Korean J. Chem. Eng., 32(10), 2087-2096 (2015)
DOI: 10.1007/s11814-015-0025-y

pISSN: 0256-1115
eISSN: 1975-7220

Using an artificial neural network to predict carbon dioxide compressibility factor at high pressure and temperature

Erfan Mohagheghian*, Habiballah Zafarian-Rigaki, Yaser Motamedi-Ghahfarrokhi**, and Abdolhossein Hemmati-Sarapardeh**,†**

*Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL, Canada
**Department of Petroleum Engineering, Amirkabir University of Technology, Tehran, Iran
(*Received 18 October 2014 • accepted 27 January 2015*)

Abstract−Carbon dioxide injection, which is widely used as an enhanced oil recovery (EOR) method, has the potential of being coupled with CO_2 sequestration and reducing the emission of greenhouse gas. Hence, knowing the compressibility factor of carbon dioxide is of a vital significance. Compressibility factor (Z-factor) is traditionally measured through time consuming, expensive and cumbersome experiments. Hence, developing a fast, robust and accurate model for its estimation is necessary. In this study, a new reliable model on the basis of feed forward artificial neural networks is presented to predict CO_2 compressibility factor. Reduced temperature and pressure were selected as the input parameters of the proposed model. To evaluate and compare the results of the developed model with pre-existing models, both statistical and graphical error analyses were employed. The results indicated that the proposed model is more reliable and accurate compared to pre-existing models in a wide range of temperature (up to 1,273.15 K) and pressure (up to 140 MPa). Furthermore, by employing the relevancy factor, the effect of pressure and temprature on the Z-factor of CO_2 was compared for below and above the critical pressure of CO_2, and the physcially expected trends were observed. Finally, to identify the probable outliers and applicability domain of the proposed ANN model, both numerical and graphical techniques based on Leverage approach were performed. The results illustrated that only 1.75% of the experimental data points were located out of the applicability domain of the proposed model. As a result, the developed model is reliable for the prediction of CO_2 compressibility factor.

Keywords: Carbon Dioxide, Compressibility Factor, Artificial Neural Network, Equations of State, Leverage Approach

INTRODUCTION

After primary and secondary oil recovery, usually more than half the original oil in place (OOIP) remains in light and medium oil reservoirs. Among enhanced oil recovery (EOR) methods, CO_2 injection has gained much popularity due to its considerable effect on oil recovery enhancement and reducing the amount of greenhouse gas emissions by its sequestration in depleted oil reservoirs [1,2].

Hence, the compressibility factor of carbon dioxide is of a high importance in the oil and gas industry for CO_2 compression, pipeline design, material balance calculations and surface facilities design for EOR and sequestration purposes [3]. Compressibility factor is normally determined through experimental procedures in PVT laboratories. This process is costly, demanding and lengthy [4], so alternatives such as equations of state (EOS) or empirical correlations are preferred for estimation of compressibility factor [5].

Equations of state usually result in a cubic equation for compressibility factor, and the root which minimizes the free Gibbs energy has to be selected as the solution. This mathematical procedure adds to the complexities involved in the engineering calculations [6]. Correlations, on the other hand, are much easier to use and provide the solution in a shorter time [4]; however, they sometimes impose additional complications on the problem, lead to errors and their accuracy, especially close to and beyond the experimental data used for the curve fitting is under question [7]. Compressibility factor should be determined with the highest possible accuracy to avoid the error propagation in the prediction of other properties such as gas density, viscosity, isothermal compressibility, thermal conductivity and gas formation volume factor for carbon storage and injection rate calculations to enhance oil recovery [8]. The exact equations or correlations for the calculation of the mentioned properties require direct knowledge of the compressibility factor.

Hence, a robust fast method is required to predict the compressibility factor of carbon dioxide quickly and precisely. Artificial neural networks (ANN's) have attracted much attention in petroleum industry for solving nonlinear complex problems and handling complicated systems since they are capable of allowing large number of weights and biases making them flexible for fitting the parameters. We developed a simple and robust predictive model based upon artificial neural networks to estimate the compressibility factor of carbon dioxide in a wide range of temperature and pressure. To the best of authors' knowledge, an ANN model for the prediction of compressibility factor of carbon dioxide, especially on a large data bank covering wide ranges of temperature and pressure, has not been proposed before. The objectives of this study are as follows:

1. Collecting a large data bank from literature sources which covers compressibility factor of carbon dioxide in a wide range of pres-

†To whom correspondence should be addressed.
E-mail: aut.hemmati@gmail.com, aut.hemmati@aut.ac.ir

sure and temperature

2. Developing a reliable, fast and accurate model based on artificial neural networks predictive modeling to estimate compressibility factor of CO_2

3. Comparing the results of the model with those of equations of state and correlations via statistical and graphical error analyses to evaluate the accuracy and reliability of the model

4. Identifying the outliers and the suspected data points by means of leverage approach

5. Investigating the effect of each individual input parameter on the Z-factor of CO_2 by means of relevancy factor

COMPRESSIBILITY FACTOR

Gas deviation or compressibility factor (Z-factor) is defined as the ratio of the actual volume of a gas at a given pressure and temperature to the ideal volume and is a measure of how far the gas deviates from ideal behavior [9]. Some thermodynamic properties such as density, isothermal compressibility and viscosity can be calculated by means of compressibility factor.

According to the law of corresponding states proposed initially by van der Waals, all gases deviate from ideality to the same extent at the same reduced pressure P_r and reduced temperature T_r [4]. This can be represented in the following mathematical form:

$$Z=f(T_r, P_r) \tag{1}$$

where

$$T_r=\frac{T}{T_c} \tag{2}$$

$$P_r=\frac{P}{P_c} \tag{3}$$

where T_c and P_c are the critical temperature and pressure, respectively. At and beyond the critical point, the liquid and gaseous phase of a pure substance like carbon dioxide cannot be distinguished. Equations of state and empirical correlations are two common methods of compressibility factor estimation.

1. Equations of State

Since 1873, many equations of state have been developed and modified for estimation of PVT properties of gas and crude oil systems. The most well-known type of EOS widely being used in petroleum industry is the van der Waals type, which is of the following general format:

$$P=\frac{RT}{v-b}-\frac{a}{v^2+uv-w^2} \tag{4}$$

In a two-parameter EOS, u and w are correlated with b, while in a three or four-parameter EOS, u and w are related to b and a third or fourth parameter. A two-parameter EOS is usually enough for the prediction of vapor pressure; however, including a third parameter in an EOS, which is mostly derived by matching the experimental data of saturated liquid density, increases the EOS flexibility and its capability for prediction of liquid and vapor density as well as vapor pressure [10].

To calculate the PVT properties, it is preferred to rewrite the above equation in terms of compressibility factor as follows:

$$Z^3-(1+B-U)Z^2+(A-BU-U-W^2)Z-(AB-BW^2-W^2)=0 \tag{5}$$

where dimensionless parameters A, B, U and W are described as functions of only pressure and temperature for a pure component as below:

$$A=\frac{aP}{(RT)^2} \tag{6}$$

$$B=\frac{bP}{RT} \tag{7}$$

$$U=\frac{uP}{RT} \tag{8}$$

$$W=\frac{wP}{RT} \tag{9}$$

where the parameters a and b are represented by:

$$a=\Omega_a\frac{R^2T_c^2}{P_c}\alpha(T_r, \omega) \tag{10}$$

$$b=\Omega_b\frac{RT_c}{P_c} \tag{11}$$

where α is the temperature dependency term and ω is the acentric factor [10]. In the following, six of the most well-known equations of states will be reviewed.

1-1. Redlich-Kwong (RK) EOS [11]

$$u=b, w=0 \tag{12}$$

$$\Omega_a=0.42747, \Omega_b=0.08664 \tag{13}$$

$$\alpha=\frac{1}{\sqrt{T_r}} \tag{14}$$

1-2. Soave-Redlich-Kwong (SRK) EOS [12]

Soave [12] replaced the term $1/\sqrt{T_r}$ in RK EOS by a more general term as follows:

$$\alpha=[1+m(1-\sqrt{T_r})]^2 \tag{15}$$

where,

$$m=(0.48+1.574\omega-0.176\omega^2)/1.18 \tag{16}$$

1-3. Peng-Robinson (PR) EOS [13]

$$u=2b, w=b \tag{17}$$

$$\Omega_a=0.457235, \Omega_b=0.077796 \tag{18}$$

The temperature dependency function is similar to Eq. (15). The authors correlated m as:

$$m=0.3796+1.485\omega-0.1644\omega^2+0.1667\omega^3 \tag{19}$$

1-4. Schmidt-Wenzel (SW) EOS [14]

$$u=(1+3\omega)b, w^2=3\omega b^2 \tag{20}$$

$$\Omega_a=[1-\eta(1-q)]^3 \tag{21}$$

where q is the smallest positive root of the following equation:

$$(6\omega+1)q^3+3q^2+3q-1=0 \quad (22)$$

then

$$\eta=\frac{1}{3(1+q\omega)} \quad (23)$$

and

$$\Omega_b=\eta q \quad (24)$$

The temperature dependency function is similar to Eq. (15). The authors correlated m as:

$$m\equiv m_1=m_0+0.01429(5T_r-3m_0-1)^2 \quad \text{for } \omega\le 0.4 \quad (25)$$

$$m\equiv m_2=m_0+0.71(T_r-0.779)2 \quad \text{for } \omega\ge 0.55 \quad (26)$$

where,

$$m_0=0.465+1.347\omega-0.528\omega^2 \quad \text{for } \omega\le 0.3671 \quad (27)$$

$$m_0=0.5361+0.9593\omega \quad \text{for } \omega>0.3671 \quad (28)$$

and for intermediate values of $0.4<\omega<0.55$

$$m=\left(\frac{0.55-\omega}{0.15}\right)m_1+2(\omega-0.4)m_2 \quad (29)$$

for supercritical compounds, α has been correlated as:

$$\alpha=1-(0.4774+1.328\omega)\ln(T_r) \quad (30)$$

1-5. Patel-Teja (PT) EOS [15]

$$u=b+c,\ w^2=bc \quad (31)$$

where,

$$c=\Omega_c\frac{RT_c}{P_c} \quad (32)$$

$$\Omega_c=1-3\eta \quad (33)$$

$$\eta=0.329032-0.076799\omega+0.0211947\omega^2 \quad (34)$$

and Ω_b is the smallest positive root of the following equation:

$$\Omega_b^3+(2-3\eta)\Omega_b^2+3\eta^2\Omega_b-\eta^3=0 \quad (35)$$

and then Ω_a can be calculated by:

$$\Omega_a=3\eta^2+3(1-2\eta)\Omega_b+\Omega_b^2+(1-3\eta) \quad (36)$$

The temperature dependency function is similar to Eq. (15). The authors correlated m as:

$$m=0.452413+1.30982\omega-0.295937\omega^2 \quad (37)$$

1-6. Lawal-Lake-Silberberg (LLS) EOS [16]

$$u=\gamma b,\ w^2=\beta b^2 \quad (38)$$

$$\gamma=\frac{1+(\Omega_w-3)Z_c}{\Omega_w Z_c} \quad (39)$$

where Z_c is the critical compressibility factor and Ω_w has been correlated with acentric factor as follows:

$$\Omega_w=\frac{0.361}{1+0.0274\omega} \quad (40)$$

$$\beta=\frac{Z_c^2(\Omega_w-1)^3+2\Omega_w^2Z_c+(1-3Z_c)\Omega_w}{\Omega_w^2Z_c} \quad (41)$$

$$\Omega_a=[1+(\Omega_w-1)Z_c]^3 \quad (42)$$

$$\Omega_b=\Omega_wZ_c \quad (43)$$

The temperature dependency function is similar to Eq. (15). The authors correlated m as:

$$m=0.14443+1.06624\omega+0.02576\omega^2-0.18074\omega^3 \quad (44)$$

2. Empirical Correlations

Empirical correlations for compressibility factor provide a direct solution and have a shorter computational time. Two correlations, which can be used for estimation of carbon dioxide compressibility factor, namely Bahadori-Vuthaluru correlation [3] and the correlations for second and third Virial coefficients [17,18], are presented in the following.

2-1. Bahadori-Vuthaluru Correlation [3]

Bahadori and Vuthaluru [3] proposed the following set of equations for calculation of carbon dioxide compressibility factor for pressures up to 50 MPa.

$$\ln(Z)=\alpha+\frac{\beta}{T_r}+\frac{\gamma}{T_r^2}+\frac{\theta}{T_r^2} \quad (45)$$

where,

$$\alpha=A_1+\frac{B_1}{P_r}+\frac{C_1}{P_r^2}+\frac{D_1}{P_r^3} \quad (46)$$

$$\beta=A_2+\frac{B_2}{P_r}+\frac{C_2}{P_r^2}+\frac{D_2}{P_r^3} \quad (47)$$

$$\gamma=A_3+\frac{B_3}{P_r}+\frac{C_3}{P_r^2}+\frac{D_3}{P_r^3} \quad (48)$$

$$\theta=A_4+\frac{B_4}{P_r}+\frac{C_4}{P_r^2}+\frac{D_4}{P_r^3} \quad (49)$$

where parameters A_1 to D_4 for pressures less than 4 MPa as well as pressures between 4 to 50 MPa can be found elsewhere [3]. Note that this correlation cannot be tested against all the experimental data in this study, and it will only be checked in the pressure range for which it has been developed.

2-2. Correlations for Second and Third Virial Coefficients [17,18]

$$Z=1+\hat{B}\frac{P_r}{T_rZ}+\hat{C}\left(\frac{P_r}{T_rZ}\right)^2 \quad (50)$$

where,

$$\hat{B}=B^0+\omega B^1 \quad (51)$$

$$B^0=0.083-\frac{0.422}{T_r^{1.6}} \quad (52)$$

$$B^1 = 0.139 - \frac{0.172}{T_r^{4.2}} \tag{53}$$

$$\hat{C} = C^0 + \omega C^1 \tag{54}$$

$$C^0 = 0.01407 + \frac{0.02432}{T_r} - \frac{0.00313}{T_r^{10.5}} \tag{55}$$

$$C^1 = -0.02676 + \frac{0.05539}{T_r^{2.7}} - \frac{0.00242}{T_r^{10.5}} \tag{56}$$

MODEL DEVELOPMENT

1. Artificial Neural Networks

ANN's are nonlinear learning mathematical models inspired from the human brain system and biological neurons which are capable of solving problems with almost any type of complexity through relating inputs and outputs; as the result they have found applications in many fields such as aerospace, electronics, medical and chemical as well as oil and gas industry. ANN's are not prejudiced in their computational procedures and can be employed for pattern recognition, optimization, function approximation and data processing [19-21].

Two popular types of ANN are feed-forward and backward. Multilayer perceptron (MLP) is the most well-known feed-forward network [22] and has been used for the model development in this study. A multilayer ANN is composed of one input layer, one output layer and one or several hidden layers in between depending on the type of problem. A widely used MLP with only one hidden layer has shown the capability of estimating almost any kind of function. Each layer consists of a different number of processing elements, called neurons, which are interconnected through links. The task of the hidden layers is to find the relations between the input and output layer [20].

To develop the model, data is randomly split into three subsets called training, validation and testing. The network is trained through training phase and learns how to make the proper connections between the input and output variables by adjusting and updating weights and biases until the difference between the desired output and the corresponding inputs is minimized. Validation subset is used to measure network generalization and stop training before

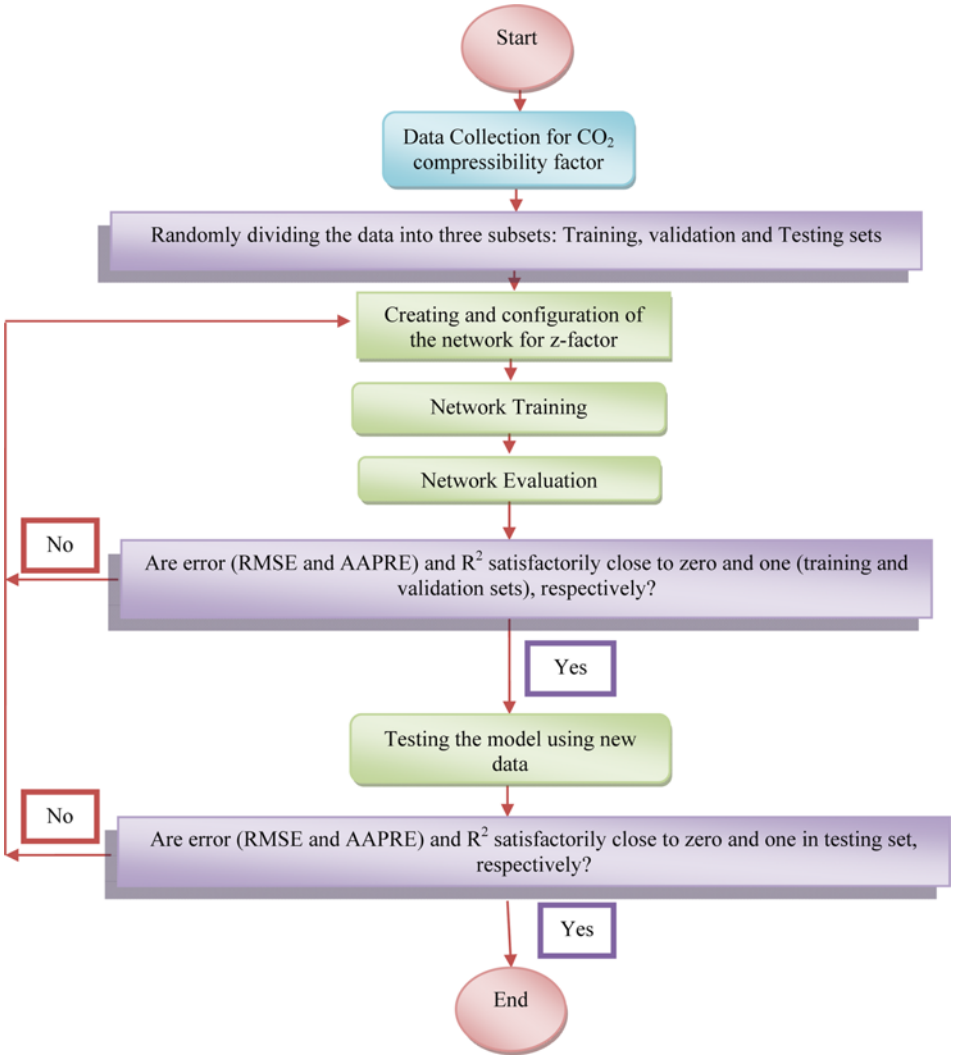

Fig. 1. Flowchart of procedure of ANN model development proposed in this study.

overfitting. Testing subset is not involved in the training phase and is used as an independent measure for evaluation of network performance [23]. Furthermore, transfer or activation functions are used to process the net output of each neuron. Various types of transfer functions have been proposed including linear, hard limit, logarithmic, log-sigmoid, tan-sigmoid and radial basis transfer functions [24].

The configuration of the network has a significant impact on the network performance and its prediction capability. The optimal configuration is usually reached by trial and error. The flowchart for the model development, which has been implemented in this study, is shown in Fig. 1.

To build and develop the model, experimental data should be collected from open literature sources and imported to the program in the proper format. The dataset should be divided randomly into three aforementioned subsets (training, validation and testing) such that data points are distributed homogeneously and different operating conditions are covered in each subset [25,26]. Then, the optimal configuration should be found by changing the number of neurons in the single hidden layer. A too small number of neurons may prevent the network from reaching the desired error, while too many neurons might lead to overfitting [23]. To obtain the optimal configuration, differnt networks are trained by means of Levenberg-Marquardt algorithm. This process has to be done iteratively by different initial guesses for weights and biases until the desired performance and error is obtained. Then each subset is evaluated by means of statistical quality criteria and finally the network with the highest performance and best prediction capability is selected.

2. Data Gathering

To develop the proposed model, 2118 data points from [27], covering temperatures from 273.15 up to 1,273.15 K and pressures up to 140 MPa as the input and experimentally derived compressibility factors for carbon dioxide as the output of the model, were collected. This dataset was randomly split into three subsets of training, validation and testing such that data points were distributed homogenously in each subset. 70% of the dataset has been used for training purpose, i.e., for constructing the model, 15% for validation, and the rest 15% for testing the prediction capability and accuracy of the model against unused data.

MODEL EVALUATION

1. Statistical Error Analysis

Several statistical parameters such as average percent relative error (APRE), average absolute percent relative error (AAPRE), root mean square error (RMSE) and coefficient of determination (R^2) have been suggested as the quality measures to assess the statistical validity and reliability of a model [28]. These parameters are as follows:

1. Average Percent Relative Error:

$$E_r=\frac{1}{n}\sum_{i=1}^{n}E_i \tag{57}$$

where E_i (percent relative error) is the relative difference between a represented/predicted value and its corresponding experimental value as below:

$$E_i=\left[\frac{Z_{i,\,exp}-Z_{i,\,rep./pred}}{Z_{i,\,exp}}\right]\times 100 \quad i=1, 2, 3, \ldots, n \tag{58}$$

2. Average Absolute Percent Relative Error:

$$E_a=\frac{1}{n}\sum_{i=1}^{n}|E_i| \tag{59}$$

3. Root Mean Square Error:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Z_{i,\,exp}-Z_{i,\,rep./pred})^2} \tag{60}$$

4. Coefficient of Determination:

$$R^2=1-\frac{\sum_{i=1}^{n}(Z_{i,\,exp}-Z_{i,\,rep./pred})^2}{\sum_{i=1}^{n}(Z_{i,\,rep./pred}-\overline{Z})^2} \tag{61}$$

where $\overline{Z}$ is the mean of the experimental data points of compressibility factor.

2. Graphical Error Analysis

To test the performance and validity of our model compared to the other investigated methods through observation, error distribution graphs and cumulative frequency plots vs. absolute percent relative errors have been used in this study. In the former technique, percent relative errors vs. an independent variable (pressure in this study) are plotted around the zero error line to check if the model follows an error trend. In the latter technique, the proportion of the data points having absolute percent relative errors below different ascending values are sketched against the absolute percent relative error in a cumulative manner for each model on the same plot to help the visual comparison. In addition, a crossplot, which is the plot of the predicted values by the proposed model vs. the experimental ones, shows how the prediction matches the reality.

3. Applicability Domain and Outlier Detection

While developing a mathematical model, it is necessary to detect the outliers. Several approaches have been proposed for outlier detection and identifying the applicability domain of a model of which Leverage approach is one of the most well-known. In this approach, residual is defined as the difference between the experimental data point and its corresponding value predicted by the model. Then to calculate hat or leverage indices, the following H matrix should be calculated as follows:

$$H=X(X^tX)^{-1}X \tag{62}$$

where X is an n×k matrix in which n stands for the number of data points and k shows the number of model parameters and X^t is the transposed of matrix X. The diagonal entries of matrix H are called hat or leverage indices in the feasible region of the problem.

Then standardized residuals (R) are plotted vs. hat indices to show the graphical representation of the applicability domain and detect the outliers. This graph is called Williams plot. The cut-off value (standardized residual equal to three) and warning leverage (H^*) equal to 3(k+1)/n (k=2, n=2118, H^*=0.0042) are used to check the validity of the model. If most of the data points lie in the range of $0\le H\le H^*$ and $-3\le R\le 3$ (i.e., within ±3 of the standard devia-

tion from the mean to ensure coverage of 99% of the normally distributed data), then it will be known that the data used for the model development as well as the model predictions are in the applicability domain, hence the developed model is statistically valid [29].

RESULTS AND DISCUSSION

An artificial neural network has been developed and used in this study to predict the compressibility factor of carbon dioxide. The collected data was split into three aforementioned subsets and the Levenberg-Marquardt back propagation algorithm was used in all the training steps to obtain the optimal configuration of the network through trial and error. The optimal network is composed of one hidden layer with 20 neurons, and transfer functions of *tansig* (tangent hyperbolic) and *purelin* in the hidden and output layer, respectively. The definitions for these transfer functions are as follows:

$$\tan\,\mathrm{sig}(n)=\frac{\exp(n)-\exp(-n)}{\exp(n)+\exp(-n)}=\frac{2}{1+\exp(-2n)}-1 \quad (63)$$

$$purelin(n)=n \quad (64)$$

Statistical parameters of the proposed model for the three different data subsets are shown in Table 1. The results proved the capability of the model for predicting compressibility factor of carbon dioxide in training and validation sets as well as the testing set. The synaptic parameters, namely weights and biases of the proposed model are also presented in Table 2. By knowing weights, biases, transfer functions and the configuration of the model, it can be used as a simple empirical correlation to predict the compressibility factor of carbon dioxide with high accuracy. According to the above, the output of the model, i.e., the compressibility factor of carbon dioxide, can be represented in the following form:

$$Z=purelin\{w_2*\tan\,\mathrm{sig}(w_1*x+b_1)+b_2\} \quad (65)$$

where w_1 and w_2 indicate the weight matrices for the hidden and output layers, respectively, b_1 and b_2 represent the biases for the hidden and output layers, respectively and x is the input vector, which is composed of reduced temperature and reduced pressure in this study.

To compare the accuracy and validity of the developed model

Table 1. Statistical parameters of the proposed model for determination of CO_2 compressibility factor

Statistical parameters	
Training set	
R^2	0.9983
AAPRE	0.6766
APRE	−0.1140
RMSE	0.0104
Number of data points	1482
Validation set	
R^2	0.9991
AAPRE	0.6233
APRE	−0.1976
RMSE	0.0077
Number of data points	318
Test set	
R^2	0.9986
AAPRE	0.7602
APRE	−0.1896
RMSE	0.0103
Number of data points	318
Total	
R^2	0.9984
AAPRE	0.6812
APRE	−0.1379
RMSE	0.0100
Number of data points	2118

Table 2. Parameters (weights and biases) of the ANN model

	Hidden layer			Output layer	
	Weights			Weight	
	P_r	T_r	Biases	w_{jk}	Bias
1	−0.2081	−4.8427	−5.0575	1.7677	5.9237
2	−3.3726	−2.3543	3.3115	0.2906	
3	−1.3967	0.3408	1.7169	2.701	
4	3.0768	2.1996	−3.0340	0.31668	
5	2.8292	−4.8768	−0.3766	−0.022098	
6	−4.6276	8.6876	4.4077	−1.6486	
7	−4.6975	−2.8499	2.9230	0.018669	
8	−4.5474	8.5928	4.3434	1.6729	
9	−1.1009	0.3375	1.5628	−4.2758	
10	−15.2857	12.8721	−2.0895	−3.9313	
11	15.2202	−12.9831	2.1164	−7.6992	
12	3.314	0.9006	1.3592	0.031707	
13	−13.1346	6.9543	−4.5769	−0.022996	
14	−6.6931	−2.7584	−10.1368	13.6972	
15	−15.3797	13.2933	−2.1767	−3.7546	
16	−9.0653	−2.8311	−11.9548	−3.2177	
17	0.4493	1.84	3.1299	2.5423	
18	−0.4095	−5.2874	−5.2853	−1.0965	
19	−8.3486	4.9147	−3.1942	0.027022	
20	4.2387	5.613	10.3835	4.0761	

Table 3. Statistical parameters for the different investigated models

Models	E_r%	E_a%	RMSE	R^2
Redlich-Kwong EOS	2.28	3.35	0.0444	0.984
Soave-Redlich-Kwong EOS	−2.22	2.23	0.0269	0.9966
Peng-Robinson EOS	0.82	1.64	0.0360	0.9906
Schmidt-Wenzel EOS	−1.23	2.07	0.0279	0.992
Patel-Teja EOS	−0.08	1.50	0.0267	0.994
Lawal-Lake-Silberberg EOS	0.49	1.13	0.0189	0.9966
Bahadori-Vuthaluru correlation	−0.92	4.38	0.0594	0.8766
Virial coefficients correlation	−2.12	3.08	0.0515	0.9665
This study	−0.14	0.68	0.0100	0.9984

against the other investigated methods, the statistical parameters mentioned above have been used. Table 3 shows the statistical parameters consisting of average percent relative error (E_r%), average absolute percent relative error (E_a%), root mean square error (RMSE) and coefficient of determination (R^2) for the six previously mentioned EOS's and two experimental correlations as well as this study. The results illustrated the better accuracy and higher prediction capability of the proposed model for the compressibility factor of carbon dioxide, since it has the lowest values for average absolute percent relative error and root mean square error as well as the closest value of R^2 to 1. The equations of state show better performance than the correlations; however, neither of them includes a significant error. The developed model is faster, easier to use, has the least error measures among all and removes the restrictions resulting from equations of state and correlations.

Afterwards, in order to visualize the comparison of all the methods with the proposed model in this study, average absolute percent relative error for each of the investigated equations of state and correlations as well as the training, validation and testing phases of this study are plotted in Fig. 2.

As can be seen from Table 3 and Fig. 2, the developed model shows better prediction capability and accuracy than all the other investigated methods. Among the correlations and equations of state, the Lawal-Lake-Silberberg equation of state has the lowest average absolute percent relative error with Patel-Teja and Peng-Robinson equations of state in the second and third position; there-

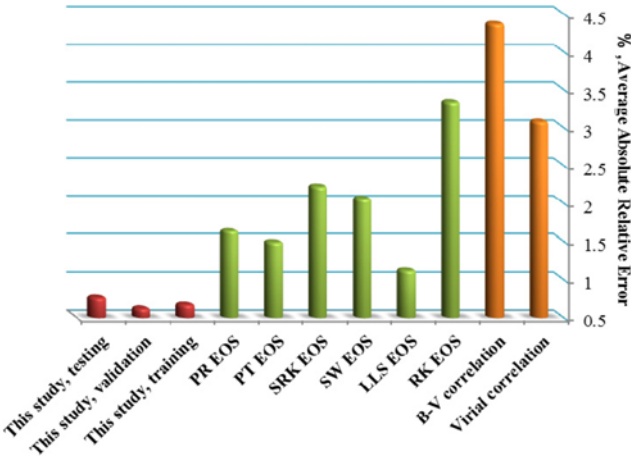

Fig. 2. Average absolute percent relative error for the different investigated models.

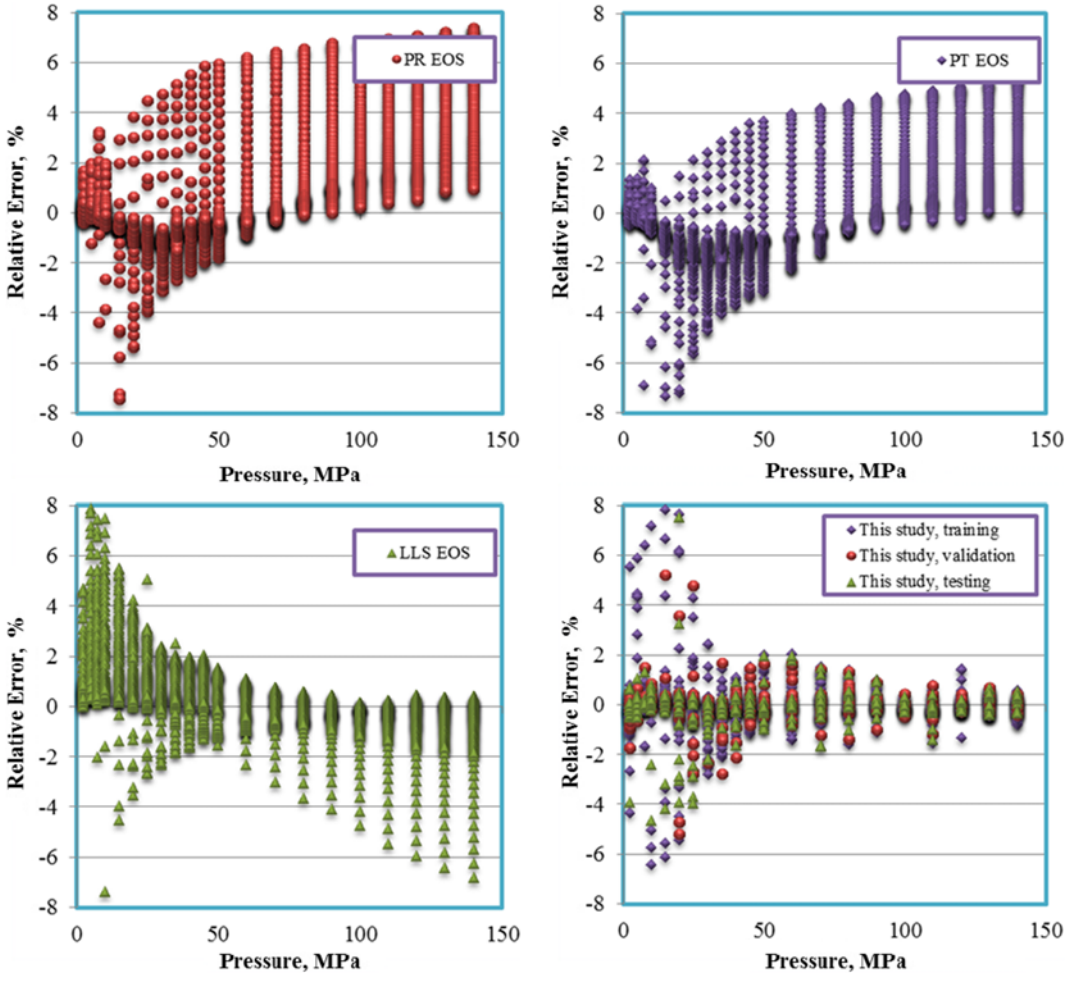

Fig. 3. Error distribution graphs for PR, PT and LLS EOS's and different phases of the proposed model.

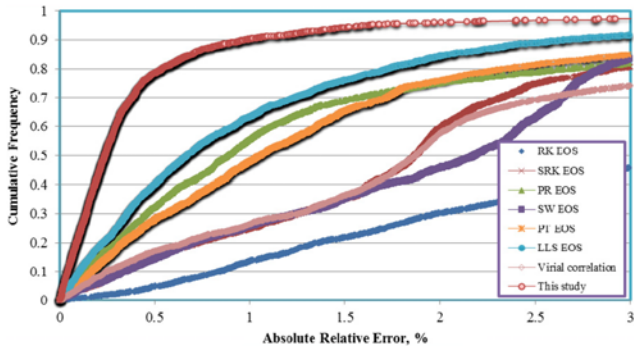

Fig. 4. Cumulative frequency vs. absolute percent relative error for the investigated methods.

fore, the error distribution graphs for these three equations of state as well as the training, validation and testing phases of this model have been plotted and shown in Fig. 3.

As can be seen from Fig. 3, in which percent relative errors for the predicted results by the selected methods against the corresponding pressures have been sketched, the developed model in this study has the closest predictions to the experimental values and thus, the least scattering and dispersion around the zero error line.

In the following, the cumulative frequency vs. the absolute percent relative error for all the investigated methods has been depicted in Fig. 4. The proposed ANN model predicts more than 70% of the data points with absolute relative error smaller than 0.5% and around 90% of the data points with absolute relative error smaller than 1%. Even the second best method, LLS EOS, predicts less than 40% of the data points with absolute relative error smaller than 0.5% and around 60% with an error of 1%. This is also an indication of the higher accuracy and validity of the developed model.

Afterwards, to observe how close the model prediction is to the experimental values and check the prediction reliability of the model, the crossplot of predicted compressibility factors vs. the experimental data points for training, validation and testing phases of the proposed model has been plotted in Fig. 5. As shown, nearly all of the points, especially for the testing phase which is unused during the model training, lie on the unit slope line, which is an indication of the accuracy and precise prediction performance of the model.

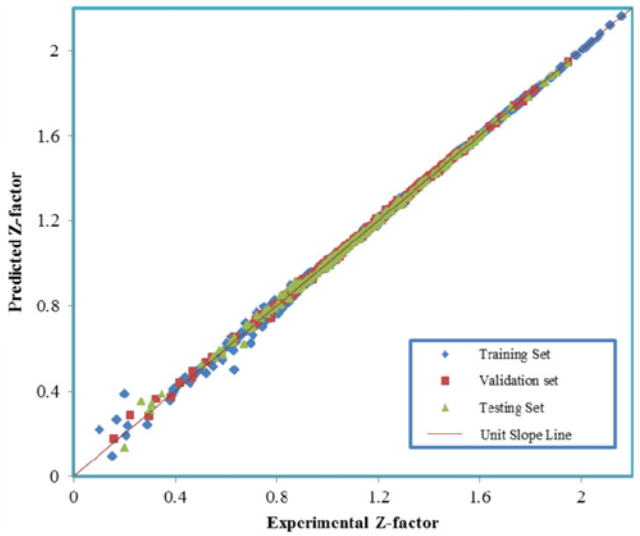

Fig. 5. Crossplot of predicted vs. experimental compressibility factors for different phases of the proposed model.

As the next step, to investigate the effect of input parameters on the output of the developed model, sensitivity analysis has been done. For this purpose, the relevancy factor (r) values with directionality have been employed to measure the degree of effect of each input on the model output as well as its sign whether for the effect to be positive or negative. The following formula is used for the calculation of the relevancy factor:

$$r(inp_i, Z) = \frac{\sum_{j=1}^{n}(inp_{i,j} - \overline{inp_i})(Z_j - \overline{Z})}{\sqrt{\sum_{j=1}^{n}(inp_{i,j} - \overline{inp_i})^2 \sum_{j=1}^{n}(Z_j - \overline{Z})^2}} \quad (66)$$

where $inp_{i,j}$ and $\overline{inp_i}$ represent the j^{th} value and the average of the i^{th} input parameter (in which inp_i is reduced temperature or pressure), respectively, Z_j and $\overline{Z}$ are the j^{th} and average value of CO_2 compressibility factor predicted by the model, respectively. T_o investigate the effect of input parameters on the output in detail, the relevancy factor for the inputs was calculated for both below and above the critical pressure of carbon dioxide (as an arbitrarily chosen pressure) to check if the physically expected trend for compressibility factor is observed. As can be seen in Fig. 6, the relevancy factor for reduced pressure is negative below the critical pressure and positive above the critical pressure, i.e., with increasing pressure, the compressibility factor decreases below the critical pressure and increases above the critical pressure. The relevancy factor for reduced temperature is positive in both regions, which means that the compressibility factor increases with increasing temperature regardless of the pressure. Care should be taken at very high pressures, as the compressibility factor reduces with increasing temperature. The absolute values of the relevancy factors indicate the absolute influence degree of the input variables and show that below the critical pressure, the effect of temperature on the Z-factor of carbon dioxide is more than pressure, but pressure has a greater effect than temperature on the Z-factor of CO_2 for pressures above the critical

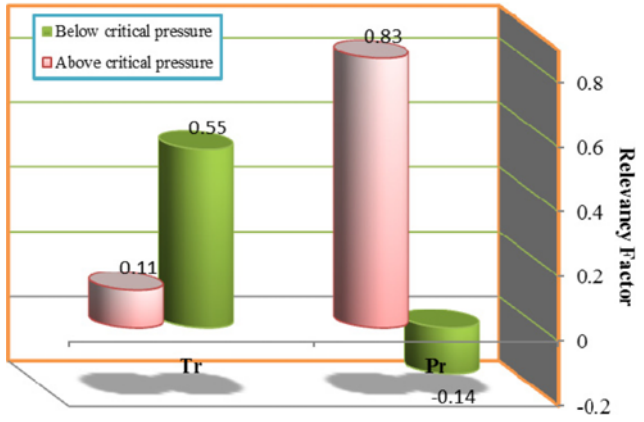

Fig. 6. Relevancy factor of the model inputs for below and above the critical pressure of CO_2.

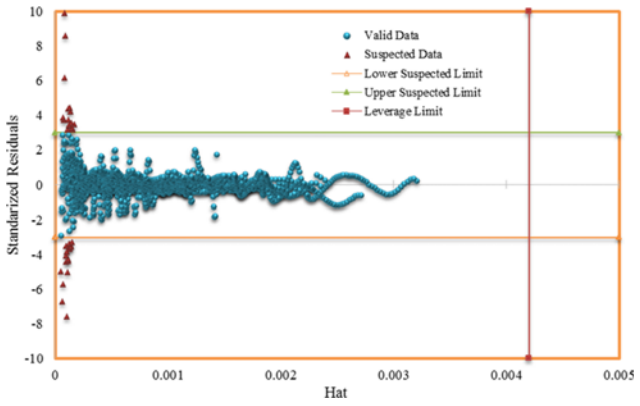

Fig. 7. Williams plot for the proposed model to identify the applicability domain.

pressure.

Finally, to check if the experimental data used for the model development are statistically reliable, the H matrix, hat indices and standardized residuals were calculated through the leverage approach and Williams plot was depicted. As can be seen in Fig. 7, nearly all the data points (except 37 points out of 2118) lie in the applicability domain of the developed ANN model (in the range of $0 \le H \le H^*$ and $-3 \le R \le 3$). This is a proof of validity and robustness of the model. The experimental data points with lesser amounts of hat indices and standardized residuals are of more reliability and hence, those 37 points can be regarded as suspected ones.

CONCLUSION

Accurate determination of CO_2 compressibility factor is of a high significance, especially for enhancing oil recovery and sequestration purposes. A fast robust ANN model has been developed which can predict the Z-factor of carbon dioxide with high accuracy and reliability. Statistical and graphical error analyses have shown the prediction capability of the proposed model and its performance compared to six widely used equations of state and two empirical correlations with respect to lower error measures and higher coefficient of determination. The relevancy factor was used to quantify the impact of the input parameters on the model output. The effect of pressure and temperature on the compressibility factor of carbon dioxide was compared for below and above the critical pressure, and it was concluded as well that the model actually follows the physically expected trend. The leverage approach was also used to determine the applicability domain of the model and the probable outliers. Based on the results, only 37 data points out of the total 2118 points are out of the applicability domain, which is an evidence of the model's robustness.

LIST OF SYMBOLS

Z : compressibility factor
T : temperature
P : pressure
Z_c : critical compressibility factor
T_c : critical temperature
P_c : critical pressure
T_r : reduced temperature
P_r : reduced pressure
R : universal gas constant
v : molar volume
a : attractive term of EOS
b : co-volume
u : EOS parameter
w : EOS parameter
A, B, U, W : dimensionless parameters of EOS
ω : acentric factor
Ω_a, Ω_b, Ω_w : EOS parameters' coefficients
α : temperature dependency term of EOS
m, m_0, m_1, m_2 : parameters correlated with acentric factor in temperature dependency term equation
η : parameter in SW and PT EOS's
q : parameter in SW EOS
c : parameter in PT EOS
γ, β, Ω_w : parameters in LLS EOS
α, β, γ, θ, A_1, …, A_4, B_1, …, B_4, …, C_1, …, C_4, …, D_1 : parameters of bahadori-vuthaluru correlation
$\hat{B}$, $\hat{C}$, B^0, B^1, C^0, C^1 : parameters of virial correlation
n : number of data points
E_i : percent relative error for data point
E_r : average percent relative error
E_a : average absolute percent relative error
RMSE : root mean square error
SD : standard deviation
R^2 : coefficient of determination
H : hat matrix
X : matrix of experimental data points
R : standardized residual
k : number of model parameters
H^* : warning leverage
w_1, w_2 : weight matrices for hidden and output layer
b_1, b_2 : biases for hidden and output layer
x : input vector
r : relevancy factor
inp_i : the i^{th} input parameter
$inp_{i,j}$: the j^{th} component of the i^{th} input parameter
$\overline{inp}_i$: average of the i^{th} input parameter
$\overline{Z}$: average of compressibility factor

REFERENCES

1. A. Hemmati-Sarapardeh, S. Ayatollahi, M.-H. Ghazanfari and M. Masihi, *J. Chem. Eng. Data*, **59**, 61 (2013).
2. A. Hemmati-Sarapardeh, S. Ayatollahi, A. Zolghadr, M.-H. Ghazanfari and M. Masihi, *J. Chem. Eng. Data*, **59**, 3461 (2014).
3. A. Bahadori and H. Vuthaluru, *Rapid estimation of carbon dioxide compressibility factor using simple predictive tool, in: SPE Asia Pacific Oil and Gas Conference and Exhibition*, Society of Petroleum Engineers (2010).
4. T. Ahmed, *Reservoir engineering handbook*, Gulf Professional Publishing (2006).
5. T. H. Ahmed, T. H. Ahmed and T. H. Ahmed, *Hydrocarbon phase*

behavior, Gulf Publishing Company (1989).
6. E. Heidaryan, J. Moghadasi and M. Rahimi, *J. Petroleum Sci. Eng.*, **73**, 67 (2010).
7. E. M. E.-M. Shokir, M. N. El-Awad, A. A. Al-Quraishi and O. A. Al-Mahdy, *Chem. Eng. Res. Design*, **90**, 785 (2012).
8. M. Mahmoud, *J. Energy Res. Technol.*, **136**, 012903 (2014).
9. A. Bahadori, H. B. Vuthaluru and S. Mokhatab, *Int. J. Greenhouse Gas Control*, **3**, 474 (2009).
10. A. Danesh, *PVT and phase behaviour of petroleum reservoir fluids*, Elsevier (1998).
11. O. Redlich and J. Kwong, *Chem. Rev.*, **44**, 233 (1949).
12. G. Soave, *Chem. Eng. Sci.*, **27**, 1197 (1972).
13. D.-Y. Peng and D. B. Robinson, *Ind. Eng. Chem. Fundam.*, **15**, 59 (1976).
14. G. Schmidt and H. Wenzel, *Chem. Eng. Sci.*, **35**, 1503 (1980).
15. N. C. Patel and A. S. Teja, *Chem. Eng. Sci.*, **37**, 463 (1982).
16. A. Lawal, E. Van der Laan and R. Thambynayagam, *Four-parameter modification of the Lawal-Lake-Silberberg equation of state for calculating gas-condensate phase equilibria, in: SPE Annual Technical Conference and Exhibition*, Society of Petroleum Engineers (1985).
17. J. O'connell and J. Prausnitz, *Ind. Eng. Chem. Process Design Development*, **6**, 245 (1967).
18. H. Orbey and J. Vera, *AIChE J.*, **29**, 107 (1983).
19. P. Tahmasebi and A. Hezarkhani, *J. Petroleum Sci. Eng.*, **86**, 118 (2012).
20. A. Ramgulam, *Utilization of artificial neural networks in the optimization of history matching*, in, The Pennsylvania State University (2006).
21. S. Mohaghegh, *Virtual intelligence and its applications in petroleum engineering*, J. of Petroleum Technology, Distinguished Author Series (2000).
22. S. Nowroozi, M. Ranjbar, H. Hashemipour and M. Schaffie, *Fuel Processing Technol.*, **90**, 452 (2009).
23. M. Lashkarbolooki, A. Z. Hezave and S. Ayatollahi, *Fluid Phase Equilib.*, **324**, 102 (2012).
24. M. T. Hagan and H. B. Demuth, *Neural networks for control*, in: American Control Conference, Proceedings of the 1999, IEEE, 1642 (1999).
25. A. Hemmati-Sarapardeh, B. Mahmoudi, S. A. Ramazani and A. H. Mohammadi, *Korean J. Chem. Eng.*, **31**, 1253 (2014).
26. M. Arabloo, M.-A. Amooie, A. Hemmati-Sarapardeh, M.-H. Ghazanfari and A. H. Mohammadi, *Fluid Phase Equilib.*, **363**, 121 (2014).
27. G. C. Kennedy, *Am. J. Sci.*, **252**, 225 (1954).
28. S. Esfahani, S. Baselizadeh and A. Hemmati-Sarapardeh, *J. Natural Gas Sci. Eng.*, **22**, 348 (2015).
29. A. Hemmati-Sarapardeh, R. Alipour-Yeganeh-Marand, A. Naseri, A. Safiabadi, F. Gharagheizi, P. Ilani-Kashkouli and A. H. Mohammadi, *Fluid Phase Equilib.*, **354**, 177 (2013).

APPENDIX A: INSTRUCTION FOR USING THE MODEL

Example: calculation of the compressibilty factor of CO_2 at 100 MPa and 473.15 K:

Solution:

First, reduced pressure and reduced temperature are calculated.

$P_r=P/P_c$; $T_r=T/T_c$;

$P_r=13.5318$ and $T_r=1.5551$

Then, change the directory of Matlab to the file in which "net" exists. The compressibility factor is calculated using the below command line in the workspace as follows:

```
Load net
x=[P_r, T_r]
Z=sim(net, x')
```

Using the aforementioned instruction for the interested pressure and temperature, the predicted compressibility factor is obtained 1.3832, while the experimental compressibility factor is 1.3793.