

PVT correlations for Pakistani crude oils using artificial neural network

Muzammil Hussain Rammay¹ · Abdulazeez Abdulraheem¹

Received: 18 May 2015 / Accepted: 17 January 2016 / Published online: 10 February 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Reservoir fluid properties such as bubble point pressure, oil formation volume factor and viscosity are very important in reservoir and petroleum production engineering computations such as outflow–inflow well performance, material balance calculations, well test analysis, reserve estimates, and numerical reservoir simulations. Ideally, these properties should be obtained from actual measurements. Quite often, however, these measurements are either not available or very costly to obtain. In such cases, empirically derived correlations are used to predict the needed properties using the known properties such as temperature, specific gravity of oil and gas, and gas–oil ratio. Therefore, all computations depend on the accuracy of the correlations used for predicting the fluid properties. Almost all of these previous correlations were developed with linear or nonlinear multiple regression or graphical techniques. Artificial neural networks, once successfully trained, offer an alternative way to obtain reliable and more accurate results for the determination of crude oil PVT properties, because it can capture highly nonlinear behavior and relationship between the input and output data as compared to linear and nonlinear regression techniques. In this study, we present neural network-based models for the prediction of PVT properties of crude oils from Pakistan. The data on which the networks were trained and tested contain 166 data sets from 22 different crude oil samples and used in developing PVT models for Pakistan crude

oils. The developed neural network models are able to predict the bubble point pressure, oil formation volume factor and viscosity as a function of the solution gas–oil ratio, gas specific gravity, oil specific gravity, and temperature. A detailed comparison between the results predicted by the neural network models and those predicted by other previously published correlations shows that the developed neural network models outperform most other existing correlations by giving significantly lower values of average absolute relative error for the bubble point, oil formation volume factor at bubble point, and gas-saturated oil viscosity.

Keywords Artificial neural network · Oil formation volume factor · Viscosity of oil · Bubble point pressure · Solution gas oil ratio

List of symbols

API	API gravity of oil
P_b	Bubble point pressure (psi)
B_{ob}	Oil formation volume factor at the bubble point pressure, RB/STB
μ_{ob}	Viscosity of oil at bubble point pressure, cp
Rs	Solution gas–oil ratio, SCF/STB
T	Reservoir temperature (°F)
γ_o	Specific gravity of oil (water = 1.0)
γ_g	Specific gravity of gas (air = 1.0)
APE	Average percent relative error
Er	Percent relative error
AAPE	Average absolute percent relative error
E_{max}	Maximum absolute percent relative error = $\max Er $
E_{min}	Minimum absolute percent relative error = $\min Er $

✉ Muzammil Hussain Rammay
hussain_petro@yahoo.com

Abdulazeez Abdulraheem
toazeez@gmail.com

¹ King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia

R	Correlation coefficient
St. dev	Standard deviation
x_i	Input parameters
i	Index for input parameters
N_i	Total number of input parameters
j	Index for hidden layer neurons
b_j	Bias for hidden layer neuron j
w_{ji}	Weights between input and hidden layer
f	Tan-sigmoid transfer function
k	Index for number of output parameters
b_k	Bias for output layer
w_{jk}	Weights between hidden and output layer
N_h	Total number of hidden layer neurons
f_L	Linear transfer function
n	Total number of data points

Introduction

During the last 70 years, engineers developed a significant number of PVT correlations due to the high importance empirical correlations for PVT properties in oil and gas engineering. These correlations are sensitive to the region and number of data points. PVT correlations accuracy varies from one region to another. Therefore, significant numbers of correlations have been developed based on the regional variation, and it is recommended in the previous studies that one should use their own region PVT correlation. A brief overview of the widely used PVT correlation is give below.

Standing (1947, 1977) presented correlations for bubble point pressure and oil formation volume factor. Standing's correlations were based on laboratory experiments carried out on 105 samples from 22 different crude oils in California. Katz (1942) presented five methods for predicting the reservoir oil shrinkage. Vazquez and Beggs (1980) presented correlations for oil formation volume factor. They divided oil mixtures into two groups, above and below 30 ° API gravity. More than 6000 data points from 600 laboratory measurements were used in developing the correlations. Glaso (1980) developed the correlation for formation volume factor using 45 oil samples from North Sea hydrocarbon mixtures. Al-Marhoun (1988) published correlations for estimating bubble point pressure and oil formation volume factor for the Middle East oils. He used 160 data sets from 69 Middle Eastern reservoirs to develop the correlation. Abdul-Majeed and Salman (1988) published an oil formation volume factor correlation based on 420 data sets. Their model is similar to that of Al-Marhoun (1988) oil formation volume factor correlation with new calculated coefficients.

Labedi (1990) presented correlations for oil formation volume factor for African crude oils. He used 97 data sets from Libya, 28 from Nigeria, and 4 from Angola to

develop his correlations. Dokla and Osman (1992) published a set of correlations for estimating bubble point pressure and oil formation volume factor for UAE crudes. They used 51 data sets to calculate new coefficients for Al-Marhoun (1988) Middle East models. Al-Yousef and Al-Marhoun (1993) pointed out that the Dokla and Osman (1992, 1993) bubble point pressure correlation was found to contradict the physical laws. Al-Marhoun (1992) published a second correlation for oil formation volume factor. The correlation was developed with 11,728 experimentally obtained formation volume factors at, above, and below the bubble point pressure. The data set represents samples from more than 700 reservoirs from all over the world, mostly from Middle East and North America.

Macary and El-Batanoney (1992) presented correlations for bubble point pressure and oil formation volume factor. They used 90 data sets from 30 independent reservoirs in the Gulf of Suez to develop the correlations. The new correlations were tested against other Egyptian data of Saleh et al. (1987), and showed improvement over published correlations. Omar and Todd (1993) presented oil formation volume factor correlation, based on Standing's (1947) model. Their correlation was based on 93 data sets from Malaysian oil reservoirs. Petrosky and Farshad (1993) developed new correlations for the Gulf of Mexico.

Kartoatmodjo and Schmidt (1994) used a global data bank to develop new correlations for all PVT properties. Data from 740 different crude oil samples gathered from all over the world provided 5392 data sets for correlation development. Almehaideb (1997) published a new set of correlations for UAE crudes using 62 data sets from UAE reservoirs. These correlations were developed for bubble point pressure and oil formation volume factor. The bubble point pressure correlation of Omar and Todd (1993) uses the oil formation volume factor as input in addition to oil gravity, gas gravity, solution gas oil ratio, and reservoir temperature. Saleh et al. (1987) evaluated the empirical correlations for Egyptian oils. They reported that Standing's (1947) correlation was the best for oil formation volume factor. Sutton and Farshad (1990a, b) published an evaluation for Gulf of Mexico crude oils. They used 285 data sets for gas-saturated oil and 134 data sets for undersaturated oil representing 31 different crude oils and natural gas systems. The results show that Glaso (1980) correlation for oil formation volume factor perform the best for most of the data of the study. Later, Petrosky and Farshad (1993) published a new correlation based on the Gulf of Mexico crudes. They reported that the best performing published correlation for oil formation volume is the Al-Marhoun (1988) correlation. McCain (1991) published an evaluation of all reservoir properties correlations based on a large global database. He recommended Standing's (1947) correlations for formation volume factor at and below the bubble point pressure.

Al-Fattah and Al-Marhoun (1994) published an evaluation of all available oil formation volume factor correlations. They used 674 data sets from published literature. They found that Al-Marhoun (1992) correlation has the least error for global data set. Also, they performed trend tests to evaluate the model’s physical behavior. Finally, Al-Shammasi (1997) evaluated the published correlations and neural network models for bubble point pressure and oil formation volume factor for accuracy and flexibility to represent hydrocarbon mixtures from different geographical locations worldwide. He presented a new correlation for bubble point pressure based on global data of 1661 published and 48 unpublished data sets. Also, he presented neural network models and compared their performance to numerical correlations. He concluded that statistical and trend performance analysis showed that some of the correlations violate the physical behavior of hydrocarbon fluid properties.

De Ghetto et al. (1994) performed a comprehensive study on PVT properties correlation based on 195 global data sets collected from the Mediterranean Basin, Africa, Middle East, and the North Sea reservoirs. They recommended Vazquez and Beggs (1980) correlation for the oil formation volume factor. Elsharkawy et al. (1994) evaluated the PVT correlations for Kuwaiti crude oils using 44 samples. Standing’s (1947) correlation gave the best results for bubble point pressure, while Al-Marhoun (1988) oil formation volume factor correlation performed satisfactorily.

Hanafy et al. (1997) published a study to evaluate the most accurate correlation to apply to Egyptian crude oils. For formation volume factor, Macary and El-Batoney (1992) correlation showed an average absolute error of 4.9 %, while Dokla and Osman (1992) showed 3.9 %. The study strongly supports the approach of developing a local correlation versus a global correlation.

Mahmood and Al-Marhoun (1996) presented an evaluation of PVT correlations for Pakistani crude oils. They used 166 data sets from 22 different crude samples for the evaluation. Al-Marhoun (1992) oil formation volume factor correlation gave the best results. The bubble point pressure errors reported in this study, for all correlations, are among the highest reported in the literature. It is also possible to increase the accuracy by updating the coefficients of Al-Marhoun correlation, but it is not necessary that the update correlation captures nonlinear behavior completely. Therefore, Mahmood and Al-Marhoun (1996) also recommended new PVT correlations for Pakistani crude oils and this recommendation is the basis for this research.

Artificial neural network

Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Typically, neural networks are adjusted, or trained, so that a particular input leads to a specific target output (MathWorks, Inc 2008).

Neural network consists of input and output neurons (elements) and layers, which are connected by further neurons and layers known as hidden layer and hidden layer neurons (Fauset 1996). These neurons are connected in a highly parallel and distributed way, so that they can map any nonlinear complex function as shown in Fig. 1. Each connection in the neural network assigns weights and layers and are connected by transfer functions. The response from each neuron is given by

$$y_j = f \left(\sum_{i=1}^{N_i} w_{ji} \times x_i + b_j \right),$$

where x_i are the input parameters, i is the index for input parameters, N_i is the total number of input parameters, j is the index for hidden layer neurons, b_j is the bias for hidden layer neuron j , y_j is the output of hidden layer neuron j , w_{ji} are weights between the input and hidden layer, f is the tan-sigmoid transfer function and

$$z_k = f_L \left(\sum_{j=1}^{N_h} w_{jk} \times y_j + b_k \right).$$

In this, k is the index for a number of output parameters, b_k is the bias for the output layer, w_{jk} are the weights between the hidden and output layer, N_h is the total number of hidden layer neurons, z_k are the outputs of the output layer, and f_L is the linear transfer function.

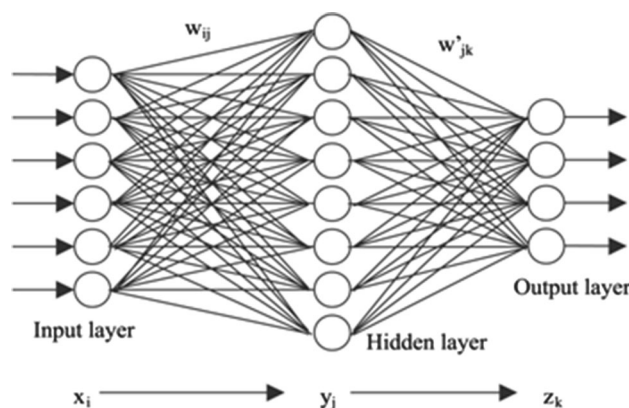


Fig. 1 Generalized ANN model architecture with input, hidden, and output layer

The work flow for the neural network design process has the following primary steps:

- Collect data.
- Create the network.
- Configure the network.
- Initialize the weights and biases.
- Train and validate the network.
- Test and use the network.

In recent years, neural networks have gained popularity in petroleum applications. Many authors have discussed the applications of neural network in petroleum engineering (Kumoluyi and Daltaban 1994; Ali 1994; Mohaghegh and Ameri 1994; Mohaghegh 1995). Few studies have been carried out to model the PVT properties using neural networks. Gharbi and Elsharkawy (1997) published neural network models for estimating bubble point pressure and oil formation volume factor for Middle East crude oils. Both neural network models were trained using 498 data sets collected from the literature and unpublished sources. The models were tested by other 22 data points from the Middle East. The results showed improvement over the conventional correlation methods with reduction in the average error for the bubble point pressure oil formation volume factor.

Varotsis et al. (1999) presented a novel approach for predicting the complete PVT behavior of reservoir oils and gas condensates using artificial neural network (ANN). The method uses key measurements that can be performed rapidly either in the laboratory or at the well site as input to an ANN. The ANN was trained by a PVT study database of over 650 reservoir fluids originating from all parts of the world. Tests of the trained ANN architecture utilizing a validation set of PVT studies indicate that, for all fluid types, most PVT property estimates can be obtained with a very low mean relative error of 0.5–2.5 %, with no data set having a relative error in excess of 5 %. Osman and Al-Marhoun (2002) developed PVT correlations using ANN for Saudi crude oils. The models were developed using 283 data sets, which were collected from Saudi Reservoirs. Gupta (2010) developed PVT correlations using artificial neural network for Indian crude oils. The models were developed using 372 data sets, which were collected from Indian reservoirs. All of the regional ANN PVT correlations outperform the previously published correlations. Mahmood and Al-Marhoun (1996) presented an evaluation of PVT correlations for Pakistani crude oils. In this study (Mahmood and Al-Marhoun 1996), they concluded that new PVT correlations for Pakistani crude oils are required and therefore this recommendation and conclusion are the basis for this research. Moreover due to the previous success of ANN in regional PVT correlations, ANN is used as an algorithm for PVT correlations of Pakistani crude oil.

Data acquisition and analysis

Data used for this work were collected from Mahmood and Al-Marhoun (1996) publication related to evaluation of PVT correlations for Pakistani crude oil. In general, this data set covers a wide range of bubble point pressure, oil FVF, solution gas/oil ratio, and gas relative density values; whereas the temperature and oil gravity belong to relatively higher values attributed to regional trends prevailing in Pakistani crude oils. Each data set contains reservoir temperature, oil gravity, total solution gas oil ratio, and average gas gravity, bubble point pressure, oil formation volume factor at the bubble point pressure, and viscosity at the bubble point pressure. The data set was randomly divided into two groups of seen data (70 % of total data) and unseen data (30 % of total data). Out of a total of 166 data points, 70 % (seen data by ANN) were used for training, validation, or cross-validation of the ANN models, the remaining 30 % (unseen data by ANN) were used to test the model to evaluate its accuracy. For viscosity at P_b data, 128 data points are available in Mahmood and Al-Marhoun's (1996) publication and these are divided in the same way as bubble point pressure and formation volume factor at P_b . A statistical description of training (seen) and test (unseen) data are given in Tables 1 and 2, respectively.

Bubble point pressure artificial neural network model

The bubble point pressure ANN model consists of four input neurons or input parameters, which are related to temperature, specific gravity of gas, API gravity of oil and solution gas–oil ratio, one hidden layer and one output neuron related to bubble point pressure. The hidden layer consists of 12 neurons, which were obtained after sensitivity runs of the number of neurons from 5 to 50. Tan-sigmoid and linear transfer functions were used in the hidden and output layer, respectively. The algorithm of ANN model for bubble point pressure is given below:

$$P_b = f_L \left(\sum_{j=1}^{N_h} w_{jk} \times \left(f \left(\sum_{i=1}^{N_i} w_{ji} \times x_i + b_j \right) \right) + b_k \right).$$

The above ANN algorithm for bubble point pressure can also be written in the following way:

$$(P_b)_N = f_L \left(\sum_{j=1}^{N_h} w_{jk} \times \left(f \left(w_{j1} \times (T)_N + w_{j2} \times (\gamma_g)_N + w_{j3} \times (\text{API})_N + w_{j4} \times (\text{Rs})_N + b_j \right) \right) + b_k \right).$$

Table 1 Statistical description of the input and output data used for training and cross-validation

Property	Max	Min	Mean	Range	St. dev	Skewness	Kurtosis
Temperature (°F)	296	182	243.87	114	25.96	−0.19	3.68
Specific gravity of gas	3.444	0.825	1.79	2.6192	0.43	0.78	3.73
API gravity of oil	43.8	29	38.64	14.8	3.64	−1.35	4.48
Solution gas–oil ratio (SCF/STB)	2496	92	513.43	2404	433.21	2.40	9.53
Bubble point pressure (psi)	4975	79	988.15	4896	909.96	1.94	7.39
B_{ob} (bbl/STB)	2.92	1.21	1.52	1.70	0.32	2.15	7.99
μ_{ob} (cp)	0.64	0.21	0.35	0.43	0.08	0.82	3.85

Table 2 Statistical description of the input and output data used for testing

Property	Max	Min	Mean	Range	St. dev	Skewness	Kurtosis
Temperature (°F)	296	182	238.12	114	26.55	−0.43	3.69
Specific gravity of gas	2.98	1.061	1.69	1.9192	0.40	1.04	4.07
API gravity of oil	56.5	29	40.05	27.5	3.43	1.66	13.66
Solution gas–oil ratio (SCF/STB)	2249	145	467.84	2104	335.20	3.39	17.68
Bubble point pressure (psi)	2885	104	926.24	2781	574.32	0.74	3.92
B_{ob} (bbl/STB)	1.91	1.20	1.37	0.71	0.14	1.69	6.80
μ_{ob} (cp)	0.58	0.22	0.36	0.36	0.07	0.62	3.94

Subscript N shows the normalized input and output parameters for ANN model. The bias values are constant and have a similar concept to constants in linear or nonlinear regression. The architecture of the ANN model for bubble point pressure is shown in Fig. 2.

To obtain weights and bias values, training of neural network was performed by Levenberg–Marquardt back-propagation algorithm. To avoid local minimum, 5000 multiple realizations with different weight and bias initialization of training were implemented and the minimum error realization was selected as the best one. After training, the optimal weights and bias values were obtained for bubble point pressure ANN and these are shown in Table 3.

All input parameters should be normalized in the range of $[-1, 1]$ before using in the ANN algorithm for bubble point. The general equation for input normalization is given below:

$$\text{Input}_{\text{norm}} = \frac{(\text{Input}_{\text{max}} - \text{Input}_{\text{min}})(x - x_{\text{min}})}{(x_{\text{max}} - x_{\text{min}})} + \text{Input}_{\text{min}},$$

$$\text{Input}_{\text{min}} = -1,$$

$$\text{Input}_{\text{max}} = 1.$$

The x_{max} and x_{min} values (ranges of input parameters) are given in Table 1. Therefore, the input parameters should be normalized by using the following equations:

$$(T)_N = \frac{2(T - 182)}{(296 - 182)} - 1,$$

$$(\gamma_g)_N = \frac{2(\gamma_g - 0.825)}{(3.444 - 0.825)} - 1,$$

$$(\text{API})_N = \frac{2(\text{API} - 29)}{(43.8 - 29)} - 1,$$

$$(\text{Rs})_N = \frac{2(\text{Rs} - 92)}{(2496 - 92)} - 1.$$

The proposed ANN model gives normalized bubble point pressure in the range $[-1, 1]$; therefore for real value of the bubble point pressure, the output value should be de-normalized by the following equation:

$$\text{Output} = \frac{(y_{\text{max}} - y_{\text{min}})(\text{Output}_{\text{norm}} - (-1))}{(1) - (-1)} + y_{\text{min}},$$

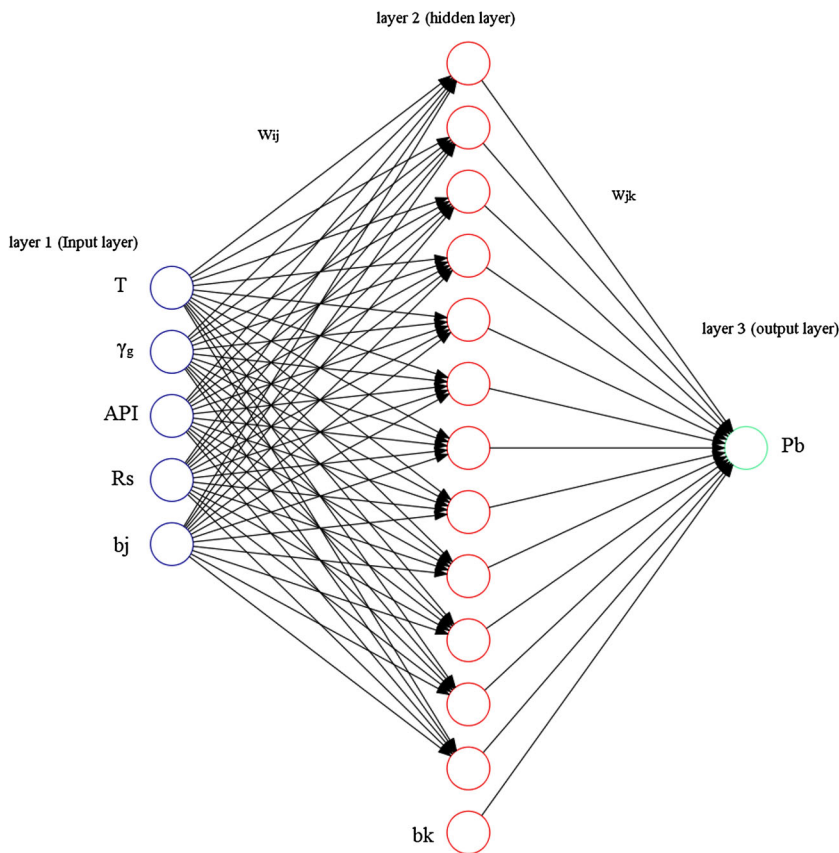
where y_{max} and y_{min} values (minimum and maximum bubble point pressures) are given in Table 1. Therefore, the output parameter should be de-normalized by using following equation:

$$P_b = \frac{(4975 - 79)((P_b)_N - (-1))}{2} + 79.$$

Oil formation volume factor of the P_b (B_{ob}) ANN model

The oil formation volume factor of the P_b ANN model consists of four input neurons or input parameters, which are related to temperature, specific gravity of gas, API

Fig. 2 Architecture of ANN model for bubble point pressure



gravity of oil and solution gas–oil ratio, one hidden layer and one output neuron related to B_{ob} . The hidden layer consists of 8 neurons, which were obtained after sensitivity runs of number of neurons from 5 to 50. Tan-sigmoid and linear transfer functions were used in the hidden and output layer, respectively. The algorithm of the ANN model for B_{ob} is given below:

$$B_{ob} = f_L \left(\sum_{j=1}^{N_h} w_{jk} \times \left(f \left(\sum_{i=1}^{N_i} w_{ji} \times x_i + b_j \right) \right) + b_k \right).$$

The above ANN algorithm for B_{ob} can also be written in the following way:

$$(B_{ob})_N = f_L \left(\sum_{j=1}^{N_h} w_{jk} \times \left(f(w_{j1} \times (T)_N + w_{j2} \times (\gamma_g)_N + w_{j3} \times (API)_N + w_{j4} \times (Rs)_N + b_j) \right) + b_k \right).$$

Subscript N shows the normalized input and output parameters for ANN model. The bias values are constant values, which have a similar concept to constants in linear or nonlinear regression. The architecture of ANN model for B_{ob} is shown in Fig. 3.

To obtain weights and bias values, training of neural network was performed by Levenberg–Marquardt back-

propagation algorithm. To avoid local minimum, 5000 multiple realizations with different weight and bias initialization of training were implemented and minimum error realization was selected as the best one. After training, the optimal weights and bias values were obtained for B_{ob} artificial neural network and these are shown in Table 4.

All input parameters should be normalized in the range of $[-1, 1]$ before using in the ANN algorithm for B_{ob} . The procedure for normalization of input parameters is the same as of bubble point ANN algorithm.

The input parameters should be normalized using the following equations for the ANN algorithm of B_{ob} :

$$(T)_N = \frac{2(T - 182)}{(296 - 182)} - 1,$$

$$(\gamma_g)_N = \frac{2(\gamma_g - 0.825)}{(3.444 - 0.825)} - 1,$$

$$(API)_N = \frac{2(API - 29)}{(56.5 - 29)} - 1,$$

$$(Rs)_N = \frac{2(Rs - 96)}{(2496 - 96)} - 1.$$

The proposed ANN model gives normalized B_{ob} in the range $[-1, 1]$; therefore for the real value of B_{ob} , the output

Table 3 Weights and bias values for P_b artificial neural network model

Weights between the input layer and the hidden layers (w_{ij})				
Hidden layer neurons (j)	Input layer neurons (i)			
	1	2	3	4
1	3.647	0.334	-0.756	-0.098
2	1.448	1.708	2.495	0.248
3	0.671	-1.191	0.348	1.427
4	1.372	0.424	2.842	2.280
5	1.928	-1.584	0.793	2.126
6	-1.201	-2.074	2.223	4.147
7	3.870	0.274	1.113	-0.742
8	5.704	0.027	1.161	0.823
9	-1.226	-0.351	0.102	-0.753
10	-1.892	-3.004	-2.224	3.219
11	3.452	-0.584	-4.716	0.833
12	2.946	0.221	0.344	0.629

Bias values for the hidden layer neurons (b_j)		Weights between the hidden layer and the output layers (w_{jk})	
Hidden layer neurons (j)	Bias (b_j)	Hidden layer neurons (j)	Output one neuron
1	-2.997	1	-0.461
2	-1.726	2	-0.216
3	-0.038	3	0.713
4	-1.939	4	-0.286
5	-0.737	5	-0.342
6	-0.601	6	0.094
7	0.337	7	0.176
8	0.951	8	1.494
9	-0.448	9	-2.163
10	-0.052	10	0.097
11	3.467	11	-0.289
12	0.764	12	-2.737

Bias values for output layer neuron (b_k)	
Output layer neuron (k)	Bias value (b_k)
1	-0.73853

B_{ob} value should be de-normalized by the following equation:

$$B_{ob} = \frac{(2.916 - 1.214)((B_{ob})_N - (-1))}{2} + 1.214.$$

Viscosity in the P_b (μ_{ob}) ANN model

The viscosity in the P_b ANN model consists of four input neurons or input parameters, which are related to temperature, specific gravity of gas, API gravity of oil and solution gas–oil ratio, one hidden layer and one output neuron

related to μ_{ob} . The hidden layer consists of 26 neurons, which were obtained after sensitivity runs of number of neurons from 5 to 50. Tan-sigmoid and linear transfer functions were used in the hidden and output layer, respectively. It is important to note that the hidden layer neurons in μ_{ob} ANN model is higher than the bubble point and B_{ob} ANN models, because nonlinearity in μ_{ob} is higher than bubble point and B_{ob} . The algorithm of the ANN model for μ_{ob} is given below:

$$\mu_{ob} = f_L \left(\sum_{j=1}^{N_h} w_{jk} \times \left(f \left(\sum_{i=1}^{N_i} w_{ji} \times x_i + b_j \right) \right) + b_k \right).$$

Fig. 3 Architecture of ANN model for B_{ob}

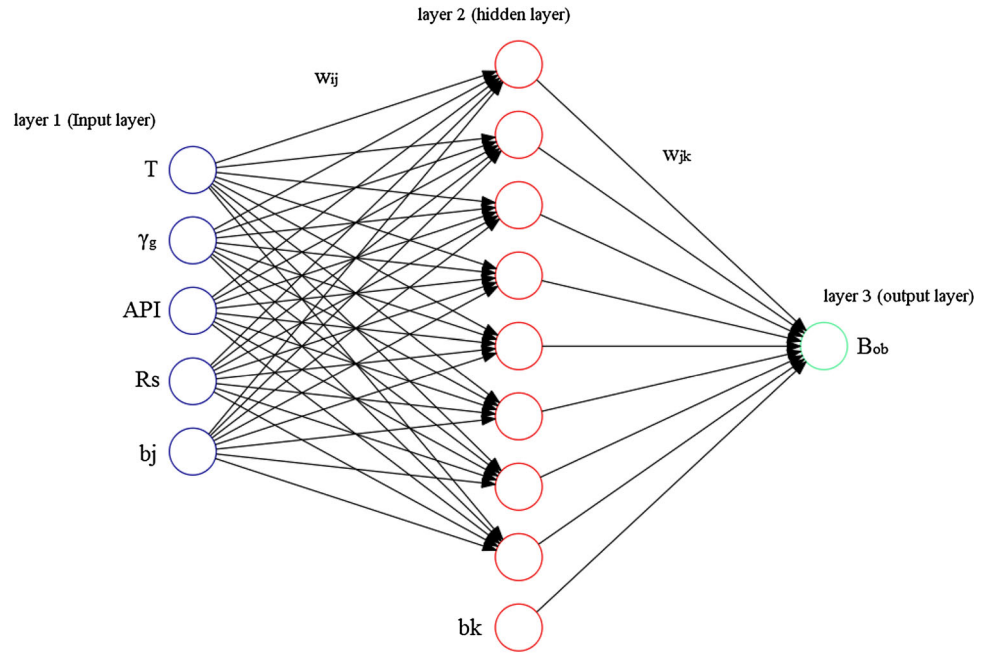


Table 4 Weights and bias values for B_{ob} artificial neural network model

Weights between the input layer and hidden layers (w_{ij})

Hidden layer neurons (j)	Input layer neurons (i)			
	1	2	3	4
1	-0.497	-0.711	-3.221	-0.921
2	0.432	-0.937	-2.129	-1.487
3	-0.126	-0.343	0.397	-1.908
4	1.298	0.477	-0.511	1.551
5	0.390	-0.720	-0.817	-1.248
6	-1.050	0.276	1.255	0.084
7	-0.090	2.205	-1.897	0.545
8	0.343	1.149	-0.246	-1.184

Bias values for hidden layer neurons (b_j)

Hidden layer neurons (j)	Bias (b_j)
1	-2.434
2	-2.131
3	0.369
4	0.231
5	0.010
6	0.153
7	2.702
8	-1.489

Weights between the hidden layer and output layers (w_{jk})

Hidden layer neurons (j)	Output one neuron
1	0.292
2	-0.550
3	-0.190
4	0.111
5	-0.479
6	-0.487
7	-0.445
8	-0.266

Bias values for the output layer neuron (b_k)

Output layer neuron (k)	Bias value (b_k)
1	0.07129

The above ANN algorithm for μ_{ob} can also be written in the following way:

$$(\mu_{ob})_N = f_L \left(\sum_{j=1}^{N_h} w_{jk} \times (f(w_{j1} \times (T)_N + w_{j2} \times (\gamma_g)_N + w_{j3} \times (API)_N + w_{j4} \times (Rs)_N + b_j)) + b_k \right).$$

Subscript N shows the normalized input and output parameters for ANN model. The bias values are constant values, which have similar concept to constants in linear or nonlinear regression. The architecture of the ANN model for μ_{ob} is shown in Fig. 4.

To obtain the weights and bias values, the training of neural network was performed by the Levenberg–Marquardt back-propagation algorithm. To avoid local

minimum, 5000 multiple realizations with different weight and bias initialization of training were implemented and minimum error realization was selected as the best one. After training, the optimal weights and bias values were obtained for μ_{ob} ANN and these are shown in Table 5.

All input parameters should be normalized in the range of $[-1, 1]$ before using in the ANN algorithm for μ_{ob} . The procedure for normalization of input parameters is the same as of bubble point pressure and B_{ob} ANN algorithms.

The input parameters should be normalized using the following equations for the ANN algorithm of μ_{ob} :

$$(T)_N = \frac{2(T - 188)}{(296 - 188)} - 1,$$

$$(\gamma_g)_N = \frac{2(\gamma_g - 0.825)}{(3.444 - 0.825)} - 1,$$

Fig. 4 Architecture of ANN model for μ_{ob}

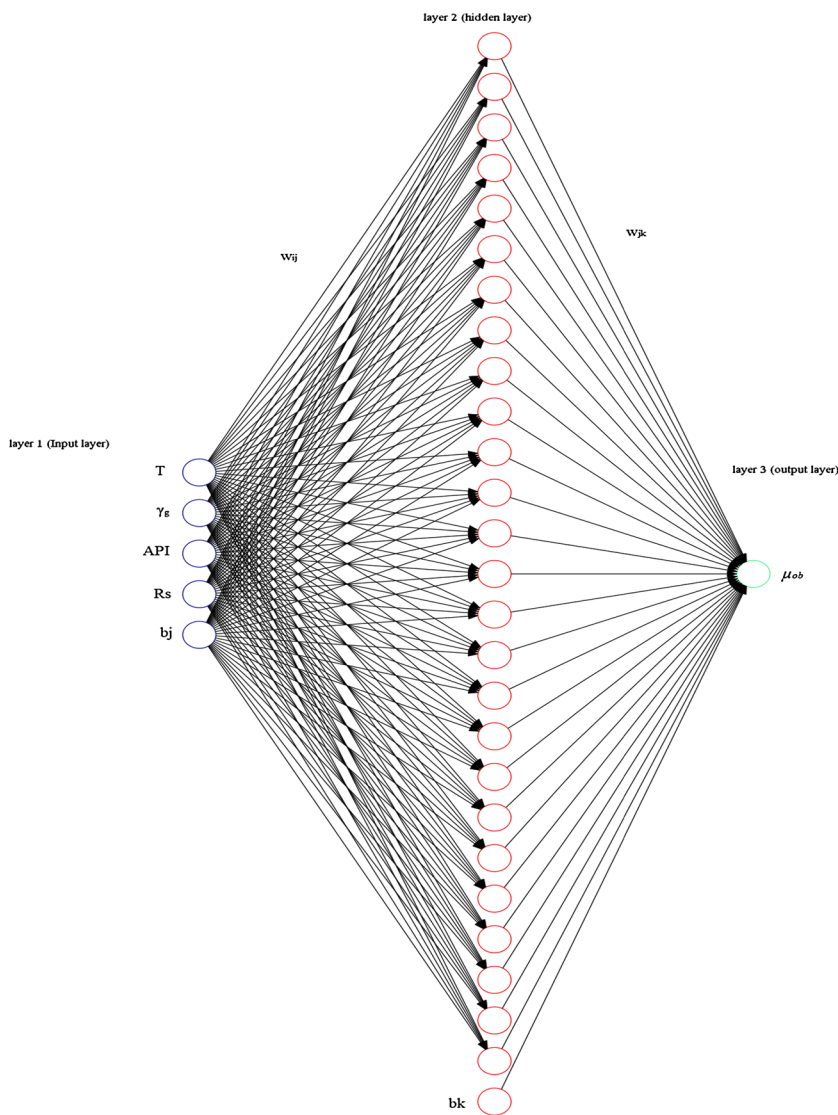


Table 5 Weights and bias values for μ_{ob} artificial neural network model

Weights between the input layer and hidden layers (w_{ij})				
Hidden layer neurons (j)	Input layer neurons (i)			
	1	2	3	4
1	-2.726	-0.166	-1.151	1.073
2	-1.622	1.182	1.594	1.712
3	-0.411	0.958	1.192	-3.193
4	1.075	-1.687	-2.430	0.457
5	1.848	-0.705	3.797	0.385
6	0.611	2.371	-0.584	2.170
7	3.312	-2.249	-1.234	0.703
8	0.768	1.796	-1.852	-2.403
9	-1.349	-3.097	-2.979	-0.231
10	2.278	1.340	1.022	2.054
11	1.923	-1.044	-4.740	-2.367
12	0.197	-1.999	-1.036	-2.363
13	0.988	-2.030	0.646	2.643
14	3.407	0.099	-1.393	-0.452
15	-6.313	-0.699	1.568	0.273
16	1.996	-1.785	-1.515	-0.991
17	-1.904	-1.389	3.804	-1.745
18	-0.044	-0.820	-3.282	-1.124
19	-1.411	-2.296	3.120	-0.946
20	-1.881	-0.526	-1.280	1.368
21	0.297	1.407	-2.016	1.526
22	3.099	1.335	0.788	3.403
23	0.375	-2.754	-1.046	-2.249
24	0.832	1.257	-0.681	3.005
25	2.547	2.403	0.883	0.151
26	1.195	-2.045	-0.810	-1.929
Bias values for the hidden layer neurons (b_j)		Weights between the hidden layer and output layers (w_{jk})		
Hidden layer neurons (j)	Bias (b_j)	Hidden layer neurons (j)	Output one neuron	
1	3.663	1	-1.458	
2	3.138	2	-0.899	
3	-2.524	3	-0.110	
4	-2.213	4	-0.526	
5	-1.706	5	-3.093	
6	-1.486	6	-1.085	
7	-0.456	7	0.616	
8	-1.073	8	0.675	
9	1.836	9	-1.455	
10	-0.910	10	0.849	
11	-0.984	11	-2.780	
12	0.618	12	-0.601	
13	-0.274	13	0.693	
14	0.611	14	0.668	
15	-0.676	15	-3.082	
16	0.440	16	1.574	

Table 5 continued

Bias values for the hidden layer neurons (b_j)		Weights between the hidden layer and output layers (w_{jk})	
Hidden layer neurons (j)	Bias (b_j)	Hidden layer neurons (j)	Output one neuron
17	-3.440	17	2.547
18	-0.884	18	1.230
19	-2.512	19	1.188
20	-2.259	20	0.875
21	-2.337	21	-0.070
22	1.967	22	-2.484
23	-3.722	23	-2.013
24	2.317	24	0.697
25	2.251	25	-0.838
26	3.162	26	0.196

Bias values for the output layer neuron (b_k)	
Output layer neuron (k)	Bias value (b_k)
1	0.316

Table 6 Statistical comparison of bubble point pressure correlations and proposed bubble point pressure ANN model

Correlations	APE	AAPE	E_{min}	E_{max}	St. dev
Standing (1947)	-43.5	49.18	0.43	391.05	68.37
Lasater (1958)	-20.61	31.31	0.04	273.65	49.36
Vazquez and Beggs (1980)	-52.07	55.31	0.16	403.99	70.3
Glaso (1980)	-24.82	32.08	0.04	247	45.64
Al-Marhoun (1988)	27.97	31.5	0.3	81.96	20.24
P_b -ANN testing data (proposed)	<i>1.437</i>	<i>4.425</i>	<i>0.197</i>	<i>34.397</i>	<i>8.151</i>
P_b -ANN all data (proposed)	<i>0.0404</i>	<i>3.5039</i>	<i>0.0005</i>	<i>53.3865</i>	<i>7.6935</i>

Italic values were used to differentiate between proposed algorithm results with the previous published model results

$$(\text{API})_N = \frac{2(\text{API} - 29)}{(43.8 - 29)} - 1,$$

$$(\text{Rs})_N = \frac{2(\text{Rs} - 92)}{(2496 - 92)} - 1.$$

The proposed ANN model gives normalized μ_{ob} in the range [-1, 1]; therefore for real value of μ_{ob} , output μ_{ob} value should be de-normalized by the following equation:

$$\mu_{ob} = \frac{(0.636 - 0.205)((\mu_{ob})_N - (-1))}{2} + 0.205.$$

Results and discussion

After training the neural network models for P_b , B_{ob} and μ_{ob} , the models become ready for testing and evaluation. To perform this, the first training data set (seen data) and the second testing data set, which were not seen by the neural network during training, were used.

To compare the performance and accuracy of the neural network model of P_b to other empirical correlations, five mostly used P_b correlations were selected. These are those of Standing (1947), Vazquez and Beggs (1980), Glaso (1980), and Lasater (1958). The statistical results of the comparison are given in Table 6. The ANN model of P_b outperforms all these empirical correlations.

To compare the performance and accuracy of the neural network model of B_{ob} to other empirical correlations, five mostly used B_{ob} correlations were selected. These are those of Standing (1947), Vazquez and Beggs (1980), Glaso (1980), and Al-Marhoun (1988, 1992). The statistical results of the comparison are given in Table 7. The ANN model of B_{ob} outperforms all these empirical correlations.

To compare the performance and accuracy of the neural network model of μ_{ob} to other empirical correlations, four mostly used μ_{ob} correlations were selected. These are those of Beggs and Robinson (1975), Chew and Connaly (1959), Khan et al. (1987), and Labedi (1992). The statistical

Table 7 Statistical comparison of B_{ob} correlations and the proposed B_{ob} ANN model

Correlations	APE	AAPE	E_{min}	E_{max}	St. dev
Standing (1947)	1.39	2.31	0.05	7.96	2.36
Vazquez and Beggs (1980)	12.84	12.84	5.99	24.83	4.37
Glaso (1980)	3.65	3.88	0.08	12.78	2.23
Al-Marhoun (1988)	2.27	2.34	0.01	13	2.55
Al-Marhoun (1992)	0.76	1.23	0.01	9.09	1.54
B_{ob} -ANN testing data (proposed)	<i>-0.062</i>	<i>0.497</i>	<i>0.011</i>	<i>3.488</i>	<i>0.740</i>
B_{ob} -ANN all data (proposed)	<i>0.0197</i>	<i>0.4143</i>	<i>0.0001</i>	<i>5.4211</i>	<i>0.7585</i>

Italic values were used to differentiate between proposed algorithm results with the previous published model results

Table 8 Statistical comparison of μ_{ob} correlations and the proposed μ_{ob} ANN model

Correlations	APE	AAPE	E_{min}	E_{max}	St. dev
Beggs and Robinson (1975)	-24.43	26.71	2.56	57.16	21.7
Chew and Connally (1959)	-3.41	12.21	1.27	25.31	13.62
Khan et al. (1987)	18.6	29.92	1.19	64.8	30.81
Labedi (1992)	-29.65	37.53	0.56	268.98	70.04
μ_{ob} -ANN testing data (proposed)	<i>-0.481</i>	<i>2.994</i>	<i>0.034</i>	<i>23.594</i>	<i>5.729</i>
μ_{ob} -ANN all data (proposed)	<i>-0.233</i>	<i>1.570</i>	<i>0.003</i>	<i>23.594</i>	<i>3.515</i>

Italic values were used to differentiate between proposed algorithm results with the previous published model results

results of the comparison are given in Table 8. The ANN model of μ_{ob} outperforms all these empirical correlations.

The proposed neural network models showed high accuracy in predicting the P_b , B_{ob} , and μ_{ob} values, and achieved the lowest relative error, lowest absolute percent relative error, lowest minimum error, lowest maximum error, and lowest standard deviation of relative error.

Conclusion

- A new ANN model was developed to predict the bubble point pressure, oil formation volume factor at P_b , and viscosity at P_b . The P_b and B_{ob} models were developed using 166 published data sets from the Pakistani crude oil samples. The μ_{ob} model was developed using 128 published data sets from the Pakistani crude oil samples.
- Out of the 166 data points, 70 % were used to train and cross-validate the ANN models for P_b and B_{ob} , and the remaining 30 % used to test the accuracy of P_b and B_{ob} models. Similarly, for the μ_{ob} model, out of the 128 data points, 70 % were used to train and cross-validate the ANN model and the remaining 30 % used to test the μ_{ob} accuracy.
- The results show that the developed models provide better predictions and higher accuracy than the published empirical correlations and have the capability to

fulfill the need of more accurate correlations for Pakistani crude oil. The present models provide predictions of P_b , B_{ob} , and μ_{ob} with an absolute average percent error of 4.4250, 0.4975, and 2.99 %, respectively, to unseen (testing) data and correlation coefficient of 0.99789, 0.997, and 0.97022, respectively, to unseen (testing) data.

Acknowledgments The authors wish to thank Dr. Al-Marhoun for his technical support related to this work and acknowledge the support provided by King Abdul-Aziz City for Science and Technology (KACST) through the Science and Technology Unit at King Fahd University of Petroleum and Minerals (KFUPM) for funding this work through Project No. 11-OIL2144-04 as part of the National Science, Technology and Innovation Plan.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1

See Figs. 5 and 6.

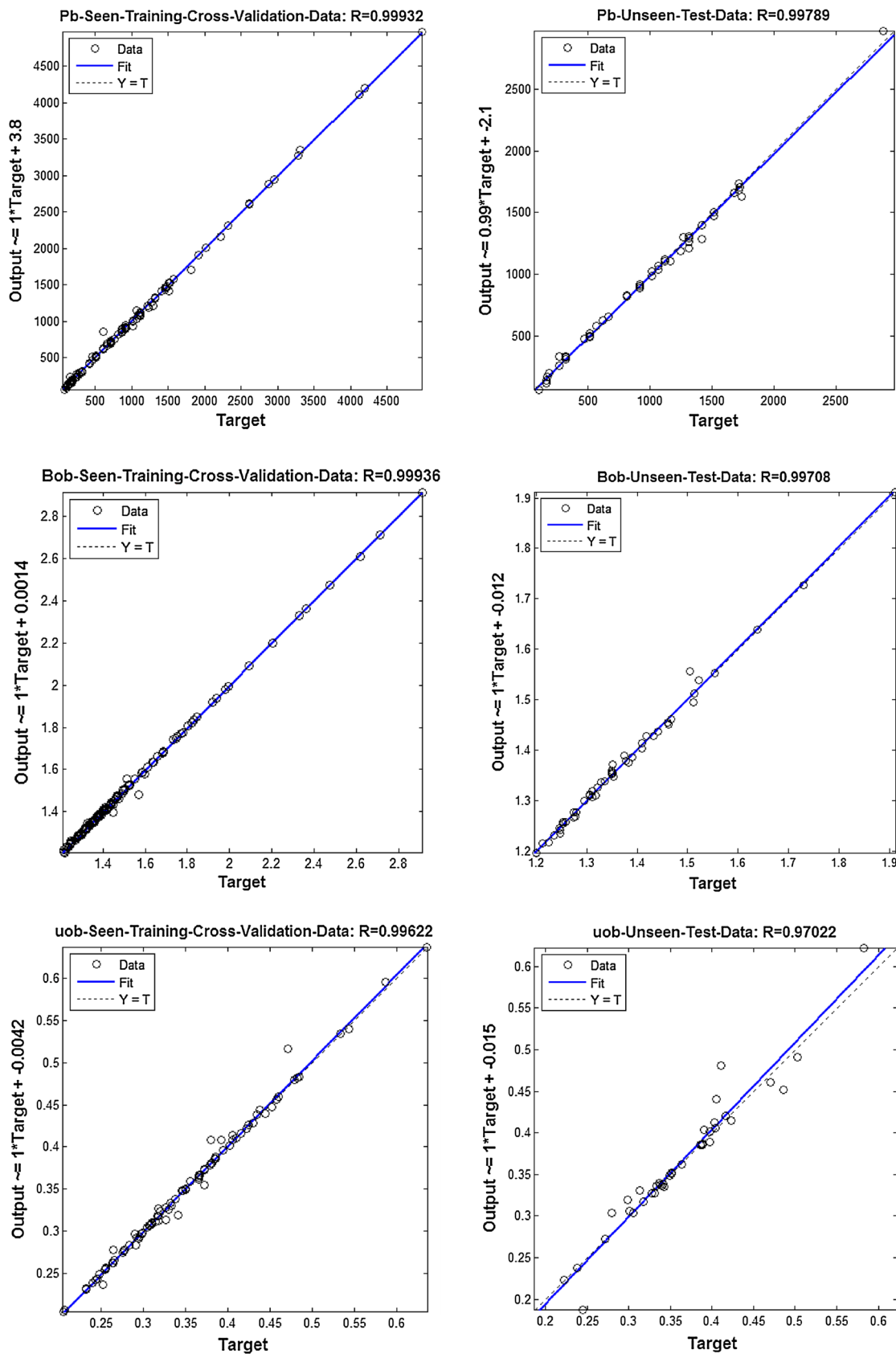


Fig. 5 Cross plots between ANN-predicted and real data

Fig. 6 Comparison plots between ANN-predicted and real data

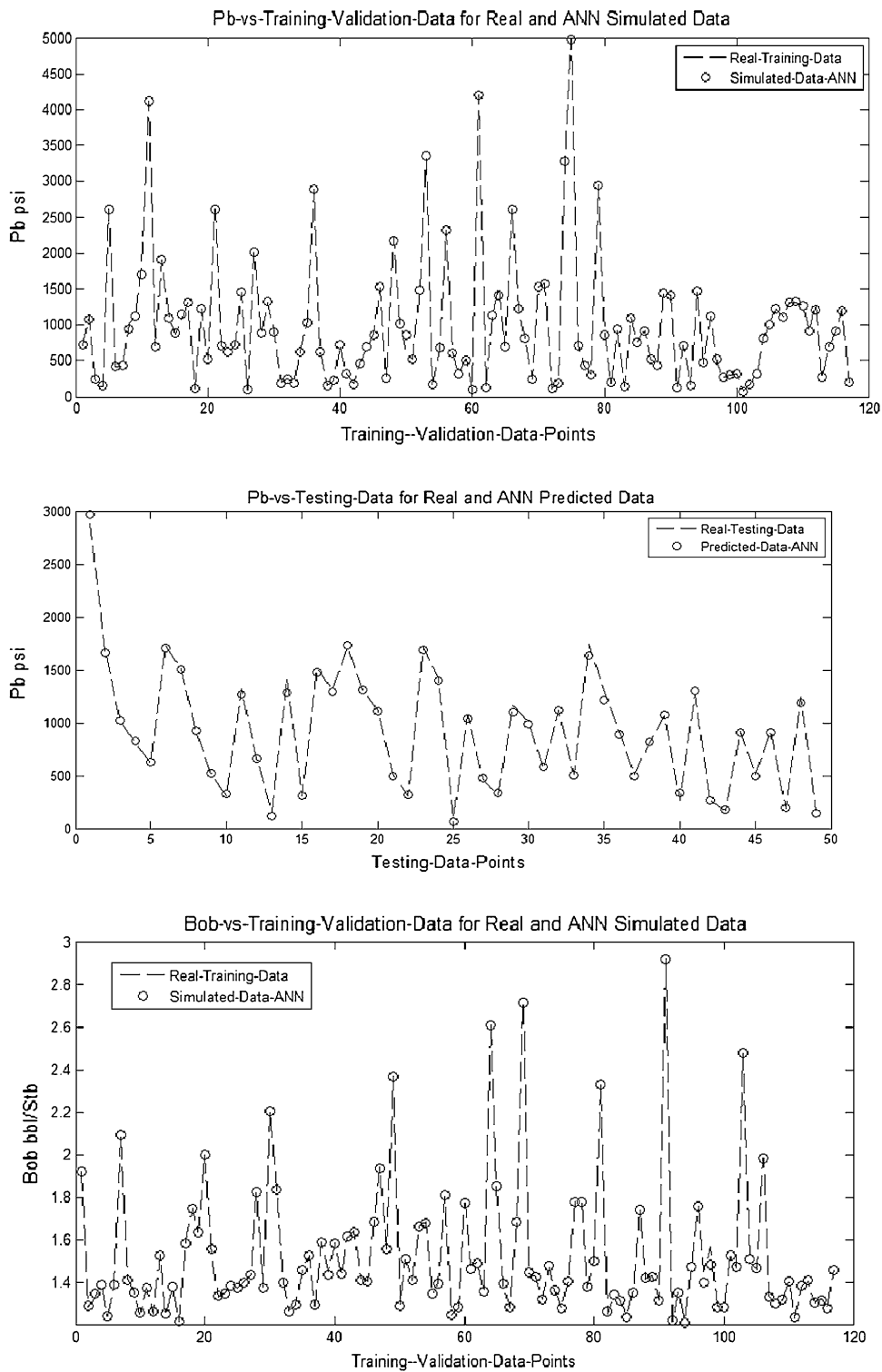
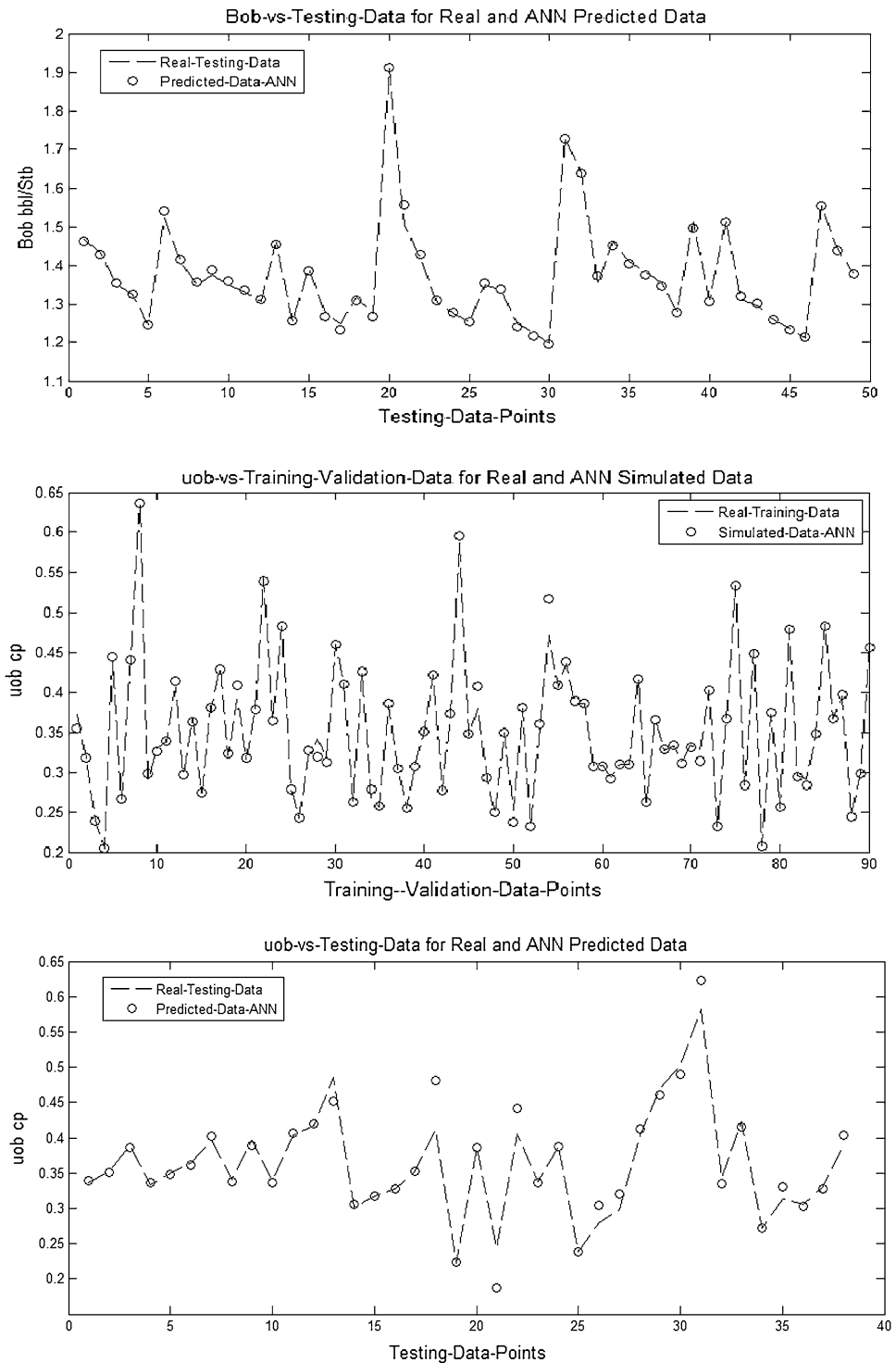


Fig. 6 continued



Appendix 2

Statistical parameters used in the study are average percentage error 'APE', average absolute percentage error (AAPE), average percentage error (APE), correlation coefficient (*R*), and standard deviation.

Relative percentage error is defined mathematically as follows:

$$Er = \frac{(X(\text{estimated}) - X(\text{measured}))}{X(\text{measured})} \times 100.$$

Absolute APE is defined mathematically as follows:

$$\text{AAPE} = \frac{1}{n} \sum_{i=1}^n |\text{Er}_i|.$$

Average percentage error is defined mathematically as follows:

$$\text{APE} = \frac{1}{n} \sum_{i=1}^n \text{Er}_i.$$

Standard deviation is defined mathematically as follows:

$$\text{Std. dev} = \left[\frac{n \sum_{i=1}^n (\text{Er}_i)^2 - (\sum_{i=1}^n \text{Er}_i)^2}{n^2} \right]^{0.5}.$$

Appendix 3

There were two transfer functions used in the proposed ANN models. These are tan-sigmoid and linear transfer functions. Tan-sigmoid function connects input layer neurons to hidden layer neurons. Linear transfer function connects hidden layer neurons to output layer neurons. Mathematically, these transfer functions are defined as follows.

Tan-sigmoid (tansig) transfer function:

$$f = \text{tansig}(n),$$

$$\text{tansig}(n) = \frac{2}{1 + e^{-2n}} - 1.$$

Linear (purelin) transfer function:

$$f_L = \text{purelin}(n),$$

$$\text{purelin}(n) = n,$$

where, n is any real input argument, f the tan-sigmoid transfer function, and f_L the linear transfer function

References

- Abdul-Majeed GHA, Salman NH (1988) An empirical correlation for FVF prediction. *J Can Pet Technol* 27(6):118–122
- Al-Fattah SM, Al-Marhoun MA (1994) Evaluation of empirical correlation for bubble point oil formation volume factor. *J Pet Sci Eng* 11:341
- Ali JK (1994) Neural networks: a new tool for the petroleum industry. In: Paper SPE 27561 presented at the 1994 European petroleum computer conference, Aberdeen, UK, 15–17 March
- Al-Marhoun MA (1988) PVT correlations for Middle East crude oils. *JPT* 40:650
- Al-Marhoun MA (1992) New correlation for formation volume factor of oil and gas mixtures. *JCPET* 31:22
- Almehaideb RA (1997) Improved PVT correlations for UAE crude oils. In: Paper SPE 37691 presented at the 1997 SPE Middle East oil show and conference, Bahrain, 15–18 March
- Al-Shammasi AA (1997) Bubble point pressure and oil formation volume factor correlations. In: Paper SPE 53185 presented at the 1997 SPE Middle East oil show and conference, Bahrain, 15–18 March
- Al-Yousef HY, Al-Marhoun MA (1993) Discussion of correlation of PVT properties for UAE crudes. *SPEFE* 8:80
- Beggs HD, Robinson JR (1975) Estimating the viscosity of crude oil system. *J Pet Technol* 9:1140–1149
- Chew J, Connally CA Jr (1959) A viscosity correlation for gas-saturated crude oils. *Trans AIME (Am Inst Min Metall)* 216:23–25
- De Ghetto G, Paone F, Villa M (1994) Reliability analysis on PVT correlation. In: paper SPE 28904 presented at the 1994 SPE European petroleum conference, London, UK, 25–27 October
- Dokla M, Osman M (1992) Correlation of PVT properties for UAE crudes. *SPEFE* 7:41
- Dokla M, Osman M (1993) Authors' reply to Discussion of correlation of PVT properties for UAE crudes. *SPEFE* 8:82
- Elsharkawy AM, Elgibaly A, Alikhan AA (1994) Assessment of the PVT correlations for predicting the properties of the Kuwaiti crude oils. In: paper presented at the 6th Abu Dhabi international petroleum exhibition and conference, 16–19 Oct
- Fauset L (1996) *Fundamentals of neural networks*. Prentice Hall, New Jersey
- Gharbi RB, Elsharkawy AM (1997) Neural-network model for estimating the PVT properties of Middle East crude oils. In: paper SPE 37695 presented at the 1997 SPE Middle East oil show and conference, Bahrain, 15–18 March
- Glaso O (1980) Generalized pressure–volume temperature correlations. *JPT* 32:785
- Gupta JP (2010) PVT correlations for Indian crude oil using artificial neural networks. *J Pet Sci Eng* 72:93–109
- Hanafy HH, Macary SA, Elnady YM, Bayomi AA, El-Batanoney MH (1997) Empirical PVT correlation applied to Egyptian crude oils exemplify significance of using regional correlations. In: paper SPE 37295 presented at the SPE oilfield chemistry international symposium, Houston, 18–21 Feb 1997
- Kartoatmodjo T, Schmidt Z (1994) Large data bank improves crude physical property correlations. *Oil Gas J* 92:51–55
- Katz DL (1942) Prediction of shrinkage of crude oils. *API, Drill Prod Pract*, pp 137–147
- Khan SA, Al-Marhoun MA, Duffuaa SO, Abu-Khamsin SA (1987) Viscosity correlations for Saudi Arabian crude oils. In: Presented at the 5th Society of Petroleum Engineers Middle East Oil Show, Bahrain, 7–10 March 1987, Paper SPE 15720
- Kumoluyi AO, Daltaban TS (1994) High-order neural-networks in petroleum engineering. In: Paper SPE 27905 presented at the 1994 SPE Western regional meeting, Longbeach, California, USA, 23–25 March
- Labedi R (1990) Use of production data to estimate volume factor density and compressibility of reservoir fluids. *J Pet Sci Eng* 4:357
- Labedi R (1992) Improved correlations for predicting the viscosity of light crudes. *J Pet Sci Eng* 8:221–234
- Lasater JA (1958) Bubble point pressure correlation. *Trans AIME (Am Inst Min Metall)* 213:379–381
- MathWorks, Inc (2008) *Neural network toolbox 6, user's guide*. MathWorks, Inc
- Macary SM, El-Batanoney MH (1992) Derivation of PVT correlations for the Gulf of Suez crude oils. In: Paper presented at the EGPC 11th petroleum exploration and production conference, Cairo, Egypt
- Mahmood MM, Al-Marhoun MA (1996) Evaluation of empirically derived PVT properties for Pakistani crude oils. *J Pet Sci Eng* 16:275
- McCain WD (1991) Reservoir fluid property correlations-state of the art. *SPEFE* 6:266

- Mohaghegh S (1995) Neural networks: what it can do for petroleum engineers. *JPT* 47:42
- Mohaghegh S, Ameri S (1994) A artificial neural network as a valuable tool for petroleum engineers. In: SPE 29220, unsolicited paper for Society of Petroleum Engineers
- Omar MI, Todd AC (1993) Development of new modified black oil correlation for Malaysian crudes. In: Paper SPE 25338 presented at the 1993 SPE Asia Pacific Asia Pacific oil and gas conference and exhibition, Singapore, 8–10 Feb
- Osman EA, Al-Marhoun MA (2002) Using artificial neural networks to develop new PVT correlations for Saudi crude oils. *Soc Pet Eng*. doi:[10.2118/78592-MS](https://doi.org/10.2118/78592-MS)
- Petrosky J, Farshad F (1993) Pressure volume temperature correlation for the Gulf of Mexico. In: Paper SPE 26644 presented at the 1993 SPE annual technical conference and exhibition, Houston, TX, 3–6 Oct
- Saleh AM, Maggoub IS, Asaad Y (1987) Evaluation of empirically derived PVT properties for Egyptian oils. In: Paper SPE 15721, presented at the 1987 Middle East oil show and conference, Bahrain, 7–10 March
- Standing MB (1947) A pressure–volume–temperature correlation for mixtures of California oils and gases. *API, Drill Prod Pract*, pp 275–287
- Standing MB (1977) Volumetric and phase behavior of oil field hydrocarbon system. Millet Print Inc., Dallas, p 124
- Sutton RP, Farshad F (1990a) Evaluation of empirically derived PVT properties for Gulf of Mexico crude oils. *SPRE* 5:79
- Sutton RP, Farshad F (1990) Supplement to SPE 1372, evaluation of empirically derived PVT properties for Gulf of Mexico crude oils. In: SPE 20277, available from SPE book Order Dep., Richardson, TX
- Varotsis N, Gaganis V, Nighswander J, Guieze P (1999) A novel non-iterative method for the prediction of the PVT behavior of reservoir fluids. In: Paper SPE 56745 presented at the 1999 SPE annual technical conference and exhibition, Houston, Texas, 3–6 October
- Vazquez M, Beggs HD (1980) Correlation for fluid physical property prediction. *JPT* 32:968