Petroleum and Coal

Article **Open Access**

IMPROVED MODELS FOR THE ESTIMATION OF PVT PROPERTIES OF CRUDE OILS

Arash Kamari [1], Mehdi Sattari [2], Amir H. Mohammadi [2], Deresh Ramjugernath [2**]*

[1] *Department of Geology, Kansas State University, 108 Thompson Hall, Manhattan, KS, 66506, USA*
[2] *Thermodynamics Research Unit, School of Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa*

Received April 10, 2019; Accepted July 1, 2019

Abstract

Reservoir fluid properties, such as oil formation volume factor and bubble point pressure, are vital parameters in many computations associated with petroleum engineering. These computations include hydrocarbon reserve estimation, and consequently, economic efficiency evaluation, fluid flow in porous media, and improved and enhanced oil recovery. Prior to the computations, the pressure-volume-temperature (PVT) properties of reservoir oil must be determined. PVT properties, in turn, are ascertained either by empirical methods, laboratory measurements, or via equations of state. The latter two methods, however, are expensive and time-consuming and require complex calculations. Therefore, it is necessary to develop an accurate and reliable model for the determination of petroleum fluid's physical properties. In this paper, a soft-computing approach is employed to develop efficient models for the calculation of bubble point pressure and oil formation volume factor properties. In pursuit of this goal, a robust mathematical algorithm, namely, gene expression programming (GEP), is applied. Moreover, an extensive databank, covering varied ranges of experimental PVT conditions from different and widespread geological and geographic regions was constructed. The databank includes data on reservoir temperature, gas oil ratio, bubble point pressure, gas gravity, and oil forma-tion volume. In addition, the GEP model results are compared to published data on oil formation volume and bubble point pressure. The results demonstrate close agreement between estimations based on the GEP model and experimental data reported in the literature for both oil formation volume and bubble point pressure properties. Moreover, the results prove that the GEP model, proposed in this study is more accurate and capable than the existing methods, for the determination of both oil formation volume factor and bubble point pressure properties.

Keywords: *Oil formation volume factor; Bubble point pressure; Gene expression programming (GEP); Empirical correlation; PVT properties; Accuracy.*

1. Introduction

The computation of data on oil reservoirs requires detailed and accurate knowledge of the reservoir fluid properties [1]. These properties, which are associated with pressure-volume-temperature (PVT) parameters, are fundamental for undertaking many types of petroleum engineering calculations. In other words, reliable solutions for many petroleum engineering problems are impossible without accurate predictions of the PVT properties of the reservoir fluids [2].

The PVT properties of primary importance for the computation of data are solution gas oil ratio (GOR), bubble point pressure (P_b), gas formation volume factor (GFVF), and oil formation volume factor (OFVF) [3]. These PVT properties are traditionally acquired experimentally by conducting laboratory tests [4]. However, measurement of PVT properties by means of laboratory experiments is expensive and complex because cores or rock samples containing petroleum reservoir fluid are generally from isolated and difficult-to-access well locations [2,5]

Empirical correlations, or models, are employed to avoid time-consuming and costly experimental measurements. A popular method to determine the PVT properties of crude oils is the use of equations of state which are based on a detailed knowledge of the composition of reservoir fluid [6].

Several PVT models have been developed for the determination of the physical properties of reservoir fluids in oil reservoirs over time. Standing [7] and Katz [8] were the first to propose models for the estimation of PVT properties, such as OFVF and bubble point pressure. Standing's [7] model is based on experimental tests conducted on 105 cores from 22 crude oils in California State. Katz's [8] method employs data on bubble point pressure, oil API gravity, gas specific gravity, reservoir temperature and solution gas oil ratio, in order to calculate oil formation volume factor.

Vazquez and Beggs [9] presented a PVT empirical correlation to estimate solution gas–oil ratio, viscosity associated with under-saturated oil reservoirs and oil formation volume factor, based on laboratory measurements of 600 cores collected from various regions of the world. Subsequently, results indicate that their empirical correlation for the estimation of OFVF has an average error of 4.7%.

Glaso [10] performed regression and graphical analyses for oil PVT properties by using data related to 45 core samples, mostly collected from North Sea region. Their results, related to oil formation volume factor and bubble point pressure, display average errors of 1.28 and 20.43 %, respectively.

Al-Marhoun [11] provided an empirical correlation for oil formation volume factor based on experimental PVT data mostly extracted from North America and the Middle East. Dokla and Osman [12] used experimental PVT data from the UAE for the development of bubble point pressure and oil formation volume factor correlations, and Petrosky and Farshad [13] developed under-saturated isothermal oil compressibility, oil formation volume factor, and solution gas–oil ratio correlations using PVT data from the Gulf of Mexico.

Arabloo *et al.* [14] implemented two constrained multivariable search techniques, including a generalized reduced gradient algorithm and successive linear programming, to develop two correlations in order to determine oil formation volume factor and bubble point pressure. To pursue their objective, they utilized experimental data associated with various geographical domains worldwide. Their results indicate that the OFVF and bubble point pressure correlations have average relative errors of 2.24 and 18.9 %, respectively. Recently, smart techniques have been increasingly employed to predict PVT properties [15-18]. However, these methods have some drawbacks: a symbolic equation is not provided; the necessity to use a large dataset as a basis for prediction and; a potential over-fitting problem, in particular for models developed by artificial neural network techniques.

The aim of this work was to propose efficient, accurate and reliable PVT models for determination of oil formation volume factor as well as bubble point pressure, as a function of reservoir temperature, crude oil API gravity and gas specific gravity, and solution gas oil ratio. To this end, a gene expression programming (GEP) [19] strategy was utilized to develop the aforementioned models. In the first step, a large dataset was collected for different oil PVT information related to various geographical regions of the world. Then, the results attained for the newly developed models were compared against actual data, and also, previously reported correlations available in literature. Additionally, the Leverage approach was employed to detect suspended and/, or outlier data points in the dataset. Finally, to evaluate the accuracy and capability performance of the developed models, and to provide a comparative study, a statistical error analysis was performed, in which error parameters and analysis, including a crossplot and error distribution plots, were sketched.

2. Oil PVT properties

A review, dating back to the early 1940s, of published methods for the prediction of oil formation volume factor (OFVF) and bubble point pressure (P_b), indicates the importance of these PVT properties from an industry point of view [14]. By definition, OFVF is the reservoir oil required to produce one barrel (1 bbl) of oil at surface conditions [20]. Additionally, in its original condition, reservoir oil contains some natural gas in solution; consequently, the pressure at which this natural gas begins to come out of solution and forms bubbles is identified as the P_b. As a result, P_b and OFVF are the most vital properties for accurate calculation of hydrocarbon reservoir recoverable reserves, the oil-water flow ratio, reservoir capacity for

production of oil, problems related to enhanced and improved oil recovery, and all other issues associated with petroleum engineering computations [9,22-24]. Therefore, developing accurate and efficient models for the determination of P_b and OFVF is a necessary.

As a result, to predict the PVT properties of petroleum reservoir fluids utilizing correlations, field measured data, like reservoir temperature (T_R), reservoir pressure (P_R), crude oil API gravity (API), gas specific gravity or gas relative density (γ_g), and solution gas oil ratio (GOR), are normally required [18]. Hence, reservoir temperature, crude oil API gravity, gas relative density, and solution gas oil ratio are considered as required variables for accurate estimation of P_b and OFVF, in line with the majority of previously published works [7,9,13,25-27] as follows:

$$P_b = f_1(T_R, \gamma_g, GOR, API) \quad (1)$$

$$OFVF = f_2(T_R, \gamma_g, GOR, API) \quad (2)$$

Knowledge of the geographical and geological conditions of reservoir oils is important for proposing PVT correlations because the chemical composition is specific for any crude oil. In other words, obtaining accurate results, by means of PVT correlations, for different crude oils, having different chemical and physical characteristics, is difficult to achieve [28]. Hence, to account for regional characteristics, PVT correlations need to be developed by using comprehensive datasets, which cover a wide range of PVT properties from almost all regions of the world. Hence, the quality and reliability of models, for estimating PVT and thermo-physical properties, are related to the applied database [29-31]. Therefore, about 755 laboratory PVT datasets, covering wide ranges of PVT experimental conditions from a wide spread of geographical and geological regions, were utilized in this study to develop and test the models for the determination of P_b and OFVF.

The dataset used for developing the P_b and OFVF models comprises reservoir temperature (in units of °F), oil formation volume factor at bubble point pressure (in units of bbl/STB), crude oil API gravity, gas gravity, and solution gas oil ratio at bubble point pressure (in units of SCF/STB), which was collected from Moghaddam *et al.* [32], Obomanu and Okpobiri [33], Bello and Villa [34], Omar and Todd [35], Dokla and Osman [12], Al-Marhoun [36], Ghetto and Villa [1], Mahmood and Al-Marhoun [28], and Ostermann *et al.* [23]. Table 1 summarizes the values of minimum, maximum, and average for reservoir temperature, gas oil ratio, oil gravity, bubble point pressure, gas gravity, and oil formation volume factor. The table confirms that the databank collected in this study covers a wide range of PVT properties from volatile oils to heavy crude oils.

Table 1. The minimum, maximum and average values associated with the PVT properties in the databank utilized for proposing the GEP models

	Unit	Min.	Max.	Avg.	Type
Oil formation volume factor, B_{ob}	bbl/STB	1.02	2.92	1.40	Output
Bubble point pressure, P_b	psi	58.02	6 613.82	1 846.05	Output
Gas gravity, γ_g	-	0.52	3.44	1.12	Input
Initial solution gas oil ratio, R_{Si}	SCF/STB	7.08	3 298.66	592.39	Input
Reservoir temperature, T_R	°F	74.00	360.93	207.17	Input
Oil gravity, API	-	6.00	56.80	34.36	Input

3. Developing the GEP models

As already mentioned, a GEP mathematical algorithm was applied, in this study, to develop two models for the determination of PVT properties of reservoir oils, (i.e. oil formation volume factor and bubble point pressure). The GEP [19] approach is a modified version of the genetic algorithm (GA), and genetic programming (GP), which is implemented for solving regression and also classification problems. It employs populations of individuals, which are chosen in keeping with fitness, and presents genetic variation utilizing one and/or more genetic operators [37]. As a comparison, the nature of the individuals is the fundamental difference between GA, GP, and GEP algorithms [19]. Consequently, the individuals in GA, GP and GEP algorithms are: the chromosomes or linear strings of fixed length; the parse trees or nonlinear entities of different shapes and sizes; and the chromosomes or genome and/or linear strings of fixed

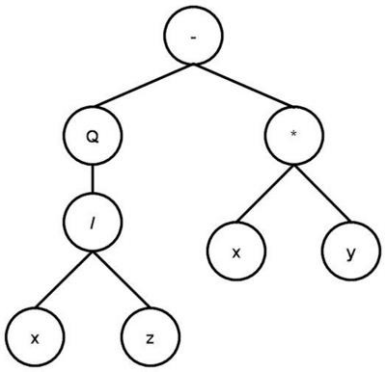
length, which are subsequently presented as nonlinear entities of different shapes and sizes, respectively [19].

In the GEP [19] algorithm, the structures of the genes allow encoding of any program for effective evolution and development of the solutions [38]. As a result, the GEP [19] mathematical algorithm employs two elements, including the expression tree (ET) and the chromosome. The chromosome has the role of an encoder for the candidate solution, which is translated into an expression tree. Each genetic chromosome involves terminals, including constants and variables; and functions structured in one and/or more genes of equal length [39]. The constants are produced by the GEP algorithm in a range selected by the employer, while the functions and variables are recognized as input data. Additionally, the gene consists of a tail made only of terminals, and a head made of functions, in addition to terminals including variables and constants [39].

The head length (h) is recognized as an input parameter for the GEP mathematical algorithm, while the tail length (t) is expressed as follows:

$$t = h(n-1)+1 \quad (3)$$

where t stands for the tail length of the gene; h shows the head length, and n is the largest arity of the functions utilized in the gene's head.



For a better understanding of the GEP procedure, Fig.1 provides an example of a two-gene chromosome composed of four functions, including -, *, / and Q, and also three terminals including x, y, and z, together with its decoded ET, and the corresponding mathematical expression, which is formulated as ($\sqrt{}$(x/z)) - (x*y).

Fig. 1. A typical two-gene chromosome with its corresponding mathematical expression; Q is the square root function

The procedure presented by Ferreira [38] was followed for developing the models, using the GEP algorithm, in order to estimate the OFVF and P_b properties as follows [38]:

I. The initialization of the population of individuals, which is based on counting the random made chromosomes of a certain number of individuals by setting various correlations, as stated;
II. The population of individuals that fits, considering fitness functions (cases);
III. The population of individuals selected, in keeping with their fitness, in order to replicate with modifications;
IV. The implementation of the same procedure, concerning confrontation of the selection environment, the genomes expression, selecting, and duplicating with modification, for the new population of individuals;
V. The repeat of the above stages for a certain number of generations, or until an optimum solution is established (convergence of the algorithm in keeping with the criteria defined).

In order to present capable, and reliable models for the calculation of OFVF and P_b, four input variables are considered. These consist of solution gas oil ratio, gas gravity, oil API gravity, and reservoir temperature. In pursuit of our goal, both datasets related to OFVF and P_b properties are randomly separated into two sub-datasets consisting of the 'Training/Learning' set and the 'Test' set.

In the development of the GEP models, 80% of the main data points related to OFVF and P_b properties, as well as their input parameters, is randomly selected for the 'Training' phase (development of the models), and 20% is assigned for the 'Test' phase (accuracy and capability evaluation), respectively. As a consequence, higher valued input variables may suppress the impact of the smaller ones during the training phase of mathematical algorithms, like in the GEP method. To overcome such an obstacle, and in order to make the GEP algorithm

perform for the estimation of both OFVF and P_b properties, all data points should be adequately processed, and well-scaled, prior to input into the GEP.

Although normalization of the data points is not necessary in the estimation process by means of GEP-based methods, better results are normally acquired after normalizing the parameters [40]. Thus, all data points related to the inputs for both OFVF and P_b properties are normalized as follows:

$$r_n = \left(\frac{r}{1.5.xr_{max}}\right)x0.8.+0.1 \quad (4)$$

where: r_n stands for the data points normalized, r indicates the actual data, and r_{max} is the maximum value of the data [41].

In the next step, the normalized data points are returned to their original values at the end of the modeling process.

4. Results and discussion

4.1. Performance evaluation

Previous research has shown that developing a correlation for bubble point pressure is more complicated than for oil formation volume factor. All of the available correlations for bubble point pressure are less rigorous and precise than those for the estimation of oil formation volume factor [14]. Hence, to obtain robust, reliable, and accurate models for both bubble point pressure and oil formation volume factor, the computational procedure related to the GEP algorithm can be used. In fact, the computational phases, associated with the GEP [38] approach, define the required parameters, which give the most accurate models for both bubble point pressure and oil formation volume factor, on the basis of the introduced variables, consisting of solution gas oil ratio, gas gravity, oil API gravity and reservoir temperature.

In order to increase the accuracy of the GEP model, compared to existing correlations, neutral genes can be added to the model (a newly developed equation), after reaching a reasonable and appropriate number of generations. This is proposed for both of the models related to OFVF and P_b properties.

Moreover, the function, in accordance with the average absolute percent relative error (AAPRE) and correlation coefficient (R2), was chosen to compute the overall fitness of the evolved programs. The program or iterative calculation was undertaken until there was no longer improvement in the precision and capability of the several proposed models, with the various functions utilized. Subsequently, the final equations for both bubble point pressure and oil formation volume factor properties were obtained as follows:

$$B_{ob} = 1 - 0.000081623\,\gamma_g\left[\sqrt{API\,T_R} + \frac{R_{Si}-4.846}{\sqrt{\gamma_g}}\right]\left(0.37658\,\gamma_g - (API - T_R)^{0.3652}\right) \quad (5)$$

$$P_b = \frac{87.3067 R_{Si}\,T_R\,\left|\gamma_g - 2.95787\right| + 7639.17}{947.493\,\gamma_g + \exp(0.000641267\,API\,T_R) + API\,T_R + 3.59953\,\gamma_g\,R_{Si}} \quad (6)$$

where: P_b denotes bubble point pressure (psi); B_{ob} stands for OFVF at bubble point pressure (bbl/STB); T_R expresses the reservoir temperature (°F); API is the crude oil API gravity; γ_g indicates the gas gravity and R_{Si} shows the solution gas oil ratio at bubble point pressure (SCF/STB).

Having developed the GEP models for the estimation of B_{ob} and P_b properties, an error analysis, including statistical error analysis, in which R2, AAPRE, average percent relative error (APRE), and root mean square (RMSE), and also a graphical error analysis, containing a parity diagram and relative error distribution plot, were performed to evaluate the prediction capability of the models.

Table 2 summarizes the statistical error parameters calculated for the B_{ob} model developed in this study. The results indicate a $R^2 = 0.93$ and an AAPRE = 3.62. The error values show that the newly developed GEP model predicts the oil formation volume factor values reliably within acceptable accuracy.

Table 2. Summarized statistical error parameters including AAPRE, APRE, RMSE and R^2 for the newly developed model for oil formation volume factor as well as the studied correlations resulted from the actual data

Method	AAPRE, %[a]	APRE, %[b]	RMSE[c]	R^2 [d]
Present study (Eq. (5))	2.17	0.18	0.07	0.93
Arabloo *et al.* model [14]	2.24	-0.04	0.07	0.94
Al-Shammasi model [25]	2.59	-0.92	0.07	0.93
Kartoatmodjo and Schmidt model [26]	2.92	-0.30	0.07	0.93
Frashad *et al.* model [41]	2.94	0.39	0.07	0.93
Al-Marhoun model [36]	3.09	-0.38	0.08	0.93
Standing model [7]	3.36	-1.98	0.08	0.93
Petrosky and Farshad model [13]	3.46	-2.35	0.08	0.93
Omar and Todd model [35]	5.03	2.08	0.12	0.85
Dindoruk and Christman model [42]	5.52	-2.94	0.14	0.83
Vazquez and Beggs model [9]	5.59	3.01	0.13	0.82
Macary and El-Batanony model [27]	9.11	-8.44	0.19	0.85
Abdul-Majeed model [43]	27.77	-27.73	0.40	0.83
Labedi model [44]	37.68	-37.64	0.68	0.93

[a] $$AAPRE\,\% = \frac{1}{n}\sum_{i=1}^{n} |\,E_i\%\,| \text{ where } E_i\% = \left[\frac{X_{\exp} - X_{\text{rep./pred}}}{X_{\exp}}\right] \times 100 \Rightarrow i = 1,2,3,...,n$$

[b] $$APRE\,\% = \frac{1}{n}\sum_{i=1}^{n} E_i\%$$

[c] $$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_{i\,\exp} - X_{i\,\text{rep./pred}}\right)^2}$$

[d] $$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(X_{(i)\exp} - X_{(i)\text{ rep./pred}}\right)^2}{\sum_{i}^{N}\left(X_{(i)\text{ rep./pred}} - averageX_{(i)rep/\,pred}\right)^2}$$

Figure 2 provides a parity diagram and a comparison between the calculated and actual values related to oil formation volume factor data. It is clear, from the figure, that most of the data points are almost placed on the line of Y=X, illustrating that there is an agreement between the newly developed model results and the actual oil formation volume factor data gathered from the literature. In order to illustrate, graphically, capability, and performance of the GEP model in estimating oil formation volume factor, the relative error percentage distribution plot is provided in Fig. 3. As can be seen in Fig. 3, a small margin of error is noticed in relation to Eq. (5).

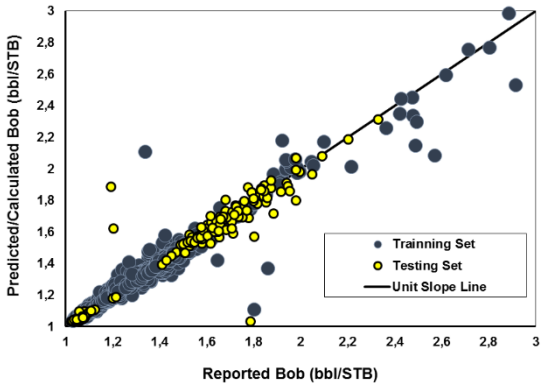


Fig. 2. Parity diagram for the estimated values by the new model and the literature-reported values of the oil formation volume factor

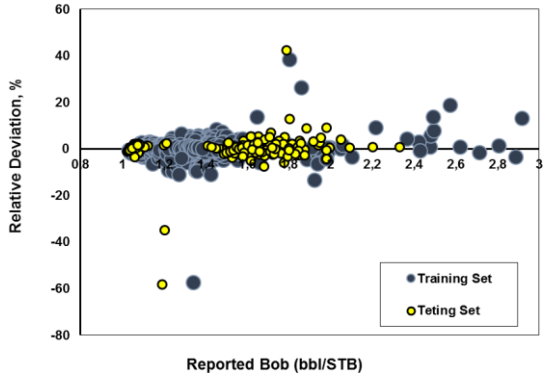


Fig. 3. Relative deviations plot of the oil formation volume factor values obtained by the newly proposed model from the available dataset values

The capability of the model developed for estimation of the oil formation volume factor was then compared with the results of 13 most widely-utilized empirical correlations available in the literature, viz. Arabloo *et al.* [14] model, Al-Shammasi [25] model, Kartoatmodjo and Schmidt [26] model, Frashad *et al.* [41] model, Al-Marhoun [36] model, Standing [7] model, Petrosky and Farshad [13] model, Omar and Todd [35] model, Dindoruk and Christman [42] model, Vazquez and Beggs [9] model, Macary and El-Batanony [27] model, Abdul-Majeed [43] model, and Labedi [44] model.

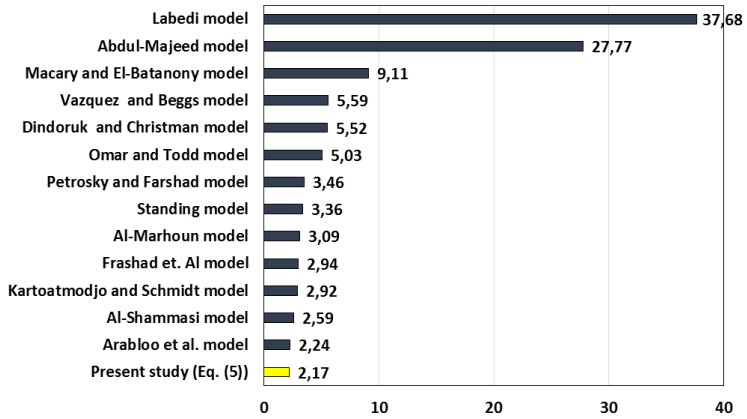


Fig. 4. Graphical comparison between the AAPRE values obtained by the newly model developed in this study for the estimation of oil formation volume factor as well as the corresponding correlations studied

A summary of the comparative study mentioned earlier, in which statistical error parameters were used, is listed in Table 2. It confirms that Eq. (5) shows a better performance for the calculation of oil formation volume factor, compared to the reviewed correlations. An acceptable AAPRE of the estimated values from the actual data was obtained. Moreover, a comparative study in terms of AAPRE is shown in Fig. 4. The bar plots drawn in Fig. 4 illustrate the accuracy of the model developed in this study for estimation of oil formation volume factor which is acceptable when compared with values calculated by other correlations listed.

Table 3 lists the statistical error parameters calculated for the P_b model proposed in the present work. The table reports that the values obtained for AAPRE, APRE, RMSE, and R^2 are 15.3%, 2.23%, 468.11, and 0.88, respectively. These values confirm the accuracy of the model for the estimation of bubble point pressure. Figure 5 is a diagram showing the parity between the calculated and reported values of bubble point pressure data, which illustrates close agreement between the newly developed model results, and the actual bubble point pressure data. Figure 6 presents the relative percentage error distribution plot for the P_b model. The figure indicates a small, existing error range, and a low scatter around the zero error line, for the bubble point pressure data and APRE obtained for the Eq. (6).

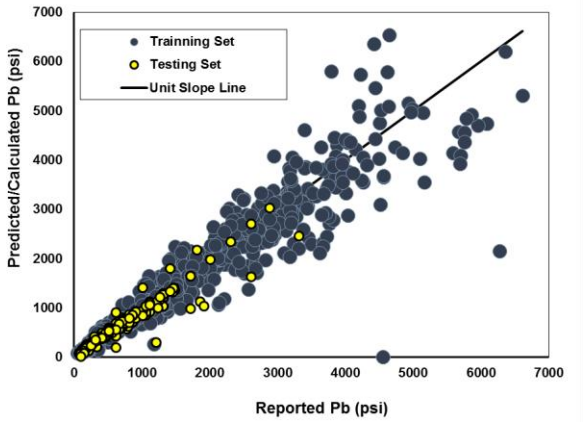


Fig. 5. Parity diagram for the estimated values by the new model and the literature-reported values of the bubble point pressure

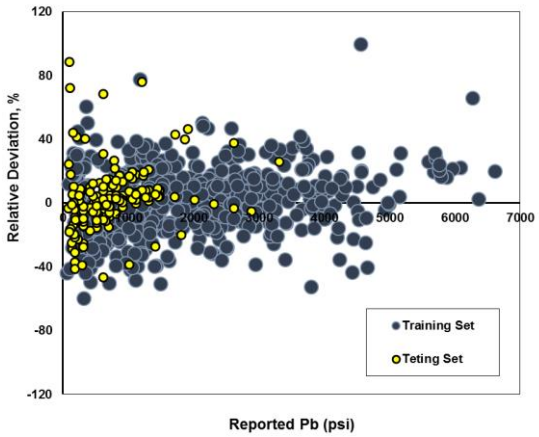
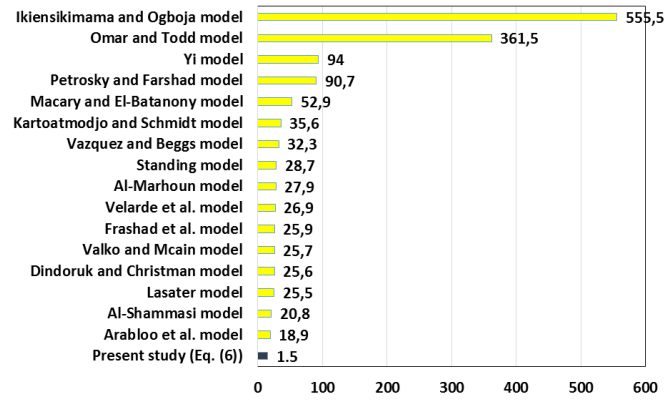


Fig. 6. Relative deviations plot of the bubble point pressure values obtained by the newly proposed model from the available dataset values

Table 3. Summarized statistical error parameters including AAPRE, APRE, RMSE and R^2 for the newly developed model for bubble point pressure as well as the studied correlations resulted from the actual data

Method	AAPRE, %	APRE, %	RMSE	R^2
Present study (Eq. (6))	15.3	2.23	468.11	0.88
Arabloo *et al.* model [14]	18.9	3.2	501.7	0.86
Al-Shammasi model [25]	20.8	−7.6	478.7	0.87
Lasater model [45]	25.5	−8.6	481.5	0.87
Dindoruk and Christman model [42]	25.6	−2.8	510.8	0.86
Valko and Mcain model [46]	25.7	0.1	584.2	0.82
Frashad *et al.* model [41]	25.9	−8.7	507.4	0.85
Velarde *et al.* model [47]	26.9	−2.1	596.6	0.82
Al-Marhoun model [36]	27.9	−4.5	550.8	0.84
Standing model [7]	28.7	−16.4	588.4	0.85
Vazquez and Beggs model [9]	32.3	−24.7	693.9	0.87
Kartoatmodjo and Schmidt model [26]	35.6	−27.2	819.7	0.84
Macary and El-Batanony model [27]	52.9	−38.3	596.4	0.85
Petrosky and Farshad model [13]	90.7	58.7	840	0.85
Yi model [48]	94	94	2 115.2	0.77
Omar and Todd model [35]	361.5	−356.0	11 387.4	0.03
Ikiensikimama and Ogboja model [49]	555.5	−555.5	5 175.9	0.40

Several of the most widely-utilized correlations, related to bubble point pressure property, were provided, including: Arabloo *et al.* [14] model, Al-Shammasi [25] model, Lasater [45] model, Dindoruk and Christman [42] model, Valko and Mcain [46] model, Frashad *et al.* [41] model, Velarde *et al.* [47] model, Al-Marhoun [36] model, Standing [7] model, Vazquez and Beggs [9] model, Kartoatmodjo and Schmidt [26] model, Macary and El-Batanony [27] model, Petrosky and Farshad [13] model, Yi [48] model, Omar and Todd [35] model, and Ikiensikimama and Ogboja [49] model. Table 3 summarizes the statistical error parameters calculated for the aforementioned correlations and the P_b model developed in this study.



As can be seen in the table, the P_b model has better performance in comparison with the reviewed methods. In order to better illustrate the comparison, AAPRE obtained for all methods is shown in Fig. 7. Figure 7 confirms that Eq. (6) is more capable, and accurate, than the other studied correlations for the estimation of bubble point pressure.

Fig. 7. Graphical comparison between the AAPRE values obtained by the new model developed in this study for the estimation of bubble point pressure as well as the corresponding studied correlations

The results and above discussion confirm that the models proposed in the present study for the evaluation of reservoir oil PVT properties, i.e., oil formation volume factor and bubble point pressure, are more reliable, and accurate than models currently available in the literature. In addition, they can have various potential applications in petroleum engineering, such as in the development of software.

4.2. Detection of outlier data points existing in the dataset

The accurate performance, and capability, of models are improved if outlier data point(s) existing in the datasets are detected and removed, because the results could be sensitive to

such data points [50-53]. Hence, the evaluation of datasets, related to both oil formation volume factor and bubble point pressure, is a requisite, since uncertainties affect the accurate performance and capability of the GEP method.

The Leverage value statistics technique was applied for the detection of outlier data points existing in the datasets, associated with both oil formation volume factor and bubble point pressure properties [50,54]. As a result, the detection of the suspended data or outliers is undertaken by means of the Williams plot, based on the H values calculated [51-52]. For more information about the Leverage approach, a detailed definition related to the computational procedure, and the equations for this technique, can be found elsewhere [51-52].

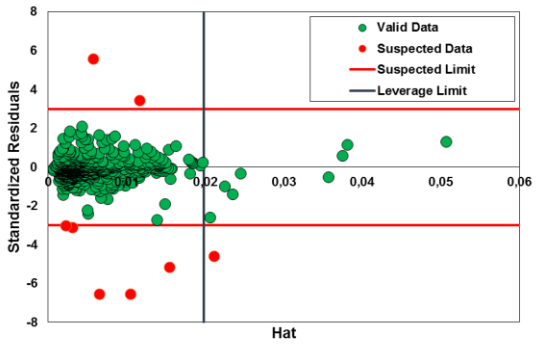


Fig. 8. Detection of the probable outlier and doubtful data of oil formation volume factor and the applicability domain of the proposed GEP model

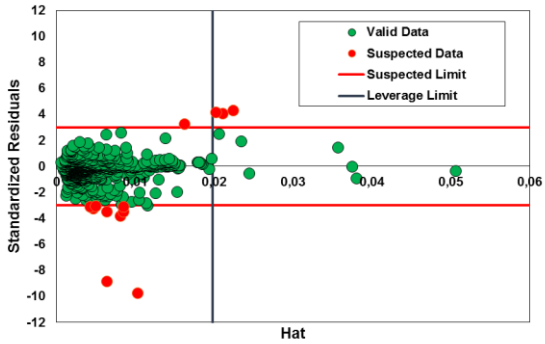


Fig. 9. Detection of the probable outlier and doubtful data of bubble point pressure and the applicability domain of the proposed GEP model

Figures 8 and 9 illustrate the Williams plots for the estimated values of oil formation volume factor and bubble point pressure, respectively, using the newly developed GEP models. As can be seen in these figures, the existence of the majority of data points in the ranges $0 \leq H \leq 0.01984$ and $-3 \leq R \leq 3$ for both of the models developed for oil formation volume factor and bubble point pressure, confirms that the applied models are statistically valid and correct in estimating these oil PVT properties. It should be noted that Figs. 8 and 9 show that there are 12 data points for the oil formation volume factor model, and 13 data points for the bubble point pressure model, compared to their corresponding actual data which are outside of the applicability domain of the GEP models, and could be viewed as outliers with doubtful values.

5. Conclusion

A novel application of the gene expression programming (GEP) method for the modeling of oil PVT properties is presented. A dataset comprising of approximately 755 laboratory PVT datasets, covering wide ranges of PVT experimental conditions from various geographical and geological regions was used for the oil formation volume factor and bubble point pressure of reservoir oil. The variables for the newly-developed models are gas gravity, oil API gravity, initial solution gas oil ratio, and reservoir temperature.

The findings indicate that the new GEP model for the calculation of oil formation volume factor has an AAPRE of 3.62% and a R2 = 0.93, and for bubble point pressure, it is 15.3% and 0.88, respectively.

In order to assess the performance and capability of the equations in estimating both oil formation volume factor and bubble point pressure properties, a comparative study was conducted, against the most widely-utilized correlations available in the open literature. The results obtained confirm that the models presented in this study are rapid to implement, accurate, and more reliable and capable than the available correlations, for the determination of both oil formation volume factor and bubble point pressure properties.

Acknowledgement

This work was based upon research supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation.

Nomenclature

GA	*genetic algorithm*
GP	*gene programming*
GEP	*gene expression programming*
ET	*expression tree*
GOR	*gas oil ratio*
OFVF	*oil formation volume factor*
T_R	*reservoir temperature*
P_b	*bubble point pressure*
B_{ob}	*oil formation volume factor at bubble point pressure*
R_{Si}	*initial solution gas oil ratio*
γ_g	*gas gravity*
API	*oil gravity*
R^2	*correlation coefficient*
RMSE	*root mean square errors*
APRE	*average percent relative error*
AAPRE	*average absolute percent relative error*

References

[1] De Ghetto G, and Villa M. Reliability analysis on PVT correlations. European Petroleum Conference 1994. Society of Petroleum Engineers.

[2] Khoukhi A. 2012. Hybrid soft computing systems for reservoir PVT properties prediction. Computers & Geosciences, 2012; 44: 109-119.

[3] Sutton RP, and Farshad F. Evaluation of empirically derived PVT properties for Gulf of Mexico crude oils. SPE Reservoir Engineering, 1990; 5(01): 79-86.

[4] El-Sebakhy EA. Data mining in forecasting PVT correlations of crude oil systems based on Type1 fuzzy logic inference systems. Computers & Geosciences, 2009; 35(9): 1817-1826.

[5] Saleh A, Mahgoub I, and Asaad Y. Evaluation of Empirically Derived PVT Properties for Egyptians Oils. Middle East Oil Show 1987.

[6] Alimadadi F, Fakhri A, Farooghi D, and Sadati H. Using a committee machine with artificial neural networks to predict PVT properties of Iran crude oil. SPE Reservoir Evaluation & Engineering, 2011; 14(01): 129-137.

[7] Standing M. A pressure-volume-temperature correlation for mixtures of California oils and gases. Drilling and Production Practice, 1947: 275-285.

[8] Katz DL. Prediction of the shrinkage of crude oils. Drilling and Production Practice, 1942: 137-147.

[9] Vazquez M, and Beggs HD. Correlations for fluid physical property prediction. Journal of Petroleum Technology, 1980; 32(06): 968-970.

[10] Glaso O. Generalized pressure-volume-temperature correlations. Journal of Petroleum Technology, 1980; 32(05): 785-795.

[11] Al-Marhoun MA. New Correlation for formation Volume Factor of oil and gas Mixtures. Journal of Canadian Petroleum Technology, 1992; 31(3): 22-26.

[12] Dokla ME, and Osman ME. Correlation of PVT properties for UAE crudes. SPE formation evaluation, 1992; 7(1): 41-46.

[13] Petrosky Jr, G, and Farshad F. Pressure-Volume-Temperature Correlations for Gulf of Mexico Crude Oils. SPE Reservoir Evaluation & Engineering, 1998; 1(05): 416-420.

[14] Arabloo M, Amooie M-A, Hemmati-Sarapardeh A, Ghazanfari M-H, and Mohammadi AH. Application of constrained multi-variable search methods for prediction of PVT properties of crude oil systems. Fluid Phase Equilibria, 2014; 363: 121-130.

[15] Boukadi F, Al-Alawi S, Al-Bemani A, and Al-Qassabi S. Establishing PVT correlations for Omani oils. Petroleum science and technology, 1999; 17(5-6): 637-662.

[16] Elsharkawy AM., Modeling the properties of crude oil and gas systems using RBF network, Asia Pacific oil & gas conference 1998, pp. 35-46.

[17] Gharbi RB, and Elsharkawy A. Neural Network Model for Estimating The PVT Properties of Middle East Crude Oils, Middle East Oil Show and Conference 1997. Society of Petroleum Engineers.

[18] Gharbi RB, and Elsharkawy AM. Universal Neural Network Based Model for Estimating The PVT Properties of Crude Oil Systems, SPE Asia Pacific Oil and Gas Conference and Exhibition 1997. Society of Petroleum Engineers.
[19] Ferreira, C., 2001. Gene expression programming: a new adaptive algorithm for solving problems. Complex Systems, 13(2): 87-129.
[20] McCain WD. The properties of petroleum fluids. PennWell Books 1990.
[21] Bandyopadhyay P, and Sharma,A. Development of a new semi analytical model for prediction of bubble point pressure of crude oils. Journal of Petroleum Science and Engineering, 2011; 78(3): 719-731.
[22] Obanijesu E, and Araromi D. Predicting bubble-point pressure and formation-volume factor of Nigerian crude oil system for environmental sustainability. Petroleum Science and Technology, 2008; 26(17): 1993-2008.
[23] Ostermann R, and Owolabi O. Correlations for the reservoir fluid properties of Alaskan crudes, SPE California Regional Meeting 1983. Society of Petroleum Engineers.
[24] Petrosky Jr G, and Farshad F. Pressure-volume-temperature correlations for Gulf of Mexico crude oils, SPE Annual Technical Conference and Exhibition 1993. Society of Petroleum Engineers.
[25] Al-Shammasi A. Bubble point pressure and oil formation volume factor correlations, SPE Middle East oil show & conference 1999, pp. 241-256.
[26] Kartoatmodjo T, and Schmidt Z. Large data bank improves crude physical property correlations. Oil and Gas Journal, 1994; 92(27).
[27] Macary S, and El-Batanoney M. Derivation of PVT correlations for the Gulf of Suez crude oils. Sekiyu Gakkai Shi, 1993; 36(6): 472-478.
[28] Mahmood MA, and Al-Marhoun MA. Evaluation of empirically derived< i> PVT</i> properties for Pakistani crude oils. Journal of Petroleum Science and Engineering, 1996; 16(4): 275-290.
[29] Kamari A, Bahadori A, Mohammadi AH, and Zendehboudi S. A reliable model for estimating a hydrate inhibitor injection rate. Journal of Natural Gas Science and Engineering, 2014; 32(23), 2837-2844.
[30] Kamari A, Gharagheizi F, Bahadori A, and Mohammadi AH. Determination of the equilibrated calcium carbonate (calcite) scaling in aqueous phase using a reliable approach. Journal of the Taiwan Institute of Chemical Engineers, 2014; 45: 1307-1313.
[31] Kamari A, Safirii A, and Mohammadi AH. Compositional Model for Estimating Asphaltene Precipitation Conditions in Live Reservoir Oil Systems. Journal of Dispersion Science and Technology, 2015; 36: 301-309.
[32] Moghadam JN, Salahshoor K, and Kharrat R. Introducing a new method for predicting PVT properties of Iranian crude oils by applying artificial neural networks. Petroleum Science and Technology, 2011; 29(10): 1066-1079.
[33] Obomanu D, and Okpobir, G. Correlating the PVT properties of Nigerian crudes. Journal of energy resources technology, 109(4): 214-217.
[34] Bello O, Reinicke K, and Patil P. Comparison of the performance of empirical models used for the prediction of the PVT properties of crude oils of the niger delta. Petroleum Science and Technology, 2008; 26(5): 593-609.
[35] Omar M, and Todd A. Development of new modified black oil correlations for Malaysian crudes, SPE Asia Pacific oil and gas conference 1993, Society of Petroleum Engineers.
[36] Al-Marhoun MA. PVT correlations for Middle East crude oils. Journal of Petroleum Technology, 1988; 40(05): 650-666.
[37] Mitchell, M., 1998. An introduction to genetic algorithms (complex adaptive systems).
[38] Ferreira C. Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence (Studies in Computational Intelligence). Springer-Verlag New York 2006, Inc., Secaucus, NJ.
[39] Teodorescu L, and Sherwood D. High energy physics event selection with gene expression programming. Computer Physics Communications, 2008; 178(6): 409-419.
[40] Alavi AH, Gandomi,AH, Mousavi M, and Mollahasani A. High-Precision Modeling of Uplift Capacity of Suction. Geomechanics and Engineering, 2010; 2: 253-280.
[41] Frashad F, LeBlanc J, Garber J, and Osorio J. Empirical PVT correlations for Colombian crude oils, SPE Latin America/Caribbean petroleum engineering conference 1996. Society of Petroleum Engineers, Port-of-Spain, Trinidad.
[42] Dindoruk B, and Christman PG. PVT properties and viscosity correlations for Gulf of Mexico oils. SPE Reservoir Evaluation & Engineering, 2004; 7(06): 427-437.
[43] Abdul-Majeed GH, Salman NH, and Scarth B. An empirical correlation for oil FVF prediction. Journal of Canadian Petroleum Technology, 1988; 27(06).

[44] Labedi RM. Use of Production Data to Estimate the Saturation Pressure Solution Gor and Chemical Composition of Reservoir Fluids, SPE Latin America Petroleum Engineering Conference 1990. Society of Petroleum Engineers, Brazil.
[45] Lasater J. Bubble point pressure correlation. J. of Petroleum Technology, 1958; 10(05): 65-67.
[46] Valko P, and McCain Jr W. Reservoir oil bubblepoint pressures revisited; solution gas –oil ratios and surface gas specific gravities. Journal of Petroleum Science and Engineering, 2003; 37(3): 153-169.
[47] Velarde J, Blasingame T, and McCain Jr W. Correlation of Black Oil Properties At Pressures Below Bubble Point Pressure-A New Approach, Annual Technical Meeting. Petroleum Society of Canada 1997.
[48] Yi X. Using wellhead sampling data to predict reservoir saturation pressure, SPE Permian Basin Oil and Gas Recovery Conference 2000, Society of Petroleum Engineers.
[49] Ikiensikimama SS, and Ogboja O. New bubblepoint pressure empirical PVT correlation, Nigeria Annual International Conference and Exhibition 2009. Society of Petroleum Engineers.
[50] Gramatica P. Principles of QSAR models validation: internal and external. QSAR & combinatorial science, 2007; 26(5): 694-701.
[51] Mohammadi AH, Eslamimanesh A, Gharagheizi F, and Richon D. A novel method for evaluation of asphaltene precipitation titration data. Chemical Engineering Science, 2012; 78: 181-185.
[52] Mohammadi AH, Gharagheizi F, Eslamimanesh A, and Richon D. Evaluation of experimental data for wax and diamondoids solubility in gaseous systems. Chemical Engineering Science, 2012; 81: 1-7.
[53] Rousseeuw PJ, and Leroy AM. Robust Regression And Outlier Detection, John Wiley & Sons, Inc. New York, NY, USA 2005, ISBN: 978-0-471-48855-2
[54] Goodall CR. 13 Computation using the QR decomposition. Handbook of Statistics, 1993; 9: 467-508.

To whom correspondence should be addressed: Professor Amir H. Mohammadi, E-mail:amir_h_mohammadi@yahoo.com and Professor Deresh Ramjugernath, Thermodynamics Research Unit, School of Engineering, University of Kwa-Zulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa, E-mail ramjuger@ukzn.ac.za