Estimation of dynamic viscosity of natural gas based on genetic programming methodology

Danial Abooali a, Ehsan Khamehchi b, *

a School of Chemical Engineering, Iran University of Science and Technology (IUST), Postal Box 16765-163, Tehran, Iran
b Faculty of Petroleum Engineering, Amirkabir University of Technology (Tehran Polytechnic), Hafez Avenue, 15914 Tehran, Iran

A R T I C L E   I N F O
Article history:
Received 29 August 2014
Received in revised form 1 November 2014
Accepted 4 November 2014
Available online

Keywords:
Natural gas
Dynamic viscosity
Correlation
Genetic programming

A B S T R A C T

Investigating the behavior of natural gas can contribute to a detailed understanding of hydrocarbon reservoirs. Natural gas, alone or in association with oil in reservoirs, has a large impact on reservoir fluid properties. Thus, having knowledge about gas characteristics seems to be necessary for use in estimation and prediction purposes. In this project, dynamic viscosity of natural gas ($\mu_d$), as an important quantity, was correlated with pseudo-reduced temperature ($T_{pr}$), pseudo-reduced pressure ($P_{pr}$), apparent molecular weight ($M_a$) and gas density ($\rho_g$) by operation of the genetic programming method on a large dataset including 1938 samples. The squared correlation coefficient ($R^2$), average absolute relative deviation percent (AARD%) and average absolute error (AAE) are 0.999, 2.55% and 0.00084 cp, respectively. The final results show that the obtained simple-to-use model can predict viscosity of natural gases with high accuracy and confidence.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Natural gas reserves are mainly found in gas reservoirs along with the gas condensates or in oil reservoirs, in the form of separated gas (such as gas cap) or dissolved in liquid crude oil. A natural hydrocarbon gas is usually defined as a homogenous mixture of light hydrocarbon and non-hydrocarbon components and also a few heavy molecules, with low viscosity and density (Ahmed, 2010). Different specifications of natural gas, especially thermodynamic properties, are the main tools discussing the conditions of gas in natural reservoirs. Many characteristics of gas extraction equipment from underground fields and reserves, such as wellhead and inside-well instruments, separators, surface operation devices, etc., are dependent on the properties of reservoir fluid, particularly gas compounds due to more sensitivity and permeation.

Knowledge about the constituents of natural gas, temperature, pressure, acidity, viscosity, density and other pressure–volume–temperature (PVT) parameters is very necessary in order to solve different problems in reservoir engineering and improve production processes performance. Some of these parameters can be measured directly, but many of them are found by means of equations of state or oil PVT analysis. Repeating laboratory measurements in various conditions are usually costly and time consuming; but the application of accurate correlations is economically advantageous and has been taken into great consideration due to their many uses in reservoir simulators. One of the vital applications of the correlations and formulas is the prediction of future properties of hydrocarbon reservoirs.

In this project, a new model was developed for estimation of natural gas dynamic viscosity ($\mu_d$) on the basis of four common gas mixture parameters: pseudo-reduced temperature ($T_{pr}$), pseudo-reduced pressure ($P_{pr}$), apparent molecular weight ($M_a$) and gas density ($\rho_g$) by operation of the genetic programming (GP) based regression method.

Viscosity is one of the most important specifications in natural gas engineering calculations. It is a fundamental property to investigate the fluid flow in different situations, particularly porous spaces. In study of the productivity rate of hydrocarbon fluids in underground reservoirs, shear tension, friction forces and especially in the industries producing and filtering fuels for combustion processes and cycles, it is extremely necessary to modify viscosity.

Density is defined as the mass of unit volume. Specific gravity ($\gamma_g$) of a gas is the ratio of gas density to density of air in which both densities are at the same temperature and pressure. In almost all cases, the density of air and gas used to calculate Specific gravity are in standard condition (i.e. $T = 60^\circ F$ and $P = 1$ atm).
Reduced temperature \( (T_r) \) is defined as the ratio of temperature of a pure compound to its critical temperature (i.e. \( T/T_c \)). For a homogenous mixture of components like natural gas, there is also a critical temperature which its value is dependent on the critical temperatures of the mixture compounds and the composition. There is another similar property for a mixture which is called “pseudo-critical temperature \( (T_{pc}) \)” and is defined as follows:

\[
T_{pc} = \sum_{i=1}^{n} x_i T_{ci}
\]

\( x_i \) and \( T_{ci} \) are molar fraction and critical temperature of compound \( i \) in the mixture. Similarly, every mixture has a critical pressure and also “pseudo-critical pressure \( (P_{pc}) \)” defined bellow:

\[
P_{pc} = \sum_{i=1}^{n} x_i P_{ci}
\]

\( P_{ci} \) is critical pressure for component \( i \) in the mixture. Pseudo-reduced pressure \( (P_{pr}) \) is pressure of the mixture components, respectively. Pseudo-critical pressure are the molar average of critical temperature and critical pressures of the mixture compounds and the composition.

In a new project, Heidaryan et al. (2013) obtained another correlation similar to Dempsey’s model. The general form of their correlation is as follows:

\[
\frac{\mu}{\mu_1} = \exp \left( \frac{1 + A_1 T_{pr} + A_2 P_{pr} + A_3 \left( 2T_{pr}^2 - 1 \right) + A_4 \left( 2P_{pr}^2 - 1 \right) + A_5 \left( 2T_{pr}^3 - 3T_{pr} \right) + A_6 \left( 4T_{pr}^3 - 3P_{pr} \right)}{A_7 T_{pr} + A_8 P_{pr} + A_9 \left( 2T_{pr}^2 - 1 \right) + A_{10} \left( 2P_{pr}^2 - 1 \right) + A_{11} \left( 4T_{pr}^3 - 3P_{pr} \right)} \right)
\]

\( \mu_p, T_{pr} \) and \( P_{pr} \) are gas dynamic viscosity, pseudo-reduced temperature and pseudo-reduced pressure, respectively.

\( \mu_1 \) is atmospheric gas viscosity modeled and proposed by Standing (1977).

There are also some other methods using the artificial neural network (ANN) technique to predict viscosity of gases (AlQuaishi and Shokir, 2011; Baniasadi and Khamenehi, 2014).

Different mathematical and statistical algorithms and methodologies can be applied in order to develop computational correlations. Genetic programing (GP) is one of the mathematical powerful methodologies with various applications which are usually used for optimizing and prediction purposes. There are some other methods such as artificial neural network (ANN), generalized regression neural networks (GRN), particle swarm optimization (PSO) (Hamedi et al., 2011), imperialist competitive algorithm (ICA), etc., applied in this area (Khishvand and Khamenehi, 2012; Kashidi et al., 2010; Khamenehi et al., 2009).

In the present project, genetic programing approach was applied to correlate natural gas viscosity with pseudo-reduced temperature, pseudo-reduced pressure, specific gravity and so on.
molecular weight. Thus, a genetic programming based multi-gene symbolic regression algorithm called “GPTIPS” (Searson et al., 2010) was applied for producing the viscosity model.

The main target of this project was to produce an accurate model for natural gas viscosity which can be applied in confident predictions. Using a large dataset including naturally real ranges of viscosity, pseudo-reduced temperature, pseudo-reduced pressure, specific gravity and molecular weight, has enhanced the applicability and accuracy of the developed model. Furthermore, the application of genetic programming for developing the gas viscosity correlation seems to be novel and innovative.

2. Materials and methods

2.1. Dataset

In this project, a large dataset including 1938 sets of experimental data of natural gas mixtures was extracted from Lee et al. (1966) research. Total data contain dynamic viscosity, pseudo-reduced temperature, pseudo-reduced pressure, apparent molecular weight, gas density and gas specific gravity. The ranges of parameters in the dataset of present study and Heidaryan et al. (2013) are shown in Table 1 in comparison with the data used in Heidaryan et al. (2013).

![Table 1](image)

Table 1. The ranges of parameters in the dataset of present study and Heidaryan et al. (2013).

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Present study</th>
<th>Heidaryan et al. (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo-reduced temperature</td>
<td>0.774104</td>
<td>1.196</td>
</tr>
<tr>
<td>Pseudo-reduced pressure</td>
<td>0.022059</td>
<td>0.02</td>
</tr>
<tr>
<td>Apparent molecular weight</td>
<td>16.043</td>
<td>53.915</td>
</tr>
<tr>
<td>Gas density (g/mL)</td>
<td>0.00038</td>
<td>0.7580</td>
</tr>
<tr>
<td>Gas specific gravity</td>
<td>0.382544</td>
<td>11.928979</td>
</tr>
<tr>
<td>Dynamic viscosity (cp)</td>
<td>0.0091</td>
<td>0.228</td>
</tr>
</tbody>
</table>

2.2. Genetic programming procedure

Genetic programming (GP) is a powerful tool for correlating and modeling projects. It was introduced in the early 1990s and has been gradually developed mostly by its innovator John Koza (1992). GP as a machine learning method evolves evolutionary computer programs in order to perform tasks. In practice, it generates a population of mathematical functions, randomly, represented as chromosomes like syntactic tree structures operating on input data. Each tree structure is also known as a gene.

If the GP process is a specification of mathematical functions, the GP will be known as “symbolic regression”. In conventional regressions, at first, the model form should be specified by the user and then the parameters of the model will be fitted. But in symbolic regression, the algorithm itself searches for the model form and then fits the parameters.

After generating the first population (parents), by weighted summation of all genes with a bias term, the overall primary model will be determined. A simple schematic of the tree structure is shown in Fig. 1.

![Fig. 1](image)

Fig. 1. A simple gene (tree structure) with operators: *, + and tanh.

Then, crossing over the best performing trees and their modification (cutting some parts of trees and exchanging cut parts between themselves) will be implemented to create a new population (children). This procedure will be iterated until the population contains functions that are able to solve the problem successfully.

When the algorithm creates a number of genes rather than one, it is called “multi-gene symbolic regression” which is a more accurate technique applied for producing a population of mathematical relations. A multi-gene consists of one or more genes which are individually usual GP trees. Thus, multi-gene approaches often give simpler functions than other models consisting of one monolithic GP tree (Searson et al., 2010). Genetic programming flowchart is shown in Fig. 2 (Khamehchi et al., 2009b).

A free open source genetic programming toolbox called “GPTIPS” has been prepared by Searson (2009) for use with MATLAB software. It was mainly written for multi-gene symbolic regression applications. So all the steps pointed previously, (generating parent genes, crossing over the best trees, mutating, producing children, etc.) are operated by GPTIPS to achieve the best correlation (Searson, 2009; Searson et al., 2010). In GPTIPS, there are some adjustable parameters like the maximum number of genes in the regressing process, primary mathematical operators, number of population and so on which should be determined before the run by user.

GPTIPS was used in this study in order to develop a non-linear correlation for gas viscosity. Input data (training and test subsets) including experimental sets of pseudo-reduced temperature, pseudo-reduced pressure, apparent molecular weight and gas density along with experimental data of gas dynamic viscosity were given to the GPTIPS program. Then, tuning parameters of the code were adjusted and modified. After running the program, the correlation was obtained with good statistical evaluation criteria and accuracy of interest.

2.3. Evaluation of correlation validity

In order to evaluate the developed model, some common statistical parameters were calculated. They are squared correlation coefficient ($R^2$), root-mean-square deviation (RMSD) and average absolute relative deviation percentage (AARD%). A low value of RMSD and AARD and high $R^2$ is preferred:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i^{exp} - y_i^{cal})^2}{\sum_{i=1}^{n} (y_i^{exp} - y^{exp})^2}$$  (8)
3. Results and discussions

By use of GPTIPS, an accurate correlation for natural gas dynamic viscosity was obtained. The model is as follows:

\[
\begin{align*}
\text{RMSD} & = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i^{\text{exp}} - y_i^{\text{cal}})^2} \\
\text{AARD(\%)} & = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i^{\text{exp}} - y_i^{\text{cal}}}{y_i^{\text{exp}}} \right| 	imes 100 \\
\text{ARD(\%)} & = \left| \frac{y_i^{\text{exp}} - y_i^{\text{cal}}}{y_i^{\text{exp}}} \right| 	imes 100
\end{align*}
\]

Where \( y_i^{\text{exp}}, y_i^{\text{cal}} \) and \( n \) are experimental, estimated, average of experimental dependent variables \((\mu_g)\) and number of samples in the dataset, respectively.

The statistical parameters of the model are given in Table 2. According to Table 2, new developed model can predict natural gas viscosity accurately. Fig. 3 shows the predicted values of \( \mu_g \) by Eq. (12) in comparison with experimental data. Absolute relative deviations of all data are shown in Fig. 4 in the form of the number of samples in each range of relative error percent. As it is shown in Fig. 4, there are some errors exceeding 5%. The average of all the errors which are higher than 5% is 7.303% and the average of errors in the range of less than 5% is 1.871%. Maximum absolute relative
error in all dataset is 23.16% and its absolute error is 0.00232 cp. On the other hand, maximum absolute error over 1938 data is 0.005749 cp and its absolute relative error is 4.406%. Total data of experimental pseudo-reduced temperature versus pseudo-reduced pressure are shown in Fig. 6. All the errors in total dataset are scattered in all ranges of temperature, pressure, pseudo-reduced temperature and pressure, density, molecular weight, viscosity, etc. However, about 12.7% of all data have the errors higher than 5% and about 1.09% have the errors larger than 10%. In other words, there are just 21 data which have the errors larger than 10%. Fig. 5 shows the cumulative frequency of errors versus absolute relative errors.

In order to evaluate the obtained correlation as well as other proposed models, a comparison has been made over test dataset and the result is presented in Table 3. In this table, Eq. (12) has been compared with models of Sanjari et al. (2011), Heidaryan et al. (2010, 2013). In order to compare the models, the ranges of $T_{pr}$ and $P_{pr}$ in the test dataset should be matched with all the considered correlations restrictions. Overall test dataset has 388 data. After deleting out of range data of $T_{pr}$ and $P_{pr}$, finally, 254 data were remained which were used in the comparison. We called this new dataset “comparing dataset”. As a result, the prediction capability of the new developed model is higher than previous relations. Furthermore, there are a lot of large errors in the study of Heidaryan et al. (2013). However, the correlations of Sanjari et al. (2011) and Heidaryan et al. (2010) seem more accurate. As another comparison, cumulative frequency of errors versus absolute relative errors is presented in Fig. 7. Due to lots of large errors, the correlation of Heidaryan et al. (2013) has not been taken into consideration in study of Heidaryan et al. (2013). For example, a sample of natural gas was considered and its viscosity was calculated by models of Heidaryan et al. (2013), Sanjari et al. (2011) Heidaryan et al. (2010) and present study. The results are given in Table 4.

Table 3 and Fig. 7 show the superiority of correlation of present project. The experimental values of $T_{pr}$, $P_{pr}$, $\mu_g$ and $M_d$ for 1938 natural gas samples as well as experimental and predicted values of $\mu_g$ are provided as supporting materials and information.

### 4. Conclusion

By application of a genetic programing approach toolbox called “GPTIPS”, an accurate model has been developed for estimation and prediction of natural gas dynamic viscosity, as function of a number of common parameters of natural gas when there are no sufficient data about composition. One of the useful applications of this kind of models is the estimation of viscosity of natural gas samples when the data about composition are not available.
A sample test showing application of new model and some other formulas.

<table>
<thead>
<tr>
<th>Sample parameters</th>
<th>Method</th>
<th>Predicted ( \mu _{i} ) (cp)</th>
<th>ARD%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu _{i} = 0.10390 ) cp</td>
<td>Heidaryan et al. (2013)</td>
<td>0.009229</td>
<td>9.11717696</td>
</tr>
<tr>
<td>( T_{W} = 0.573843414 )</td>
<td>Heidaryan et al. (2010)</td>
<td>0.035570459</td>
<td>19.53644918</td>
</tr>
<tr>
<td>( P_{W} = 9.661213448 )</td>
<td>Sanjari et al. (2011)</td>
<td>0.108553041</td>
<td>4.86336921</td>
</tr>
<tr>
<td>( \rho _{W} = 0.4999 ) g/mL</td>
<td>Present study</td>
<td>0.103673047</td>
<td>2.18433616</td>
</tr>
<tr>
<td>( M_{w} = 41.2916 ) g/mol</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Nomenclature

| AARD % | average absolute relative deviation |
| ANN | artificial neural network |
| ARD% | absolute relative deviation |
| GP | genetic programing |
| GRN | generalized regression neural networks |
| ICA | imperialist competitive algorithm |

References


Rasouli, H., 2009b. Intelligent system for continuous gas lift operation and design with unlimited gas supply. Fluid Phase Equilib. 264, 90.


