasts will be more accurate for the purpose of planning. Consequently, the present study seeks to obtain a model for predicting monthly oilfield production volume. In this study, a mathematical model is developed for predicting the output of crude oil in an oil field in the Niger Delta area of Nigeria.

2. RESEARCH METHODOLOGY

The research was conducted mainly by using extensive literature study, and also with the aid of secondary design instrument where data were collected from the database of final results from laboratory analysis of production records in one of the oil and gas producing industry in Nigeria for two years.

The framework of this study was the identification of most significant parameters that can influence the output of oil wells production in Nigeria, formulating suitable and most appropriate forecast model with the aid of the significant parameters identified as important features that can assist in predicting future oil production.

First step taken in this study was visiting the oil fields (oil flow stations) for information gathering. For example, various wells production volumes, reservoir behaviors, water content of wells production are variables that determine the rate of production of an oil well. The parameters taken were analyzed before an appropriate model was chosen for the study.

The Multiple Linear Regression analysis method was eventually chosen for its ability to accommodate several variables and its high level of accuracy, unlike others that can only analyze limited numbers of variables.

3. MULTIPLE LINEAR REGRESSION MODEL

Consider a response variable (Y) which depends on a number of variables (X). A multi linear regression model can be stated as [17]

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + e \quad (1)$$

X is called the independent or predictor variable and e is the random error. The variables β are called the regression coefficients and they determine the contribution of each of the predictor variables. A number of assumptions are required for multiple linear regression modeling. The first is that the random error is normally distributed and the second is a statement of independence of the random errors. Fitting data to a multiple linear model is done by using the least square method. Details of the least square

method can be found in any good statistic text for example, see [16]. However, a number of statistical software are available for multiple linear regression analysis. Examples are SPSS, Minitab and Statgraphics.

The residual of a multi linear regression model is the difference between the observed value and the fitted value

$$Residual = Y - \bar{Y} \quad (2)$$

4. PREDICTOR VARIABLES FOR OILFIELD OUTPUT

Six factors were identified as the most significant factors that can affect the prediction of the oil field output (net availability); they were identified as listed below:

- a) Net oil produced (bopd) previous month in the oilfield,
- b) Actual Gross oil produced (bpd) in the oilfield previous,
- c) Basic Sand and Water contained in oil produced in the oilfield previous month,
- d) Volume of Gas produced by the wells in the oilfield in previous month,
- e) Gas Oil Ratio of previous month production in the oilfield, and
- f) The total numbers of well stings (conduit) produced in previous month.

5. DATA OBTAINED

The data obtained from the oilfield for the factors affecting the output of the oil field as well as the actual output (net availability) is shown in Table 1 for a two year period spanning from February 2009 to January 2012.

The regression coefficients of the multi linear model can be obtained by using a set of normal equations for the basic form of the model given by equation (1) using the method of least squares. Using matrix approach as suggested in [17], the regression coefficients for the multi linear model regression model can be obtained.

6. SIX VARIABLE MODEL

The multi linear regression model was solved using SPSS software and the regression coefficients obtained are as follows:

$$\begin{aligned} \beta_0 &= -9624.492, \beta_1 = -0.859, \beta_2 0.969, \beta_3 \\ &= 29960.455, \beta_4 - 0349, \beta_5 \\ &= 7257.375, \beta_6 1533.954 \\ \beta_6 &= -1533.954 \end{aligned}$$

Table 1: Parameters data from the oilfield in Niger Delta Region, Nigeria

Months	Gross (bpd)	IPSC Net (bopd)	BS&W	Gas (mscf/d)	GOR (mscf/stb)	No of Conduits	Net Availability (bopd)
T	X1	X2	X3	X4	X5	X6	Y
Feb.'09	204,324	122,683	0.43	99,530	0.81	92	117,850
March '09	204,324	122,683	0.43	99,530	0.81	92	117,844
April '09	204,324	122,683	0.43	99,530	0.81	92	117,844
July '09	17,020	12,507	0.42	11,897	0.95	4	4,318
Aug. '09	17,020	12,507	0.42	11,897	0.95	4	4,318
July '10	5,346	4,725	0.12	2,796	0.40	17	4,725
Aug. '10	34,176	26,670	0.12	19,256	0.63	17	25,508
Sept. '10	34,846	30,500	0.12	25,544	0.63	18	29,358
Oct. '10	58,124	47,959	0.17	45,221	2.75	32	47,959
Nov. '10	90,402	68,348	0.24	61,522	1.79	40	57,518
Dec. '10	109,037	89,660	0.18	73,862	1.66	48	64,966
Jan. '11	112,249	91,141	0.19	64,589	1.40	47	91,141
Feb. '11	135,333	113,472	0.16	80,520	1.43	55	93,943
March '11	143,265	112,630	0.21	82,419	1.47	58	90,116
April '11	144,174	113,861	0.21	84,258	1.49	58	91,558
May '11	143,425	112,703	0.21	81,635	1.46	58	90,389
June '11	141,933	110,157	0.22	78,598	1.44	58	87,822
July '11	142,668	104,575	0.27	75,715	1.46	58	82,365
Aug. '11	135,036	102,248	0.24	72,253	1.44	58	76,810
Sept. '11	139,810	104,418	0.25	75,423	1.46	61	59,943
Oct. '11	141,219	103,111	0.27	27,244	1.50	62	60,416
Nov. '11	156,029	112,960	0.28	45,838	0.81	75	107,898
Dec. '11	156,819	113,030	0.28	64,318	1.17	77	110,535
Jan. '12	155,074	113,594	0.27	55,416	0.49	75	113,161

Substituting the values of the regression coefficients to the multi linear regression model in Equation 1, the model for predicting the oil field output was obtained as shown in Equation 3:

$$Y = -9624.492 - 0859 \times_1 - 0.969 \times_2 + 29960.455 \times_3 + 0.349 \times_4 - 7257.375 \times_5 + 1533.945 \times_6 \quad (3)$$

7. TEST FOR MODEL ADEQUACY

It is important to test whether the model developed provides information for predicting the output of the oil field. This was examined by carrying out an analysis of variance. The analysis of variance for the full regression model was carried out with SPSS software. For the multiple linear regression model developed. The null and alternative hypothesis were developed as:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$$

H_1 : At least one of the regression coefficients is non-zero.

The result obtained shows that $F > F_{0.05}$ since the value of F obtained from the analysis of variance is 42.901 whereas $F_{0.05} = 2.66$ at $v_1 = 2$ and $v_2 = 16$ degrees of freedom. The result indicates that the utility of the model for predicting oil field output is very impressive. The multiple coefficient of determination R^2 obtained for the model equal to 0.938. This indicates a very good fit between the model and its six variables.

8. HYPOTHESIS TESTING ON REGRESSION COEFFICIENTS

The utility of the full regression model has been confirmed in predicting the oil field output using the six independent variables. Further analysis was carried out in order to determine the significance of each of the regression coefficients to the final model. A

screening procedure was used to determine which coefficients were less important and so could be removed from the model. At the first stage using all the six variables, the P value of X_3 is equal 0.510. Since $\alpha = 0.1$ and the P-value is greater than 0.1, we delete the variable X_3 from the model. After the screening procedure, variables X_2 and X_6 were determined to be the most important contributory variables to oil field output prediction.

9. TWO VARIABLE MODEL

The resultant two variables model for predicting the oil field output obtained is:

$$Y = -2754.228 + 0.449X_2 + 705.162X_6 \quad (4)$$

10. COMPARISON OF TWO MODELS

The two models developed were tested for accuracy. The absolute values of the mean errors of the two models were calculated. Computations reveal that whereas the absolute mean error of the two variables model was 7.49, that of the six variables model was 39.37. This shows the superiority of the two variables model over the six variables model. It was observed that the coefficient of determination of the six variable models was 0.938 compared to 0.924 for the two variable models. It is to be noted however that this metric cannot be used to compare the two models as the R^2 has been shown to increase with addition of more variables to a linear model. Figure 1 shows the normal probability plot of the residuals. The plot is approximately linear; therefore, it is safe to assume that the error terms are normally distributed.

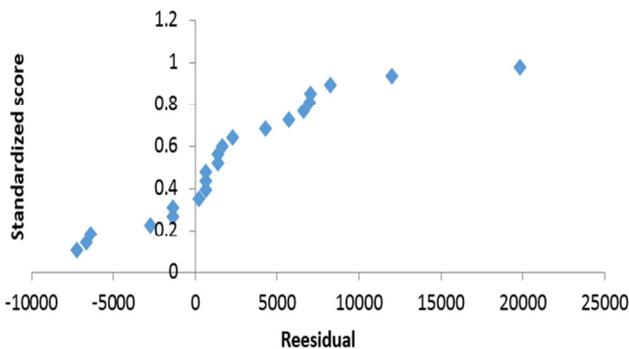


Figure 1: Normal probability plot of residuals of two variable models

11. CONCLUSION

In this paper a multiple linear regression model for predicting the output of an oil field in the Niger Delta area of Nigeria has been developed. The model takes

into account two variables which affect oil output. The developed model has been subjected to statistical analysis and has been found to be useful in predicting oil field output. The present model will be very useful to the oilfield manager for decision making in the oil field.

12. REFERENCES.

- [1] Akuru, U.B and Okoro, O.I, "A Prediction on Nigeria's Oil Depletion Based on Hubbert's Model and the Need for Renewable Energy, ISRN Renewable Energy, 2011, pp 1- 6.
- [2] Chen, Y., Ding, X., Liu, H. and Yan, Y. "Comparison of oilfield predicting model", *Engineering*, vol. 5, 2013, pp 637-641.
- [3] Wang, J Feng, I, Zhao, L, Snowden, S and Wang, X "A comparison of two typical multicyclic models used to forecast the world's conventional oil production, *Energy Policy* Vol. 39, 2011, pp 7616 - 7621.
- [4] Zhu, S.C and Zhao, X "Application of Artificial Neural Network in the Prediction of Output in Oilfield," International Joint Conference on Artificial Intelligence, 2009, pp 155 - 158
- [5] Tang, X., Zhang, B., Hook, M. and Feng, L, "Forecast of oil reserves and production in Daqing oilfield of China" *Energy*, vol. 25, 2011, pp 3097 - 3102
- [6] King, R.R., Buldyrev, S.V., Dokholyan, N.V., Havlin, S., Lopez, E., Paul, G., and Stanley, H.E, "Using Percolation Theory to Predict oilfield performance" *Physica A*, vol. 314, 2012, pp 103 - 108.
- [7.] Ken, B.A and James . A.C. "Discussion of Multicyclic Hubbert Modeling as a Method for Forecasting Future Petroleum Production". *Energy Fuels*, vol. 25, no 4, 2011, pp 1578-1584.
- [8] Hinge, M and Yonghe, W. "A Grey Forecasting Model Based on BP Neural Network for Crude Oil Production and Consumption in China". *Third International Symposium on information processing*, 2010, pp 105 - 109.
- [9] Salami, A.W., Mohammed, A.A, Adeyemo, J.A., and Olanlokun, O.K., "Modeling of Reservoir inflow for Hydropower Dams Using Artificial Neural Network", *Nigerian Journal of Technology*, Vol. 34, no 1, pp 28-36.
- [10] Ghazwan, J, "Application of Neural network to optimize oilfield production", *Asian Transactions on Engineering*, vol. 2, no 3, 2012, pp 10-23.
- [11] Izni, G.M and Radzuan, R., "A Study on Prediction of Output in Oilfield Using Multiple Linear Regressions". *International Journal of Applied Science and Technology*, Vol. 1, no 4, 2011, pp 21 - 27
- [12] Na, W.B., Su, Z.W., Zhang, P., " Research on oilfield production forecast based on least square fitting and

- improved BP neural network”, *Applied Mechanics and materials*, vol. 333 – 335, 2013, pp 1456 – 1460
- [13] Guo, L and Deng, X., “Application of Improved multiple linear regression method for oilfield output forecasting”, *International Conference on Information Management, innovation Management and Industrial Engineering*, vol. 1, 2009, pp 133-136
- [14] Zhang, Y., Jiang, G.J., Wang,C.L and Zhao, X.W., "The Development and Application for Oil Output Prediction Software System Based on Life Model", *Applied Mechanics and Materials*, Vol. 40-41, pp 317-321.
- [15] Makinde F. A., Orodu O. D., Ladipo A. O. and Anawe P. A. L, “Cumulative Production forecast of an oil well using Simplified “Hyperbolic-Exponential “ decline models”, *Global Journal of Researches in Engineering*, Vol. 12, Issue 2, 2012, pp 14-25
- [16] Montgomery, D.C and Runger, G.C, “*Applied Statistics and Probability for Engineers*” 2003, John Wiley and Sons
- [17] Brown, S.H, “Multi linear Regression Analysis: A matrix Approach with MATLAB”, *Alabama Journal of Mathematics*, 2009, pp 1 – 3.