Estimation of PVT Properties Using Artificial Neural Networks and Comparison of Results with Experimental Data

Masoud Mehrizadeh

Department of Petroleum Engineering, Khazar University, Baku, Azerbaijan
E-mail: mmehrizadeh@khazar.org

Abstract

The importance of pressure, volume and temperature (PVT) properties, bubble pressure, gas-to-oil solubility ratio and oil volume factor have made it necessary to precisely determine these properties for the calculation of reservoir performance. In the absence of laboratory measurements to determine the PVT properties of crude oil, two methods, used commonly, include the equations of state and the experimental relations of PVT. The equation of state is based on the information concerned with the fluid composition details of the reservoir, which is very expensive and time-consuming to determine. Whereas, PVT relationships are based on data obtained from easily measured ground layers. These data include reservoir pressure, reservoir temperature, and the specific gravity of oil and gas. Recent Studies show that artificial neural networks have a great ability to predict the PVT properties. Due to the training capability in neural networks, these networks were rapidly applied in engineering and were widely used in petroleum engineering. Estimation of porosity and permeability of reservoirs, prediction of outflow generated by oil wells, estimation of oil recovery, prediction of asphaltene deposition and estimation of PVT properties are the most important applications of artificial neural networks in petroleum engineering. By preparing and collecting more than 1000 PVT data related to southern Iran reservoirs, 577 data were selected to be used in the project and were randomly divided into two parts, 486 data for network training and 91 data for testing. The three-layer structure (one hidden layer with 6 neurons) was selected as the best structure and the batch back-propagation training method as the best learning algorithm. The results of the network were in a good agreement with experimental data, which the average relative error using training set in estimation of the volume factor and densities of oil were 0.557 and 0.509% respectively and using test data were 1.032 and 1.104% respectively.

Keywords: neuron, neural network, error backpropagation, neural network application, PVT properties, supersaturated reservoirs.

Introduction

PVT properties such as oil formation volume factor (B_O) and density (ρ_O) are used extensively in almost all petroleum engineering issues, especially in the reservoir engineering sector. For example, the application of these properties can be seen in issues such as material balance, well testing, reservoir simulation, etc (Al-Marhoun & Osman, 2002).

Actually, these properties should be experimentally determined in the laboratory, but usually experimental data are not always available and require a great deal of time and investment. There are many empirical correlations that estimate the different PVT properties. These relationships are obtained using linear or nonlinear regressions and sometimes predict these properties in the form of graphs. These relationships and correlations mainly are not accurate due to using limited empirical data extracting them. These relations and correlations are not accurate due mainly to applying limited empirical data out of which they are extracted.

Researchers have recently used the idea of artificial neural networks to determine the PVT properties. Applying this method, albeit of having the limitations mentioned in the paragraph above, has an important learning advantage, which means one can train the network and update the network whenever new data is available.

Subject literature and previous studies

Gharbi and Elsharkawy (1999) presented a neural network model for predicting PVT properties of Middle Eastern crude oil. The data used to train this network (498 data) include the most collected ones to develop the model for predicting PVT properties from Middle Eastern crude oil resources. The model can predict bubble pressure and oil volume factor as a function of dissolved gas to oil ratio, gas specific gravity, oil specific gravity and temperature. In this study, the network results are compared with the results of the empirical relations for the Middle East. The empirical relationships whose results are compared with the network results include; Al-Marhoun, Standing and GlasØ.

Marhoun and Osman (2002) developed a PVT neural network to estimate bubble point pressure. The model design uses 803 recorded data collected from Malaysia, Middle East, Gulf of Mexico and Colombia. Network input data include gas to oil ratio, API °, relative gas density and reservoir temperature. This model is more accurate than other experimental models (Goda *et al.,* 2003).

Goda et al. (2003) presented a model that predicts bubble point pressure and oil volume factor through two interconnected neural networks. The first network has four layers which reservoir temperature, API °, relative gas density and gas to oil ratio are the input to first layer. Their network has two hidden layers, each has 10 neurons, that are activated by the Log Sigmoid transfer function. The output layer has a neuron whose transfer function is of pure linear type. This network uses 160 data from the Middle East region for training and another 20 data for network testing.

The second network also has four layers. The purpose of the network is to estimate the oil volume factor. The input layer has five neurons for the five input data types including reservoir temperature, API °, relative gas density, gas-to-oil ratio and finally the bubble point pressure predicted from the first grid. Each of its two hidden layers has 8 neurons with the Log Sigmoid transfer function. The output layer has a neuron with the oil volume factor output data. The main difference between this network and previous PVT predicting networks is that other networks use four common input data to predict the oil volume factor, but in addition to the four data, this model uses the bubble point pressure estimated by another neural network (Moghadassi *et al.,* 2008).

Moghadasi et al. (2008) presented a new model for predicting reservoir PVT properties. The data set used to design this model is extracted from Perry's Chemical Engineers' Handbook. Different training schemes such as SCG, LM and RP have been used/applied for the back-propagation algorithm. The LM algorithm with 60 neurons in the hidden layer with the lowest mean square error (MSE) of 0.000606 was selected as the best network (Laugier & Richon, 2003).

Neural network structure

Preparing and collecting the data required for neural network were one of the most difficult parts of this research. After collecting more than 1000 experimental PVT data from over 55 samples of under-saturated ($P_{res} < P_{bubble}$) reservoirs in Iran, 577 data on oil formation volume factor and density were selected. These data were

then randomly divided into two sections, out of which 486 were used for network training and the rest (91) for network testing.

Reviewing numerous/wide volume of books and articles; pressure (P), temperature (T), oil API degree, relative coefficient of dissolved gas (R_s) and specific gas density (γ_g) were selected as network input parameters.

The first step to design a neural network is to determine the number of layers, neurons of each layer and the transfer function of each layer. It is always clear that the number of neurons in the first layer is the same as the parameters input to the neural network and it is a function of the problem type. Therefore, the first layer in this case study consists of 5 neurons. The first layer neurons, i.e. the input information, transfer their values directly to the second layer.

Table 1. Overview of total data used

Number of Points	PVT Property	Minimum	Maximum	Mean
577	Pressure, psi	535	9300	3764.12
577	Temperature, °F	116	277	215.145
577	Solution GOR, SCF/STB	66.8	1998.39	922.305
577	S.G. of Total Gas Evolved	0.77	1.61	1.056
577	Stock Tank Oil Gravity, °API	7.24	42.89	26.99
577	FVF, bbl/STB	1.039	2.408	1.593
577	Oil Density, lb/ft^3	33.14	62.085	44.368

To reduce the range of input variables in order to make more stable in proposed neural network, all the input variables as well as the target variables according to eq. 1 have become dimensionless and then are used in the network.

$$X_{new} = \frac{X_{old} - \min(X_{old})}{\max(X_{old}) - \min(X_{old})} \tag{1}$$

The number of neurons in the output layer is also equal to the number of target parameters.

To determine the network structure, all possible modes have been investigated and for each case, the Average Absolute Percent Relative Error and relative square root

mean square error (RMS) have been calculated for train and test data. The following equations illustrate the relationships used:

$$E_i = \left| \frac{X_{exp} - X_{est}}{X_{exp}} \right| \tag{2}$$

$$E_a = \frac{1}{n} \sum_1^n E_i \tag{3}$$

$$RMS = [\frac{1}{n} \sum_1^n E_i^2]^{1/2} \tag{4}$$

$$r = \sqrt{1 - \sum_1^n (X_{exp} - X_{est})^2 / \sum_1^n \left| X_{exp} - \overline{X} \right|} \tag{5}$$

$$\overline{X} = \frac{1}{n} \sum_1^n X_{exp} \tag{6}$$

RMS or the relative root mean square error represents the distribution of data around the deviation from zero error, and r or the Correlation Coefficient is a measure of the success rate of the proposed relationship in reducing the standard deviation. After investigating the results, two modes of one hidden layer with 6 neurons and two hidden layers (each of 5 neurons) were selected and different modes were investigated on these two structures to determine the optimal state. For instance, several transfer functions were used and the weights obtained in the previous step were used/applied/considered as the initial weights in the next step.

Finally, the three-layer structure (one hidden layer with 6 neurons) was chosen as the final structure with Tanh transfer function for the hidden layer and linear transfer function for the output layer. Batch back propagation was Also identified as the best network training method.
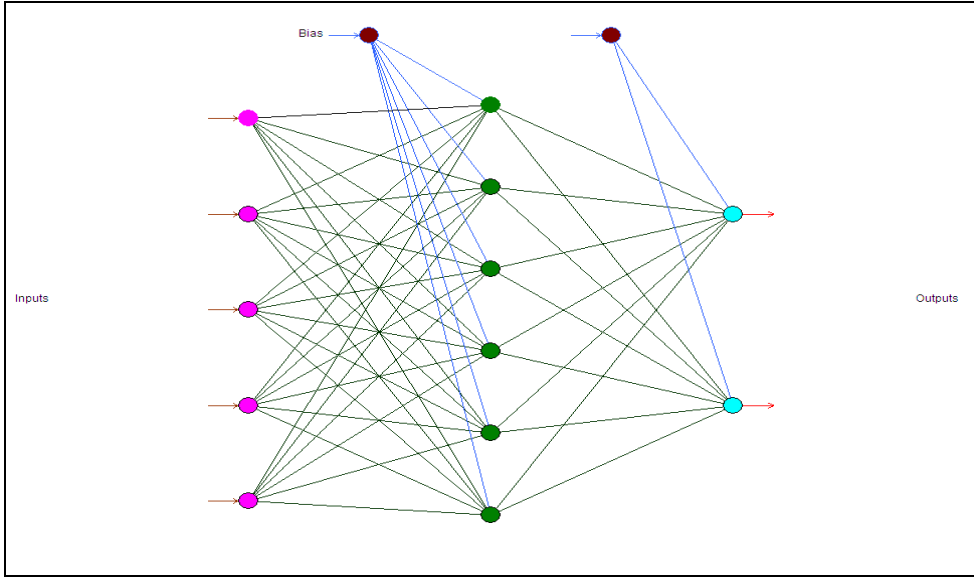
Figure 1. Schematic of proposed network

Table 2 Proposed Network's Weights and Bias Values

Connections	Weights	Connections	Weights	Connections	Weights
N1L1-N1L2	-1.1444	N3L1-N6L2	1.1901	B1-N5L2	0.6623
N1L1-N2L2	-0.0814	N4L1-N1L2	0.7074	B1-N6L2	0.3032
N1L1-N3L2	0.2933	N4L1-N2L2	0.4247	N1L2-N1L3	0.7084
N1L1-N4L2	0.2293	N4L1-N3L2	-0.6211	N1L2-N2L3	-0.3885
N1L1-N5L2	0.0806	N4L1-N4L2	0.7755	N2L2-N1L3	-0.5457
N1L1-N6L2	-0.3875	N4L1-N5L2	0.2433	N2L2-N2L3	-0.1033
N2L1-N1L2	0.7034	N4L1-N6L2	0.3784	N3L2-N1L3	0.3992
N2L1-N2L2	0.3061	N5L1-N1L2	-0.0271	N3L2-N2L3	0.152
N2L1-N3L2	0.176	N5L1-N2L2	0.6449	N4L2-N1L3	0.1444
N2L1-N4L2	0.7632	N5L1-N3L2	0.4846	N4L2-N2L3	-0.7827
N2L1-N5L2	0.2602	N5L1-N4L2	0.1442	N5L2-N1L3	0.3189
N2L1-N6L2	0.1391	N5L1-N5L2	0.8289	N5L2-N2L3	-0.0003
N3L1-N1L2	1.238	N5L1-N6L2	-0.5597	N6L2-N1L3	0.247
N3L1-N2L2	-0.9579	B1-N1L2	1.6673	N6L2-N2L3	-0.4815
N3L1-N3L2	0.5877	B1-N2L2	0.6059	B2-N1L3	0.2195
N3L1-N4L2	0.9374	B1-N3L2	0.1885	B2-N2L3	-0.1374
N3L1-N5L2	0.3539	B1-N4L2	0.3618		

Result

After selecting the optimal structure for the neural network and its training, the second part of the data (91 data) was used to test estimation accuracy of the network.

Table 3 presents the error values of the network results. Figures 2, 3, 4, and 5 also show the scatter plots comparing experimental results with those estimated by the network:

Table 3. Network Results Error Values

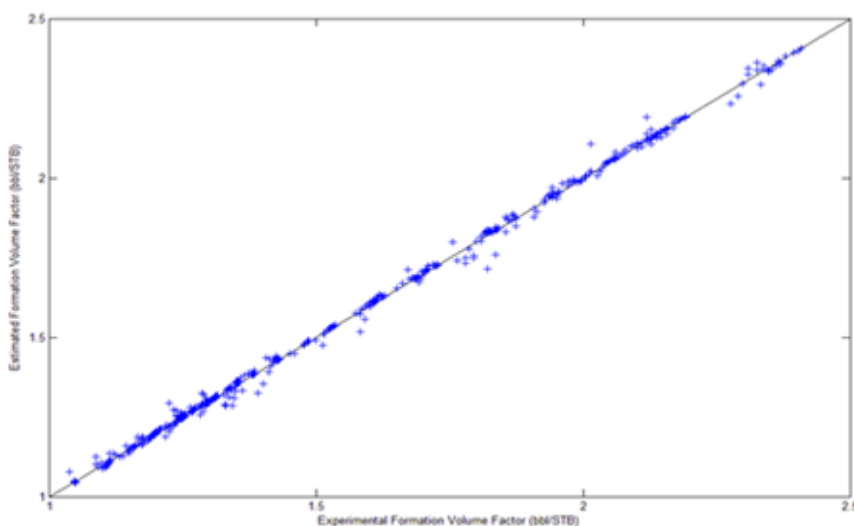|	E_{min} (%)	E_{max} (%)	E_a (%)	RMS	Correlation Coefficient
$B_o(Train)$	9.36841E-05	5.84041065	0.556592298	1.020049926	0.999590131
$\rho_o(Train)$	0.000652503	7.098559348	0.508632407	1.047043715	0.982429677
$B_o(Test)$	0.001333529	5.138218685	1.032430012	1.480137336	0.99909827
$\rho_o(Test)$	0.004537934	5.971239206	1.104391183	1.538169325	0.956828755

Figure 2. Comparison between experimental values of and values estimated by the network for data used in training
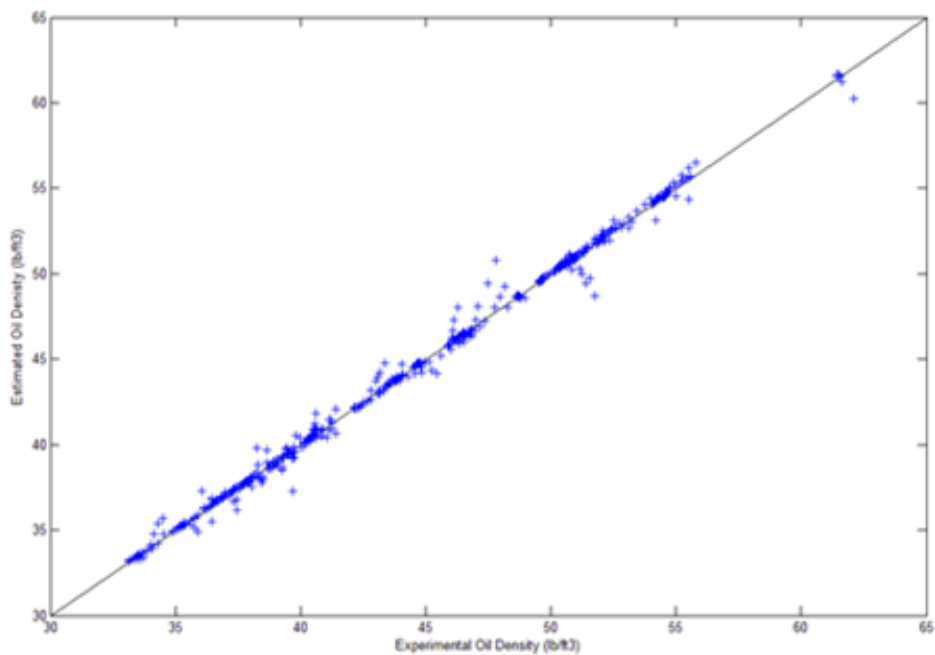
Figure 3. Comparison between the experimental values and the values estimated by the network for the data used in training
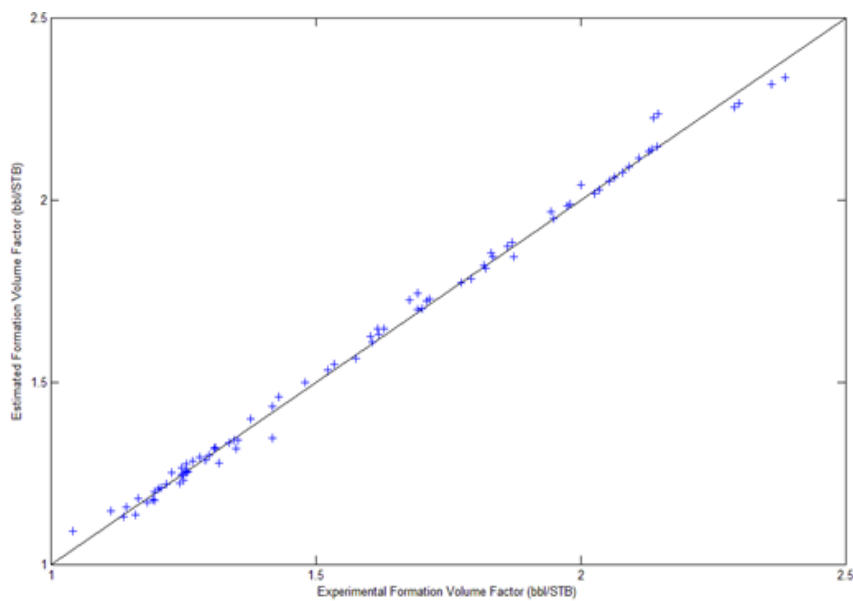

Figure 4. Comparison between experimental values and values estimated by the network for the data used in testing
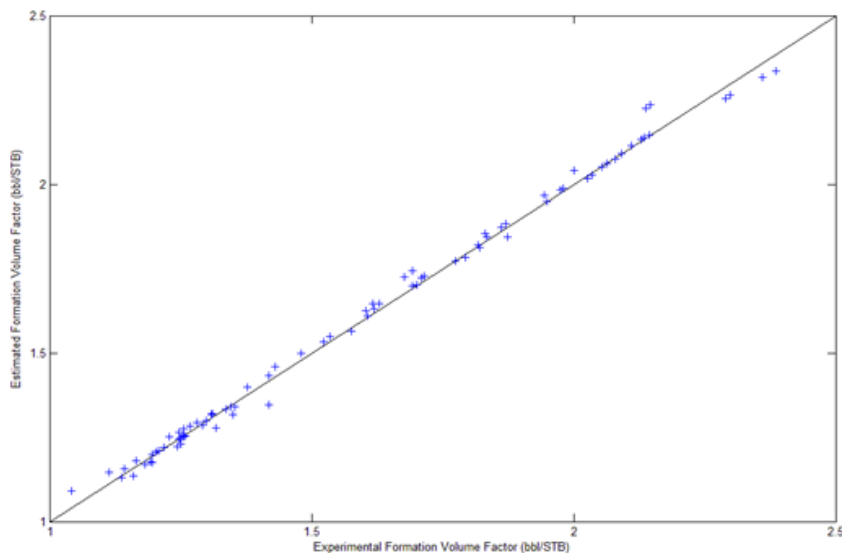
Figure 5. Comparison between experimental values and values estimated by the network for the data used in testing

Comparison with empirical relationships

To evaluate the accuracy of the presented network compared with the empirical relationships published, several relationships are selected and the results are compared. To use the empirical equations and due to this fact that the samples are under-saturated, i.e. their pressure is above the bubble pressure, the volume factor and density are calculated in two steps. First, by/through using the empirical relation, their value at the bubble point is determined, then by defining the Coefficient of Isothermal Compressibility, that value is calculated at the desired pressure: The first step goes through applying the empirical relation by which their value at the bubble point is determined and following that by defining the Coefficient of Isothermal Compressibility, the value is calculated at the desired pressure.

$$B_0 = B_{ob}EXP[c_o(p_b - p)] \tag{6}$$

$$B_0 = B_{ob}EXP[c_o(p - p_b)] \tag{7}$$

As can be seen, the above relationships need to calculate the value of that property at the bubble point, determine the bubble pressure and determine the coefficient of isothermal compressibility. They are, therefore, used to calculate a property in the supersaturated sample.

Many researchers have proposed empirical relationships to determine the PVT properties of which the most reliable relations have been attempted to be used while comparing   the proposed network. (In cases where all 3 relationships were not available from one researcher, the relationship by another researcher was assumed to be relevant).

The relationships used are as follows (Laugier, S. & Richon, 2003; Danesh, 1998; William, 1990):

Ahmed's Correlations:

$$B_{ob} = F + a_1 T + a_2 T^2 + a_3/T + a_4 P + a_5 P^2 + a_6/P + a_7 R_s + a_8 R_s^2 + a_9/R_s$$

$$F = a_{10} + (R_s^{a11} \times API^{a12}/\gamma_g^{a13})$$

$$a_1 = -4.5243973 \times 10^{-4} \qquad\qquad a_2 = 3.9063637 \times 10^{-6}$$

$$a_3 = -5.5542509 \qquad\qquad a_4 = -5.7603220 \times 10^{-6}$$

$$a_5 = -3.9528992 \times 10^{-9} \qquad\qquad a_6 = 16.289473$$

$$a_7 = 3.8718887 \times 10^{-4} \qquad\qquad a_8 = 7.0703685 \times 10^{-8}$$

$$a_9 = -1.4358395 \qquad\qquad a_{10} = -0.12869353$$

$$a_{11} = 0.023484894 \qquad\qquad a_{12} = 0.015966573$$

$$a_{13} = 0.021946351$$

$$B_o = B_{ob} \times EXP[D[EXP(aP) - EXP(aP_b)]]$$
$$D = [4.588893 + 0.0025999 \times R_s]^{-1}$$
$$a = -0.00018473$$
$$\rho_o = \rho_{ob} EXP[B(EXP(aP) - EXP(aP_b))]$$

$$B = -(4.588893 + 0.0025999 \times R_s)^{-1}$$

The Al-Marhoun equation is used to calculate the bubble pressure.

$$P_b = a \times R_s^b \times \gamma_g^c \times \gamma_o^d \times T^e$$

$$a = 5.38088 \times 10\text{-}3 \qquad\qquad b = 0.715082$$

$$c = \text{-}1.87784 \qquad\qquad d = 3.1437$$
$$e = 1.32657$$

$$\gamma_O = \frac{141.5}{°API + 131.5}$$

$$\rho_{ob} = \frac{62.4 \times \gamma_g + 0.0136 \times R_s \times \gamma_g}{0.972 + 0.000147 \left[R_s \left(\dfrac{\gamma_g}{\gamma_o} \right)^{0.5} + 1.25 \times T \right]^{1.175}}$$

The Standing relation is used to calculate the density at the bubble point.

Standing's Correlations:

$$B_{ob} = 0.9759 + 0.000120 \left[R_s \left(\frac{\gamma_g}{\gamma_o} \right)^{0.5} + 1.25 \times T \right]^{1.2}$$

$$P_b = 18.2 \left[\left(R_s / \gamma_g \right)^{0.83} 10^a - 1.4 \right]$$

The Vesquez-Beggs relation is used to calculate Co.
$$C_o = \frac{-1433 + 5 \times R_s + 17.2 \times T - 1180 \times \gamma_g + 12.61 \times °API}{10^5 \times P}$$

Glaso's Correlations:

$$B_{ob} = 1.0 + 10^A$$

$$A = -6.58511 + 2.91329 \times Log(B_{ob}^*) - 0.27683 \left(Log(B_{ob}^*) \right)^2$$

Here the Standing equation was used to calculate the bubble pressure and the Vesquez-Beggs relation to calculate Co.

After computing the results of the empirical relationships, they were compared with the results from the neural network which are shown in the following tables: (In each case, the relative error of absolute mean, minimum error, maximum error, and root relative squared error were compared).

Table 4. Comparison of Neural Network Error with Other Experimental Equations to Esimate Oil Volume factor (Training Data)

	RMS	E_a (%)	E_{min}	E_{max} (%)
ANN	1.020	0.5565	0.000094	5.8404
Ahmed	8.649	7.535	0.019	18.258
Standing	4.266	3.138	0.007	16.089
Glaso	3.034	2.291	0.003	14.298

Table 5. Comparison of neural network error with Ahmed empirical equation to estimate oil density (training data)

	E_{max} (%)	E_{min} (%)	E_a (%)	RMS
ANN	7.0985	0.00065	0.508	1.047
Ahmed	14.463	0.021	4.132	4.946

Table 6. Comparison of Neural Network Error with Other Experimental Equations to estimate Oil Volume factor (Test Data)

	E_{max} (%)	E_{min} (%)	E_a (%)	RMS
ANN	5.1382	0.0013	1.0324	1.480
Ahmed	17.792	0.216	7.337	8.508
Standing	16.089	0.110	3.035	4.173
Glaso	14.298	0.027	2.185	3.080

Table 4. Comparison of neural network error with Ahmed empirical equation to
estimate oil density (test data)

	E_{max} (%)	E_{min} (%)	E_a (%)	RMS
ANN	5.971	0.00453	1.1043	1.538
Ahmed	13.780	0.354	4.070	4.895

The results obtained indicate the high accuracy of the proposed network with respect to these experimental relationships.

Conclusion

1. A new model was presented to determine the oil volume factor and density for Iran's under-saturated reservoirs. The proposed model is based on artificial neural networks and has been developed using 577 empirical data from over 55 Iranian under-saturated samples.

2. From 577 data, 486 data were used for network training and 91 data for network testing.

3. The proposed network has an input layer with 5 neurons representing the network inputs and one hidden layer with 6 neurons and an output layer with 2 neurons representing the network outputs. Batch back propagation algorithm was also used for network training.

4. The results showed that the network is capable of estimating oil volume factor and oil density with high accuracy and provides better results than existing empirical relationships. The average network error in determining these properties was about 0.5% for training data and about 1% for test data.

References

Al-Marhoun, M.A., & Osman, E.A. (2002). Using Artificial Neural Networks to develop New PVT Correlations for Saudi Crude Oils. SPE 78592.

Danesh, A. (1998). PVT and Phase Behaviour of Petroleum Reservoir Fluids. Elsevier Science B.V., Netherlands.

Gharbi, B.C., & Elsharkawy, M. (1999). Neural Network Model for Estimating the PVT Properties of Middle East Crude Oils, SPE Reservoir Eval. & Eng. 2(3).

Goda, H.M., Shokri, E.M., Fattah, K.A., & Sayyouh, M.H. (2003). Prediction of the PVT Data using Neural Network Computing Theory; SPE 85650.

Laugier, S., & Richon, D. (2003). Use of artificial neural networks for calculating derived thermodynamic quantities from volumetric property data, Fluid Phase Equilib. 210, 247-255.

Moghadassi, A.R., Parvizian, F., Hosseini, S.M., & Fazlali, A.R. (2008). A New Approach for Estimation of PVT Properties of Pure Gases Based on Artificial Neural Network Model.

William, D.M. (1990). The Properties of Petroleum Fluids. PennWell Publishing Company, USA.