*Article*

# Modeling Natural Gas Compressibility Factor Using a Hybrid Group Method of Data Handling

**Abdolhossein Hemmati-Sarapardeh [1], Sassan Hajirezaie [2], Mohamad Reza Soltanian [3], Amir Mosavi [4,5], and Shahaboddin Shamshirband [6,7,*]**

[1] Department of Petroleum Engineering, Shahid Bahonar University of Kerman, Kerman, Iran; aut.hemmati@aut.ac.ir
[2] Department of Civil and Environmental Engineering, Princeton University, Princeton, United States; sassanh@princeton.edu
[3] Department of Geology, University of Cincinnati, Cincinnati, OH 45221, United States; soltanma@uc.edu
[4] Faculty of Health, Queensland University of Technology, Brisbane QLD 4059, Australia; amirhosein.mosavi@qut.edu.au
[5] School of the Built Environment, Oxford Brookes University, Oxford OX3 0BP, UK; a.mosavi@brookes.ac.uk
[6] Department Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam
[7] Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam
* Correspondence: shahaboddin.shamshirband@tdtu.edu.vn

**Abstract:** A Natural gas is increasingly being sought after as a vital source of energy, given that its production is very cheap and does not cause the same environmental harms that other resources, such as coal combustion, do. Understanding and characterizing the behavior of natural gas is essential in hydrocarbon reservoir engineering, natural gas transport, and process. Natural gas compressibility factor, as a critical parameter, defines the compression and expansion characteristics of natural gas under different conditions. In this study, a simple second-order polynomial model based on the group method of data handling (GMDH) is presented to determine the compressibility factor of different natural gases at different conditions, using corresponding state principles. The accuracy of the model evaluated through graphical and statistical analyses. The results show that the model is capable of predicting natural gas compressibility with an average absolute error of only 2.88%, a root means square of 0.03, and a regression coefficient of 0.92. The performance of the developed model compared to widely known, previously published equations of state (EOSs) and correlations, and the precision of the results demonstrates its superiority over all other correlations and EOSs.

**Keywords:** Natural gas; gas compressibility factor; group method of data handling (GMDH); big data; Equation of state; Correlation

---

## 1. Introduction

The increasing demand for oil and coal as energy and the technological and environmental concerns associated with its production and consumption have drawn attention toward natural gas. The natural gas consumption generates less pollutants and greenhouse gases [1, 2]. Understanding the behavior of natural gas is important to all reservoir and chemical engineering calculations that deal with gas as one of the main phases. Among the parameters of evaluating the behavior of natural gas, the gas compressibility factor is essential for determining the natural gas's phase behavior. Gas compressibility represents the proportion of volume a given amount of gas at a specific pressure and temperature to the ideal volume of it at the common conditions. Gas compressibility makes the difference between ideal gas and real gas. The following relationship is generally used to calculate gas compressibility:

$$z = \frac{V_{\text{real}}}{V_{\text{ideal}}} = \frac{V}{(nRT)/P} \quad (1)$$

Where R represents the universal gas constant and z is the gas compressibility factor. $V_{real}$ (V) and $V_{ideal}$ denote real and ideal gas volumes, respectively. T and P and n represent the gas temperature, pressure, and some moles, respectively.

At low pressure and temperature conditions, gas molecules have fewer interactions and collisions, and behavior can be considered ideal. However, at high temperature and pressure, the collisions between the molecules become critical and need to be taken into account when making predictions for gas expansion or contraction [2, 3].

There are various techniques to measure the compressibility factor. One main way is by performing compression-expansion experiments. Overall, the experimental measurement of compressibility factor appeared to be an accurate approach compared to all other approaches. However, they are generally slow, cumbersome, and costly. Also, it is reported not feasible to conduct an experiment for every single condition considering various pressure and temperature at which the compressibility is needed. Using equations of state (EOS) is another approach to determine the compressibility factor. When utilizing EOS, the reservoir characteristics are being employed. Generally, these equations come from the following form when the gas compressibility factor is the target PVT parameter:

$$Z^3 + a \times Z^2 + b \times Z + c = 0 \quad (2)$$

where a, b, and c represent the empirical constants of composition functions for temperature, pressure, and gas. Furthermore, Z denotes the gas compressibility factor. Even though these equations are advantages and their implementation can facilitate the measurement of other gas properties such as enthalpy, entropy, and Gibbs free energy, they are usually implicit higher-order equations that require intense computations. Besides the complex computations, the binary interaction coefficients used in some EOS's need to be measured by conducting experiments that may not be practical. Further, it has been shown that these equations are not suitable for predicting hydrocarbon gas properties [4].

Empirical correlations are another source of determining gas compressibility factor, which is easy and fast to use but is generally associated with erroneous predictions [5-7]. A minor estimation error in compressibility factor of correlations would lead to false prediction of formation, density and the amount of gas. Therefore, development of fast, user-friendly, and accurate models to predict the compressibility factor is critical.

Several researchers have attempted to develop methods to estimate the compressibility factor. For instance, Katz and Standing [8] developed a graphical approach of the basis of pseudo-reduced. Van der Waals [9] was one of the pioneers of EOS methods by taking into account the intramolecular forces and volume of molecules. Using Van der Waals EOS for determining gas compressibility factor leads higher accuracy compared to the empirical approach introduced by Katz and Standing. Other authors who contributed to the development of reliable EOS's are Peng–Robinson [10], Lawal–Lake–Silberberg [11], Patel–Teja [12], and Soave–Redlich–Kwong [13].

A general expression for the PVT relationship of fluids has the following form [2]:

$$P = \frac{RT}{v - b} - \frac{a}{v^2 - uv - w^2} \quad (3)$$

An expression for gas compressibility factor can be written by rewriting the above equation and implementing the equation for gas compressibility factor as follows:

$$Z^3 - (1 + B - U)Z^2 + (A - BU - U - W^2)Z - (AB - BW^2 - W^2) = 0 \quad (4)$$

Where A, B U, and W are dimensionless parameters that can be determined from the current pressure, temperature, and composition.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2019 doi:10.20944/preprints201905.0263.v1



In order to develop faster methods, several authors have developed correlations that can be explicitly used to address the problem. In 1973, Hall and Yarborough [14] transformed the graphical chart of Katz and Standing into a relatively simple correlation by fitting their correlation to the chart and determining the correlation coefficients. Brill and Beggs [15] also employed Katz and Standing chart and developed a correlation to estimate gas compressibility factor. Dranchuk [16] used an EOS developed by Benedict–Webb–Rubin [17] and proposed a gas compressibility correlation in 1974. Abu-Kassem joined Dranchuk in 1975 to develop an analytical equation for reduced gas density that can be efficiently utilized to determine gas compressibility factor [18]. In 1975, Gopal collected multiple correlations for the gas compressibility factor at various conditions [19]. Kumar [6] introduced a novel model for gas compressibility factor to be used by Shell. In 2010, Heidaryan et al. used multiple regression analysis [20], and in the same year, Azizi employed genetic programming [21]. A comprehensive study of the mentioned methods was conducted by Sanjari and Lay [22] in which the performance of the methods mentioned above have been investigated.

Recently, the usage of intelligent models in the oil and gas industry has attracted much attention. These models have been used to determine, oil and gas thermodynamic properties, reservoir formation properties and miscibility conditions required for gas injection processes [23-32]. These models take both input and output values to get trained and later can make predictions. Even though the original, intelligent models were considered black box models, there have been numerous modifications to these models to make them transparent and usable methods, and their performance has significantly improved over the past few years. Intelligent models have been used in many reservoir engineering calculations. There are also some intelligent models that were developed specifically for predicting natural gas properties [23-25]. We have already developed two intelligent models for predicting natural gas compressibility factor using the same data bank of this study [1, 2]. However, they are a black box, and their usage generally needs software.

In this study, a novel supervised approach of GMDH proposed as a robust model to estimate the compressibility factor. To do this, a comprehensive data bank of compressibility factor at wide ranges of temperature, pressure, and composition was used. Several statistical quality measures and graphical techniques were used to assess and evaluate the performance of the proposed model. These statistical and graphical methods include root mean square error (RMSE), average absolute percent error (AAPRE), regression coefficient ($R^2$), average percentage relative error (APRE), crossplot, and error distribution curves. Additionally, the performance of previously published well-known correlations and EOSs was investigated and compared to the proposed model. references.

## 2. Data Acquisition

The reliability of any intelligent model is dependent on the data bank that has been utilized during the training and testing stages of model development. Here, various range of pressure, temperature, and gas composition conditions were used to ensure the development of a valid model that can determine the gas compressibility factor. Two dimensionless parameters of pseudo-reduced for temperature and pressure are defined to be used in the developed model for predicting gas compressibility. These parameters are calculated from the current pressure and temperature, and the pseudo critical pressure and temperature described as:

$$P_{\mathrm{pr}} = \frac{P}{P_{\mathrm{pc}}}$$
$$T_{\mathrm{pr}} = \frac{T}{T_{\mathrm{pc}}} \quad (5)$$

where $P_{pc}$ and $T_{pc}$ represent the pseudo critical pressure and temperature, respectively. In addition, $P_{pr}$ and $T_{pr}$ are the pseudo-reduced pressure and temperature.

For a gas with multiple components, $P_{pc}$ and $T_{pc}$ are calculated from the critical temperature and pressure of the individual components as follows:

$$P_{pc} = \sum_{i=1}^{n} y_i P_{ci} \quad (6)$$
$$T_{pc} = \sum_{i=1}^{n} y_i T_{ci} \quad (7)$$

In these equations, $T_{ci}$ and $P_{ci}$ stand for the critical temperature and pressure of component i, and $y_i$ describes the mole fraction of the component *i*.

The data sets in this article were collected from various literature sources [17, 33-38]. Further the Table 1 represents the statistical details of the data bank used. As the table demonstrates, the pressures, temperatures, and compositions comprise a comprehensive range, ensuring that the developed model based on this data set would be a reliable predictor for various types of natural gases at different conditions.

Table 1: The statistical parameters for the data used for z-factor modeling

| Property | Max. | Min. | Average |
|---|---|---|---|
| $C_1$ | 97.48 | 17.27 | 71.18 |
| $C_2$ | 28.67 | 0 | 3.86 |
| $C_3$ | 13.16 | 0 | 1.44 |
| i-$C_4$ | 2.23 | 0 | 0.21 |
| n-$C_4$ | 3.10 | 0 | 0.36 |
| i-$C_5$ | 2.85 | 0 | 0.18 |
| n-$C_5$ | 0.79 | 0 | 0.10 |
| $C_6$ | 2.68 | 0 | 0.20 |
| $C_{7+}$ | 8.17 | 0 | 0.64 |
| $MwC_{7+}$ | 150 | 0 | 50 |
| $Sgc_{7+}$ | 0.90 | 0 | 0.31 |
| $H_2S$ | 73.85 | 0 | 13.92 |
| $CO_2$ | 54.46 | 0 | 6.00 |
| $N_2$ | 25.15 | 0 | 1.83 |
| Pressure ,psi | 7026 | 154 | 2820 |
| Reservoir temperature, °F | 300 | 40 | 147 |
| Tpr | 1.96 | 0.97 | 1.46 |
| Ppr | 10.19 | 0.17 | 3.75 |
| Z-factor | 1.241 | 0.40 | 0.86 |

## 3. Evaluating the model performance

There are multiple statistical parameters that are employed to assess the performance of a model. The parameters that were used in this work include APRE, AAPRE, RMSE, and $R^2$. A simple presentation of the mentioned parameters is presented here:

1. APRE ($E_r$%):

$$E_r = \frac{1}{n}\sum_{i=1}^{n} E_i \qquad (8)$$

where *Ei* stands for the relative variation of predicted value from an experimental value expressed as Percentage Relative Error:

$$E_i = \left[\frac{(O)_{\exp} - (O)_{\text{rep./pred}}}{O_{\exp}}\right] \times 100 \Rightarrow i = 1,2,3,\ldots,n \qquad (9)$$

2. AAPRE:

$$E_a = \frac{1}{n}\sum_{i=1}^{n} | E_i | \qquad (10)$$

3. *RMSE*:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(O_{i\exp} - O_{i\,\text{rep./pred}}\right)^2} \qquad (11)$$

4. $R^2$:

 doi:10.20944/preprints201905.0263.v1



$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(O_{i\,\exp} - O_{\mathrm{i\ rep./pred}}\right)^2}{\sum_{i=1}^{n} \left(O_{\mathrm{i\ rep./pred}} - \overline{O}\right)^2} \quad (12)$$

In these formulas, $\overline{O}$ is the mean value of experimental data output.

Another approach to evaluate the performance of a model and compare it with other models is the usage of graphical error analysis. The graphical approaches used in this study are cross-plot, frequency vs. absolute relative error, error distribution and trend analysis curves. Crossplots are utilized to assess the performance of a model in which the estimated data by the model are plotted against the experimental values, and one can observe the accuracy of the model depending on how close the trend is to a unit-slope line that crosses the origin. Further, the cumulative frequency of data points against the absolute relative error is plotted to quantify the number of data points that can be accurately predicted by the model. Besides, the error distribution curve was plotted to evaluate the error trend of the model when an independent variable is increased. The proximity of data points to the zero-error line tests the precision of that model. Finally, a trend analysis is performed to investigate whether or not the developed model can accurately estimate the trend of gas compressibility factor at different pressures.

## 4. Model Development

GMDH, every two independent parameters are coupled with a quadratic polynomial expression and form $\binom{M}{2}$ new variables as follows:

$$z_i^{GMDH} = ax_i + bx_j + cx_ix_j + dx_i^2 + ex_j^2 + f \quad (14)$$

And the new matrix can be represented by the new variables as follows:

$$v_z = (z_1, z_2, \ldots z_n) \quad (15)$$

In the next step, the least square method is utilized to reduce the difference between the actual data and the model predictions as presented below:

$$\delta_j^2 = \sum_{i=1}^{Nt} (y_i - z_i^{GMDH})^2 \quad j = 1, 2 \ldots, \binom{M}{2} \quad (16)$$

In this equation, the quantity of data points in the training set is shown by Nt. In the next step, the general matrix is written as follows:

$$Y = A^T X \quad (17)$$

Writing the general matrix in the above form helps with offering a general formulation to determine the unknown quadratic polynomial coefficients as shown below:

$$A^T = YX^T (XX^T)^{-1} \quad (18)$$

In the later stage, the data set is divided into subsets of testing and training, the model coefficients are obtained during the training stage, and the testing set is utilized to determine the best combination of the two independent variables based on the following condition:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2019 doi:10.20944/preprints201905.0263.v1



$$\delta_j^2 = \sum_{i=Nt+1}^{N} (y_i - z_i^{GMDH})^2 \prec \varepsilon \quad j = 1,2\ldots,\binom{M}{2} \tag{19}$$

The combined variables will be stored if this criterion is met, otherwise the algorithm eliminates this combination of two variables and the iteration will continue. More information about this modeling approach can be found in our previous work [46].

## 5. Results

A The GMDH has been successfully implemented as an evolutionary intelligent model to predict the natural gas's compressibility. The inputs of the model were gas composition, pressure, and temperature as shown in Table 1. In addition to $C_1$-$C_6$ and $C_{7+}$ components, $H_2S$, $CO_2$, and $N_2$ gases were considered as the components. The pressure and temperature values are used to calculate the pseudo-reduced pressure and temperature as discussed previously. During the data acquisition stage, a wide range of input parameters was considered as shown in Table 1. The pressure ranges from 154 to 7026 Psia, reservoir temperature ranges from 40 to 300 $^0$F, and the compressibility factor values cover a wide range of 0.4 to 1.24. The distribution of input and output data is illustrated in Figure 1. In addition to the data distribution, normal distribution curves were plotted in this figure. Pseudo reduced pressure and temperature were the chosen parameters for this figure due to their impact on any reservoir fluid properties. As can be seen, all three data sets follow a relatively normal shape, especially the gas compressibility factor data. The mean values for the pseudo-reduced temperature, pseudo-reduced pressure, and gas compressibility factor are 1.5, 3.9 and 0.9, respectively. The bin size in all cases is 40. A schematic flowchart of the model is illustrated in Figure 2. In order to assess the accuracy of the developed model, various statistical and graphical methods such as average absolute relative error and error distribution curve were employed as will be discussed in this section.

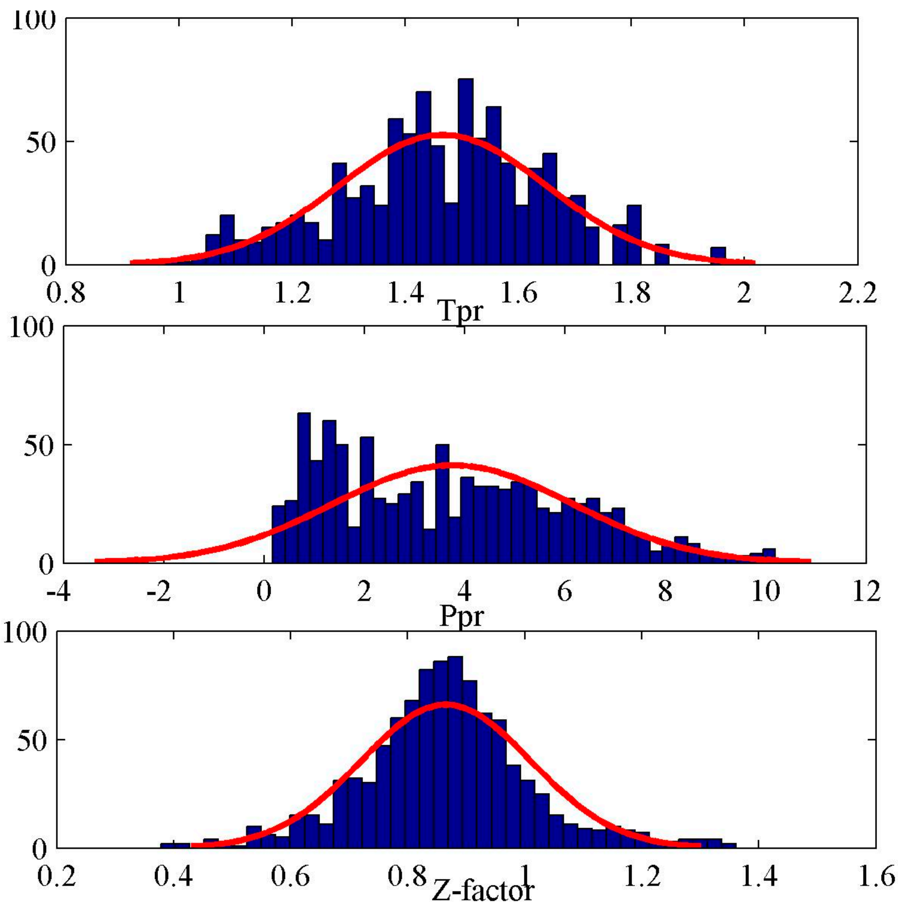


Fig. 1: Distribution of the input and output data sets.

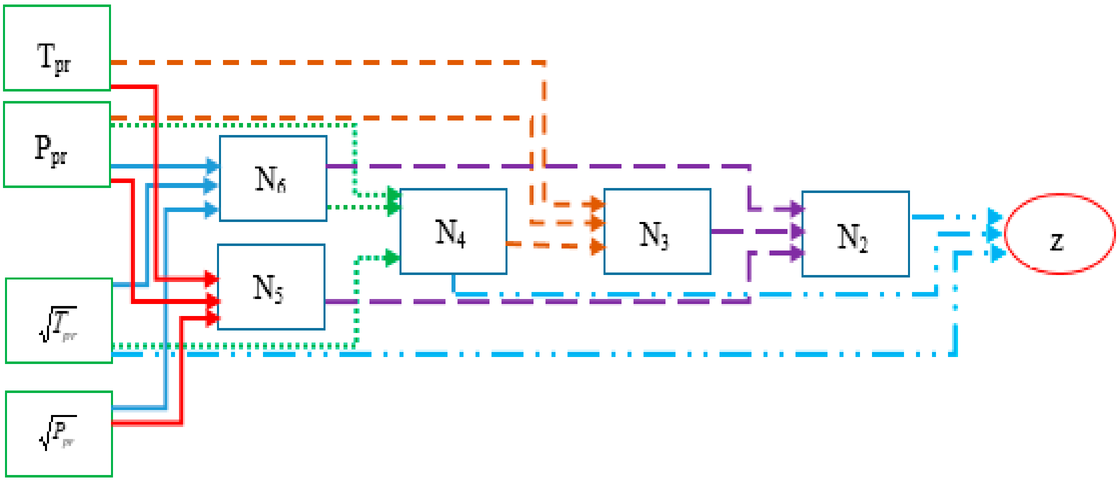


Fig. 2: A schematic flowchart of the proposed GMDH for predicting z-factor

After optimizing the model, the genome and nodal expressions were obtained as follows:

$$z = -0.268213 - 1.09959 N_2 \sqrt{T_{pr}} + 0.434222 T_{pr}^2 - 1.02209 N_4 + 0.397836 N_4^2 + 2.58207 N_2 \tag{20}$$

$$N_6 = 2.28818 + 1.46235 P_{pr} - 1.05874 P_{pr} \sqrt{T_{pr}} + 3.50204 \sqrt{T_{pr} P_{pr}} - 0.667833 T_{pr} - 4.84879 \sqrt{P_{pr}} \tag{21}$$

$$N_5 = 1.88105 - 0.466011 T_{pr} P_{pr} + 1.5661 T_{pr} \sqrt{P_{pr}} - 0.265557 T_{pr}^2 + 0.863307 P_{pr} - 2.90311 \sqrt{P_{pr}} \tag{22}$$

$$N_4 = -2.14451 - 0.0929261 P_{pr} N_6 + 0.0120013 P_{pr}^2 - 4.51872 N_6 \sqrt{T_{pr}} + 1.89131 T_{pr} + 5.85344 N_6 \tag{23}$$

$$N_3 = -0.461086 + 0.0344126 T_{pr}^2 - 0.048752 P_{pr} + 0.00674994 P_{pr}^2 + 2.4872 N_4 - 1.11543 N_4^2 \tag{24}$$

$$N_2 = -0.182453 - 17.0122 N_5 + 18.4918 N_5 N_3 + 17.1725 N_6 - 18.6856 N_6 N_3 + 1.22045 N_3 \tag{25}$$

where $T_{pr}$ and $P_{pr}$ represent the pseudo-reduced temperature and pressure, respectively, and $N_2$-$N_6$ represent the virtual variables or nodes of the model. It should be noted that the equations are second order polynomials that can be easily used to estimate the gas compressibility factor. As can be seen in the first equation, gas compressibility factor can be calculated by having the pseudo-reduced temperature, and the virtual parameters $N_2$ and $N_4$ and. $N_4$ can be calculated by knowing pseudo-reduced pressure and by calculating the virtual parameter $N_6$. $N_6$ is a simple function of temperature and pseudo-reduced pressure and can be directly calculated. In order to calculate $N_2$, the virtual parameters $N_3$ and $N_5$ are needed in addition to $N_6$. $N_3$ can be calculated after calculating the virtual parameter $N_4$. Finally, $N_5$ can be directly calculated similarly to $N_6$ by having pseudo-reduced pressure and temperature. Statistical quality measures and graphical techniques were applied to assess the performance of the developed model, as well as to compare the model with the results of five of the well-known EOSs namely van der-Waals [47] EOS, Lawal-Lake-Silberberg [48] EOS, Peng-Robinson [10] EOS, Soave-Redlich-Kwong [13] EOS, and Patel-Teja [12] EOS as well as ten empirical correlations namely, Dranchuk-Purvis-Robinson [49], Dranchuk-Abou-Kassem [18], Beggs-Brill [50], Shell Oil Company [6], Gopal [51], Hall-Yarborough [14], Sanjari and Lay [22], Heidaryan *et al*. [20], Azizi *et al*. [21], and Kamari et al. [52]. Table 2 presents the results of a statistical assessment of the GMDH model and previously published correlations and EOSs. As can be observed in this table, the developed model demonstrates the most accurate performance compared to other models when comparing the root mean square, the average absolute percentage relative error, and regression coefficient. The Hall Yarborough correlation was found to be the next most accurate model followed by the Patel-Teja EOS model based on the statistical information presented in this table.

Table 2: Statistical error analyses for the correlations, EOSs, and the GMDH model

| Method | AARE % | RMSE | $R^2$ | AAPRE % |
|---|---|---|---|---|
| Dranchuk-Abou-Kassem Correlation | 4.21 | 0.0992 | 0.5748 | 8.18 |
| Kamari et al. | -1.10 | 0.0629 | 0.8248 | 5.92 |
| Dranchuk-Purvis-Robinson Correlation | 4.66 | 0.0555 | 0.9064 | 4.77 |
| Hall-Yarborough Correlation | 1.46 | 0.0429 | 0.8924 | 3.59 |
| Beggs-Brill Correlation | 4.08 | 0.0785 | 0.7763 | 6.53 |
| Shell Oil Company Correlation | 4.36 | 0.0596 | 0.7819 | 6.24 |
| Gopal Correlation | 6.12 | 0.0910 | 0.7371 | 6.26 |
| Azizi *et al*. Correlation | 4.26 | 0.0792 | 0.7723 | 6.25 |
| Heidaryan *et al*. Correlation | 3.61 | 0.0762 | 0.7788 | 5.80 |
| Sanjari and Lay Correlation | 0.66 | 0.0697 | 0.8119 | 5.67 |
| van der-Waals EOS | 0.31 | 0.0696 | 0.7711 | 6.42 |
| Peng-Robinson EOS | -5.34 | 0.0599 | 0.8916 | 6.10 |
| Patel-Teja EOS | -1.18 | 0.0447 | 0.8806 | 4.15 |
| Soave-Redlich-Kwong EOS | -3.14 | 0.0493 | 0.8938 | 4.82 |
| Lawal-Lake-Silberberg EOS | -2.66 | 0.0453 | 0.8942 | 4.43 |
| The GMDH model, Train | -0.21 | 0.0351 | 0.9177 | 2.89 |
| The GMDH model, Test | -0.10 | 0.0340 | 0.9171 | 2.85 |
| The GMDH model, Total | -0.19 | 0.0349 | 0.9176 | 2.88 |

Noticing the APRE and AAPRE values of the models in this study reveals that, while the APRE value of some of the models is not smaller than those of others, their AAPRE is. The definition of APRE can be used to explain this observation. As a matter of fact, APRE (Average Percentage Relative Error) is a relative value and should not be used by itself to approve or reject a model. For instance, an APRE value close to zero would be obtained if half of the data points are overestimated by a model and the remaining half are underestimated, which would in return present a false assessment of the model performance. On the other hand, if most of the data points are estimated accurately by a model and the remaining few data points are either underestimated or overestimated, a positive or negative APRE value would be obtained, respectively, which again cannot testify the model performance. As an example, Van der Waals model has a much smaller APRE value than that of Dranchuk-Purvis-Robinson. However, it is considered to be a less accurate model than the latter one

Another way to illustrate the performance of the models and to compare them more comprehensively is by using graphical analyses. Several graphical analyses were employed in this work to evaluate the performance of the most popular models from both quantity and quality standpoints. Figure 3 presents the cross-plot of the developed model in which the calculated data is plotted against the measured data for both testing and training sets. The location of the majority of the data points on the y=x line supports the accurate predictions of the developed GMDH model. This is true for both of the training and testing data sets. The error distribution of the model is shown in Figure 4. As the figure shows, most of the data points are near the zero percent error line. This indicates that the GMDH model does not have a systematic error trend as the gas compressibility factor increases (most of the previously published models suffer from an error trend). The figure indicates that the error of the testing set is smaller than the training set. The distribution of the relative error of predictions is plotted in Figure 5. In addition to the distribution of data points, the normal distribution is plotted indicated by the red line. The figure indicates that the relative error of predictions accurately follows the normal distribution and that most of the data points (predictions) have a relative error close to zero demonstrating the accuracy of the mode in predicting gas compressibility factor. Figures 6 and 7 depict the statistical information reported in Table 2 for a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2019 doi:10.20944/preprints201905.0263.v1



graphical demonstration of the performance of the models. Further, these figures indicate that models with a smaller RMSE do not necessarily have a smaller average absolute percent relative error. This means that when performing statistical analyses, care should be taken to avoid the misinterpretation of the results by focusing on only one statistical parameter.

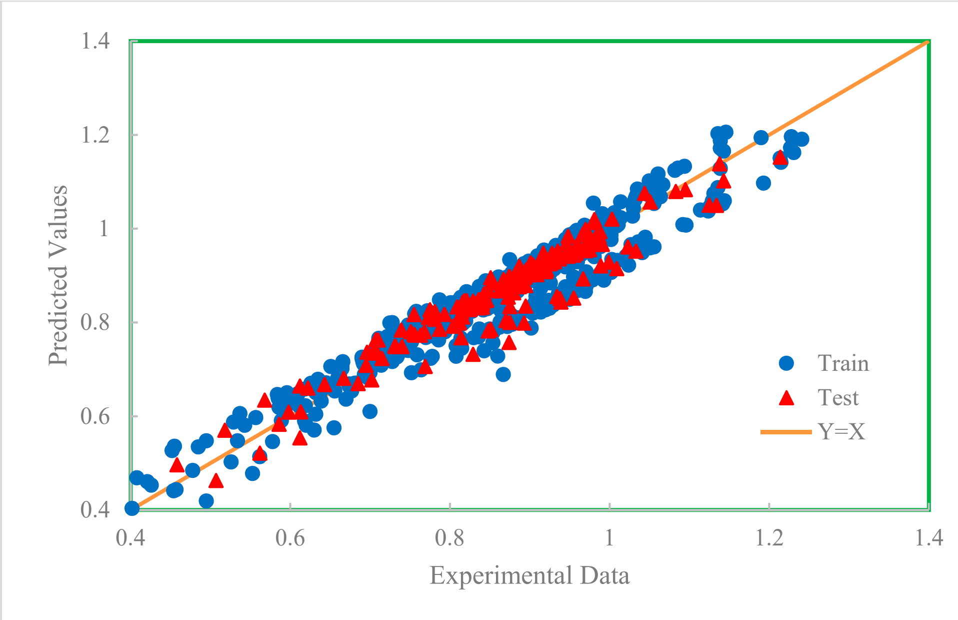


Fig. 3: Crossplot of the predicted z-factors versus experimental data

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2019 doi:10.20944/preprints201905.0263.v1



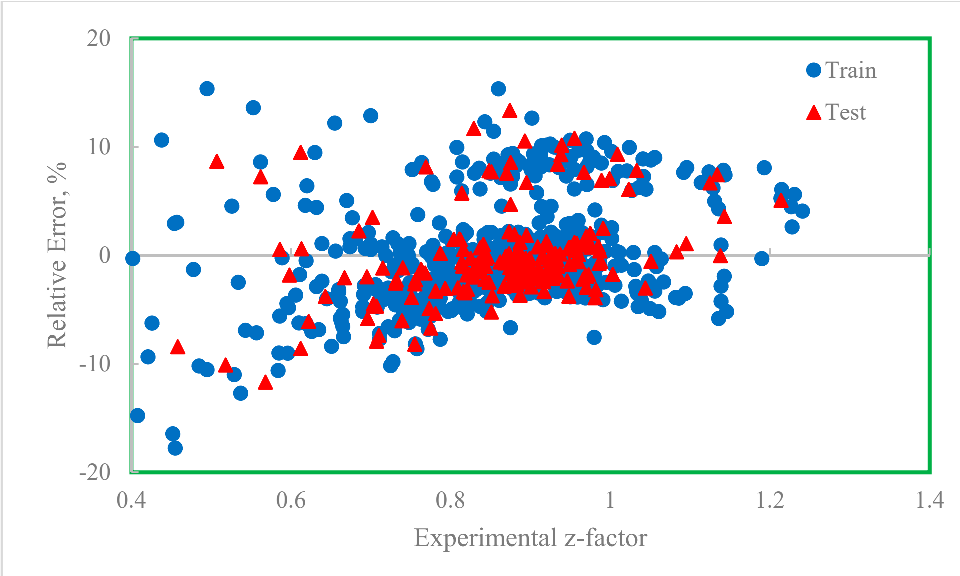


Fig. 4: Error distribution curve of the proposed GMDH model versus experimental z-factor

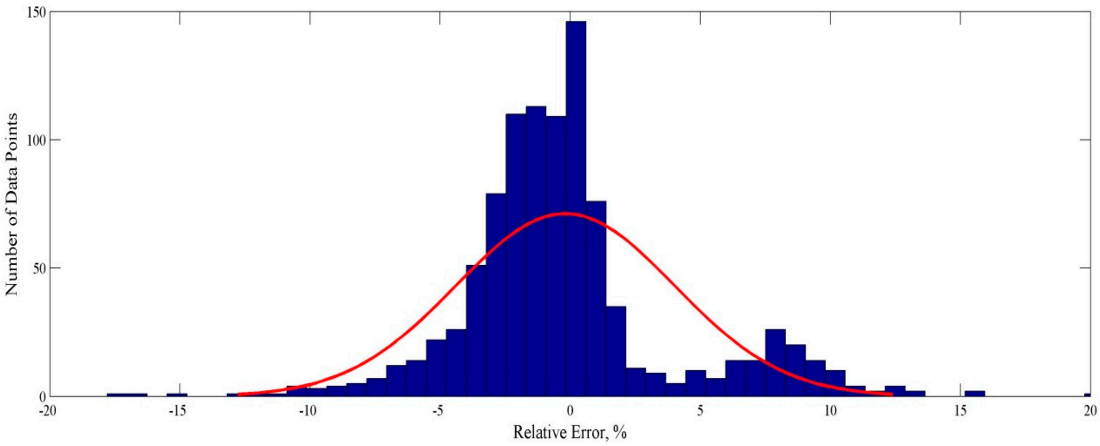


Fig. 5: Distribution of relative error of the proposed GMDH model

 doi:10.20944/preprints201905.0263.v1



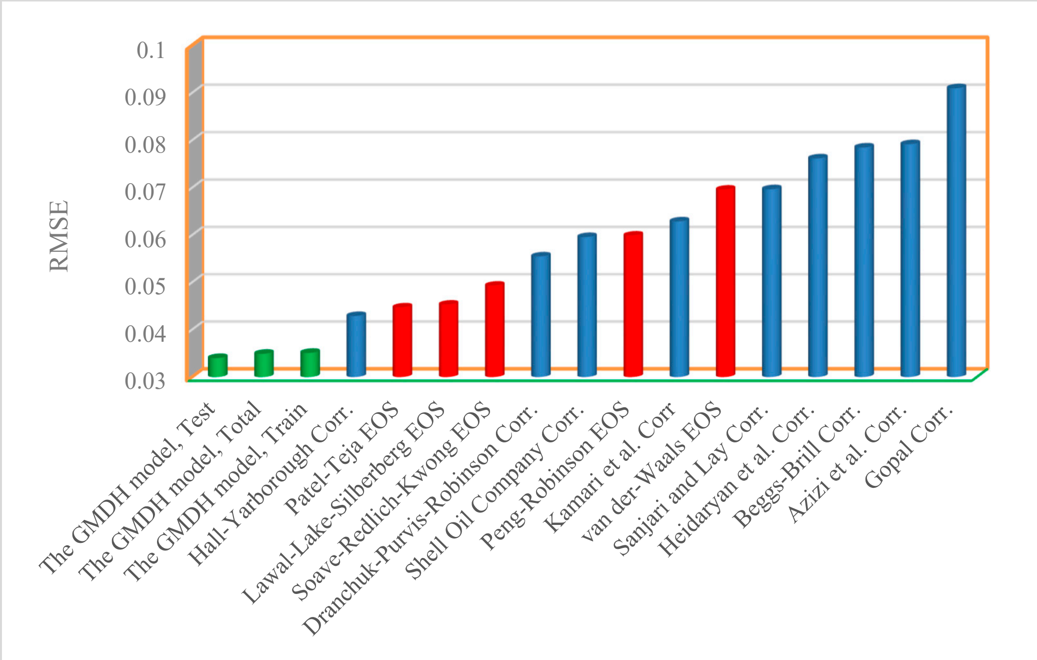


Fig. 6: RMSE of the existing correlations, EOSs, and GMDH models for predicting z-factor

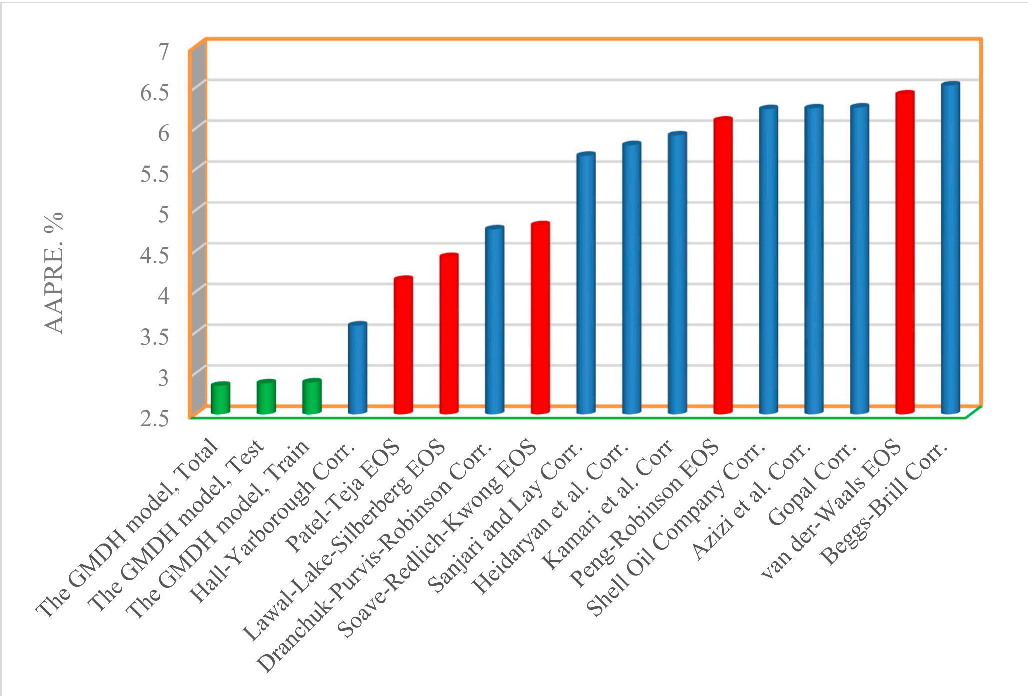


Fig. 7: Average absolute percent relative error of the existing correlations, EOSs, and GMDH models for predicting z-factor

As can be seen in Figure 7, Beggs-Brill, Van der Waals, Gopal, Azizi, and Shell Oil Company models have the highest AAPRE values meaning that their predictions are less accurate than other models. It can be observed that the proposed model in this work is distinctly superior to the previously published models by an AAPRE of only 2.89%, 2.85%, and 2.88% for the training, testing and total data sets, respectively. A better vision of the superiority of the suggested model can be observed in Figure 6 in which the developed model has a much smaller RMSE value than all of the published models.

 doi:10.20944/preprints201905.0263.v1



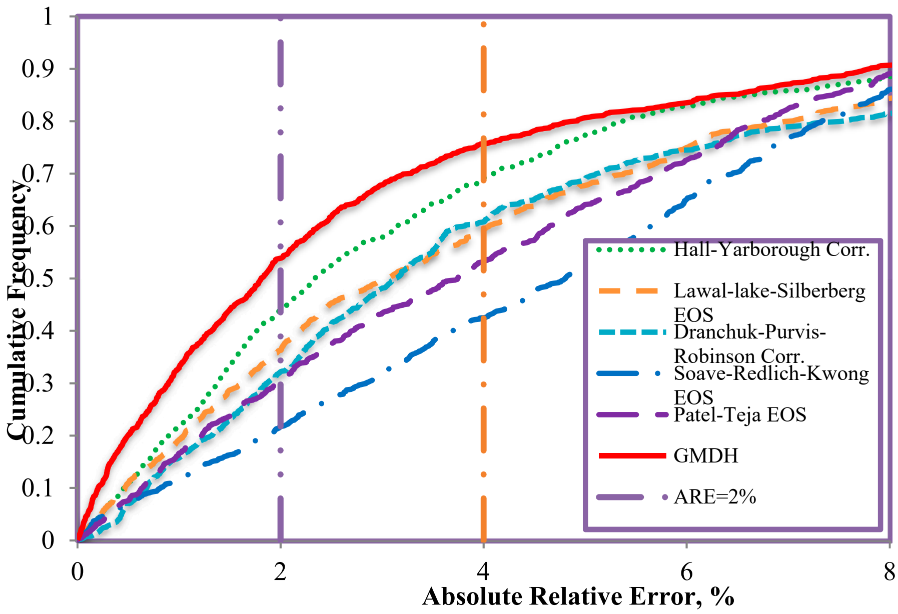


Fig 8. Cumulative frequency versus absolute relative error for existing models as well as the proposed GMDH model for predicting gas compressibility factor

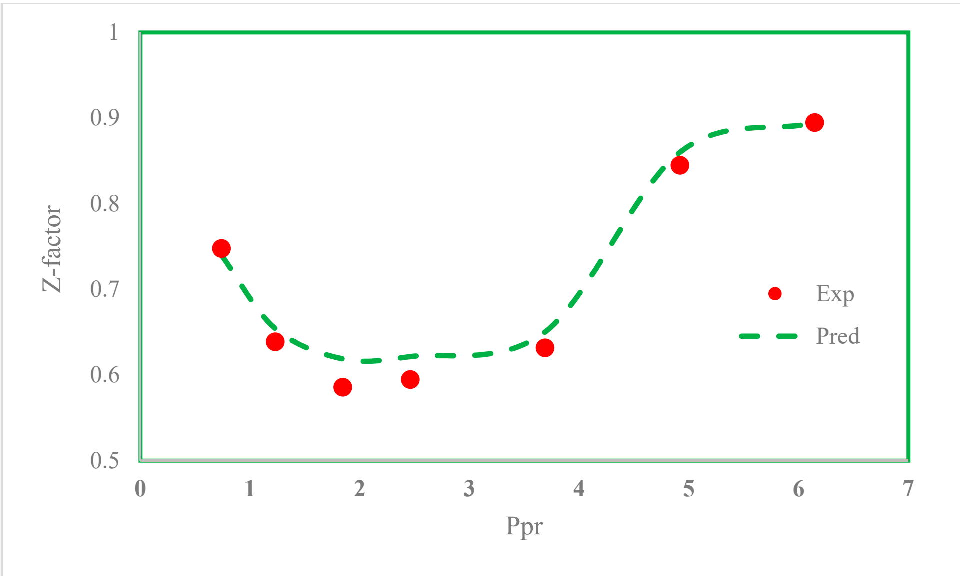


Fig. 9: Variation of z-factor as a function of Ppr at Tpr=1.19 for a gas sample

In order to compare the accuracy of the developed GMDH model with the previously published correlations, cumulative frequency plots of the models are illustrated in Figure 8. This figure helps to achieve a better quantitative evaluation of the developed model when compared to the previously published models. This figure shows that the GMDH model can predict nearly 52% of the data with an absolute relative error of 2%. More importantly, the model can predict 75% of the data points with an absolute relative error of only 4%. It can be observed that the model estimates the gas compressibility factor with the smallest absolute relative error for any number of data points included in the predictions. This verifies the consistency of the developed GMDH model in accurately estimating gas compressibility factor within a wide range of reservoir conditions. The figure also shows that Hall-Yarborough correlation is the next accurate model that can estimate gas compressibility factor with small error values when any number of data points are considered. Another important finding from this figure is the comparison between Lawal-lake-Silberberg EOS model and Dranchuk-Purvis-Robinson correlation. While Lawal-lake-Silberberg EOS is more accurate in estimating gas compressibility factor for up to 50% of the data points, Dranchuk-Purvis-Robinson correlation becomes the superior model when more than 50% of the data points are included. This trend changes again when 80% or more of the data points are included.

Finally, the predicted compressibility values by the GMDH model were plotted in Figure 9 against the experimental data at different pseudo-reduced pressure values and constant pseudo-reduced temperature of 0.7 to confirm the capability of the developed model in accurately estimating natural gas compressibility factor at different conditions. The experimental trend in this figure indicates that by increasing the pressure, gas compressibility factor first decreases and then increases. This trend has been accurately estimated by the developed GMDH model as shown by the dashed line in the same figure.

## 5. Conclusions

In this work, GMDH was used to predict the compressibility factor of natural gas. The results showed that the developed model's performance is consistently more accurate than the previously published well-known correlations at different temperature and pressure conditions. This was confirmed by measuring the root mean square, average absolute percent relative error, and regression coefficient to be 0.03, 2.88%, and 0.92, respectively. The Hall Yarborough correlation was determined as the second most accurate correlation for estimating the natural gas compressibility factor. In addition, the error distribution curve analysis indicated that the presented model in this study does not have an error trend when predicting very low and very high compressibility factor values. The experimental trend of gas compressibility factor showed that by increasing the pressure, the Z-factor first decreases and then increases. This trend was perfectly shown by the developed GMDH model in this work.

The results of this study show that most of the previously published correlations have been developed based on limited data sets and are only able to estimate the compressibility factor within limited ranges of pressure and temperature conditions. While, the proposed GMDH model can accurately predict the natural different gas compressibility factor at low and high temperature and pressure conditions.

**Conflicts of Interest:** Authors declare conflicts of interest or state.

## References

1. Kamari, A., et al., *Prediction of sour gas compressibility factor using an intelligent approach.* Fuel processing technology, 2013. **116**: p. 209-216.

2. Shateri, M., et al., *Application of Wilcoxon generalized radial basis function network for prediction of natural gas compressibility factor.* Journal of the Taiwan Institute of Chemical Engineers, 2015. **50**: p. 131-141.

3. Firoozabadi, A., *Thermodynamics of hydrocarbon reservoirs*. 1999: McGraw-Hill New York.

4. Elsharkawy, A.M., *Efficient methods for calculations of compressibility, density and viscosity of natural gases.* Fluid Phase Equilibria, 2004. **218**(1): p. 1-13.

5. Kamyab, M., et al., *Using artificial neural networks to estimate the z-factor for natural hydrocarbon gases.* Journal of Petroleum Science and Engineering, 2010. **73**(3-4): p. 248-257.

6. Kumar, N., *Compressibility factors for natural and sour reservoir gases by correlations and cubic equations of state.* 2004, Texas Tech University.

7. Sanjari, E. and E.N. Lay, *Estimation of natural gas compressibility factors using artificial neural network approach.* Journal of Natural Gas Science and Engineering, 2012. **9**: p. 220-226.

8. Standing, M.B. and D.L. Katz, *Density of natural gases.* Transactions of the AIME, 1942. **146**(01): p. 140-149.

 doi:10.20944/preprints201905.0263.v1



9. Van Der Waals, J.D. and J.S. Rowlinson, *On the continuity of the gaseous and liquid states*. 2004: Courier Corporation.

10. Peng, D.-Y. and D.B. Robinson, *A new two-constant equation of state.* Industrial & Engineering Chemistry Fundamentals, 1976. **15**(1): p. 59-64.

11. Lawal, A., *Application of the Lawal-Lake-Silberberg equation-of-state to thermodynamic and transport properties of fluid and fluid mixtures*, in *Technical Report TR-4-99*. 1999, Department of Petroleum Engineering, Texas Tech University Lubbock.

12. Patel, N.C. and A.S. Teja, *A new cubic equation of state for fluids and fluid mixtures.* Chemical Engineering Science, 1982. **37**(3): p. 463-473.

13. Soave, G., *Equilibrium constants from a modified Redlich-Kwong equation of state.* Chemical engineering science, 1972. **27**(6): p. 1197-1203.

14. Hall, K.R. and L. Yarborough, *A new equation of state for Z-factor calculations.* Oil Gas J, 1973. **71**(25): p. 82.

15. Beggs, D.H. and J.P. Brill, *A study of two-phase flow in inclined pipes.* Journal of Petroleum technology, 1973. **25**(05): p. 607-617.

16. Dranchuk, P., R. Purvis, and D. Robinson. *Computer calculation of natural gas compressibility factors using the Standing and Katz correlation*. in *Annual Technical Meeting*. 1973. Petroleum Society of Canada.

17. Simon, R. and J.E. Briggs, *Application of Benedict-Webb-Rubin equation of state to hydrogen sulfide-hydrocarbon mixtures.* AIChE Journal, 1964. **10**(4): p. 548-550.

18. Dranchuk, P. and H. Abou-Kassem, *Calculation of Z factors for natural gases using equations of state.* Journal of Canadian Petroleum Technology, 1975. **14**(03).

19. Gopal, V., *Gas z-factor equations developed for computer.* Oil and Gas Journal, 1977. **75**(8): p. 8-13.

20. Heidaryan, E., J. Moghadasi, and M. Rahimi, *New correlations to predict natural gas viscosity and compressibility factor.* Journal of Petroleum Science and Engineering, 2010. **73**(1-2): p. 67-72.

21. Azizi, N., R. Behbahani, and M. Isazadeh, *An efficient correlation for calculating compressibility factor of natural gases.* Journal of Natural Gas Chemistry, 2010. **19**(6): p. 642-645.

22. Sanjari, E. and E.N. Lay, *An accurate empirical correlation for predicting natural gas compressibility factors.* Journal of Natural Gas Chemistry, 2012. **21**(2): p. 184-188.

23. Hajirezaie, S., et al., *A smooth model for the estimation of gas/vapor viscosity of hydrocarbon fluids.* Journal of Natural Gas Science and Engineering, 2015. **26**: p. 1452-1459.

24. Dargahi-Zarandi, A., et al., *Modeling gas/vapor viscosity of hydrocarbon fluids using a hybrid GMDH-type neural network system.* Journal of Molecular Liquids, 2017. **236**: p. 162-171.

25. Hajirezaie, S., et al., *Development of a robust model for prediction of under-saturated reservoir oil viscosity.* Journal of Molecular Liquids, 2017. **229**: p. 89-97.

26. Hemmati-Sarapardeh, A., et al., *On the evaluation of density of ionic liquid binary mixtures: modeling and data assessment.* Journal of Molecular Liquids, 2016. **222**: p. 745-751.

27. Kamari, A., A. Safiri, and A.H. Mohammadi, *Compositional model for estimating asphaltene precipitation conditions in live reservoir oil systems.* Journal of Dispersion Science and Technology, 2015. **36**(3): p. 301-309.

28. Rostami, A., et al., *New empirical correlations for determination of Minimum Miscibility Pressure (MMP) during N2-contaminated lean gas flooding.* Journal of the Taiwan Institute of Chemical Engineers, 2018. **91**: p. 369-382.

29. Kamari, A., et al., *Characterizing the CO2-brine interfacial tension (IFT) using robust modeling approaches: A comparative study.* Journal of Molecular Liquids, 2017. **246**: p. 32-38.

30. Hajirezaie, S., et al., *Numerical simulation of mineral precipitation in hydrocarbon reservoirs and wellbores.* Fuel, 2019. **238**: p. 462-472.

31. Dashtian, H., et al., *Convection-diffusion-reaction of CO2-enriched brine in porous media: A pore-scale study.* Computers & Geosciences, 2019.

32. Hajirezaie, S., S.A. HEMMATI, and S. Ayatollahi, *Soft Computing Model for Prediction of Sour and Natural Gas Viscosity.* 2014.

33. Robinson Jr, R. and R. Jacoby, *Better compressibility factors.* Hydrocarbon Processing, 1965. **44**(4): p. 141-145.

34. Buxton, T.S. and J.M. Campbell, *Compressibility factors for lean natural gas-carbon dioxide mixtures at high pressure.* Society of Petroleum Engineers Journal, 1967. **7**(01): p. 80-86.

35. McLeod, W.R., *Applications of molecular refraction to the principle of corresponding states*. 1968, The University of Oklahoma.

36. Wichert, E. and K. Aziz, *Calculate Zs for sour gases.* Hydrocarbon Processing, 1972. **51**(5): p. 119-&.

37. Whitson, C.H. and S.B. Torp. *Evaluating constant volume depletion data*. in *SPE Annual Technical Conference and Exhibition*. 1981. Society of Petroleum Engineers.

38. Elsharkawy, A.M. and S.G. Foda, *EOS simulation and GRNN modeling of the constant volume depletion behavior of gas condensate reservoirs.* Energy & Fuels, 1998. **12**(2): p. 353-364.

39. Ivakhnenko, A.G., *Polynomial theory of complex systems.* IEEE transactions on Systems, Man, and Cybernetics, 1971(4): p. 364-378.

40. Shankar, R., *The GMDH.* Master's Thesis. University of Delaware, 1972.

41. Sawaragi, Y., et al., *Statistical prediction of air pollution levels using non-physical models.* Automatica, 1979. **15**(4): p. 441-451.

42. Madala, H.R., *Inductive Learning Algorithms for Complex Systems Modeling: 0*. 2018: cRc press.

43. Atashrouz, S., E. Amini, and G. Pazuki, *Modeling of surface tension for ionic liquids using group method of data handling.* Ionics, 2015. **21**(6): p. 1595-1603.

44. Atashrouz, S., M. Mozaffarian, and G. Pazuki, *Modeling the thermal conductivity of ionic liquids and ionanofluids based on a group method of data handling and modified Maxwell model.* Industrial & Engineering Chemistry Research, 2015. **54**(34): p. 8600-8610.

45. Ivakhnenko, A. and G. Krotov, *Multiplicative and additive nonlinear Gmdh algorithm with factor degree optimization.* AVTOMATIKA, 1984(3): p. 13-18.

46. Hemmati-Sarapardeh, A. and E. Mohagheghian, *Modeling interfacial tension and minimum miscibility pressure in paraffin-nitrogen systems: Application to gas injection processes.* Fuel, 2017. **205**: p. 80-89.

47. Van der Waals, J.D., et al., *The Continuity of the Liquid and Gaseous States*. 1873.

*Preprints* (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 May 2019 doi:10.20944/preprints201905.0263.v1



48. Lawal, A., *Application of the Lawal-Lake-Silberberg Equation-of-State to Thermodynamic and Transport Properties of Fluid and Fluid Mixtures*. 1999, Technical Report TR-4-99, Department of Petroleum Engineering, Texas Tech University, Lubbock, Texas.

49. Dranchuk, P.M., R.A. Purvis, and D.B. Robinson. *Computer calculation of natural gas compressibility factors using the Standing and Katz correlation*. in *Annual Technical Meeting*. 1974. Petroleum Society of Canada.

50. Brill, J. and H. Beggs, *University of Tulsa:" Two phase flow in pipes,"*. Intercomp Course, The Hague, 1973.

51. Gopal, V., *Gas z-factor equations developed for computer*. Oil and Gas Journal (Aug. 8, 1977), 1977: p. 58-60.

52. Kamari, A., et al., *A corresponding states-based method for the estimation of natural gas compressibility factors.* Journal of Molecular Liquids, 2016. **216**: p. 25-34.