CSCanada
Advances in Petroleum Exploration and Development
Vol. 5, No. 2, 2013, pp. 24-30
DOI:10.3968/j.aped.1925543820130502.1168

ISSN 1925-542X [Print]
ISSN 1925-5438 [Online]
www.cscanada.net
www.cscanada.org

Prediction of Oil Formation Volume Factor Using an Intelligent Tool: Artificial Neural Network

Azubuike, I. I[a]; Ikiensikimama, S. S[a],*

[a] Department of Petroleum & Gas Engineering, University of Port Harcourt, Port Harcourt, Nigeria.
* Corresponding author.

Received 5 April 2013; accepted 12 June 2013

Abstract

The Oil Formation Volume Factor parameter is a very important fluid property in reservoir engineering computations. Ideally, this property should be obtained from actual measurements. Quite often, this measurement is either not available, or very costly to obtain. In such cases, empirically derived correlations are used in the prediction of this property. This work focuses on the use of an intelligent tool known as an artificial neural network (ANN) to address the inaccuracy of empirical correlations used for predicting oil formation volume factor. The new intelligent model was developed using 448 published data from the Middle East, Malaysia, Africa, North Sea, Mediterranean basin, Gulf of Persian fields and 160 data set collected from the Niger Delta Region of Nigeria. The data set was randomly divided into three parts of which 60% was used for training, 20% for validation, and 20% for testing. Both quantitative and qualitative assessments were employed to evaluate the accuracy of the new intelligent model to the existing empirical correlations. The ANN intelligent model outperformed the existing empirical correlations by the statistical parameters used with a lowest rank of 0.6313 and better performance plot.

Key words: Oil formation volume factor; Empirical correlation; Artificial neural network; Back propagation; Statistical analysis

Azubuike, I. I, & Ikiensikimama, S. S (2013). Prediction of Oil Formation Volume Factor Using an Intelligent Tool: Artificial Neural Network. *Advances in Petroleum Exploration and Development*, *5*(2), 24-30. Available from: URL: http://www.cscanada.net/index.php/aped/article/view/j.aped.1925543820130502.1168
DOI: http://dx.doi.org/10.3968/j.aped.1925543820130502.1168

INTRODUCTION

Reservoir fluid behavior is one of the important parameters that needs to be understood for proper characterization of the fluids, material balance calculation, and economics of reservoir management. Among those PVT properties is the Oil Formation Volume Factor (FVF), which is defined as the volume of reservoir oil that would be occupied by one stock tank barrel oil plus any dissolved gas at the bubble point pressure and reservoir temperature. Precise prediction of FVF is very important in reservoir and production computations.

PVT properties are always determined experimentally based on actual samples collected from either well bore or at the surface. Such samples may be very expensive to obtain. In case where laboratory data are not available or unreliable[1, 2, 3] etc., derived the correlations to estimate the oil formation volume factor.

The correlations were developed using linear and non-linear regression or graphical techniques. The correlations are accurate within the range of data that were used to develop them[4].

To address the complexity and inaccuracy of the correlations, a new predictive tool was developed in this study to estimate the oil formation volume factor using artificial neural networks (ANNs). The ANNs are biologically inspired, non – algorithmic, non – digital, massively parallel distributive, adaptive information processing systems. They resemble the human brain in acquiring knowledge through learning process and in storing knowledge in interneuron connection strengths tasks[5, 6]. A new model was developed using 608 data sets: 372 sets were used to train the model, 118 sets to validate the relationship established during the training process, and another 118 sets were used to test the accuracy of the model.

The Theory that inspires neural network systems is drawn from many disciplines: primarily from neuroscience; engineering, and computer science;

secondarily from psychology, mathematics, and physics. These sciences are working toward the common goal of building intelligent system[6, 7]. Artificial neural network initially grew from the full understanding of some ideas and aspects about how biological systems work, especially the human brain. Neural systems are typically organized in layers. Layers are made up of a number of interconnected nodes (artificial neurons), which contain activation functions. Patterns are presented to the network via the input layer, which communicates to one or more hidden layers where the actual processing is done through a system of fully or partially weighted connections (Figure 1). The hidden layers then linked to the output layer. Neural network contains some sort of learning rule that modifies the weights of the connections according to the input patterns[8].

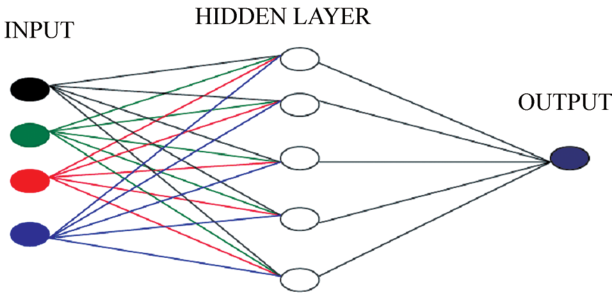

Figure 1
Schematic of an Artificial Neural Network with One-Hidden Layer

They are also called parallel-distributed processing system, which depicts the parallel operations of the nodes or neurons in a network processing system. Sometimes they are referred to as adaptive system, because the values of these connections can change so that the network can perform more effectively and efficiently.

The advantage of ANN over the conventional correlations is that, neural networks have large degrees of freedom for fitting parameters, and thus, capture the systems' non-linearity better than regression methods. They are also superior to the regression models in that they could be further trained and refined when additional data become available and hence improve their prediction accuracy. On the other hand, it is impossible to make any further change in a linear or non linear regression model as soon as a model development is over[9, 10, 11, 12].

Many investigators recognized that the neural network can serve the petroleum industry to create more accurate PVT models[13]. Gharbi, and El- Sharkawy (1997) Published two articles in this area[14]. The first article use the neural system to estimate the PVT data for Middle East crude oil reservoirs, while the second article was oriented to developing a universal neural network for predicting PVT properties for any oil reservoir.

Artificial neural network technique was in the field of PVT in order to estimate the formation volume factor at the bubble point pressure[4]. Their model was developed using 803 published data gathered from Malaysia, Middle East, Gulf of Mexico, and Colombia. They designed a three layer network; the input layer has four neurons covering the input data of gas-oil ratio, API gravity, relative gas density, and reservoir temperature; one hidden layer with five neurons and a single neuron for the formation volume factor in the output layer. This model showed a higher accuracy than the empirical correlations with an absolute average percent error of 1.789%, a standard deviation of 2.2053% and correlation coefficient of 0.988.

A work based on neural network using Matlab 7.5 to predict both bubble point pressure, and oil formation volume factor with the aid of two separated networks[7]. The data used was a set of 160 measured data points collected from the Middle East region; 140 points were used for training, and 20 for testing. The network performed better than empirical correlation with average relative error percent of 0.030704 and correlation coefficient of 0.9981.

Research on an Artificial Neural Network (ANN) was presented for an estimation of PVT properties of compounds[11]. The data set was collected from Perry's Chemical Engineers' Handbook. Different training schemes for the back propagation learning algorithm,; Scaled Conjugate Gradient (SCG), Levenberg-Marquardt (LM) and Resilient back Propagation (RP) methods were used. The accuracy and trend stability of the trained networks were tested. The LM algorithm with sixty neurons in the hidden layer proved to be the best suitable algorithm with the minimum Mean Square Error (MSE) of 0.000606. The ANN's capability to estimate the PVT properties is one of the best estimating method with high performance.

Recently, publication was made on an artificial neural network models to predict oil formation volume factor for different API gravity ranges[15]. The new models were developed using combination of 448 published data from the Middle East, Malaysia, Africa, North Sea, Mediterranean basin, Gulf of Persian fields and 1389 data set collected from the Niger Delta Region of Nigeria. Models were developed for four different API gravity classes: heavy oils for API $\leq$ 21, medium oils for 21 < API $\leq$ 26, blend oils for 26 < API $\leq$ 35 and light oils for API > 35. This models performed better than conversional empirical correlation developed to predict the same fluid property.

From the foregoings, works have been done in the areas of ANN for oil FVF prediction such as references [4, 7, 9, 11, 12, 14, 16, 17, 18]. However, most of these works concentrated on regional data. Attempt was made on using a universal ANN model for oil FVF using data from Malalysia, Middle East, Gulf of Mexico, and Colombia regions[4, 7, 14]. It was observed that some of the existing data in the literature from areas such as North Sea, Mediterranean basin, Gulf of Persia and Niger-

Delta Field were not included in[4] work who used data from references [14, 7]. Therefore, this study is centered on developing a general intelligent ANN model for predicting FVF, using published data from the Middle East, Malaysia, Africa (other countries expect Nigeria) and the Niger Delta Region of Nigeria. This work will also, evaluate and compare the accuracy of the ANN intelligent model to those of the existing empirical correlations.

1. DATA ACQUISITION AND ANALYSIS

Data used for this work were collected from published sources and Niger-Delta region of Nigeria. The Niger-Delta data sets were validated and it was assumed that the published data was validated by the authors. After the validation of the Niger-Delta data and the combination of the published data, 608 data sets were gotten. The 608 data set comprises; 93 data set from[19] , 195 data set from[20] , 160 data set from[21] and 160 data set from Niger-Delta. The data sets used in this work are similar to that used in[15] expect that 160 data set from the Niger Delta data pool was used. This work differ from that of[15] in that in this work a universal ANN model for oil FVF is sought for while that of[15] is for different API classes. Each data set contains reservoir temperature, oil gravity, total solution gas oil ratio, average gas gravity, and oil formation volume factor. Of the 608 data points, 372 were used to train the ANN models, the remaining 118 to cross-validate the relationships established during the training process and 118 to test the model to evaluate its accuracy and trend stability. The description of training and test data are given in Table 1 and Table 2, respectively.

Table 1
Summary of the Maximum and Minimum Values of Training Data for Oil Formation Volume Factor Neural Network

Parameter	GOR scf/stb	Gas relative density, lb/ft^3	API^o	Temperature ^{0}F	FVF bbl/stb
Maximum	3298.66	1.789	56.8	341.0	2.887
Minimum	8.61	0.501	6.0	74.0	1.017

Table 2
Summary of Maximum and Minimum Values of Test Data for Oil Formation Volume Factor Neural Network

Parameter	GOR scf/stb	Gas relative density, lb/ft^3	API^0	Temperature ^{0}F	FVF bbl/stb
Maximum	1700.0	1.789	50.90	302.0	2.047
Minimum	18.82	0.566	7.30	80.0	1.041

2. NEURAL NETWORK ARCHITECTURE

Matlab (7.5 version) neural network module was used to build the network using back-propagation algorithm with the Levenberg-Marquardt procedure for the optimization steps[22,15]. Back propagation Neural Network (BPNN) is a multi-layered network, and information flows from the input to the output through at least one hidden/middle layer. Each layer contains neurons that are connected to all neurons in the neighboring layers. The connections have numerical values (weights) associated with them. During the training phase, the weights are adjusted according to the generalized delta rule. Training is completed when the network is able to predict the given output. A three layers network was used in this work. A Levenberg- Marquardt algorithm was used to train the three-layer network. The first layer consists of four neurons representing the input values of reservoir temperature, solution gas oil ratio, gas specific gravity and API oil gravity. The second (hidden) layer consists of 25 neurons, and the third layer contains one neuron representing the output values of the oil formation volume factor. The data were divided into two groups: training group (490 data sets) and testing group (118 data sets). The training group is further split into two groups: the first (372 data sets) was used to train the network; the second set (118 data sets) was used to test the error during the training, this is called cross validation. Cross validation gives the ability to monitor the general performance of the network and prevent the network from over fitting the training data. In a BPNN, the input activity is transmitted forward while the error is propagated backwards. The neurons in the BPN use a transfer function that is sigmoid or S shaped. A key feature of the sigmoid function is that it has a minimum value of 0 and a maximum value of 1 and is differentiable everywhere with a positive slope. The derivative of the transfer function is required to calculate the error that is backpropagated and it is also easy to calculate.

3. QUANTITATIVE AND QUALITATIVE SCREENING

To compare the performance and accuracy of the new model to other empirical correlations, two forms of analysis were performed which include quantitative and qualitative. For quantitative screening method, statistical error analysis was used. The statistical parameters used for the assessment were percent mean relative error (E_r), percent mean absolute error (E_a), percent standard deviation relative (S_r), percent standard deviation absolute (S_a) and correlation coefficient (R).

For correlation comparison, a new approach of combining all the statistical parameters mentioned above

(E_r, E_a, S_r, S_a and R) into a single comparable parameter called Rank was used[23]. A brief description of the method follows. The use of multiple combination of statistical parameters in selecting the best correlation can be modeled as a constraint optimization problem with the function formulated as,

$$\text{Min } Z_i = \sum_{j=1}^{m} S_{i,j}\, q_{1,j} \quad (1)$$

subject to

$$\sum_{i=1}^{n} S_{i,j} \quad (2)$$

with

$$0 \le S_{ij} \le 1 \quad (3)$$

Where $S_{i,j}$ is the strength of the statistical parameter j of correlation i and q_{ij}, the statistical parameter j corresponding to correlation i. j = E_r, E_a, ….R^1 ,where R^1 = (1-R) and Z_i is the rank, RK (or weight) of the desired correlation. The optimization model outlined in equations 1 to 3 was adopted in a sensitivity analysis to obtain acceptable parameter strengths. The final acceptable parameter strengths so obtained for the quantitative screening are 0.4 for E_a, 0.2 for R, 0.15 for S_a, 0.15 for S_r, and 0.1 for E_r. Finally, equation 1 was used for the ranking. The correlation with the lowest rank was selected as the best correlation for that fluid property. It is necessary to mention that minimum values were expected to be best for all other statistical parameters adopted in this work except R, where a maximum value of 1 was expected. Since the optimization model (Equations 1 to 3) is of the minimizing sense a minimum value corresponding to R must be used. This minimum value was obtained in the form (1-R). This means the correlation that has the highest correlation coefficient (R) would have the minimum value in the form (1-R). In this form the parameter strength was also implemented to 1-R as a multiplier. Ranking of correlations was therefore made after the correlations had been evaluated against the available database.

For qualitative screening, performance plots were used. The performance plot is a graph of the predicted versus measured properties with a 45° reference line to readily ascertain the correlation's fitness and accuracy. A perfect correlation would plot as a straight line with a slope of 45°.

4. TREND ANALYSIS

Several authors[18,24,4,] had put forth the argument that the development of new empirical PVT correlations should always be followed with a check for the trend of the correlation model against the physical laws. This is because correlations exist[25] though performing well with the data set used for its development but, contradicted some of the physical laws or trend/patterns.

The trend analysis is to check ANN intelligent model against the physical laws. The following physical laws exist such that the oil FVF is an increasing function of reservoir temperature, gas gravity, solution gas oil ratio (GOR) and API gravity. For the trend test only one variable in the correlation was varied while the rest were held constant. Four points were selected, which covered the whole range of the reservoir properties. These values were taken as the means of the entire test data set grouped into four categories; heavy (API ≤ 21), medium (21 < API ≤ 26), blend (26 < API ≤ 35) and light (API > 35) oil.

5. RESULTS AND DISCUSSION

The trained ANN intelligent model was tested with 118 data points that were not previously used during training and validation. These data were randomly selected by the MATLAB tool to test the accuracy and stability of the model. The performance of the ANN intelligent model was compared with field data and the predictions from other empirical correlations such as references [1, 2, 21, 26, 19, 23]. These predictive correlations were carefully selected, having been developed specifically for the prediction of FVF in various regions that were considered in this study.

The results of the assessment as presented in Figure 2 gives statistical accuracies for all the oil FVF correlations examined. From this figure the ANN intelligent model ranked best with E_a of 0.9691 and correlation coefficient (R) of 0.9939.

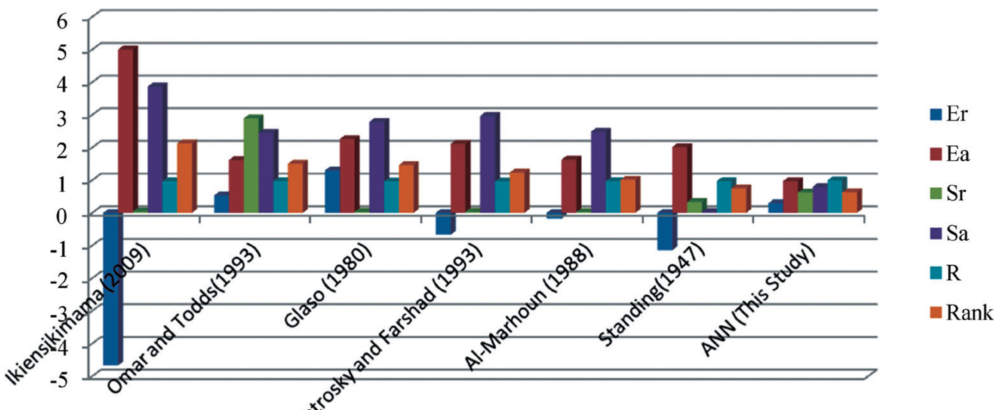

Figure 2
Comparison of the Statistical Accuracy for Different Correlations

Figures 3-9 illustrated cross plots of the predicted versus experimental FVF values. A cross plot is graph of predicted versus measured properties with a 45^0 reference line to readily ascertain the correlation's fitness and accuracy. Compared to other cross plots, Figure 9 shows the most tight cloud of points around the 45^o line indicating the excellent agreement between the experimental and the calculated data values. Again, this indicates the superior performance of the ANN intelligent model compared to other empirical correlations.

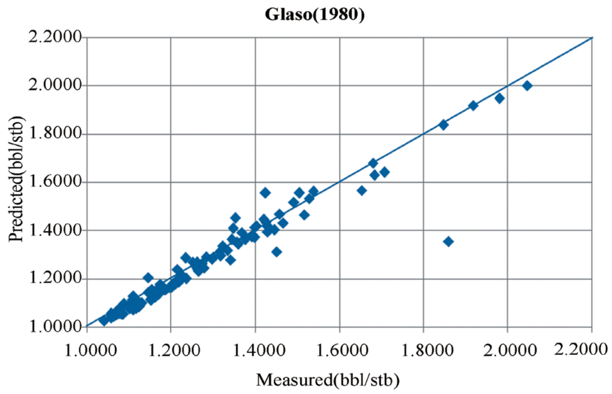

Figure 3
Cross Plot of Glaso Correlation

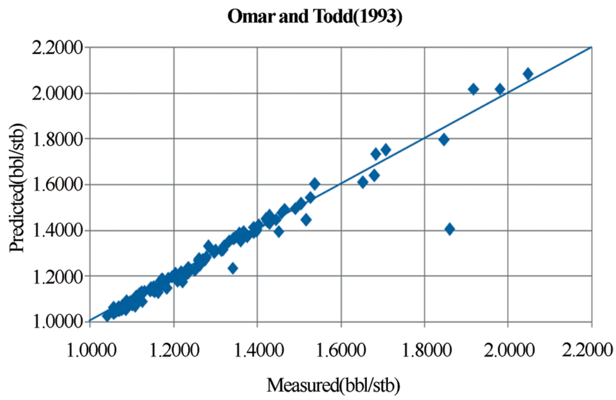

Figure 4
Cross Plot of Omar and Todd Correlation

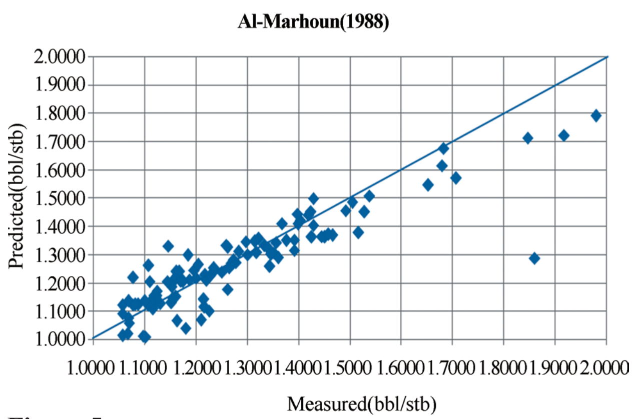

Figure 5
Cross Plot of Al-Marhoun Correlation

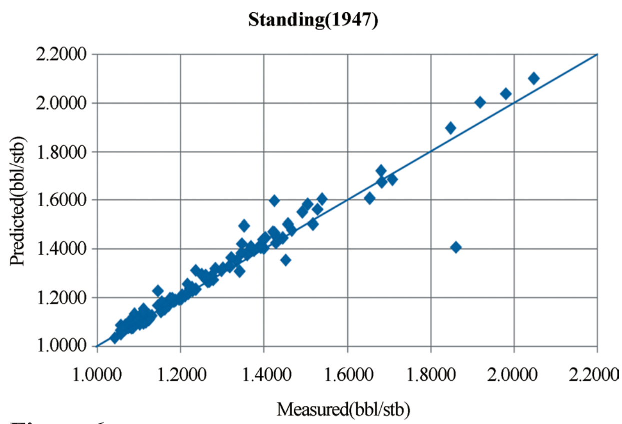

Figure 6
Cross Plot of Standing Correlation

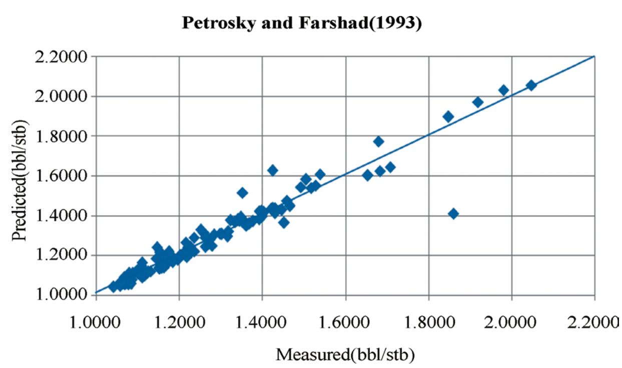

Figure 7
Cross Plot of Petrosky and Farshad Correlation

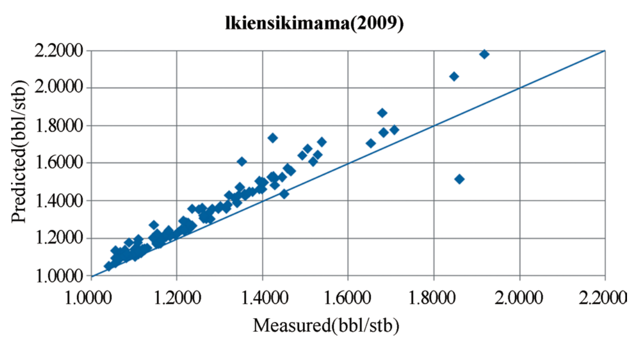

Figure 8
Cross Plot of Ikiensikimama Correlation

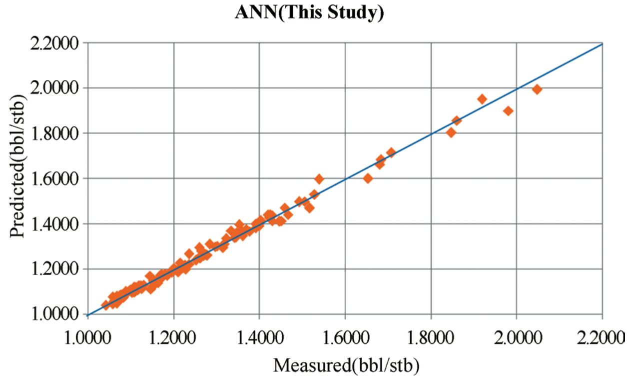

Figure 9
Cross Plot of ANN (This Study) Correlation

The model was found to be physically correct. Figures 10 to 13 showed that FVF increases with increase in reservoir temperature, solution gas oil ratio, gas gravity and oil API. The stability of the model indicated that the ANN intelligent model does not over fit the data, which implies that it was successfully trained.

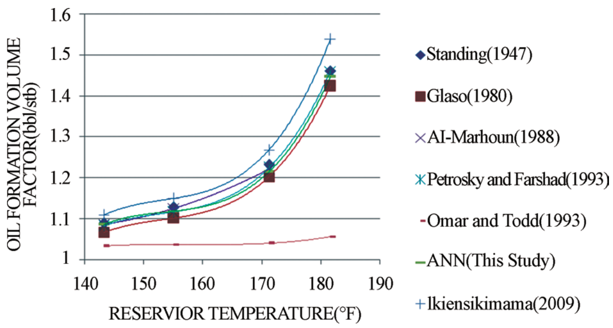

Figure 10
Temperature Trend for Oil Formation Volume Factor

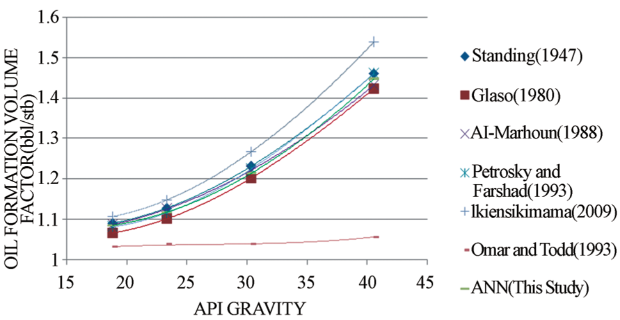

Figure 11
API Gravity Trend for Oil Formation Volume Factor

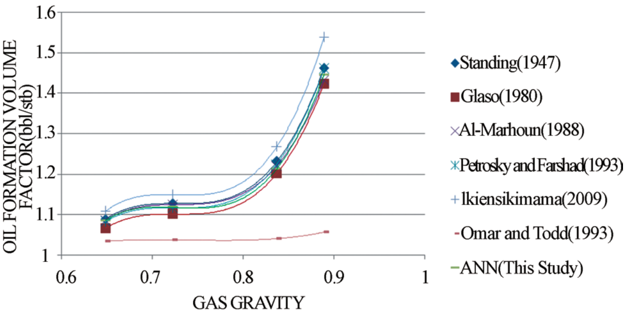

Figure 12
Gas Gravity Trend for Oil Formation Volume

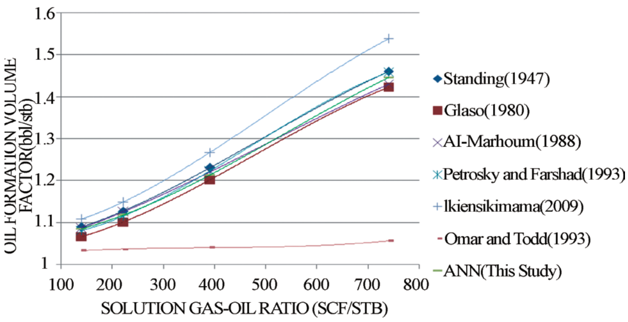

Figure 13
Solution Gas-Oil Ratio Trend for Oil Formation Volume Factor

CONCLUSIONS

The newly developed artificial neural network intelligent model for predicting crude oil formation volume factor was found to be better than the empirical correlations. The new intelligent tool outperformed the existing correlation by the statistical parameters used. It also shows a lowest rank of 0.6313 and better performance plot as compared to the existing empirical correlations for those regions where the data was used.

The trend analyses performed on the developed network shows that the ANN intelligent Model obeyed the physical laws and also, does not over fit the data, which implies that it was successfully trained.

REFERENCES

[1] Standing M. B. (1947). A Pressure-Volume-Temperature Correlation for Mixtures of California Oils and Gases. *Drill & Prod. Pract.*, *API*(1947), 275-87.

[2] Glaso, O. (1980). Generalized Pressure-Volume Temperature Correlations. *JPT*, *5*, 785.

[3] Beggs, H. D, & Vazquez, M. E. (1980). Correlation for Fluid Physical Property Prediction. *JPT* (June 1980), 968.

[4] Osman, E. A., Abdel-Wahhab, O. A., & Al-Marhoun, M. A. (2001, March). Prediction of Oil Properties Using Neural Networks. SPE Paper 68233 Presented at the SPE Middle East Oil Show Conference, Bahrain.

[5] Ali, J. K. (1994, March). Neural Networks: A New Tool for the Petroleum Industry. SPE Paper 27561 Presented at the European Petroleum Computer Conference, Aberdeen, U.K.

[6] Buscema, M (2002). A Brief Overview and Introduction to Artificial Neural Networks. *Substance Use & Misuse*, *37*(8-10), 1093-1149.

[7] Shokir, E. M., Goda, H. M., Sayyouh, M. H., & Fattah, K. A. (2004). Modeling Approach for Predicting PVT Data. *Engineering Journal of the University of Qatar*, *17*, 11-28.

[8] Kay, A. (2001). *Artifical Neural Networks*. Computer World 35, February.

[9] Deng, A. D. (2007). *Prediction of PVT Oil Properties Using Artificial Neural Network* (Master's thesis). University of Ibadan, Department of Petroleum Engineering, Ibadan, Nigeria.

[10] Gharbi, R. B., & Elsharkawy, A. M. (1997, April). Universal Neural-Network Model for Estimating the PVT Properties of Crude Oils. Paper SPE 38099 Presented at the SPE Asia Pacific Oil & Gas Conference, Kuala Lumpur, Malaysia.

[11] Moghadassi, A. R., Parvizian, F., Hosseini S. M., & Fazlali, A. R. (2009). A New Approach for Estimation of PVT Properties of Pure Gases Based on Artificial Neural Network Model. *Braz. J. Chem. Eng.*, *26*(1), 199-206.

[12] Omole, O., Falode, O. A., & Deng, A. D. (2009). Prediction of Nigerian Crude Oil Viscosity Using Artificial Neural Network. *Petroleum and Coal*, *151*(3), 181-188.

[13] Al-Marhoun, M. A., & Osman, E. A. (2002, October). Using Artificial Neural Networks to Develop New PVT Correlations for Saudi Crude Oils. Paper SPE 78592 Presented at the 10th Abu Dhabi International Petroleum Exhibition and Conference (ADIPEC), Abu Dhabi, UAE.

[14] Gharbi, R. B., Elsharkawy, A. M. (1997, March). Neural-Network Model for Estimating the PVT Properties of Middle East Crude Oils. Paper SPE 37695 Presented at the SPE Middle East Oil Show and Conference, Bahrain.

[15] Azubuike, I. I., & Ikiensikimama, S. S. (2013). Forecasting Oil Formation Volume Factor for API Gravity Ranges Using Artificial Neural Network. *Advances in Petroleum Exploration Development*, *5*(1), 14-21.

[16] Elsharkawy, A. M. (1998, October). Modeling the Properties of Crude Oil and Gas Systems Using RBF Network. Presented at the *SPE Asia Pacific Oil & Gas Conference*, Perth, Australia.

[17] Varotsis, N., Gaganis V., Nighswander, J., & Guieze P. (1999, October). A Novel Non-Iterative Method for the Prediction of the PVT Behavior of Reservoir Fluids. Paper SPE 56745 Presented at the 1999 SPE Annual Technic Conference and Exhibition, Houston, Texas.

[18] Al-Shammasi, H. Y. (2001). A Review of Bubblepoint Pressure and Oil Formation Volume Factor Correlations. *SPE Reservior Evaluation & Engineering*, 146-149.

[19] Omar, M. I., & Todd, A. C. (1993, February). Development of Modified Black oil Correlation for Malaysian Crudes. Presented at the 1993 SPE Asia Pacific Oil and Gas Conference, Singapore.

[20] De-Ghetto, G., Paone, F., & Alikhan, A. A. (1994, October). Reliability Analysis on PVT Correlation. Presented at the European Petroleum Conference, London, U.K.

[21] Al-Marhoun, M. A. (1988). PVT Correlations for Middle East Crude Oils. *Journal of Petroleum Technology*, *40*(5), 650-666.

[22] MATLAB. (2004). *Neural Network Toolbox Tutorial*.

[23] Ikiensikimama, S. S. (2009). *Reservoir Fluid Property Correlations*. Port Harcourt: Advances in Petroleum Engineering, hi Ikoku Petroleum Engineering Series, and IPS Publications.

[24] Al-Yousef, H. Y., & Al-Marhoun, M. A. (1993). Discussion of Correlation of PVT Properties for UAE Crudes. *SPE Formation Evaluation*, *3*, 80-81.

[25] Dokla, M., & Osman, M. (1992). Correlations of PVT Properties for UAE Crudes. *SPE Formation Evaluation*, *7*(1), 41-46.

[26] Petrosky, J., & Farshad, F. (1993, October). Pressure Volume Temperature Correlation for the Gulf of Mexico. Paper SPE 26644 Presented at the *1993 SPE Annual Technical Conference and Exhibition*, Houston, TX.