New tools to determine bubble point pressure of crude oils: Experimental and modeling study

Mohammad Ali Ahmadi a,*, Sohrab Zendehboudi b, Lesley A. James b, Ali Elkamel c, Maurice Dusseau d, Ioannis Chatzis e, Ali Lohi f

a Department of Petroleum Engineering, Petroleum University of Technology, Ahwaz, Khuzestan, Iran
b Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John’s, NL, Canada
c Department of Chemical Engineering, University of Waterloo, Waterloo, ON, Canada
b Department of Earth and Environmental Sciences, University of Waterloo, ON, Canada N2L 3G1
def Department of Chemical Engineering, Ryerson University, Toronto, ON, Canada

A R T I C L E   I N F O

Article history:
Received 8 April 2014
Received in revised form
30 July 2014
Accepted 12 August 2014
Available online 2 September 2014

Keywords:
experimental study
bubble point pressure
gene expression programming
crude oil
predictive tools

A B S T R A C T

Adequate Knowledge of reservoir fluid characteristics (e.g., bubble point pressure) plays a crucial role while conducting modeling/simulation of production processes in petroleum reservoirs. Although many efforts have been made to obtain proper correlations for prediction of bubble point pressure (BPP) of reservoir fluids, there is still relatively high magnitude of error with the developed predictive tools available in the literature. To fill this lacuna, a robust and effective technique, called gene expression programming (GEP), is employed to determine BPP of crude oil samples as a function of temperature, oil composition, molecular weight of C7, and specific gravity of C7. The GEP method is built based on the experimental (or real) data used for training and testing phases in order to develop an appropriate correlation. The previous predictive methods are also reported in this study and employed to calculate BPP as a function of independent parameters when the same data bank is utilized. Comparing the outputs obtained from the previous models with the BPP values predicted by the GEP technique, it was found that the GEP approach exhibits higher accuracy and lower uncertainty on the basis of statistical analysis in terms of coefficient of determination (R²) and mean squared error (MSE). Great precision attained in this study through using GEP recommends linking reservoir simulator packages with the GEP tool when thermodynamic properties such as BPP are required for modeling and optimization purposes.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In general, all calculations in chemical/petroleum engineering are strongly affected by fluid thermodynamic properties that might be determined through experimental works, analytical/theoretical models, empirical equations and statistical correlations. In other words, modeling simulation of hydrocarbon reservoirs, operation and design of surface facilities, determination of inflow performance, estimation of oil/gas in place, and analysis of well testing and material balance data are strongly dependent on the fluid PVT properties like density and bubble point pressure (BPP) which are important parameters in order to make a proper plan for reservoir development (Kloubek, 1972; Holcomb and Outcalt, 1999; Mishchuk et al., 2000; Fainerman et al., 2004; Sun et al., 2005; Fainerman et al., 2006; Yazaydin and Martin, 2007; Bandyopadhyay and Sharma, 2011; Dixit et al., 2012; Adeleke et al., 2013; Li and Yang, 2013; Simjoo et al., 2013).

BPP is defined as the maximum pressure at which the first gas bubble evolves from the corresponding liquid phase (Farasat et al., 2013). BPP, which is obtained from some routine laboratory methodologies such as Constant Composition Expansion (CCE) (Zheng et al., 2000; Shen et al., 2001), is commonly used to compute other important PVT characteristics including oil viscosity (μo), oil density (ρo) and oil formation volume factor (Bfo). Thus, accurate prediction of this property appears to be vital to determine the subsequent parameters that are required for reservoir simulation and design of equipment involved in oil production processes (Bandyopadhyay and Sharma, 2011; Ahmadi et al., 2014a).

Although, available experimental procedures result in reliable data, they are still taken into account as lengthy and costly methods (Baker et al., 2003). In addition, the laboratory tests are significantly affected by the quality and quantity of the fluid samples collected particularly when the pressure near the wellbore goes below BPP (Proett and Chin, 2000; Dong et al., 2007;
Nnochiri and Lawal, 2010; Deisman et al., 2013). Considering various drawbacks reported here, development of simple, systematic and accurate ways to determine BPP seems necessary. Numerous techniques such as equation of states (EOSs), empirical relationships, and intelligent models have been developed to obtain the value of BPP. However, there are still some issues with them. For instance, prediction accuracy of EOS models is highly dependent on the type of fluid, mixing rules used, type of EOS selected for computation (Sarkar et al., 1991; Wang and Gmehling, 1999; Wang and Pope, 2000, 2001; Guo et al., 2001; Fires et al., 2001; Nikookar et al., 2008). Standing (1947) also introduced an equation to correlate BPP to gas solubility, temperature, oil gravity and gas gravity. Assuming the Henry’s law is valid, Lasater (1958) introduced a model to forecast the saturation pressure (or BPP) of oil samples where no non-hydrocarbon impurities are present in the mixtures. Moreover, a graphical model was obtained based on the North Sea data by Glaso to find the values of BPP, oil formation volume factor ($B_o$), total formation volume factor ($B_r$) and oil viscosity ($\mu_o$). This model considers a correction factor if the oil mixture contains some gaseous impurities including H$_2$, N$_2$ and H$_2$S (Glaso, 1980). Velarde et al. (1997) developed a predictive correlation through combining available correlations where a great number of data points were used. Based on the data collected from the oil samples (taken from the Middle East reservoirs), Gharbi and Elsharkawy (1999) applied a conventional smart technique, known as Artificial Neural Network (ANN), to predict important PVT properties including BPP and gas/oil ratio (GOR). Following this study, Gharbi et al. (1999) constructed a multi-layer perceptron ANN system to estimate BPP when a larger volume of data is employed. Given three different PVT databases, El-Sebakhy et al. (2007) obtained a mathematical relationship to determine parameters BPP and $B_o$ using support vector regression technique. Regardless of the advantages reported for the ANN models, this type of predictive tools has its inherent limitations and constrains due to complexities, ambiguities and nonlinear behaviors of reservoir parameters (Ahmadi, 2012; Ahmadi and Shadizadeh, 2012; Ahmadi et al., 2013a; Ahmadi and Ebadi, 2014; Ahmadi et al., 2014b, 2014c). For instance, they do not provide any linear or non-linear equations to estimate target functions. In addition, obtaining the optimum values for the parameters of the ANN models is not an easy task such that it takes fairly long time through the trial and error procedure if the optimization technique is not available. Therefore, researchers made attempts to get aids from different methods like Adaptive Neuro-Fuzzy Inference System (ANFIS), and Support Vector Machine (SVM). For instance, the first technique has been used to forecast the reservoir characteristics and operational conditions (Ahmadi et al., 2013a, 2013b; Ahmadi and Ebadi, 2014). Support Vector Machine (SVM) modeling has been also conducted to predict BPP when some handy PVT or and thermodynamic parameters are considered as inputs (Farasat et al., 2013).

The main objective of the research reported in this paper is to introduce a user friendly, effective and quick correlation to calculate BPP of crude oil samples. To gain this goal, a novel Artificial Neural Network in the form of gene expression programming (GEP) is applied through an extensive statistical manner. The GEP technique introduced in this study uses a large number of the real data collected from various sources to attain a non-linear equation in terms of main thermodynamic conditions and properties such as temperature and density (Hoffman et al., 1953; Jacoby and Berry, 1958; Vogel and Yarborough, 1980; Williams et al., 1980; Hong, 1982; Li et al., 1985; Coats and Smart, 1986; Drohm et al., 1988; Jhaveri and Youngren, 1988; Riemens et al., 1988; Ahmed, 1989; Pedersen et al., 1989; Agarwal et al., 1990; Danesh et al., 1991, 1992; Pedersen et al., 1992; Moharam and Fahim, 1995; Wu and Rosenegger, 1999; Wu and Rosenegger, 2000; Elsharkawy, 2003). The real data are divided into different groups to implement training, testing and validation stages. To confirm the capability as well as superiority of the new correlation obtained from GEP modeling, the previous available correlations are employed to estimate the target variable (e.g., BPP). The performance comparison is statistically performed in this study. Further information on the modeling methodology, last works and effectiveness of the new BPP equation is provided in the following sections.

2. Methodology description

2.1. Experimental procedure

First, the sample is placed in a container which is equipped with pressure gauges. Checking the container pressure is conducted through the pressure sensors. The sample is then pressurized to guarantee single phase condition. As a next stage, the quality of the sample is determined by measuring the GOR at laboratory conditions in order to detect whether gas may have been lost through leaking valves. A sight glass is put in the flow line to detect the presence of water phase while transferring the sample to the PVT cell. The transferred sample is brought to reservoir conditions in the PVT cell. Thus, a single stage flash is conducted. The gas and liquid are collected in an equipment which has been built in the mechanical lab. Both phases are in contact to achieve the thermodynamic equilibrium. After that, measurements

**Nomenclature**

**Abbreviations**

- BPP: bubble point pressure
- GEP: gene expression programming
- CCE: Constant Composition Expansion
- EOS: equation of state
- ANN: Artificial Neural Network
- ANFIS: Adaptive Neuro-Fuzzy Inference System
- SVM: Support Vector Machine
- GOR: gas oil ratio
- GA: genetic algorithms
- ET: expression tree

**Variables**

- $T$: temperature
- Vol./Inter.: the mole percentage ratio of volatile components (e.g., C$_1$ and N$_2$) to intermediate components (e.g., C$_2$–C$_6$, CO$_2$ and H$_2$S)
- Mw: molecular weight
- SG: specific gravity
- $N$: number of the data points
- $y_{exp}$: experimental values of the output
- $y_{pre}$: predicted values of the output

**Subscripts**

- exp: experimental
- pre: predicted
of the pressure, temperature and volume are feasible. Gas and liquid are then analyzed to determine their compositions, employing the gas chromatograph. Based on the data resulted from the GC analysis and also the GOR data, the overall composition is computed. A portion of the reservoir fluid is added to the second PVT cell equipped with a sight glass. The test begins with recording the volume of the cell content at 20 °C and the reservoir temperature. The magnitude of thermal expansion is obtained from the volume measurement. This type of measurements at various pressures permits to find the bubble point pressure (BPP). Equilibrium condition is attainable through rocking the cell. The measurement errors can be determined (or detected) if the Y-function is plotted.

2.2. Genetic Programming

Genetic Programming (GP) is a part of genetic algorithms (GAs) with a great ability to automatically develop computer programs (Gandomi et al., 2011a; Ahmadi et al., 2014d). Genetic Programming (GP) inspired from the principle of Darwinian natural assortment for evolutionary algorithms like GA. The theoretical concept of GP was introduced by Koza (1992) for the first time after various tests on symbolic regression case were carried out. The main distinction between the GP and original GA arises from the demonstration format (type) of the final solution. The outcomes of the current GP systems are computer-based programs, being exhibited in the form of tree topologies through a functional programming language (Koza, 1992), while original genetic algorithm generates a string of numbers to display the final result. Referring to the GP technique, the developed individuals are parse trees that are assorted in length all the time in the favor of fixed-length binary strings. To develop the superior values for a set of the defined model variables, conventional optimization methodologies (e.g., GA) are mostly employed to obtain the optimal values of the model parameters. On the other hand, the GP technology comes of the current GP systems are computer-based programs, the demonstration format (type) of the final result.

The original GP model looks like a tree-based Genetic Programming (GP). Each member in the tree-based GP represents a tree topology which is made of functions and terminals. The functions and terminals in the tree are chosen from various sets of proposed/introduced functions and terminals. Providing more clarification, the function within the tree may contain the basic math operations, Boolean logic functions or/and other mathematical functions such as +, −, ×, ÷, AND, OR, NOT. Moreover, the terminals include the proper arguments for the functions. They can also contain numerical constants, logical constants, variables, and so on. To generate computer program in a tree-like topology, the functions and terminals are picked randomly and constructed together (Gandomi et al., 2011a; Ahmadi et al., 2014d). Fig. 1 depicts a simple tree structure of Genetic Programming (GP) (Alavi et al., 2011).

It is important to note that the gene expression programming (GEP) is a linear branch of the conventional GP. The linear branch of GP demonstrates an explicit difference between the genotype and the phenotype of an individual. Hence, the individuals appear as linear strings in the GEP model (Oltean and Grosan, 2003; Gandomi et al., 2011a; Ahmadi et al., 2014d). More information on the GP technique is found in these documents (Oltean and Grosan, 2003; Javadi and Rezania, 2009; Torres et al., 2009; Alavi et al., 2011; Gandomi et al., 2011a; Ahmadi et al., 2014d).

The conversion corresponds to the first position in the K-expression. This is related to the solution of the ET and pores in the string one by one (Gandomi et al., 2011b). This GEP statement is expressed through a mathematical relationship as given below (Gandomi et al., 2011a):

\[
\sqrt{\frac{\text{X}}{\text{Y}}} \cdot \frac{\text{X}}{\text{Y}}
\]  

in which, X and Y are the variables; “÷” is the element separator to simplify reading. Eq. (1) is normally called a Karva notation or a K-expression (Gandomi et al., 2011a, 2011b; Ahmadi et al., 2014d). A K-expression or an expression tree (ET) in the GEP model is presented in the form of a diagram as shown in Fig. 2 (Gandomi et al., 2011a).

As explained before, the GEP genes contain a steady length which is initially set for a given case. Hence, the length of the genes in the GEP is not a critical parameter. However, the size of the ETs is crucial in the GEP modeling (Gandomi et al., 2011a, 2011b).

![Fig. 1. A classic LISP structure in the GP model in the form of a parse tree that represents the algebraic expression \((X/Y)*\) by a two-gene chromosome.](image-url)
2011b). It is important to note that a particular number of suspended components exists in the GEP, not demonstrating appropriate elements for the genome mapping. Thus, the length of the gene of the GEP model might be the same as or longer than the length assigned for a K-expression. GEP utilizes a head–tail structure to assure the truthfully of a randomly collected genome. Due to this fact, each GEP gene contains a head and a tail while the head may consist of both function and terminal symbols, although the tail might have only terminal symbols (Ferreira, 2001, 2006; Gandomi et al., 2011a, 2011b; Ahmadi et al., 2014d).

The similar methodology that has been proposed by Ferreira (2001) was employed in this study. The input parameters include temperature, oil composition, molecular weight of C\textsubscript{7}+ and specific gravity of C\textsubscript{7}+. It is clear that the output variable is the bubble point pressure in this modeling technique. The computational algorithm is briefly presented here as follows:

1. The first stage is initialization of the population or randomly generation of the chromosomal structures of the individuals through setting several correlations in the form of pars trees using the operators (e.g., /, C\textsubscript{0}, +, *).

2. Next stage is to determine the fitness magnitude of every individual of the generated population by an objective function (OF) which is given below (Ferreira, 2001, 2006)

\[
OF = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i^{exp} - y_i^{pre}}{y_i^{exp}} \right| \times 100
\]

in which, \(N\) stands for the number of the data points utilized in the GEP model. \(y_i^{exp}\) and \(y_i^{pre}\) represents the experimental and predicted values of the output parameter (e.g., \(P_b\)), respectively.

3. Then, it is time to choose the individuals to act as appropriate parents for substitute, which were assessed based on the fitness values. In this study, the tournament technique was employed to offer a suitable assortment of the population in every generation.

4. The genetic operators that include mutation, duplicate, and inversion are used for gene reproduction. It is important to note that the mutation stage has significant contribution to proficient alteration of populations of the individuals. The replication process also makes copies of the individual chromosomes selected in step 3. On the other hand, the inversion operator is responsible to produce new individuals through adaptation of the heads of randomly chosen genes.

5. The transposable components of GEP are parts of the genome that are usually activated and flown to another spot in the chromosome. This stage is known as transportation and insertion.

6. Recombination is the last important stage in the GEP method. This step is carried out in three ways; namely, one-point recombination, two-point recombination, and gene recombination, resulting in generation of new chromosomes (Ferreira, 2006). Having new generation, the methodology is repeated until the stopping criterion (e.g., maximum number of generations) is met.

A graphical demonstration of the GEP technique is presented in Fig. 3 for further clarification (Ferreira, 2001, 2006).

It is worth noting that the optimum values of the model parameters are obtained by using the genetic algorithm (GA) method. More details on the GEP technique have been provided by Ferreira (2001, 2006).

### 2.4. Data collection

A large database was used in developing the GEP mode using the real data for the purpose of this study. The objective is to obtain a non-linear equation in terms of main thermodynamic properties such as temperature and density. The collected data were taken from these sources (Hoffman et al., 1953; Jacoby and

![Fig. 2. A particular Karva program in the gene expression programming (GEP) scheme, referring to the algebraic expression \(\sqrt[4]{xy} / \sqrt{y}\)](image)

![Fig. 3. A simple schematic of the GEP methodology (Modified from Ferreira [2001]).](image)
3. Results and discussion

The data sets borrowed from Farasat et al. (2013) (including four main categories) are used to predict the bubble point pressure (BPP) (Farasat et al., 2013). The data bank consists of temperature, BPP and mole fractions of mixture (reservoir fluid) components (e.g., nitrogen, methane, ethane, propane, etc.) as listed in Table 1.

This section summarizes the previous predictive methods for BPP. First of all, the BPPs predicted by the Bandyopadhyay and Sharma’s (2011) model versus the experimental BPP data are plotted in Fig. 4. As clear from Fig. 4, the reported model is able to determine the saturation pressure or BPP with a correlation coefficient ($R^2$) of 0.9278. Based on Fig. 4, it is also concluded that the predicted BPP values show considerable deviations from the line with slope 1 at high ranges of the target function. This implies that the Bandyopadhyay and Sharma’s (2011) model exhibits fairly high error and uncertainty where the BPP holds elevated values. This statement is confirmed when the relative error (deviation) is plotted for all estimated BPPs with respect to the real data (see Fig. 5).

Another conventional method to estimate BPP of the crude oils is SRK equation of state (EOS). The prediction performance of the SRK EOS is presented in Fig. 6 when a comparison is made between the experimental BPP data and SRK predicted values. According to Fig. 6, the coefficient of determination ($R^2$) is higher than 0.9 in terms of the data points closeness to the line $Y=X$, though it is noticed that a number of points are considerably deviated from the fitted line. Fig. 7 demonstrates the relative

<table>
<thead>
<tr>
<th>Variables</th>
<th>Min</th>
<th>Max</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bubble point pressure, psia</td>
<td>313</td>
<td>6880</td>
<td>2283.2</td>
</tr>
<tr>
<td>Temperature, °F</td>
<td>128</td>
<td>324</td>
<td>177.3</td>
</tr>
<tr>
<td>Hydrogen sulfide, mol fraction</td>
<td>0</td>
<td>3.68</td>
<td>0.14</td>
</tr>
<tr>
<td>Carbon dioxide, mol fraction</td>
<td>0</td>
<td>9.11</td>
<td>1.09</td>
</tr>
<tr>
<td>Nitrogen, mol fraction</td>
<td>0</td>
<td>1.67</td>
<td>0.36</td>
</tr>
<tr>
<td>Methane, mol fraction</td>
<td>5.63</td>
<td>74.18</td>
<td>33.10</td>
</tr>
<tr>
<td>Ethane, mol fraction</td>
<td>0.84</td>
<td>12.45</td>
<td>7.35</td>
</tr>
<tr>
<td>Propane, mol fraction</td>
<td>0.43</td>
<td>11.87</td>
<td>6.33</td>
</tr>
<tr>
<td>Butanes, mol fraction</td>
<td>0.95</td>
<td>8.40</td>
<td>4.58</td>
</tr>
<tr>
<td>Pentanes, mol fraction</td>
<td>0.40</td>
<td>6.65</td>
<td>3.27</td>
</tr>
<tr>
<td>Hexanes, mol fraction</td>
<td>0</td>
<td>6.65</td>
<td>3.20</td>
</tr>
<tr>
<td>Heptanes-plus, mol fraction</td>
<td>10.72</td>
<td>83.20</td>
<td>40.63</td>
</tr>
<tr>
<td>Molecular weight of C7+</td>
<td>134</td>
<td>324</td>
<td>230.9</td>
</tr>
<tr>
<td>Specific gravity of C7+</td>
<td>0.743</td>
<td>0.942</td>
<td>0.861</td>
</tr>
</tbody>
</table>

*Table 1* Range of the parameters used in the GEP techniques while predicting bubble point pressure (BPP) (Farasat et al., 2013).
deviation of the SRK outputs from the corresponding real BPPs. As seen from Fig. 7, the maximum error percentage is around 80 while employing the SRK model to forecast BPP.

The Peng–Robinson (PR) model can be also employed to predict BPP (or saturation pressure) of the reservoir fluids. The scatter plot of the PR model outputs versus the laboratory BPP is shown in Fig. 8. As it is seen in Fig. 8, the correlation coefficient ($R^2$) obtained for the PR model is higher than that for the previous methods reported in this study such that the Peng–Robinson model has superior predictive performance, compared to the SRK, and Bandyopadhyay and Sharma (2011) approaches. To examine the PR model in terms of error prospect, the relative error distribution of the PR model in terms of error prospect, the relative error distribution of the Elsharkawy (2003) method is reported in this study such that the Peng–Robinson model has

As discussed earlier in the text, no integrate model (or correlation) with adequate accuracy is found in the literature for prediction of the saturation pressure (or BPP) of the crude oil samples. A new correlation was developed to determine the BPP of crude oil samples. To achieve reasonable accuracy, the GEP principle is aimed to correlate the BPP parameter with variables such as oil composition and temperature, using the data bank of Farasat et al. (2013). Hence, initially it is assumed that the BPP is a function of some important parameters as shown below

$$BPP = f(Temperature, Oil\ composition, Molecular\ weight\ of\ C_7+,\ Specific\ gravity\ of\ C_7+)$$

(4)

Conducting gene expression programming (GEP) modeling, a new correlation was developed to determine the BPP of crude oil samples. The GEP model introduces an equation which appears to be integrate, vigorous and effective if the input parameters includes temperature, oil composition, molecular weight of $C_7+$, and specific gravity of $C_7+$. For the sake of simplicity, a new term, called Vol./Inter., was defined in the GEP model as the mole percentage of volatile component such as $C_1$ and $N_2$ dividing by that of intermediate components that include $C_2$–$C_6$, $CO_2$ and $H_2S$. It is believed that one can get an estimation of BPP with high confidence and precision through using the developed correlation. The model resulted from the GEP implementation is presented as follows:

$$BPP = A \times \frac{m_{C_7+}}{C_7+} - \frac{B}{C_7+} + C \times \frac{Vol./Inter.}{D \times SG_{C_7+} + E \times SG_{C_7+}} + \frac{F}{\text{MW}_{C_7+}} + T \times 505.967 + \frac{2.45789 \times 10^6}{T} + \frac{G}{Vol./Inter.}.$$  

(5)

In Eq. (5), Vol./Inter. represents the mole percentage ratio of volatile components (e.g., $C_1$ and $N_2$) to intermediate components (e.g., $C_2$–$C_6$, $CO_2$ and $H_2S$), $\text{MW}$ is molecular weight, and $SG$ stands for
the specific gravity, and \( T \) is the symbol for temperature \( (^\circ F) \). In addition, the BPP (saturation pressure) is in psi.

The variables \( A, B, C, D, E, F \) and \( G \) included in Eq. (5) are also defined with the following relationships:

\[
A = 1488.46 - SG_{C7^+} \times 1087.64 - \frac{601.512}{SG_{C7^+}} + 1670.58 + T \\
\times 0.40771 \frac{6345.73}{T} + 8.72863 \\
\times \text{Vol./Inter.}
\]

\[
B = \frac{4132.53}{\text{Vol./Inter.}} \frac{6.1977 \times 10^7}{\text{Mw}_{C7^+}} + T \times 3181.39 + \frac{1.0693 \times 10^8}{T}
\]

\[
C = \frac{20964.7}{\text{Vol./Inter.}} \frac{1.76162 \times 10^6}{\text{Mw}_{C7^+}} - T \times 19.1963 - \frac{21.5427}{\text{Vol./Inter.}} \\
+ \frac{12851.5}{T} - \frac{3477.49 + SG_{C7^+} \times 9532.73}{T}
\]

\[
D = 1.05345 \times 10^{-6} - \frac{1.50748 \times 10^8}{\text{Mw}_{C7^+}} - T \times 676.048 \\
- \frac{1.20252 \times 10^7}{T} - \frac{12178.2}{\text{Vol./Inter.}}
\]

\[
E = \frac{641357}{\text{Vol./Inter.}} \frac{1.02891 \times 10^8}{\text{Mw}_{C7^+}} - T \times 301.034 \\
- \frac{1.91446 \times 10^6}{T} - \frac{3873.62}{\text{Vol./Inter.}}
\]

\[
F = 2.51471 \times 10^{-08} + T \times 25332.5 \\
+ \frac{3.47893 \times 10^8}{T} - \frac{244573}{\text{Vol./Inter.}}
\]

\[
G = 20125.6 - T \times 23.9271 - \frac{703193}{T}
\]

The digits reported for the coefficients in the developed correlation have been determined through conducting a systematic sensitivity analysis towards optimizing the parameters of the GEP approach.

Simple functions (e.g., exponential, power, \( +,-,^* \)) were employed to construct the correlation shown by Eq. (5). The required parameters such as the number of head size, chromosome, gene, population, mutation rate and inversion rate were determined to be 8, 40, 8, 2000, 0.044 and 0.1, respectively.

Fig. 12 depicts the correlation between outcomes of the developed GEP correlation and the corresponding experimental BPP data. According to Fig. 12, most of the GEP outputs follow the straight line with a high correlation coefficient (e.g., 0.9864). It is also clear that there is a very good agreement between the predicted BPP values and the laboratory data. Another demonstration of gained results is depicted in Fig. 13. Showing the robustness of the correlation in another way, Fig. 13 presents the predicted and actual BPP values versus the data index, implying very low differences between the calculated and real data. The relative error percentage of the BPP outputs determined from the GEP model is plotted versus the values of experimental saturation pressure (see Fig. 14). Based on Fig. 14, the maximum relative deviation is not more than 30%, even for the low range of the BPP data. Given the coefficient of determination and relative error percentage, it was...
found that the developed GEP correlation for estimation of BPP is superior than other thermodynamic and/or conventional models in terms of integrity and accuracy. Results obtained from the statistical analysis are summarized in Table 2, based on the comparison made between the predictive tools. According to Table 2, it is concluded that the developed GEP correlation for estimation of BPP is superior to other thermodynamic and/or conventional models in terms of accuracy.

To validate the performance and accuracy of the introduced GEP correlation, the values of BPP for five different crude oils (taken from different Iranian oil fields) were determined. The compositions of two of the crude oil samples are tabulated in Table 3. The results obtained from the conventional models and the correlation proposed in the current study are reported in Fig. 15. As illustrated in Fig. 15, the GEP correlation results exhibit much more closeness to the experimental BPP data, compared to other methods.

To evaluate the correctness of the developed GEP model, the Analysis of Variance (ANOVA) technique was used to conduct a systematic parametric sensitivity analysis based on the input parameters involved in the target function. This stage helps to determine the importance (weight) of the input variable in the bubble point pressure. According to this analysis, Fig. 16 demonstrates relative importance of the inputs. As clear from panel (a) of the figure, the mole percentages of C1 and C7+ considerably affect the magnitude of BPP. Combined effect of the concentration of volatile and intermediate components on the output is presented.

---

### Table 3

Composition of the two crude oil samples for BPP experiments.

<table>
<thead>
<tr>
<th>Sample 1</th>
<th>Components</th>
<th>Flashed oil (mol%)</th>
<th>Flashed gas (mol%)</th>
<th>Reservoir oil (mol%)</th>
<th>Sample 2</th>
<th>Components</th>
<th>Separated oil (mol%)</th>
<th>Separated gas (mol%)</th>
<th>Reservoir oil (mol%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H2S</td>
<td>0.000</td>
<td>0.230</td>
<td>0.163</td>
<td></td>
<td>H2S</td>
<td>0.000</td>
<td>0.13</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>0.000</td>
<td>0.206</td>
<td>0.146</td>
<td></td>
<td>N2</td>
<td>0.000</td>
<td>0.91</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>CO2</td>
<td>0.000</td>
<td>2.686</td>
<td>1.904</td>
<td></td>
<td>CO2</td>
<td>0.000</td>
<td>6.71</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>0.042</td>
<td>65.696</td>
<td>46.574</td>
<td></td>
<td>C1</td>
<td>0.27</td>
<td>19.52</td>
<td>1.88</td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td>0.189</td>
<td>7.847</td>
<td>5.618</td>
<td></td>
<td>C2</td>
<td>2.35</td>
<td>15.55</td>
<td>3.45</td>
<td></td>
</tr>
<tr>
<td>iC4</td>
<td>0.357</td>
<td>1.434</td>
<td>1.120</td>
<td></td>
<td>iC4</td>
<td>1.12</td>
<td>2.41</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td>nC4</td>
<td>2.122</td>
<td>3.883</td>
<td>3.370</td>
<td></td>
<td>nC4</td>
<td>5.70</td>
<td>6.63</td>
<td>5.77</td>
<td></td>
</tr>
<tr>
<td>iC5</td>
<td>2.753</td>
<td>1.262</td>
<td>1.696</td>
<td></td>
<td>iC5</td>
<td>3.85</td>
<td>2.08</td>
<td>3.70</td>
<td></td>
</tr>
<tr>
<td>nC5</td>
<td>4.140</td>
<td>1.592</td>
<td>2.333</td>
<td></td>
<td>nC5</td>
<td>5.57</td>
<td>2.30</td>
<td>5.30</td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td>10.128</td>
<td>1.580</td>
<td>4.068</td>
<td></td>
<td>C6</td>
<td>6.30</td>
<td>1.74</td>
<td>5.92</td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td>7.817</td>
<td>0.783</td>
<td>2.830</td>
<td></td>
<td>C7</td>
<td>74.84</td>
<td>0.63</td>
<td>68.67</td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td>4.812</td>
<td>0.256</td>
<td>1.582</td>
<td></td>
<td>Total</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td>4.833</td>
<td>0.009</td>
<td>1.413</td>
<td></td>
<td>GOR: 237 SCF/STB</td>
<td>1233.12 SCF/STB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td>5.421</td>
<td>0.000</td>
<td>1.578</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C11</td>
<td>6.010</td>
<td>0.000</td>
<td>1.749</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C12</td>
<td>5.1376</td>
<td>0.000</td>
<td>14.954</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOR:</td>
<td>500.00</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Fig. 15.** Comparison between predictive performances of different models versus experimental BPP data.

**Fig. 16.** Relative importance of all variables affecting the value of BPP; (a) all single components, (b) combined term to consider the mole percentage of light and intermediate components.
in the panel (b) of Fig. 16 in terms of weight percentage, clearly showing high contribution of this lumped parameter to BPP.

4. Conclusions

Accurate determination of the bubble point pressure (BPP) of the crude oils is important to chemical and petroleum engineering in industrial and research sectors. Provided the laboratory data, the gene expression programming (GEP) was used to develop an efficient correlation for estimation of BPP with high precision and low uncertainty. The developed correlation relates BPP to other properties such as temperature, specific Gravity of \( C_7 \), molecular weight of \( C_7 \), and oil composition. The available predictive tools were applied on the same data points used in the current study. According to the results attained from the GEP model, an acceptable match was noticed between the predicted magnitudes of BPP and the real BPP data, despite the fact that the common correlations such as SRK, Peng–Robinson, Elsharkawy (2003) were not able to give adequate prediction of BPP such that fairly low coefficient of determination \( (R^2) \) and elevated values for relative error were obtained. The intelligent correlation proposed for BPP parameter looks user friendly, robust and precise, with broad (and high) potentials in petroleum and chemical industries. Furthermore, the predictive model is an interesting alternative to examine the performance of the commercial reservoir simulators (e.g., PVti in ECLIPSE package) which are being used for history matching and calculating reservoir fluid characteristics. It is worth noting that the developed GEP correlation is trustable with high confidence when the required bubble point pressure (BPP) data are unavailable.

Acknowledgements

The authors would like to acknowledge the National Iranian Oil Company and Petroleum University of Technology (PUT) for support throughout this study.

References


Kloubek, J., 1972. Measurement of the dynamic surface tension by the maximum bubble pressure method. 3rd. Factors influencing the measurements at high
frequency of the bubble formation and an extension of the evaluation to zero age of the surface. J. Colloid Interface Sci. 41 (1), 7–16.