International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878, Impact Factor (2014): 3.05

Assessment of Petroleum Reservoir Recovery Factor Using Complexity Scoring and Artificial Neural Network

Samuel A. Afari[1], Kwame Sarkodie[2], Wilberforce N. Aggrey[3], Anthony Morgan[4]

[1]Department of Petroleum Engineering, Kwame Nkrumah University of Science and Technology,P.M.B.Kumasi, Ghana
[2]Department of Petroleum Engineering, Kwame Nkrumah University of Science and Technology, P.M.B. Kumasi, Ghana
[3]Department of Petroleum Engineering, Kwame Nkrumah University of Science and Technology, P.M.B. Kumasi, Ghana
[4]Department of Petroleum Engineering, Kwame Nkrumah University of Science and Technology, P.M.B. Kumasi, Ghana

Abstract: ***The Recovery factor (R.F) is an important parameter needed for assessing the commercial viability of a petroleum reservoir. This parameter is however very difficult to determine as existing models require large number of parameters that must be known accurately. This paper demonstrates the use of complexity scoring approach and artificial neural network in predicting the recovery factor of petroleum reservoirs. Various oilfields with known complexity parameters and recovery factors were considered. These complexity parameters were then scored based on carefully defined criteria. The scored parameters were then used to train a carefully designed artificial neural network (ANN) which was then validated and tested for RF prediction. Results show good prediction of RF with this approach.***

Keywords: Recovery factor, Artificial Neural Network (ANN), complexity scoring, structural complexity, MATLAB

1. Introduction

Oil and gas (hydrocarbons) are very important energy sources worldwide. One objective of the petroleum industry when an oilfield is first discovered is to obtain an accurate estimate of the recoverable hydrocarbon volume in place before any capital is invested in production and development.

A field's potential is determined by its Recovery Factor (RF) value; defined as the fraction of the total volume of hydrocarbons originally in place (HOIP) that can be recovered over the course of a field's economic life. This signifies that the higher the RF of an oil and gas field, the more efficient and effective the crude oil has been produced from the reservoir. Optimizing the production of an oil or gas field must be the interest of both governments and operators, consequently maximizing recovery levels at the appropriate economic and political moment may prove too tempting to most; thus, the employment of RF as a measure of comparison to measure the performance of a reservoir. It is important to comprehend what it is, how it is calculated and the uncertainty inherent to the parameters from which it is derived.

Recovery factor, a seemingly trivial calculation has inherent uncertainty and can vary due to many factors. Many models have been developed to predict recovery factors. However most of these models are specific to the type and nature of reservoirs, fluids and wells. A common approach used to determine recovery factor through is via material balance calculations (Dake, 1998). An analytical model for predicting recovery factors in gas reservoirs has also been developed by Conelson (Conelson, 1974). Aguilera provides a good review of recovery factor prediction approaches in naturally fractured reservoirs (Aguilera, 1999). A modern approach is via reservoir simulators. These approaches also have the inherent problems of requiring several input parameters that must be accurately known making most of them impractical to use at the early stages of reservoir development when recovery factor is most needed for investment decision making.

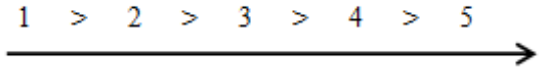
This paper demonstrates the use of complexity scoring approach and artificial neural network in predicting recovery factors. The two most important advantages of this approach is that it does not need accurate knowledge of its parameters and can be used even at the early stages of field development once the basic inputs can be "roughly" estimated.

1.1 Complexity Scoring

The scoring approach that was used by Wickens and Kelly (Wilkens & Kelly, 2010) was adopted in this work. The approach is based on the four main reservoir geological and fluid properties that have been studied to most affect recovery factor. The parameters are:

- Vertical Reservoir heterogeneity
- In-Situ Oil Viscosity
- Structural complexity
- SOIIP Areal density

Following the scoring approach, a value of 1 to 5 is assigned (scored) to the various parameters in order of their decreasing effects on recovery factor. That is, an assigned value of 1 to a particular parameter will show a higher influence of that parameter on the recovery factor. The trend decreases from 1 to 5 with a score of 5 bearing the lowest effect on the recovery factor.



Decreasing effect of parameter score on recovery factor

A scoring criterion was thus designed for each of the factors considered to affect recovery factor. Table 1 shows how

viscosity, STOIP areal density and Vertical reservoir heterogeneity are score.

Table 1: Summary of scoring system

Parameter	*Range*				
Score	1	2	3	4	5
Viscosity, cp	0 to 1	1 to 10	10 to 100	100 to 1000	$>10^3$
Vert. Res. Het. (k_{max}/k_{min})	1 to 10	10 to 100	100 to 1000	10^3 to 10^4	$>10^4$
STOIP Areal density, million Sm^3/km^2	>4.5	4.5 to 2	2 to 1	1 to 0.5	<0.5

Similarly compartmentalization/faulting (field geology) which are difficult to quantify similarly are scored according to their effects on recovery factor. Description of field geology that was used in defining and assigning a score to it as outlined in Table 2.

Table 2: Structural complexity scoring system

Score	*Description of Structural Complexity*
1	Excellent reservoir quality
2	Good reservoir quality with minor faults
3	Faulted complex reservoir with evenly distributed major and minor faults.
4	Highly faulted reservoir with major faults and relatively small minor faults
5	Highly faulted/ fractured reservoir compartmentalization.

1.2 Artificial Neural Network

Artificial neural networks are information processing systems that are a rough approximation and simplified simulation of the biological neuron network system. Artificial neural networks have the ability to recognize complex patterns quickly with a high degree of accuracy, it makes no assumptions about the nature and distribution of the data and they are not biased in their analysis. In addition, artificial neural networks have non-linear tools and as such are good at predicting non-linear behaviors. Neural networks form a broad category of computer algorithms that solve several types of problems, including pattern classification (Smith, et al., 1992), functions approximation (Zainuddin & Pauline, 2007), filtering (El-Hawary, 1994), optimization (Anant , Parag , & Mukesh , 2012) and automatic control (Bhongade, Gupta, & Tyagi, 2010). In the realm of petroleum engineering and geoscience, the concept has been applied to a number of problems (Li, Chan C.W, & Nguayen, 2013)-(Raeesi, Mordzedeh, Ardejani, & Rahimi,M., 2005).

The starting point for most neural networks is a model neuron, as shown in Figure 3. This neuron consists of multiple inputs and a single or multiple output. Each input is modified by a weight, which multiplies with the input value. The neuron will combine these weighted inputs and, with reference to a threshold value and activation function, use these to determine its output. Similar to how the brain learns by training, a neural network can be trained to recognize pattern after which it can be used to make predictions. Readers are referred to (Neocleous & Schizas, 2002) on the details of the concept.

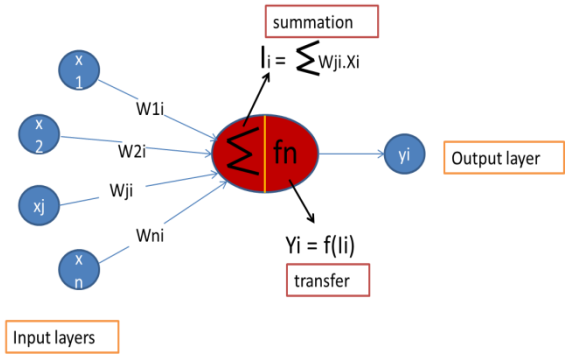


Figure 1: Basic structure of a neural network

2. Methodology

Figure 2 summarizes the methodology adapted in this work

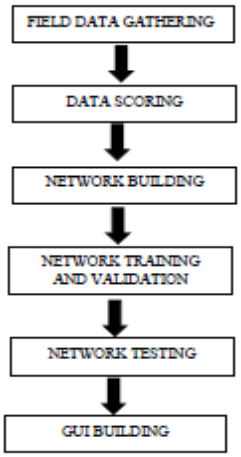


Figure 2: Summary flow-chart of research procedure

Viscosity, vertical reservoir heterogeneity, areal density, structural compartmentalization/ faulting and recovery factor data from various oilfields across the globe were gathered for the studies. The fields were not restricted particular to but consisted of fields various sizes, fluid properties and geological characteristics. This was adopted to prevent restrictions to the final model as much as possible. In all, sixty-two (62) fields were considered in this study. Table 3 shows descriptive statistics of the data collected.

Table 3: Summary statistics for properties form the 62 oilfields

Property/ Descriptive statistic	*Viscosity*	*Areal STOIP density*	*Vertical reservoir heterogeneity*	*Structural complexity*
Mean	73.03 cp	29.32 (x 10^6 sm3/km2)	2550.784	2.16
Standard Deviation	251.56 cp	50.83 (x10^6 sm3/km2)	13075.52	1.404834
Variance	62261.85cp^2	2542.17x(10^6 sm3/km2)^2	168211820.6	1.941727
Skewness	4.07	2.36	7.09	1.06

The complexity scoring approach adopted from Wilkins and Kelly (Wilkens & Kelly, 2010) was applied to the dataset. Below shows three examples of fields with their actual data (Tables 4 and 5) values and their score data (Table 6).

Table 4: Fluid and reservoir data for three field examples

Field	Recovery factor (%)	Viscosity, (cp)	Areal density (sm^3/km^2)	Vertical reservoir heterogeneity score, (k_{max}/k_{min})
Auk	18	0.9	1.26	10000
Fulmar	49	0.42	11.43	2000
Claire	14	6	18.1	2.5

Table 5: Geological/structural description of three field examples

Field	*Structural Complexity*
Auk	Vuggy, fractured dolomite reservoir with mostly secondary porosity and complex faults
Fulmar	Kimmeridge sandstone with large scale coarsening upwards sequence of excellent reservoir quality
Claire	Devonian-carboniferous fractured sandstones divided into nine fault-bounded segments

Table 6: Field score for three examples

Field	Viscosity score	Areal density score	Vertical reservoir heterogeneity score	Structural complexity score
Auk	1	3	5	3
Fulmar	1	1	2	1
Claire	2	1	1	5

MATLAB Neural network toolbox version 2013b was used to develop the artificial neural network. The Neural Network toolbox has the ability to design, implement, visualize, and simulate neural networks. Networks can automatically be created with default architectures or one can fully change it to suit their own network. Graphical user interfaces and command line functionality can be used. MATLAB was chosen due to its readily availability, user friendly nature and customizable nature.

The MATLAB function code takes information such as the variables, number of hidden units and training rule as arguments. Thus, the basic topology of the artificial neural network (the number of hidden layers, number of hidden units, the hidden unit transfer function, and training algorithm) is specified for the data to be inputted for the training to begin.

The precise number of hidden layers and hidden units required for a modeling task remains an open question (Zang, 2000). Thus, the structure of the neural network was developed based on trial and error. The network was initially created with one (1) hidden layer of ten (10) hidden units with gradient descent learning rule. Different architectures were then examined varying the various components of the neural network architecture on the training set to find the optimal architecture.

The total number of data set (number of fields) used was sixty two(62), with four data points each accounting for viscosity, structural compartmentalization/faulting, vertical reservoir heterogeneity and STOIIP areal density. Out of the 62 data set, 70% was used for training, 15% used for validation and 15% was saved to test the neural network.

MATLAB provides a displayed window which enables you to monitor the training progress and perform plot buttons to evaluate the performance of the network.

Errors between the target and the neural network output can be decreased by utilizing a supervised learning. With the intention of validating the neural model, this work investigates the least mean square error (LMSE) method, which is based on reducing the mean square error and is used to adjust the weights and biases. The mean square error is a performance function used in data training. The MSE can be defined as:

$$\mathrm{MSE} = \frac{1}{Q}\sum_{i}^{q}[e\,(k)]^2 = \frac{1}{Q}\sum_{I=1}^{q}[t(k) - a(k)]^2 \qquad (1)$$

MAPE is applied to calculate the average error between the actual measure and the prediction. The equation of MAPE can be defined:

$$\mathrm{MAPE} = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{|x(t)-\hat{x}(t)|}{x(t)}\right] x100 \qquad (2)$$

It is one complete presentation of the entire training data to the network during the training process. The number of epoch

has a significant effect on the output of the network, a very high or low value could cause either over fitting of the network training set or a poor performance/high MSE. The cross-validation stopping approach was used to control number of iteration using the performance plots as diagnostic tools. Over fitting sets in at the point where the MSE of training drastically reduces and MSE of validation increases, and at this point the training is stopped and the network saved.

After training of the network, a cross plot was generated with the target and the outputs. The ideal results will be for all the points to fall on the unit slope but this will largely depend on the accuracy of the model. The main factor that determines the accuracy of the model is the correlation coefficient (R). The closer the root mean square is to zero, the more accurate the results. An "R" value closer to 1 means that there is a strong correlation and the network is able to predict the expert's evaluation. Therefore for very accurate results, the two points have to have appreciable values closer to 0 and 1 respectively. Also, other plots such as the training, validation, testing and overall are generated.

3. Results and Discussion

Twelve different training algorithms varied over three hidden layer transfer functions (logsig, tansig and purelin) and one to twenty hidden unit sizes. In total 720 networks were generated. After many sets of neural network created, there were many outcomes as with regard to the performance of the network. The network topology and rules that gave the best performance is described in Table 7 below:

Table 7: Best network Attributes

Number of inputs	4
Number of hidden layer	1
Number of Hidden layer neurons	4
Training algorithm	Powell-Beale conjugate-gradient back propagation
Transfer function	Log-sigmoid function

Regression plots for this network show the target (known recovery potential) versus the network's output or predicted recovery factors. It is desired that all points fall on the unit slope line indicating there was no difference between the target and the prediction. This network yielded a test performance (correlation coefficient, R) of 96.086% with training and validation performances as 91.76% and 96.22% respectively. According to Zang (Zang, 2000), a very good fit to training sample but poorer fit to test sample is a symptom of over fitting and poor generalization. Thus lower training (compared to testing) correlation coefficient is desirable as it indicates the absence of over fitting but good generalization performance. The overall coefficient of correlation given by this network was 93.43% depicting a good relation between the predicted and targeted recovery factors. This is because an R value closer to one (1) means there is a strong correlation and the network is able to predict the target recovery factors. This gives the network a high degree of accuracy to predict other fields.

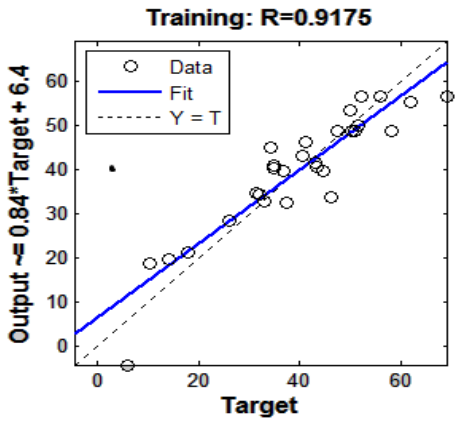


Figure 3: Regression plot for training dataset

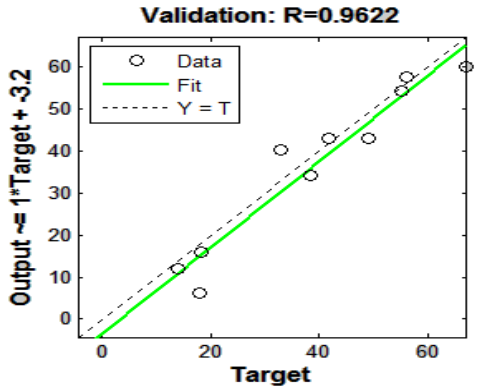


Figure 4: Regression plot for validation dataset

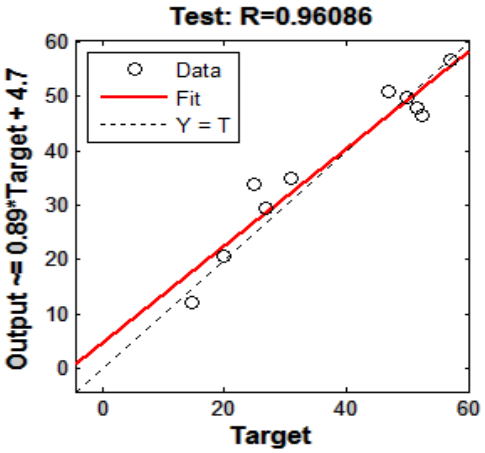


Figure 5: Regression plot for test dataset

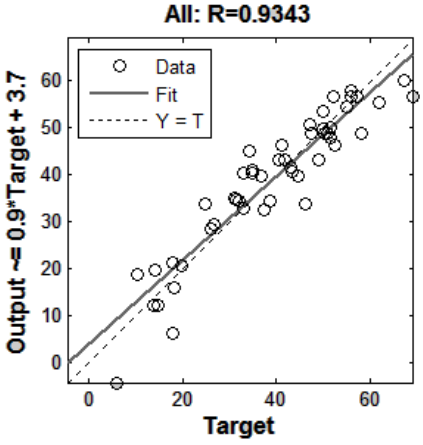


Figure 6: Regression plot for all dataset

A Mean square error performance plot was also generated, showing the performance of the network on testing, training and validation data sets. The dependent variable here is the mean squared error and the independent variable is the number of epochs or training iterations. As the number of epoch increases it is desired that the error also decreases till it flattens off at an acceptable level. From the **Figure 7**, training was optimal only at the 5th epoch while validation was optimal only at the 9th epoch, slightly higher than the training performance. This is an indication of the ease with which the network recognized the pattern within the dataset after a few iteration. It also indicates that the procedure is memory economic while still providing good prediction performance.

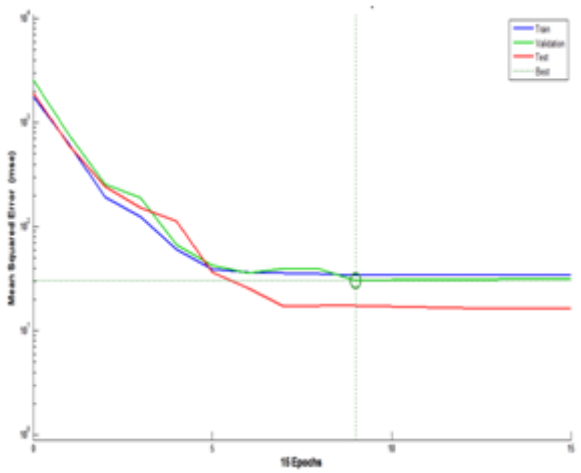


Figure 7: Means square error performance plot

A histogram analysis of the training, validation and testing errors can be seen in Fig. 7. It could be seen that the errors are normally distributed about the zero point. This indicates a mean error close to zero depicting good network prediction performance.

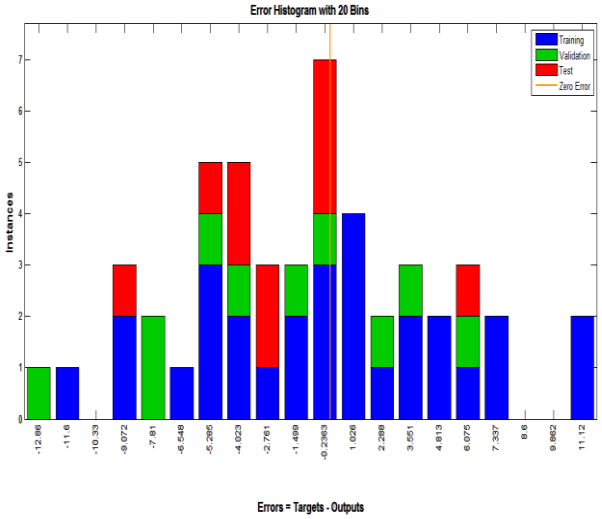


Figure 8: Error histogram with 20 bins

The final network, weights and biases have been saved and spiced into a user friendly software we call Recovery Potential Predictor (RP^2).With knowledge of the four main inputs (actual or scored), one can use this software to estimate a fields recovery factor in a fast, economical and reliable manner. Figure 8 shows the graphical user interface for the RP^2 software.

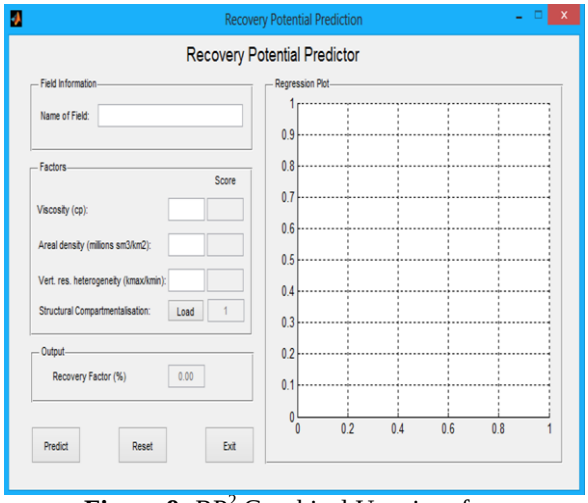


Figure 9: RP^2 Graphical User interface

4. Conclusion

In this study, we have developed an approach to forecast the recovery potential of oil fields with unknown recovery potential using complexity scoring and artificial neural network. The network shows very good prediction accuracy with a test correlation coefficient of about 96 % using dataset from sixty-two (62) fields. The procedure also proves to be memory economic with optimal training and validation performances reached only at the 5th and 9th epochs respectively.

The procedure would be very useful in estimation recovery factors with limited and roughly estimated reservoir properties. This makes it applicable even at the early stages of field development where accurate data is most of the times unavailable. It should be mentioned at this stage that the procedure is not a panacea and that accurate data should always be sought for recovery factor estimation.

References

[1] L. Dake, Fundamentals of reservoir engineering, Amsterdam: Elsevier Science B.V, 1998.
[2] D. Conelson, "Analytical Prediction of Natural Gas Reservoir Recovery Factors," Journal of Canadian Petroleum Technology, 1974.
[3] R. Aguilera, "Recovery Factors And Reserves In Naturally Fractured Reservoirs," Journal of Canadian Petroleum Technology, 1999.
[4] L. Wilkens and R. Kelly, "Rapid Assessment of Potential Recovery Factor;A new correlation demonstrated on U.K and U.S fields," in SPE Annual Technical Conference and Exhibition (ATCE), Florence,Italy, 2010.
[5] J. Smith, L. Breedveld, M. Derksen, G. Kateman, H. Balfoort, J. Snook and J. Hofstraat, "Pattern classification with artificial neural networks: classification of algae, based upon flow cytometer data," Analytica Chimica Acta, vol. 258, no. 1, pp. 11-25, 1992.
[6] Z. Zainuddin and O. Pauline, "Function Approximation Using Artificial Neural Networks," INTERNATIONAL JOURNAL OF SYSTEMS APPLICATIONS, ENGINEERING & DEVELOPMENT, vol. 1, no. 4, pp. 173-178, 2007.

[7] F. El-Hawary, "Artificial neural network for additive noise filtering techniques," in OCEANS '94. 'Oceans Engineering for Today's Technology and Tomorrow's Preservation.' Brest, 1994.
[8] B. Anant , D. Parag and S. Mukesh , "Artificial Neural Networks based Methodologies for Optimization of Engine Operations," International Journal of Scientific & Engineering Research, vol. 3, no. 5, pp. 1-5, 2012.
[9] S. Bhongade, H. Gupta and B. Tyagi, "Artificial neural network based automatic generation control scheme for deregulated electricity market," in IPEC, 2010 Conference Proceedings, IEEE, Singapore, 2010.
[10] X. Li, Chan C.W and H. Nguayen, "Application of Neural decison tree approach for prediction of petroleum production," Journal of Petroleum science and engineering, vol. 104, pp. 11-16, 2013.
[11] G. Moadi, M. Mohadesi and Moradi M.R, "Prediction of wax disapperance temperature using artificial neural networks," Journal of petroleum science and technology, vol. 108, pp. 74-81, 2013.
[12] M. Raeesi, A. Mordzedeh, F. Ardejani and Rahimi,M., "Classification and identification of hydrocarbon reservoir lithofacies and their heterogeneity using seismic attribute, logdata and artificial neural network," Journal of petroleum science and engineering, vol. 82, pp. 151-165, 2005.
[13] C. Neocleous and C. Schizas, "Artificial Neural Network Learning;a comparative review," in Methods and application of artificial intelligence, Thessaloniki, Springer Berlin Heidelberg, 2002, pp. 300-313.
[14] P. Zang, "Neural Networks for Classification: A Survey," ieee transactions on systems, man, and cybernetics—part c: applications and reviews, pp. 451-462, 2000

Authors' Profiles

Samuel A. Afari received his BSc. and MSc. degrees in Petroleum Engineering from the Heriot Watt University, UK (2012) and Kwame Nkrumah University of Science and Technology (KNUST), Ghana (2010), respectively. He currently serves as an Assistant lecturer at the department of petroleum engineering of KNUST and been teaching and active in research since 2012, with a number of publication to his credit.

Kwame Sarkodie is also currently an Assistant Lecturer at the department of petroleum engineering since 2012. He holds a BSc and MSc in Petroleum engineering from the Kwame Nkrumah University of science and technology, Ghana and Heriot watt university, UK respectively. He has a number of publications to his name.

Wilberforce Aggrey holds a BSc. and MSc. in petroleum engineering from the Kwame Nkrumah University of science and technology, Ghana and Heriot watt university, UK respectively. He is currently an assistant lecturer at the department of petroleum engineering of KNUST. Before coming to the department, he serves as a teaching assistant at the Oil and gas engineering department of the All Nations University, Ghana between 2009 and 2011. He also worked briefly with Shell Ghana ltd. Mr. Aggrey currently has a number of publications to his name.

Anthony Morgan is currently a teaching assistant at the department of petroleum engineering of KNUST. He holds a BSc. in Petroleum engineering with first class honors from the Kwame Nkrumah University of science and technology, Ghana.