

Fuzzy Sequential forward search for Oil Formation Volume Factor Predictive Tool Factor for Niger Delta Crude Oil

K. K. Salam^{1*}, D. Araromi¹, S. S. Ikiensikimama² & A. O. Arinkoola¹

¹ Petroleum Engineering Unit, Department of Chemical Engineering,
Ladoke Akintola University of Technology, Ogbomosho, Nigeria

² Department of Petroleum & Gas Engineering, University of Port
Harcourt, Port Harcourt, Nigeria

Kazeem Kolapo SALAM (R. Eng)

*Petroleum Engineering Unit,
Department of Chemical Engineering,
P. M. B. 4000,*

*Ladoke Akintola University of Technology(LAUTECH),
Ogbomlso. Oyo State.
Nigeria.*

+234-805-706-6800

+234-703-438-4645.

emails: kaykaysalam@yahoo.co.uk, kaykaysalam@gmail.com

ABSTRACT

Accurate prediction of fluid properties are essentials for all reservoir engineering calculations such as estimation of reserves, well testing analysis and in numerical reservoir simulation. Oil formation volume factor is one of the properties that can either be gotten from empirical or experimental method.

This work focuses on the use of fuzzy sequential forward techniques to develop a oil formation volume factor model using 1,316 data obtained from 45 different oil fields in the Niger Delta, Nigeria. The data set was randomly divided into two parts with 750 used for training and 566 for testing. The model developed has the lowest Root Mean Square Error (RMSE) of 0.0784 when compared with published correlation used for prediction. The accuracy of the developed model was tested with crossplot and statistical analysis. The model developed outperformed the existing correlations when subjected to further statistical analysis.

Keyword: Formation volume factor, Fuzzy inference system, correlation, Artificial Intelligence, Root mean square error

1. INTRODUCTION

Accurate prediction of fluid properties are essentials for all reservoir engineering calculations such as estimation of reserves, well testing analysis and in numerical reservoir simulation [1, 2]. Any small error in

measurement of these properties results in inappropriate reservoir estimation and poor performance of the reservoir model. One of these properties is Formation Volume Factor(FVF), which is defined as the volume of reservoir oil that would be occupied by one stock tank barrel plus dissolved gas at the bubble point pressure and reservoir temperature[3]

Formation Volume Factor of oil can be obtained by either experimental method or estimated using empirically derived Pressure Volume Temperature (PVT) correlations. Experimental method gives accurate results under controlled conditions however, the results are heavily dependent on the validity of the reservoir fluid samples especially when reservoir has been depleted below bubble point pressure [4]. Empirical correlations are mostly developed using either multiple regression techniques or graphical techniques with each correlation developed for a certain range of fluid characteristics and geographical areas with similar fluid compositions. However, the accurate prediction of the correlations are not known beforehand and the correlations often perform poorly outside the range of data used [3, 4]. In the last half a century, a lot of correlations have been developed to predict FVF and other reservoir physical properties. A comprehensive review can be found in the works of (Al-Shammasi, 1999, Goda *et al*, 2003)[5, 6].

Artificial neural network (ANN) is a computer-based algorithm consisting of a pool of simple processing units which communicate by sending signals to each other over a large number of weighted connections to learn the behavior of a data population by self-tuning its parameter in such a way that the trained ANN matches the employed data accurately [7]. Graupe, 2007 highlighted some of the advantages of ANN [8]. ANN has been widely used in solving petroleum engineering problems and some of the work related to reservoir fluid properties is the work of Garbi and Elsharkawy who in 1997 used ANN to solve two problems related prediction of PVT properties [9]. Other works on PVT properties prediction using ANN are Goda *et al*, 2003, Osman *et al*, 2001, Al-Marhoun and Osman, 2002 [6,3,1]. In other areas of petroleum engineering, Goda *et al*, 2007 used ANN for prediction of Irreducible water saturation in Australian Hydrocarbon basin [10], Dewpoint pressure prediction for gas condensate reservoir [2], Permeability prediction [11] and so on.

ANN suffers some limitations despite its wide application in solving petroleum engineering problem. The summary of the shortcomings are highlighted in Cubillos, 2001, Olatunji *et al*, 2011 [12,13]. Olatunji *et al*, 2011 demonstrated that Sensitivity Based Linear Learning Methods (SBLLM) is an improvement that was able to solve the instability problem in conventional feed forward ANN when they used the SBLLM model to predict PVT properties of crude oil from data collected from three different sources. The SBLLM model was more stable when compared with ANN and the existing empirical correlation but the correlation coefficient of the SBLLM is less than that of ANN [13]. Fuzzy logic which is also an AI tool is capable of overcoming some of the limitations encountered in the use of ANN for solving problems[14]. The fuzzy system uses linguistic variable classification such as LOW, HIGH and MEDIUM to incorporate the imprecision inherent in many real systems including human reasoning; thus providing the ability to

capture knowledge from data that is inherently imprecise and maintaining a high level of performance in the presence of uncertainty[15]. This is important because most measured data are inherently imprecise because of individual differences, measurement errors and noise. Other key advantageous feature of this approach lies in its ability to arrive at transparent model that is easily interpretable.

Salam *et al* (2011) demonstrated superior performance of below bubble point viscosity model developed through Fuzzy logic techniques when compared their model with widely used empirical correlations in the industry[16]. In another research, Olatunji *et al*[17] used Type-2 fuzzy logic to predict Pressure Volume Temperature (PVT) properties of crude oil system using data from the work of Olatunji *et al* [13]. The developed model was compared with ANN model and three existing correlations used to predict the PVT properties. The fuzzy model outperform the ANN and three correlations. In other area especially where data collection are imprecise and characterized with error, Fuzzy logic have been proved reliable and stable in such environment [18,19]. This work will develop a prediction tool using Fuzzy logic using data collected across fields in Niger delta region in Nigeria and the model will be compared with existing correlations used in the region.

2. DATA GATHERING

The data used for this study were obtained from conventional PVT reports that derived the various fluid properties through differential liberation process from different oil fields in the Niger Delta. A total of 245 PVT field reports were used to extract 1316 datasets for Fuzzy prediction of the undersaturated oil FVF. Table 1 shows the PVT input data range, average and standard deviation of the data used for this study.

	Min	Max	Average	Stad dev.
FVF above bubble point, bbl/STB	1.022	3.2636	1.4887	0.3896
FVF at bubble point, bbl/STB	0.134	3.2705	1.4758	0.412
Gas gravity, (air=1)	0.564	1.29	0.7461	0.136
Gas Oil Ratio, SCF/STB	19	2948.8	861.76	645.15
Oil gravity,	0.766	0.964	0.8565	0.0465
Bubble point Pressure, psia	67	6560	2928.9	1226
Initial Pressure, psia	115	8451	3794.1	1237.3
Temperature, °F	104.4	264	172.79	29.76
API gravity, °API	15.2	53.23	34.19	8.8018
Compressibility factor, psia ⁻¹	3.4*10 ⁻⁷	0.0017	1.656*10 ⁻⁵	7*10 ⁻⁵

3. FEATURE SELECTION OF THE MODEL

Sequential Forward Search (SEQSRCH) is a data optimization tool that selects influence of parameter or combination of parameters on model prediction by building an ANFIS model for selected parameter, trains it for one epoch number and report it's RMSE value. The parameter with lowerest RMSE value will have much influence on the model to be developed. SEQSRCH performs an exhaustive search within the available inputs to select the set of inputs that have most pronounced influence on the output. The variables are Gas gravity, API gravity, Oil gravity, Temperature, Above bubble point pressure, Gas oil ratio, bubble point Formation volume factor.

Table 2: SEQSRCH results of parameters used for model development

Parameter	RMSE of Training	RMSE of Testing
A	0.0896	0.1141
B	0.0975	0.3744
C	0.1527	0.6607
D	0.1244	0.596
E	0.1319	0.6604
F	0.1241	0.5954
G	0.1629	0.7222
H	0.1262	0.5985
(AB)	0.0884	0.8291
(AC)	0.0893	0.1108
(AD)	0.0891	0.1051
(AF)	0.1157	0.5733
(FE)	0.1183	0.5971
(FG)	0.1102	0.5839
(FH)	0.0998	0.1272
(ABC)	0.0864	0.6241
(ABD)	0.0874	2.203
(FHE)	0.0999	1.1259
(FHG)	0.1039	0.8919
(ABF)	0.0876	2.1528
(ABCD)	0.0843	0.7035
(EFHG)	0.0944	1.3304
(ABCF)	0.0857	1.4954

Table 2 show the effect of individual variable on the proposed ANFIS model developed for one epoch number. Considering the RMSE values of the eight parameters subjected to SEQSRCH, the other of decreasing influence on ANFIS model are A, B, F, D, H, E, C and H. When the tool selected two parameters, seven different combinations were formed with their decreasing influence on ANFIS model for training and testing listed as; (AB), (AD), (AC), (FH), (FG), (FE) and (AF) respectively. Five

combinations were proposed by the SEQSRCH when ANFIS model was developed using three parameters and they are presented in decreasing order of their influence on the model as: (ABC), (ABD), (ABF), (BHE), and (FHG) respectively. When the ANFIS model was trained and tested with four different parameters selected by the model, three different combinations were proposed namely: (ABCD), (ABCF) and (EFHG).

The three parameter combination were categorised into the the following: Type 1 which comprises of A - Above bubble point pressure, B - Formation volume factor at bubble point, C - Oil gravity, and D - Gas oil ratio (ABCD); Type 2 comprises of A - Above bubble point pressure, B - Formation volume factor at bubble point, C - Oil gravity, and F - gas gravity.three variable chosen are Oil gravity, (ABCF) and Type 3 comprises of E - API gravity, F - gas gravity, G - Temperature and H - compressibility factor Formation volume factor at bubble point, and Gas-oil ratio (EFHG). Type 1 group has the lowerest RMSE values for both training and testing the model at one epoch number. The tool recommended Type 1 parameter combination as the best combination for the model.

3.1. ADAPTIVE NETWORK FUZZY LOGIC INFERENCE SYSTEM

Neuro-fuzzy model adopted in this work is based on the Adaptive Network based Fuzzy Inference System (ANFIS) [20]. The final output of the system is the weighted average of all rule outputs, computed as shown in Figure 1.

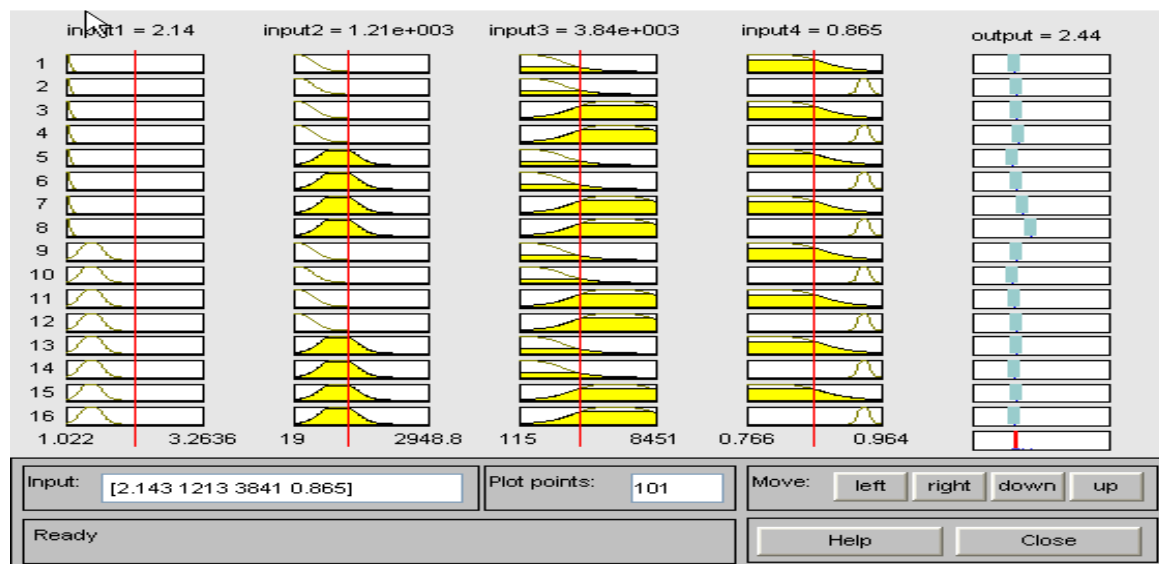


Fig. 1. Rule view of fuzzy inference system.

A four input layer network in the frame of Sugeno type model was used to implement ANFI network. The architecture of the fuzzy model is shown in Figure 2.

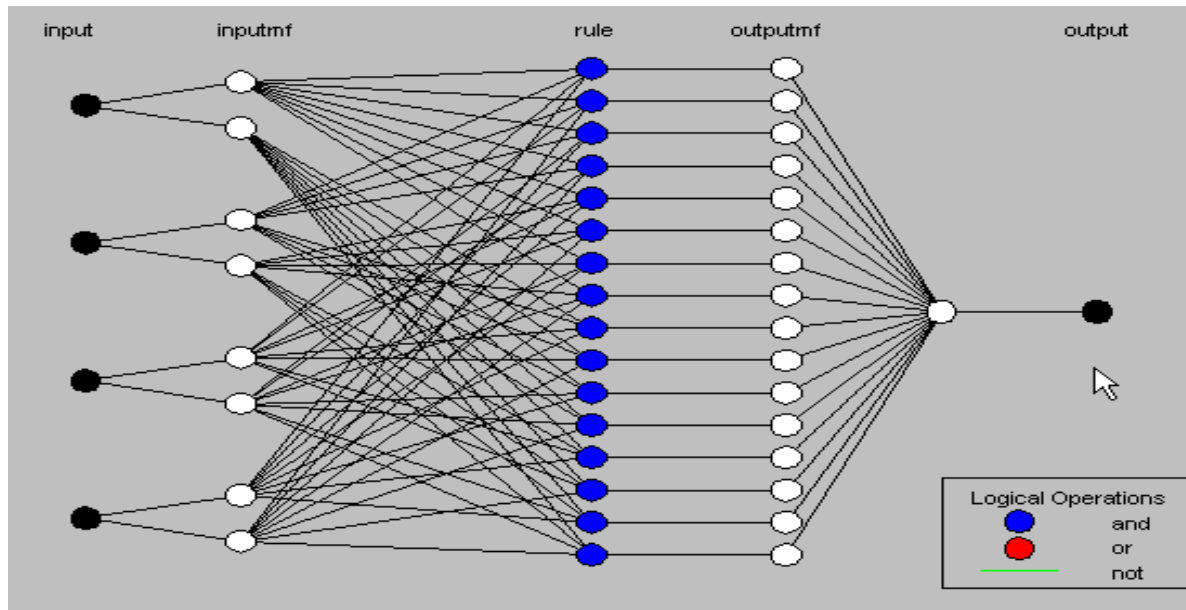


Figure 2: Schematic of fuzzy logic used for this work.

3.2. DATA PARTITIONING AND IMPLEMENTATION PROCESS

ANFIS used in this work is data-driven. This is because, the fuzzy rules were generated directly from the input-output data sets. In this case, the neuro-fuzzy model learnt the salient features in the data and automatically adjusted the system parameters so as to reduce the error between the calculated output and the measured output. In MATLAB scatter partition implemented as GENFIS2 was used to generate the inference system. The choice was based on its ability to model uncertainty of real measurements very well and also is well suited than ANFIS which can extract rules only where the premises are connected through differentiable t-norms to generates as many rules as the product of the number of linguistic terms per variable [21]. The model was run on a Duo core (1.6GHz, 2GB RAM) PC. The dataset was partitioned into 750 randomly selected data points for training and 566 randomly selected for checking the model.

4. RESULTS AND DISCUSSION

After development of ANFIS model trained with 750 random data point from a total of 1316 data points. The other 566 data that were not used during training of the ANFIS model were used to test the stability and accuracy of the model. There will be comparison between the existing empirical models and the developed model using the following statistical information: Absolute Absolute Percentage Error (AAPE),

Minimum and Maximum Average Absolute Percentage Relative Error ($E_{min}\%$ and $E_{max}\%$), Root Mean Square Error (RMSE) and correlation coefficient (R) residual error, r-square values, AAPE and RMSE. The statistical information are determined using the formula for calculation of AAPE, AAPRE, RMSE and R are reported in the work of Hemmati and Kharrat, 2007 [4].

4.1. Residual Errors

Figure 3 and 4 show the trend of reduction in error generated by the model for training and testing of the ANFIS for first 100 epoch numbers. The error generated by training data reduced from a RMSE of 0.0843 to 0.0447 and that of testing data reduced from a value of 0.7035 to a value of 0.045. The result recorded for training and testing error are in line with the value gotten when the parameters are combined and their maximum RMS error value are combined.

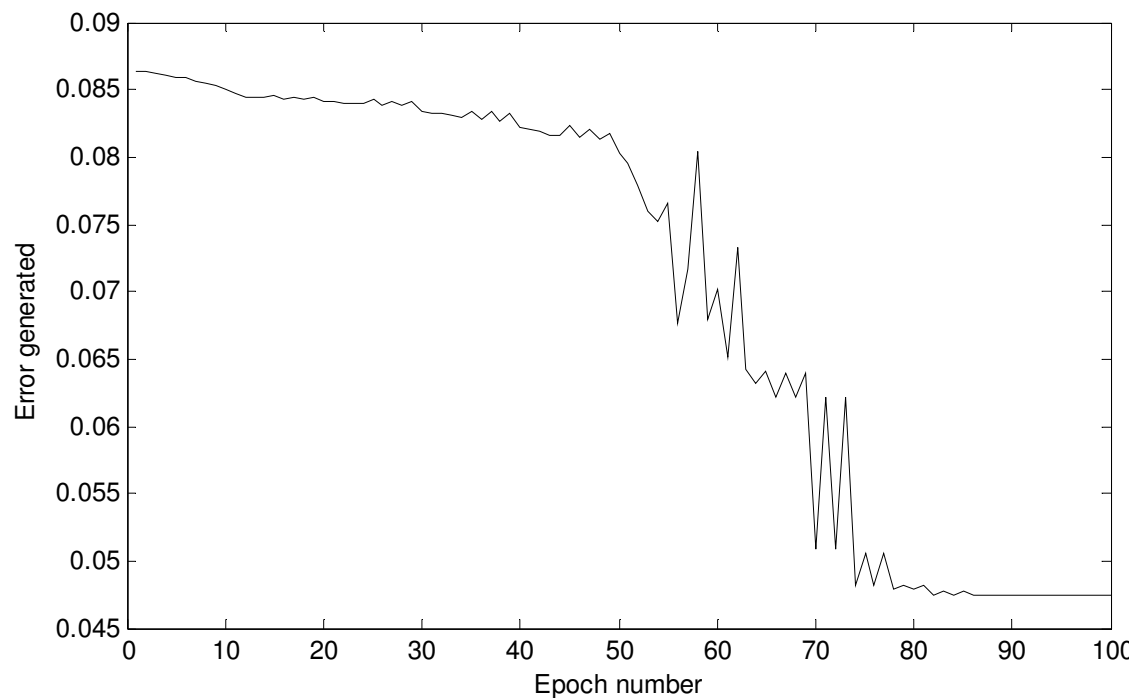


Fig. 3. Error trend of the training data for 100 epoch number

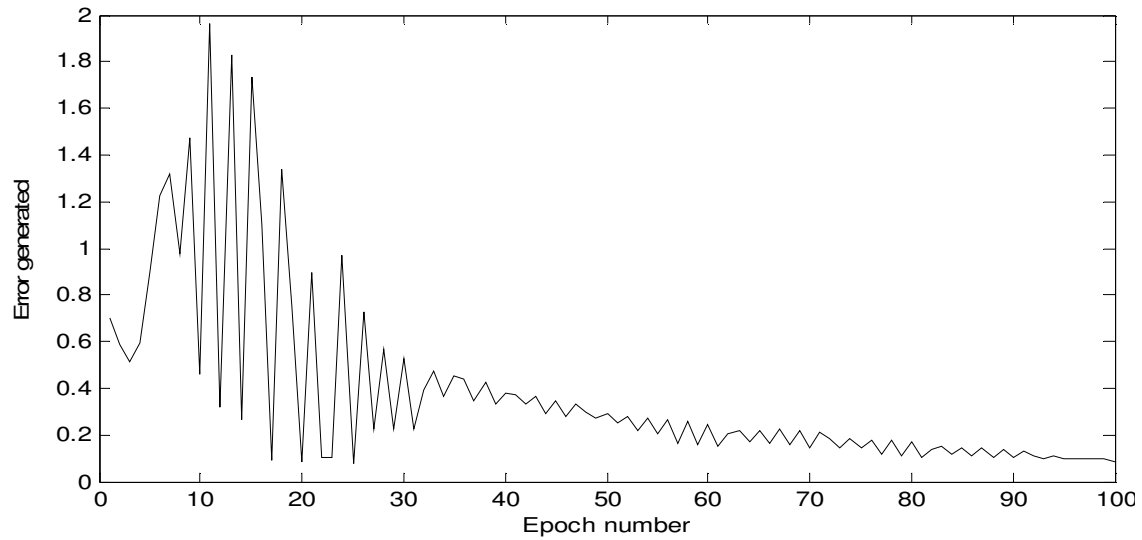


Fig. 4. Error of testing data for 100 epoch number

Figures 5 – 10 show the residual error distribution of the the training, testing and whole data and also of the empirical correlations about the zero line to check the error trend of the models. From the residual error of the training model (Figure 9), the values reduces from -0.7302 to 0.2218. It was also observed that highest residual was more pronounced between data 200 and 230.

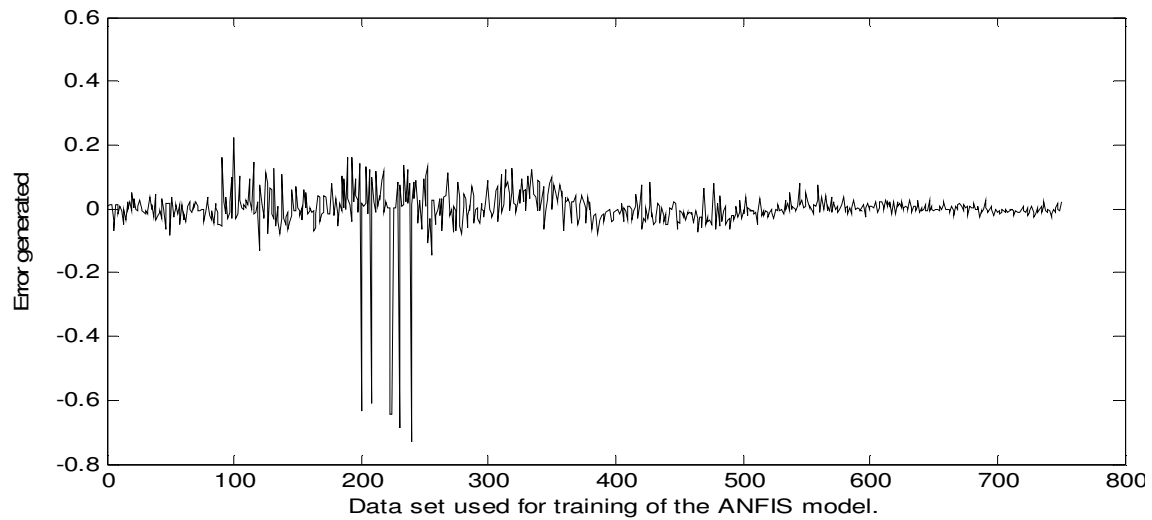


Fig. 5. Error Analysis for training data set.

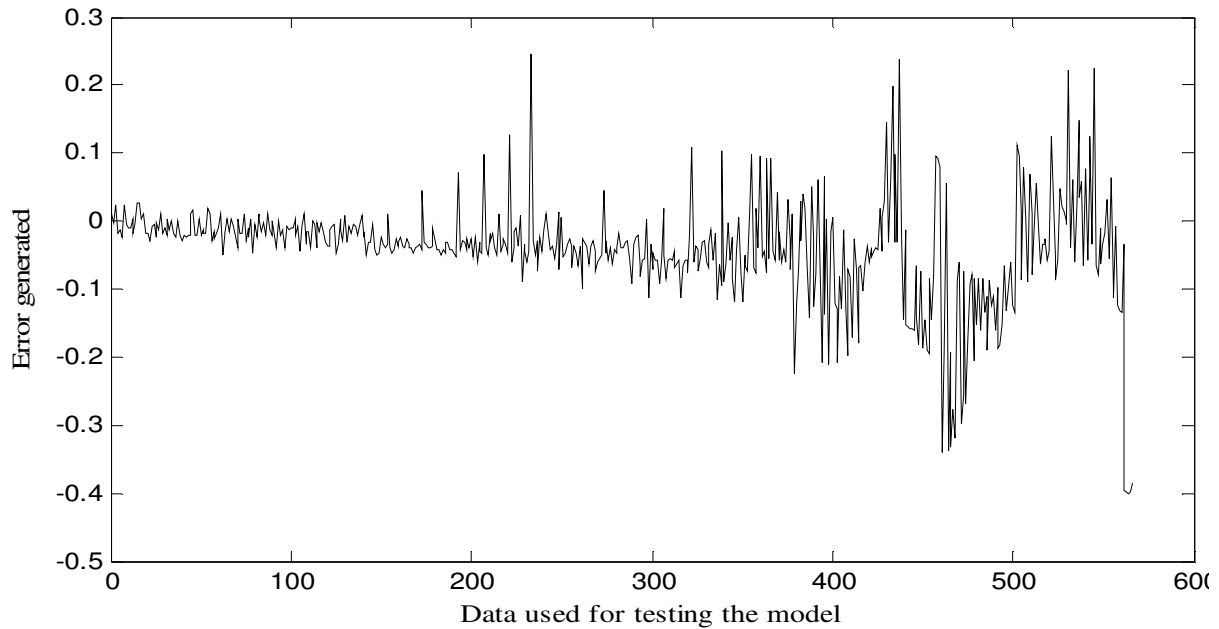


Fig. 6. Error Analysis of data used in testing the model

This could be as a result of imprecise measurement of the parameters and are catered for during **fuzzification** stage. The residual of testing model are between -0.40 and 0.2459. For the whole data used for the model development, the residual error are from -0.7302 to 0.2459.

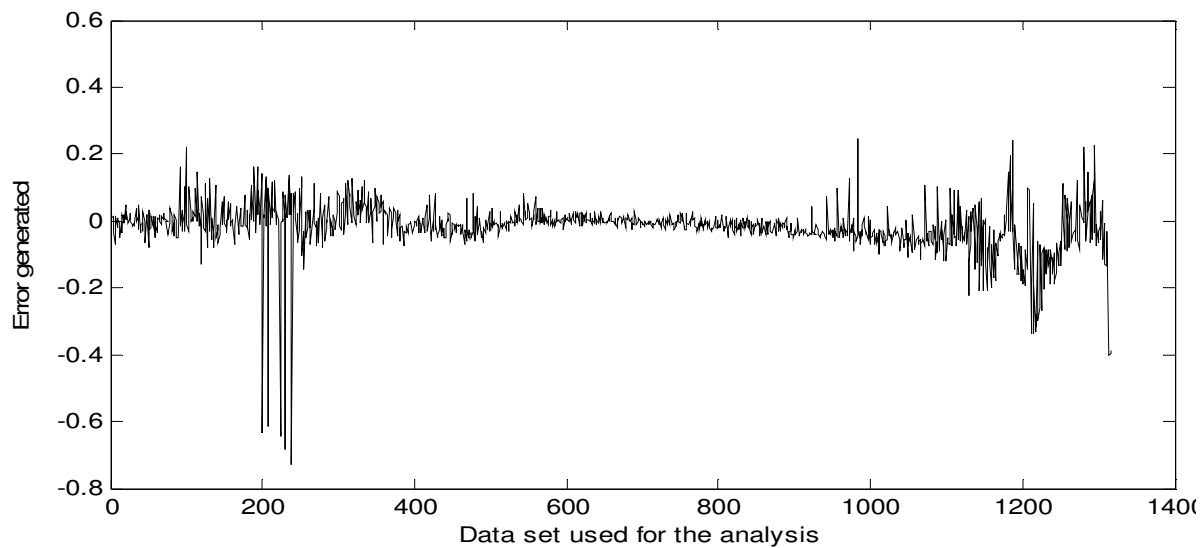


Fig. 7. Error Analysis of the ANFIS model using all the data

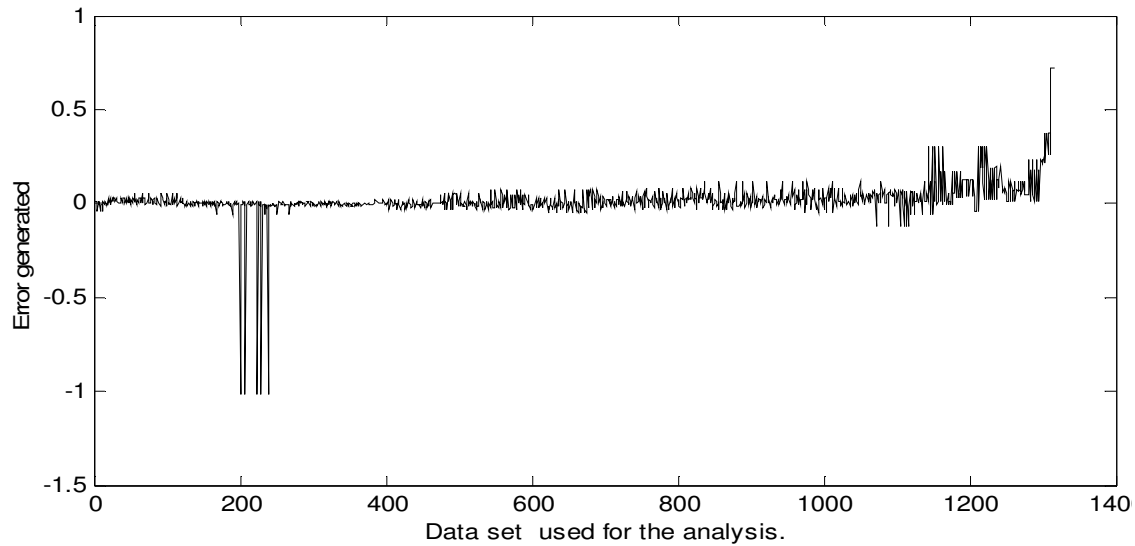


Fig. 8. Error analysis for AI marhoun model [22]

The residual analysis were also conducted on the predictions of three empirical correlations used for determination of FVF which are: Al marhoun (1988)[22], Almarhoun (1992) [23] and Vasquez and Beggs (1980) [24]. The residual of Almarhoun [22] are from -1.0154 to 0.7180. The effect of imprecise measurement of data is pronounced in the outcome of the residual error because the highest residual without 200 and 230 data points are -0.0156 which was further reduced to -1.0154 when the imprecise data points are considered.

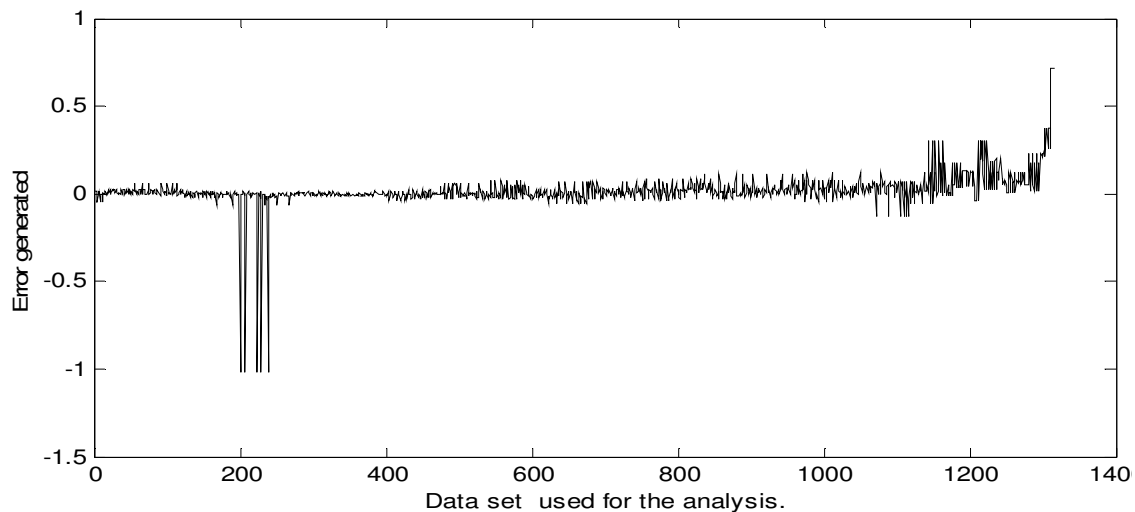


Fig.9 . Error analysis for AI marhoun model [23]

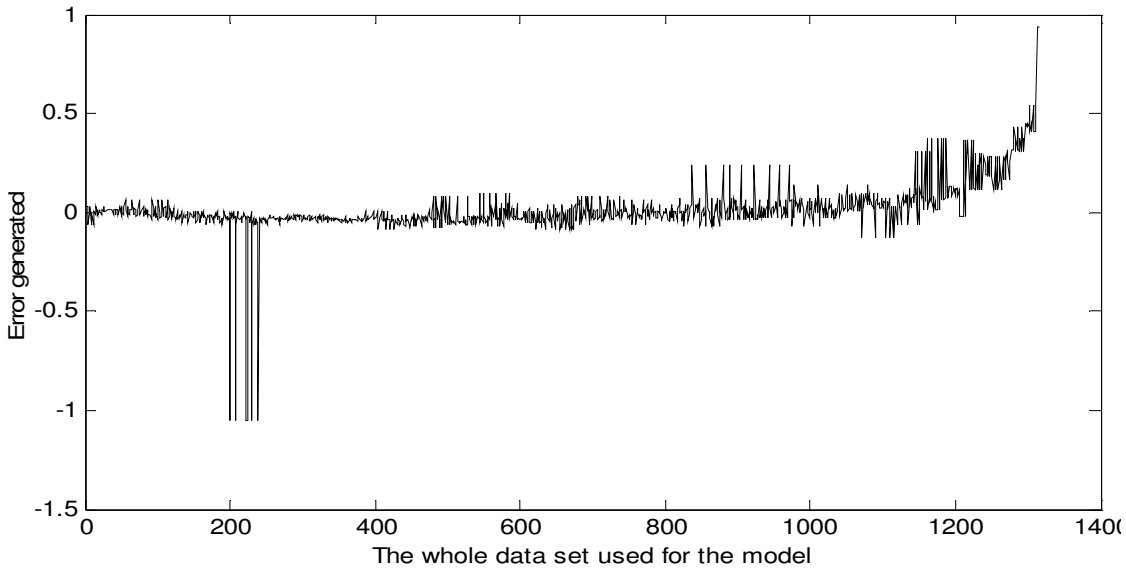


Fig. 10. Error Analysis of Van and Beggs model [24]

The above imprecision in data is pronounced in the model of Almarhoun (1980) [23] and Vasquez and Beggs (1980) [24]. The error of Almarhoun (1992) [23] are from -1.0154 to 0.5930 and that of Vasquez and Beggs (1980) [24] are from -1.0489 to 0.9355. From the outcome it was discovered that the ANFIS model at all stage outperform the empirical correlations of Al marhoun (1988) [22] , Almarhoun (1992) [23] and Vasquez and Beggs (1980)[24] based on its error values that range between -0.7302 to 0.2218 for training, -0.40 to 0.2459 for testing data and -0.7302 to 0.2459 for the whole data set.

4.2. Crossplot

Crossplot was plotted for the entire ANFIS model for training, testing and whole data as well as the three empirical correlations used for comparison in this study. The cross plot of experimental and estimated bubble point oil FVF were shown from Figure 11 all through 16. The closer the data point to the straight line on each graph, the better the model under investigation. This is interpreted by the correlation coefficient. The correlation coefficient of the ANFIS model has the highest value of 0.985 which show that they is a close agreement between the predicted and experimental data.

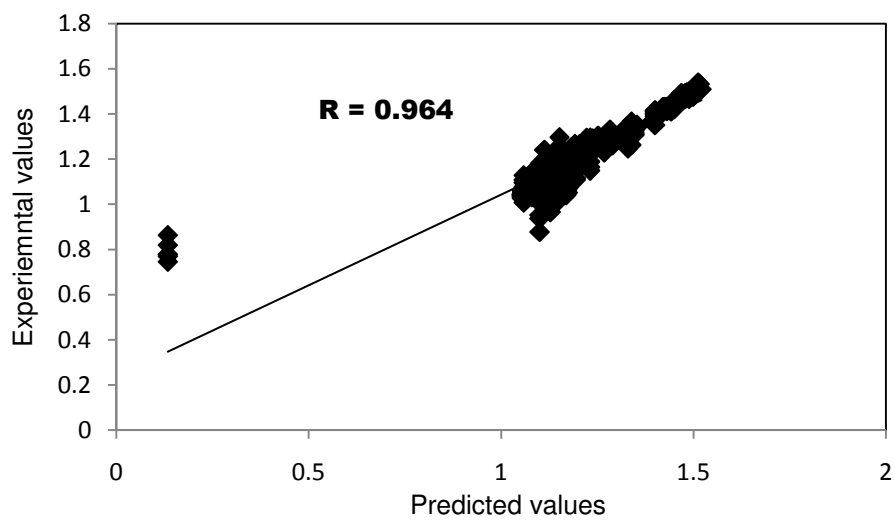


Fig. 11. Crossplot of Training data against predicted values

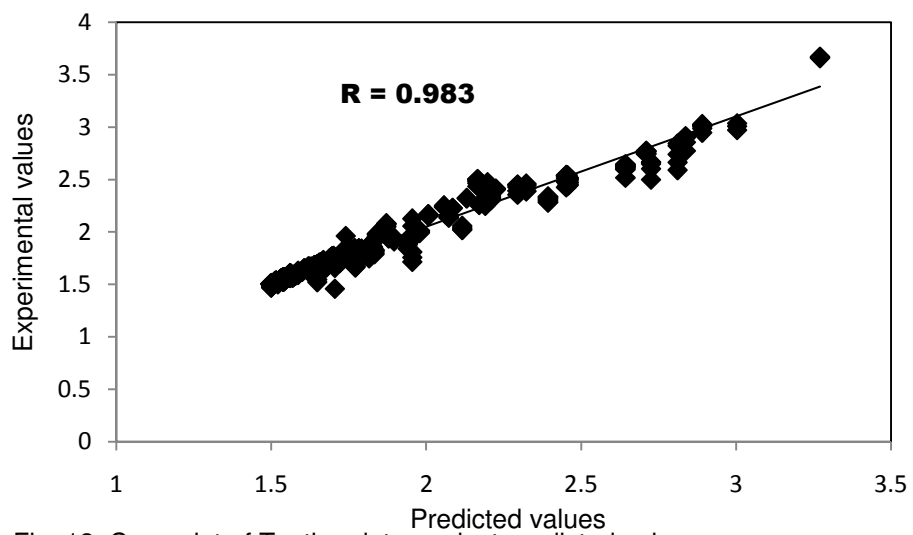


Fig. 12. Crossplot of Testing data against predicted values

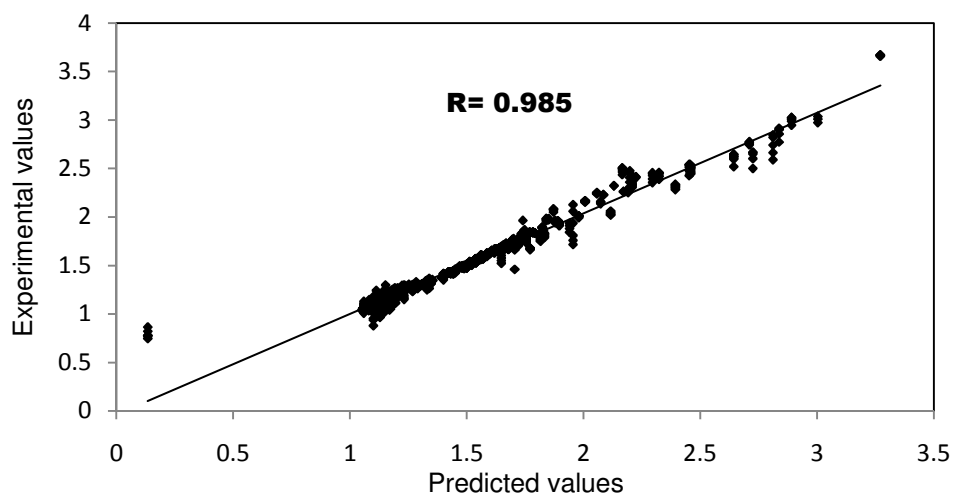


Fig. 13. Crossplot of whole data against predicted values

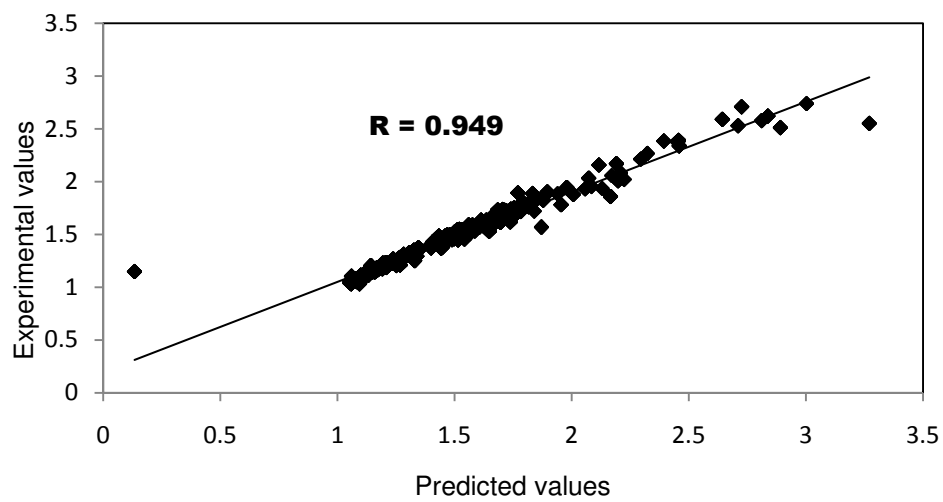


Fig.14. Crossplot of whole data against Almarhoun model [22]

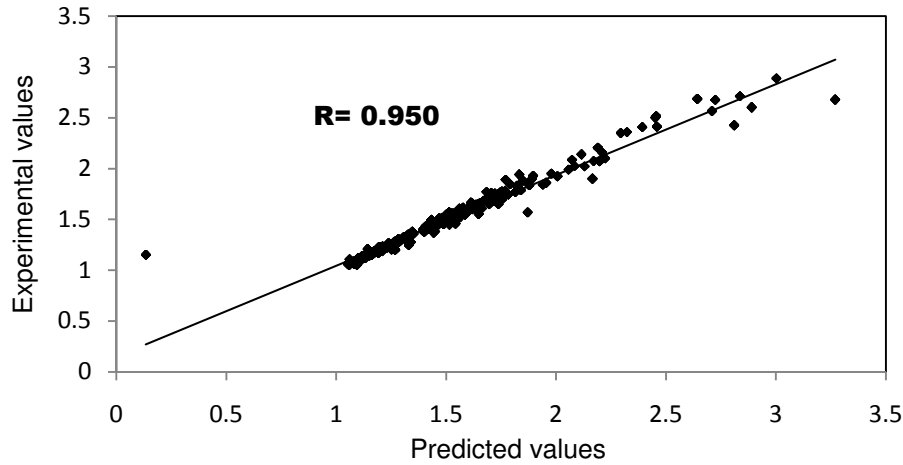


Fig. 15. Crossplot of whole data against Almarhoun model [23]

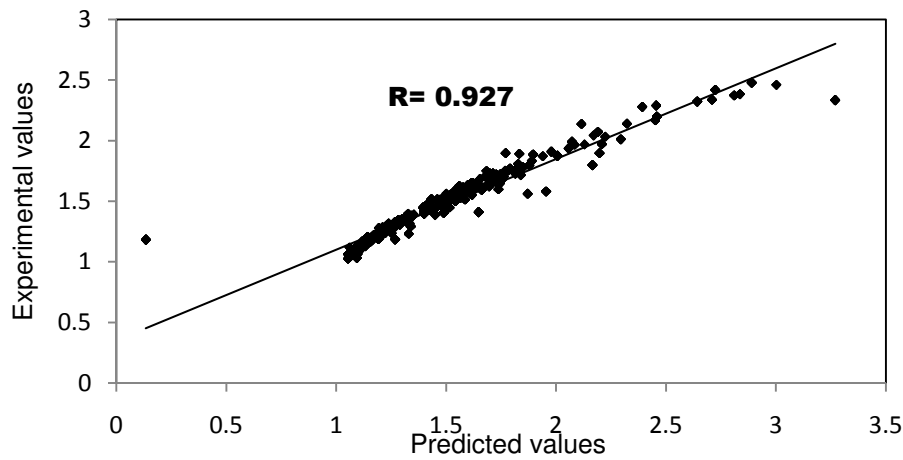


Fig. 16. Crossplot of Vasquez and Beggs model [24]

4.3. Statistical Analysis

The developed model was compared with existing empirical correlation using statistical analysis in order to evaluate the performance, suitability, and accuracy of the model. The analysis made use of the following statistical tools: Absolute Absolute Percentage Error (AAPE), Minimum and Maximum Average Absolute Percentage Relative Error (Emin% and Emax%), Root Mean Square Error (RMSE) and correlation coefficient (R).

Table 3: Comparison of Proposed model and existing empirical correlation

Model	Emin (%)	Emax (%)	AAPE	RMSE	R
Training	4.90E-06	0.7302	6.2176	0.0714	0.964
testing	3.04E-05	0.4001	3.0107	0.0868	0.983
whole	4.90E-06	0.7302	4.8384	0.0784	0.985
Almarhoun (1992)[23]	2.01E-06	1.0154	5.5576	0.103	0.949
Almarhoun (1998)[22]	1.29E-05	1.0169	5.2989	0.0951	0.950
Vasquez and Beggs (1980)[24]	1.66E-05	1.0489	7.0672	0.1367	0.927

The values of testing and whole data sets show in table 3 a high accuracy in predicting OFVF with their correlation coefficient values of 0.983 and 0.985 while the values from the three empirical correlations of Al marhoun (1988)[22] , Almarhoun (1992)[23] and Vasquez and Beggs (1980)[24] were 0.975, 0.974 and 0.963, respectively. The correlation coefficient of testing data is greater than training data, this could be attributed to the fact that there was a pronounced error between 200 and 230 data set which falls under training data set. The model is able to mimic the data for testing despite the uncertain nature of the data at the points mentioned. A similar trend of testing data greater than training data was also noticed in the work of Olatunji et al, (2011)[17]. A correlation coefficient value closer to 1 indicates that there is a good agreement between the experimental and predicted OFVF values. Also shown in table 3 are AAPE, RMSE, Emin, and Emax of the model developed as well as the three empirical correlations. The lower value of AAPE signifies a better accuracy of the correlation.

5. CONCLUSION

A model for the prediction of below point oil formation volume factor (OFVF) is developed through Fuzzy network and the following are deduced from the developed model that:

1. The trend of the error tend to reduce from 0.8 to 0.0023 for 100 epoch number which uphold better prediction of the model.
2. The model recorded the lowest value of Minimum AAPRE (Emin%), Maximum AAPRE (Emax%) , AAPE and RMSE values of 4.90E-06, 0.7302, 4.8384 and 0.0784 when compared with three existing correlations.
3. The prediction from the model gives a better prediction and accuracy than the published correlations with correlation coefficient of 0.985.
4. Error analysis conducted on all the model show that the developed model consistently predicted below bubble point oil formation volume factor with the lowest error.

REFERENCES

- [1] Al-Marhoun MA, Osman EA. Using Artificial Neural Network to develop new PVT correlation for Saudi Crude Oil. SPE Paper No. 78592 presented at the 10th Abu Dhabi International Petroleum Exhibition and conference, Abu Dhabi, United Arab Emirate, October 13-16. 2002; 1-10.
- [2] Meisam KA, Farhang JF, Yaser A. Dewpoint Pressure Estimation of Gas Condensate Reservoir, Using Artificial Neural Network. SPE Paper No. 107032 presented at the 2007 SPE Europe/EAGE Annual Conference and Exhibition, London, United Kingdom, June 11-14. 2007; 1-10.
- [3] Osman EA, Abdel-Wahhab OA, Al-Marhoun MA. Prediction of Oil PVT Properties Using Neural Networks. SPE Paper No. 68233 SPE Middle East Oil Show, Bahrain, Kingdom of Bahrain, March 17-20. 2001; 1-15.
- [4] Hemmati MN Kharrat R. A Correlation Approach for Prediction of Crude-Oil PVT Properties. Paper No. 104543, 2007 15th SPE Middle East Oil & Gas Show and Conference, Bahrain, Kingdom of Bahrain, March 11-14. 2007; 1-15.
- [5] Al-Shammasi AA.. Bubble Point Pressure and Oil Formation Volume Factor Correlations. SPE Paper No. 53185 presented at the SPE Middle East Oil Show, Bahrain, Kingdom of Bahrain, February 20-23. 1999; 1-17.
- [6] Goda HM, El-M Sohir EM, Fattah, KA, Sayyouch MH. Prediction of PVT Data using Neural Nstwork Computing Theory. SPE Paper No. 85650 presented at the 27th Annual SPE International Technical Conference and Exhibition, Abuja, Nigeria, August 4-6. 2003; 1-14.
- [7] Obanijesu EO, Araromi DO. Predicting Bubble-point Pressure and Formation-volume Factor of Nigerian Crude Oil System for Environmental Sustainability, Journal Petroleum Science and Technology, 2008; 26 (7): 1993-2008
- [8] Graupe D. Principles of Artificial Neural Networks Advanced Series on Circuits and Systems, 2nd Ed. Vol. 6 . World Scientific Publishing, Singapore. 2007.
- [9] Garbi RB, Elsharkawy AM. Universal Neural Network Model for estimating the PVT properties for Crude Oil System, SPE Paper No. 38099 presented at the SPE Asia Pacific Oil & Gas Conference, Kuala Lumpur, Malaysia. April 14-16. 1997; 1-10.
- [10] Goda MH, Meier MH, and Behrenbruch P. Use of Artificial Intelligence Technique for Predicting Irreducible Water Saturation. SPE Paper No. 109886 presented at the Asia Pacific Oil and Gas Conference, Jakarta, Indonesia, October 30- November 1. 2007; 1-10.
- [11] Singh S. Permeability Prediction Using Artificial Neural Network: A Case Study of Uinta Basin. SPE Paper No. 99286 presented at the SPE International Student Contest at the SPE Annual Technical Conference and Exhibition, Dallas, Texas, October 9-12. 2005; 1-8.
- [12] Cubillos F, Callejas H, Lima EL Vega MP. Adaptive Control Using a Hybrid Neural Model: Application to a Polymerisation Reactor. Brazilian Journal of Chemical Engineering. 2001; 18(1): 111-120.

- [13] Olatunji SO, Selamat A, AbdulRaheem A, Omatu S. Modeling the correlations of crude oil properties based on sensitivity based linear learning method. *Engineering Applications of Artificial Intelligence*. 2011; 24: 686-696.
- [14] Aguilar J, Cerrada M. Fuzzy classifier system and generic programming on system identification. *Proceedings of the Generic and Evolutionary Computation Conference*, San Francisco, July 7–11. 2001; 1245–1251.
- [15] Roffel B, Betlem B. *Process Dynamics and Control: Modelling for Control and Prediction*. 2nd ed. John Wiley and Sons Ltd. 2006.
- [16] Salam KK, Araromi DO, & Ikiensikimama SS. Neuro-Fuzzy Modeling for the Prediction of Below-Bubble-Point Viscosity. *Petroleum Science and Technology*. 2011; 29(17): 1741-1752.
- [17] Olatunji SO, Selamat A, AbdulRaheem A. Predicting correlations properties of crude oil systems using type-2 fuzzy logic systems. *Expert Systems with Applications*. 2011; 38: [10911–10922](#).
- [18] Wei J, Zhao J. Pricing decisions with retail competition in a fuzzy closed-loop supply chain. *Expert Systems with Applications*. 2011; 38: 11209–11216.
- [19] Samanta B. Prediction of chaotic time series using computational intelligence. *Expert Systems with Applications* 2011; 38: 11406–11411.
- [20] Jang JS. Anfis: Adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics*. 1993; 23(3): 665–685.
- [21] Abraham A. *Neuro-Fuzzy Systems: State-of-the-Art Modeling Techniques*. In *Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence*, Springer, Berlin, J. Mira, A. Prieto (Eds.). 2001.
- [22] Al-Marhoun MA. PVT Correlations for Middle East Crude Oils. *JPT*. 1988; 650.
- [23] Al-Marhoun MA. New Correlation for formation Volume Factor of oil and gas Mixtures, *JCPT*. 1992; 22.
- [24] Vazquez M and Beggs HD. Correlation for Fluid Physical Property Prediction. *JPT*. 1980; 968.

