Petroleum and Coal

Article **Open Access**

IMPROVED METHODS FOR THE DETERMINATION OF VISCOSITIES OF DEAD, SATURATED, AND UNDER-SATURATED RESERVOIR 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, USA
[2] Thermodynamics Research Unit, School of Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa

Received June 4, 2019; Accepted July 31, 2019

Abstract

The viscosity of crude oil is an important parameter governing fluid flow in both porous media and pipelines. The experimental measurement of reservoir oil viscosity is time consuming, difficult, and costly. Therefore, it is necessary to develop reliable models to determine viscosity. In this study, three models for dead, saturated, and under-saturated reservoir oils viscosities were developed, based on the gene expression programming (GEP) computational scheme. A dataset comprising more than 1000 sets of experimental pressure-volume-temperature (PVT) measurements drawn from Iranian oil fields were utilized in the model development. The dataset contains variables, which are regarded as significant and identifiable, namely, the gravity of oil, reservoir temperature, solution gas oil ratio, and bubble point pressure. The models developed are compared to a number of empirically-derived correlations. A comparison of results, using the GEP-based models for dead, saturated, and under-saturated viscosities was also undertaken with respect to experimental data. There are suitable qualitative and quantitative agreements between the models results and experimental data. The average absolute relative deviations for the dead, saturated and under-saturated oil viscosity models are 17.42, 13.55, and 1.47 %, respectively. The models are simple to use and enable fairly rapid estimations.

Keywords: *Viscosity; Pressure region; Gene expression programming (GEP); Empirical correlation; Enhanced oil recovery (EOR).*

1. Introduction

Reservoir oil viscosity, simply put, is a fluid's resistance to flow through porous media. The calculation of crude oil viscosity is necessary for a wide range of engineering interventions. For example, it is required for the design of enhanced oil recovery (EOR) processes, the evaluation of fluid flow rate through porous media, the estimation of hydrocarbon reserves, equipment design, prediction of reservoir performance, and the development of reservoir and production simulation software [1-7]. Furthermore, this parameter plays a key role in estimating the deposition of wax in transportation pipelines [8].

The oil production capacity of hydrocarbon reservoirs depends on the viscosity of the oil, with low viscosities allowing for higher production yields [9]. Therefore, in the area of thermal EOR, most processes, including steam injection, hot water injection, and steam assisted gravity drainage, are closely associated with reducing the viscosity of heavy oil. Thus, it is clear that the advancement of reliable methods, including laboratory measurements and empirically derived models, are required to enhance the measurement and understanding of crude oil viscosity in oil reservoirs.

While the viscosity of reservoir oil is influential, it is itself dependent on a number of thermophysical factors, including, reservoir pressure, reservoir temperature, solution gas–oil ratio, bubble point pressure, the gravity of gas and oil, and the composition of the oil mixture [10-11].

Normally, the measurement of reservoir oil viscosity is undertaken by conducting experimental measurements that simulate reservoir conditions under particular temperature conditions. Nevertheless, experimental measurement of reservoir oil viscosity at different temperatures is sometimes prohibitive, due to the high price of sampling equipment and the related tests [10]. Therefore, empirically derived methods, or modelling, can be useful in estimating viscosity when experimental data on the specific reservoir is not available.

In line with their pressure regime, there are several empirical and semi-empirical correlations available for dead, saturated, and under-saturated oil, which will be described in the next section. In addition, there are some models for the prediction of viscosity on the basis of corresponding states method [12-14]. However, the corresponding states-based method has not gained wide acceptance, because it requires both complex calculations/computations, and the composition of fluid, for viscosity estimation [7, 15].

In addition to the corresponding states method, smart techniques, like neural networks and support vector machine approaches, are applied to determine reservoir oil viscosities. Obanijesu and Omidiora [16] developed an artificial neural network model for the determination of viscosities of Nigerian crudes. Hemmati-Sarapardeh *et al.* [17] proposed a smart model based on a least squares version of the support vector machines' mathematical scheme, for the estimation of Iranian oil field viscosities. There is agreement between the experimental data and the results obtained by these recent models. But these methods have drawbacks: they sometimes have an over-fitting problem; they normally require large databases and; they do not provide a symbolic equation for future use [18-20].

This paper describes three innovative models, based on the gene expression programming (GEP) [21] computational scheme. They are designed to determine, in a consistent manner, the dead, saturated, and under-saturated viscosities of crude oils. In order to develop such models, more than 1000 sets of experimental PVT data, drawn from Iranian oil fields, were used. Many empirically-derived correlations were used to compare the results of the models, and to demonstrate the accuracy and simplicity of the models. In addition, the Leverage methodology was applied to detect suspended and outlier data points in the datasets of dead, saturated and under-saturated crude oils viscosities.

2. The viscosities of reservoir oils

By conducting laboratory tests on bottom hole cores or surface recombined samples, the viscosities of reservoir oils can be determined isothermally, at reservoir temperature and at different reservoir pressures [16].

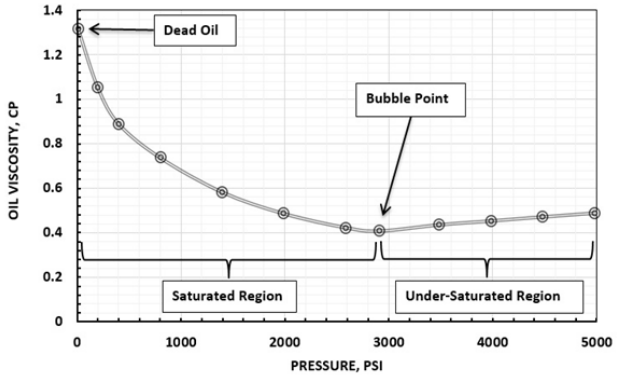


Fig. 1. A representative trend plot of viscosity versus pressure illustrating regions related to dead, saturated, and under-saturated oils viscosities

The crude oils viscosities depend on the reservoir pressure and temperature. The viscosity, at above bubble point pressure, increases and, below bubble point pressure, reduces, in relation to changes in pressure [10]. Thus, a specific correlation is required for each pressure regime due to differences in the qualities of oil in the various regions [17]. As a result, reservoir oil viscosity is determined in three pressure regimes: in the regimes above and below the bubble point pressure, and in dead oil (gas-free reservoir oil) (Fig. 1).

Different empirical correlations have been developed over the years for dead, saturated and, under-saturated crude oils. In this study, several empirical correlations related to dead oils [2, 7, 15, 22-32], saturated oils [2, 7, 15, 23, 26, 27, 29, 31, 33-35] and under-saturated oils [2, 15, 22, 27, 29, 31, 34, 36, 37] were used for comparison to the models developed. Tables 1-3 list these correlations, as well as their applicability ranges, data origin, and formulas.

Table 1. List of the previously reported dead oil viscosity models investigated in this study as well as their applicability ranges and formulas

Method	Origin of Data	T (K)	API	μ_{od} (cP)	Formula
Beal [22]	US	310-394	10.1-52	0.865-1550	$\mu_{od}=\left(0.32+1.8\times\frac{10^7}{API^{4.53}}\right)\left(\frac{360}{T+200}\right)^x;x=e^{[2.302585\left(0.43+\frac{8.33}{API}\right)]}$
Beggs and Robinson [23]	-	294-419	16-58	-	$\mu_{od}=10^x-1;x=e^{[3.0324-0.02023API]}\ T^{-1.163}$
Glaso [24]	North Sea	283-422	20-48	0.6-39	$\mu_{od}=\left(\frac{3.141\times10^{10}}{T^{3.44}}\right)\log(API)^{[0.313\log(T)-36.447]}$
Kaye [25]	Offshore California	334-412	7-41	-	$\mu_{od}=10^{[T^{-0.65}10^{(2.203-0.0254API)}]}-1,API\le12;\ \mu_{od}=10^{[T^{-0.65}10^{(2.305-0.03354API)}]}-1,API>12;$
Al-Khafaji *et al.* [26]	-	289-422	15-51	-	$\mu_{od}=\frac{10^{(4.9563-0.00488T)}}{(API+\frac{T}{30}-14.29)^{2.709}}$
Petrosky [27]	Gulf of Mexico	319-415	25-46	0.72-10.25	$\mu_{od}=\frac{2.3511\times10^7}{T^{2.10255}}\log(API)^{[4.59388\log(T)-22.82792}$
Egbogah and Ng [28]	-	288-353	5-58	-	$\mu_{od}=10^x-1;x=10^{[1.8653-2.5086\times10^{-2}\ API-0.56441\log(T)]}$
Labedi [2]	Libya	311-425	32-48	0.66-4.79	$\mu_{od}=10^{9.224}\ API^{-4.7013}\ T^{-0.6739}$
Kartoatmodjo and Schmidt [29]	Worldwide	300-433	14-59	0.5-586	$\mu_{od}=\frac{1.6\times10^9}{T^{2.8177}}\log(API)^{[5.7526\log(T)-26.9718]}$
Bennison [30]	North Sea	277-422	11-20	6.4-8396	$\mu_{od}=10^{[-0.8021API+23.8765]}\ T^{[0.31458API-9.21592]}$
Elsharkawy and Alikhan [15]	Middle East	311-422	20-48	0.6-33.7	$\mu_{od}=10^x-1;x=e^{[2.16924-0.02525API-0.68875\log(T)]}$
Hossain *et al.* [31]	Worldwide	273-375	7-22	12-451	$\mu_{od}=10^{[-0.71523API+22.13766]}\ T^{[0.269024API-8.268047]}$
Naseri *et al.* [7]	Iran	314-421	17-44	0.75-54	$\mu_{od}=10^{[11.2699-4.2699\log(API)-2.052\log(T)]}$
Alomair *et al.* [32]	Kuwait	293-433	10-20	1.78-11360	

Table 2. List of the previously reported saturated oil viscosity models investigated in this study as well as their applicability ranges and formulas

Method	Source of data	Solution GOR (SCF/STB)	Saturation pressure (MPa)	μ_{od}, cP	Formula
Chew and Connally I [33]	US	51-3544	0.91-38.92	0.37-50	$\mu_{ob} = A\mu_{od}^{B}; A = 0.2 + \frac{0.8}{10^{0.00081R_s}}; B = 0.43 + \frac{0.57}{10^{0.00072R_s}}$
Chew and Connally II [33]	US	51-3544	0.91-38.92	0.37-50	$\mu_{ob} = A\mu_{od}^{B}; A = 10^{[R_s(2.2\times10^{-7}R_s - 7.4\times10^{-4})]};$ $B = 0.68\times10^{[-8.62\times10^{-5}R_s]} + 0.25\times10^{[-1.1\times10^{-3}R_s]} + 0.062\times10^{[-3.74\times10^{-3}R_s]}$
Chew and Connally III [33]	US	51-3544	0.91-38.92	0.37-50	$\mu_{ob} = A\mu_{od}^{B}$ $A = 0.987583 - 0.1746773\times10^{-2}R_s + 0.2067531\times10^{-5}{R_s}^2 - 0.1310529\times10^{-8}{R_s}^3 + 0.3229416\times10^{-12}{R_s}^4$ $B = 0.9900216 - 0.112183\times10^{-2}R_s + 0.1427879\times10^{-5}{R_s}^2 - 0.9440539\times10^{-9}{R_s}^3 + 0.2312365\times10^{-12}{R_s}^4$
Beggs and Robinson [23]	-	20-2070	0.91-36.30	-	$\mu_{ob} = A\mu_{od}^{B}; A = \frac{10.715}{(R_s+100)^{0.515}}; B = \frac{5.44}{(R_s+150)^{0.338}}$
Al-Khafaji *et al.* [26]	-	0-2100	-	-	$\mu_{ob} = A\mu_{od}^{B}; X_1 = \log(R_s)$ $A = 0.247 + 0.2824X_1 + 0.5657X_1^2 - 0.4065X_1^3 + 0.0631X_1^4$ $B = 0.894 + 0.0546X_1 + 0.07667X_1^2 - 0.0736X_1^3 + 0.01008X_1^4$
Khan *et al*. [34]	Saudi Arabia	24-1901	0.74-29.75	0.13-77.4	$\mu_{ob} = \frac{0.09\gamma_g^{0.5}}{R_{sb}^{\frac{1}{3}}[\frac{T+459.67}{459.67}]^{4.5}(1-\gamma_o)^3}$; if $P < P_b \rightarrow \mu_o = \frac{\mu_{ob}\, e^{[-2.5\times10^{-4}(P-P_b)}}{(\frac{P}{P_b})^{0.14}}$
Petrosky [27]	Gulf of Mexico	21-1855	10.85-65.85	0.21-7.4	$\mu_{ob} = A\mu_{od}^{B}; A = 0.1651 + \frac{0.6165}{10^{(6.0866\times10^{-4}R_s)}}; B = 0.5131 + \frac{0.5109}{10^{(1.1831\times10^{-3}R_s)}}$
Labedi [2]	Libya	13-3533	0.41-43.83	0.115-3.72	$\mu_{ob} = \frac{10^{(2.344-0.03542API)}\mu_{od}^{0.6447}}{P_b^{0.426}}$; at $P < P_b \rightarrow \mu_o = \frac{\mu_{ob}}{1-(10^{-3.876}P_b^{05423}API^{1.1302}(1-\frac{P}{P_b})}$
Kartoatmodjo and Schmidt [29]	Worldwide	2.3-572	0.10-41.74	0.1-6.3	$\mu_{ob} = -0.06821 + 0.9824X_2 + 4.034\times10^{-4}X_2^2;$ $X_1 = 0.43 + 0.5156\times10^{[-8.1\times10^{-4}R_s]}; X_2 = [0.2001 + 0.8428\times10^{[-8.45\times10^{-4}R_s]}]\mu_{od}^{X_1}$
Elsharkawy and Alikhan [15]	Middle East	10-3600	0.69-25.51	0.05-21	$\mu_{ob} = A\mu_{od}^{B}; A = \frac{1241.935}{(R_s+641.026)^{1.12410}}; B = \frac{1768.841}{(R_s+1180.335)^{1.06622}}$
Hossain *et al*. [31]	Worldwide	19-493	0.83-43.24	3.6-360	$\mu_{ob} = A\mu_{od}^{B}$ $A = 1 - 0.001718831R_s + 1.58081\times10^{-6}{R_s}^2$; $A = 1 - 0.002052461R_s + 3.47559\times10^{-6}{R_s}^2$
Naseri *et al*. [7]	Iran	255-4116	2.89-40.68	0.11-18.15	$\mu_{ob} = 10^{1.1145}P_b^{-0.4956}\mu_{od}^{0.9961}$
Bergman and Sutton [35]	Worldwide	6-6525	0.45-71.02	0.21-4277	$\mu_{ob} = A\mu_{od}^{B}; A = \frac{1}{1+(\frac{R_s}{344.198})^{0.855344}}; B = \frac{0.617677}{1+(\frac{R_s}{567.953})^{0.819326}}$

Table 3. List of the previously reported under-saturated oil viscosity models investigated in this study as well as their applicability ranges and formulas

Method	Origin of data	P (MPa)	P_b (MPa)	μ_{ob} (cP)	μ_o (cP)	Formula
Beal [22]	USA	-	-	0.142-127	0.16-315	$\mu_o = \mu_{ob} + [0.001(P - P_b)](0.024\mu_{ob}^{1.6} + 0.038\mu_{ob}^{0.56})$
Vazquez and Beggs [36]	Worldwide	0.87-65.50	-	-	0.117-148	$\mu_o = \mu_{ob}\, e^{[(5.50318\times10^{-5}+3.77163\times10^{-5}\mu_{ob}^{0.278})(P-P_b)]}$
Khan *et al*. [34]	Saudi Arabia	-	0.74-33.05	0.13-77.4	0.13-71	$\mu_o = \mu_{ob}\, e^{[9.6\times10^{-5}(P-P_b)]}$
Petrosky [27]	Gulf of Mexico	11.03-70.67	10.85-65.86	0.211-3.546	0.22-4.09	$\mu_o = \mu_{ob} + 1.3449 \times 10^{-5}(P - P_b)10^{X_2}; X_1 = \log(\mu_{ob}); X_2 = -1.0146 + 1.3322X_1 - 0.4876X_1^2 - 1.15036X_1^3$
Labedi [2]	Libya	-	0.41-43.84	0.115-3.72	-	$\mu_o = \mu_{ob} + \frac{\mu_{od}^{0.9036}P_b^{0.6151}}{10^{(2.488+0.01976\,API)}}(\frac{P}{p_b} - 1)$
Orbey and Sandler [37]	-	5.10-100.00	-	0.217-3.1	0.225-7.3	$\mu_o = \mu_{ob}\, e^{[\alpha(P-P_b)]}; \alpha = 6.89 \times 10^{-5}$
Kartoatmodjo and Schmidt [29]	Worldwide	0.17-41.47	0.17-32.92	0.168-184.86	0.168-517.03	$\mu_o = 1.0081\mu_{ob} + 1.127 \times 10^{-3}(P - P_b)(-6.517 \times 10^{-3}\mu_{ob}^{1.8148} + 0.038\mu_{ob}^{1.59})$
Elsharkawy and Alikhan [15]	Middle East	8.87-68.94	-	-	0.2-5.7	$\mu_o = \mu_{ob} + \frac{10^{-2.0771}(P - P_b)\mu_{od}^{1.19279}}{\mu_{ob}^{0.40712}P_b^{0.7941}}$
Hossain *et al*. [31]	Worldwide	2.07-23.44	0.83-43.24	3.6-360	3-517	$\mu_o = \mu_{ob} + [0.004481\,(P - P_b)](0.555955\,\mu_{ob}^{1.068099} - 0.527737\,\mu_{ob}^{1.063547})$

To develop these types of correlations, the common reservoir fluid data, including the oil gravity (API), solution gas oil ratio (GOR), bubble point pressure, and reservoir temperature, were used. Comparative correlations were used in this study to detemine the most significant variables to apply for each of the three models. Reservoir temperature and crude oil API gravity were found to be essential for the estimation of dead oil viscosity. Dead oil viscosity and saturated pressure were essential parameters needed for saturated reservoir oil viscosity determination. Reservoir pressure, as well as bubble point pressure and viscosity at bubble point were essential variables for accurate estimation of under-saturated reservoir oil viscosity.

As a result, the reservoir parameters used to develop the GEP-based models for the determination of dead, saturated, and under-saturated crude oil viscosities are as follows:

$$\mu_{od} = f_1(T_R,\ API) \quad (1)$$

$$\mu_{ob} = f_2(\mu_{od},\ P) \quad (2)$$

$$\mu_o = f_3(\mu_{ob},\ P,\ P_b) \quad (3)$$

The above equations were applied to more than 1000 sets of experimental data of Iranian oil fields, comprising the gravity of oil, reservoir temperature, solution gas oil ratio, and bubble point pressure. A rolling ball viscometer (Ruska, series 1602) was used to measure the Iranian reservoir oil viscosities at different pressures above and below saturation pressure.

Tables 4-6 list the ranges of the applied variables used to develop the GEP-based models for the estimation of dead, saturated, and under-saturated oil viscosities, respectively. It is worth noting that the range of data presented in Tables 4-6 includes almost all of the data available, related to Iranian oil fields.

Table 4. Applicability ranges and units of the applied variables for developing the dead oil viscosity model

Reservoir property	Unit	Min.	Max.	Avg.	SD	Type
Temperature	°F	50.27	290.26	176.11	42.84	Input
Oil API gravity	-	17.30	43.56	29.32	7.00	Input
Dead oil viscosity	cP	0.55	69.50	7.41	11.44	Output

Table 5. Applicability ranges and units of the applied variables for developing the saturated oil viscosity model

Reservoir property	Unit	Min.	Max.	Avg.	SD	Type
Saturation pressure	psi	158.09	5 701.43	1 705.64	7.50	Input
Dead oil viscosity	cP	0.55	37.18	4.55	5.05	Input
Saturated oil viscosity	cP	0.18	25.58	1.92	2.59	Output

Table 6. Applicability ranges and units of the applied variables for developing the under-saturated oil viscosity model

Reservoir property	Unit	Min.	Max.	Avg.	SD	Type
Bubble point pressure	psi	729.53	5 115.47	1 135.39	7.81	Input
Bubble point viscosity	cP	0.18	18.16	1.62	2.19	Input
Pressure	MPa	5.03	86.18	25.38	11.63	Input
Under-saturated oil viscosity	cP	0.18	31.00	1.84	2.97	Output

3. Methodology

Gene expression programing, which was first proposed by Ferreira [21], is regarded as one of the most reliable artificial intelligence techniques. Along with the use of evolutionary algorithms, computation-based models can be built. A search and tuning method was applied in the GEP mathematical approach. This is employed for selecting, and it enables an evolutionary process to occur [38]. Furthermore, a number of soft-computing models have been applied in the GEP algorithm, in order to solve a specific symbolic regression problem based on some fitness functions.

In the GEP algorithm, individuals are recognized as fixed length linear strings, and then are presented as nonlinear entities of various size and shape [39]. The individuals are composed of multi-genic chromosomes, so that each GEP gene includes two distinct sections, head and tail. The head domain involves symbols from a function set and a terminal set, and the tail section involve symbols composed of only a terminal set as follows [40]:

$$t = h \times (n - 1) + 1 \quad (4)$$

where: *t* and *h* are symbol sets (function and terminal), and *n* denotes the gene number.

To illustrate the performance of the GEP approach mathematically, a simple GEP-based model including a chromosome composed of two genes connected together by a multiplication fitness function is presented as follows:

$$\left(\frac{x}{y}\right) * (r - z) \quad (5)$$

where: x, y, r and z are recognized as input variable to predict the target parameter; and - , * and – denote the fitness functions of division, multiplication and deduction, respectively. It should be noted that there are more fitness functions in the GEP mathematical approach including +, ln, log, power, and root square.

There are certain important steps to follow in order to perform a symbolic regression by means of a gene expression programing algorithm. These steps require randomly generating an initial population related to the chromosomes by means of Karva language and; translating the chromosomes into soft-computing models. In order to complete the models, it is necessary to select proper fitness functions; apply these fitness functions to evaluate the performance of the generated computer programs or models; check the final computational model in terms of accuracy and simplicity; use the most suitable individual population selected to reproduce the chromosomes; propose a new, generated soft-computing model, as a consequence of reproducing the chromosomes and, finally; repeat the above-mentioned steps until a more reliable and accurate soft-computing symbolic model is obtained [38]. In order to develop the model, three distinct datasets for dead, saturated, and under-saturated oil viscosities were prepared. In order to evaluate the capability of the models developed, 80% of the experimental data points in each database was used to train the models, and 20% of the data was taken to test the models' predictive ability.

4. Results and discussion

The results will be discussed in three sections, in line with the nature of the three models that were developed, namely: dead oil, saturated, and under-saturated oil viscosity.

Normally, in the GEP-based symbolic regression approach, the most effective input variables with respect to the target parameter, are selected automatically from the set of all of the independent predictors. In this study however, the algorithm in each case was fixed on the most significant variables, obtained from a comprehensive review of the literature, as described in Section 2 above.

4.1. Dead oil viscosity model

Reservoir temperature and crude oil API gravity are required for the estimation of dead oil viscosity. Consequently, using the GEP algorithm, the best GEP-based model, based on the independent predictors (reservoir temperature and oil gravity), was obtained after the GEP evolution process, which is given as follow:

$$\mu_{od} = \frac{614.82\ API \times T - 63529.0\ T + 2.0359 \times 10^7}{T \times API^3 - 482088} \quad (6)$$

where: μ_{od} denotes the dead oil viscosity (cP); *T* represents reservoir temperature (°F) and finally, *API* stands for dead oil gravity.

Table 7 (top section), summarizes the statistical error factors, including average absolute relative deviation (AARD), root mean square error (RMSE) and R-squared error (R^2), for the model associated with dead oil viscosities. The table reports reasonable values for these error parameters, for example, AARD = 17.29%, RMSE = 1.82, and R^2 = 0.97.

Table 7. Summarized statistical error factors including AARD, RMSE and R^2 for the models developed for dead, saturated and under-saturated oil viscosities

Developed model	AARD [a] %	RMSE [b]	R2 [c]
Dead oil viscosity model			
Training	17.40	1.95	0.974
Testing	16.86	1.16	0.900
Overall	17.29	1.82	0.973
Saturated oil viscosity model			
Training	14.06	0.53	0.966
Testing	11.53	0.26	0.961
Overall	13.55	0.49	0.966
Under-saturated oil viscosity model			
Training	1.51	0.10	0.999
Testing	1.29	0.06	0.998
Overall	1.47	0.09	0.999

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

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

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

In order to graphically illustrate the performance of the model, a comparison between the calculated values of dead oil viscosity using Eq. (6), and the actual values from the dataset used is demonstrated in terms of a parity plot, or crossplot, as well as a relative error distribution (Fig. 2).

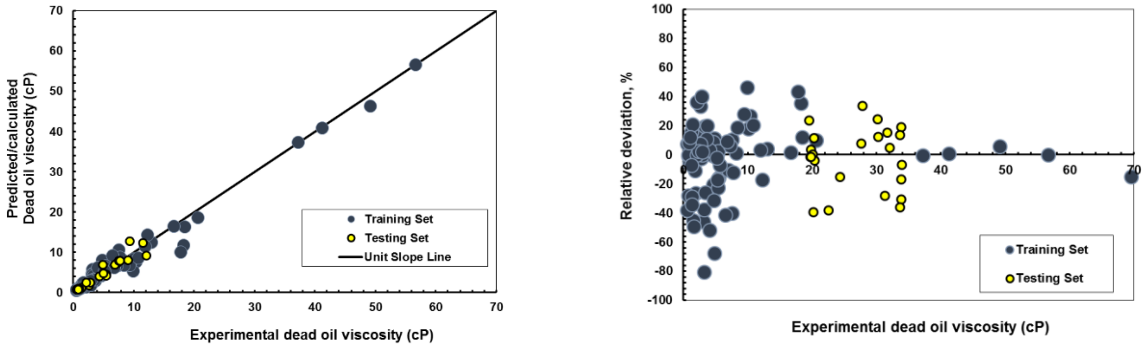


Fig. 2. Parity plot and relative deviation distribution plot for the dead oil viscosity model

Furthermore, in order to evaluate and demonstrate the accuracy and capability of the model for dead oil viscosity, a comprehensive comparative study was conducted by applying the most reliable empirically-derived correlations available in the literature [2, 7, 15, 22-32]. Table 8 reports the results of the comparative study conducted for the estimation of dead oil viscosities. It is clear from the table that the model proposed for dead oil viscosities is more accurate, than the empirically-derived correlations reviewed, in terms of all error factors investigated.

Table 8. Summarized statistical error factors including AARD, RMSE and R^2 for the model developed for dead oil viscosity as well as the studied empirical correlations results from the actual data

Method	AARD %	R^2	RMSE
Beal [22]	891.2	0.1088	83.14
Beggs and Robinson [23]	216.7	0.0376	245.05
Glaso [24]	33.4	0.9270	3.84
Kaye [25]	52.0	0.2268	10.19
Al-Khafaji *et al.* [26]	29.9	0.7283	6.04
Petrosky [27]	41.6	0.8695	4.18
Egbogah and Ng [28]	55.6	0.9208	3.26
Labedi [2]	177.9	0.3910	14.95
Kartoatmodjo and Schmidt [29]	36.8	0.9065	4.50
Bennison [30]	70.9	0.6689	12.25
Elsharkawy and Alikhan [15]	72.9	0.9065	13.25
Hossain *et al.* [31]	68.9	0.6000	16.24
Naseri *et al.* [7]	27.5	0.8233	3.88
Alomair *et al.* [32]	72.4	0.8275	6.36
This study	17.2	0.9730	1.82

The comparative study shows that the work of Naseri *et al.* [7] for the determination of dead oil viscosity of Iranian crudes is second to the model proposed in this study, according to the AARD results, which are equal to 27.5 %, as shown in Fig. 3.

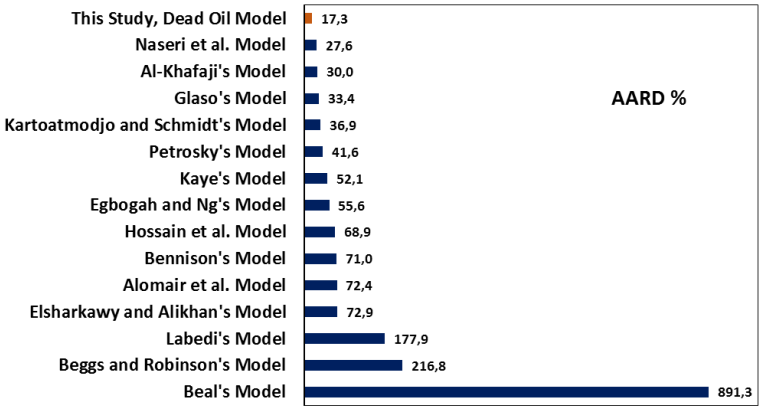


Fig. 3. Graphical comparison between the AARD % values obtained by the model developed in this work for the determination of dead oil viscosity as well as the studied comparative methods

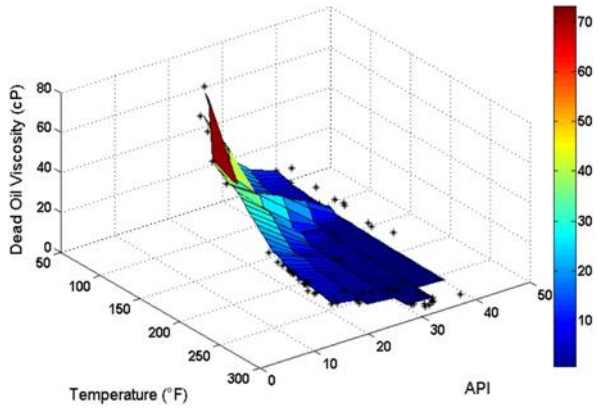


Fig. 4. 3D plot of change of dead oil viscosity versus change in temperature and oil API gravity

The smoothness of the model for dead oil viscosity is represented by means of a trend analysis, which illustrates the changes of viscosity versus input variables. To this end, a 3D plot of changes to dead oil viscosity, oil API gravity, and reservoir temperatures is presented. Fig. 4 shows that dead oil viscosity decreases with an increase in temperature and oil API gravity. The results in this figure confirm the general trend of change [41] of dead oil viscosity versus oil API gravity and temperature, which is captured using the model.

4.2. Saturated oil viscosity model

In developing a reliable model for the estimation of saturated oil viscosity of Iranian crudes using GEP methodology, pressure at bubble point as well as dead oil viscosity were found to be the most effective input variables or predictors (see section 2). As a result, the encoded GEP algorithm was run until there was no significant improvement in the efficiency, simplicity and accuracy of the equations obtained. Consequently, the final model for the determination of saturated oil viscosity which was obtained is as follow:

$$\mu_{ob} = \frac{3.5752\,\mu_{od} + 1.9812}{145.0377\,P_b - 0.1379\,\mu_{od} + 9.4204} + 0.71308^{\mu_{od}{}^{7.3}} + 0.083617 \qquad (7)$$

where: μ_{ob} is the saturated oil viscosity (cP); μ_{od} expresses the dead oil viscosity (cP); P_b stands for the bubble point pressure (psi).

The middle section of Table 7 lists the statistical error parameters calculated for Eq. (7). The results show that the equation proposed in this study, for the determination of saturated oil viscosity, has an acceptable accuracy with an AARD %, RMSE and R^2 equal to 13.55, 0.49 and 0.96, respectively.

Fig. 5 illustrates the estimated saturated oil viscosity versus the experimental data points. The top view is a crossplot, and the bottom view demonstrates the relative error distribution plot of saturated oil viscosity. As can be seen, there is satisfactory agreement between the estimated and experimental saturated oil viscosity data.

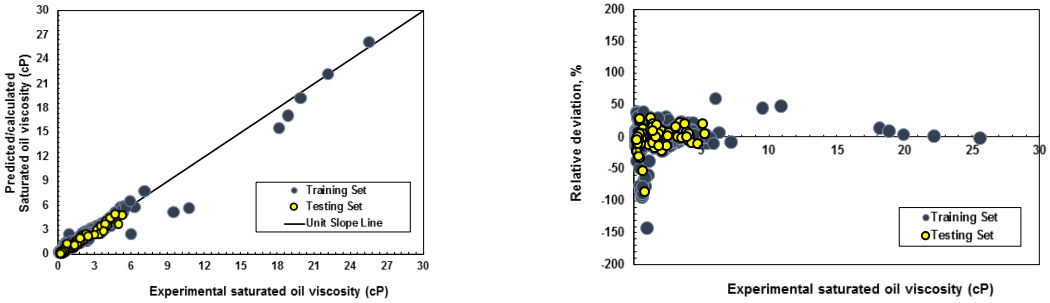


Fig. 5. Applicability ranges and units of the applied variables for developing the saturated oil viscosity model

Moreover, the results obtained were compared with empirical correlations available in open literature [2, 7, 15, 23, 26, 27, 29, 31, 33-35]. Table 9 summarizes the calculated AARD %, RMSE and R-squared values for this model along with the comparative methods studied. The table clearly reveals that Eq. (7) is the most accurate with regard to all of error factors. Additionally, Fig. 6 graphically shows an AARD % analysis of the reviewed models, which confirms the low deviation of Eq. (7) in comparison with them. As a result, it is confirmed that an increase in pressure, ultimately to the pressure at bubble point, leads to a reduction of saturated oil viscosity [17], and Fig. 7 confirms this change of saturated reservoir oil viscosities.

Table 9. Summarized statistical error factors including AARD, RMSE and R^2 for the model developed for saturated oil viscosity as well as the studied empirical correlations results from the actual data

Method	AARD %	R^2	RMSE
Chew and Connally I [33]	29.6	0.8238	1.18
Chew and Connally II [33]	39.6	0.7999	1.46
Chew and Connally III [33]	Out of range	0	Out of range
Beggs and Robinson [23]	31.9	0.5320	1.35
Al-Khafaji *et al.* [26]	20.7	0.7992	1.32
Khan *et al.* [34]	Out of range	0	Out of range
Petrosky [27]	27.3	0.7399	1.14
Labedi [2]	248.6	0.4283	3.70
Kartoatmodjo and Schmidt [29]	25.7	0.7895	1.11
Elsharkawy and Alikhan [15]	24.4	0.6576	1.20
Hossain *et al.* [31]	Out of range	0	Out of range
Naseri *et al.* [7]	52.3	0.5672	1.38
Bergman and Sutton [35]	26.7	0.7339	1.15
This study	13.5	0.9660	0.49

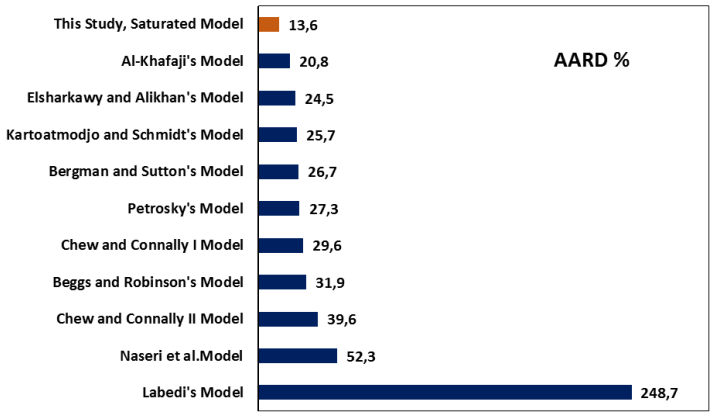


Fig. 6. Graphical comparison between the AARD % values obtained by the model developed in this work for the determination of saturated oil viscosity as well as the studied comparative methods; the obtained AARDs for Chew and Connally III [33], Khan *et al.* [34], and Hossain *et al.* [31] methods are out of range

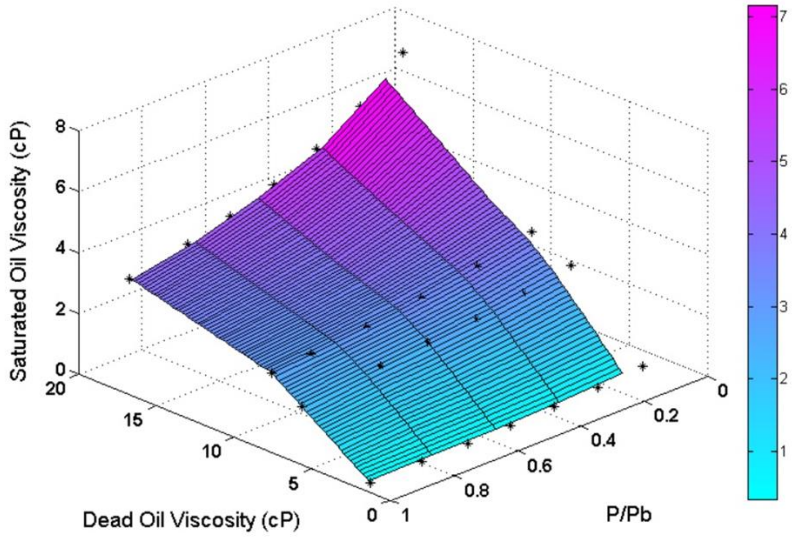


Fig. 7. 3D plot of change of saturated oil viscosity versus change in pressure ultimately to the pressure at bubble point and dead oil viscosity

4.3. Under-saturated oil viscosity model

Previously published research, mentioned in section 2, recommends pressure, saturated oil viscosity, and bubble point pressure, as the most effective reservoir parameters for the estimation of under-saturated oil viscosity. Therefore, these variables were selected as significant parameters in the development of the model. The final model, for the determination of under-saturated reservoir oil viscosity using the GEP algorithm which was obtained is as follows:

$$\mu_o = \frac{0.01115\,P}{P_b} + \frac{1.1989\times10^{-8}\ (P\times\mu_{ob})^{\,2} + 7.9372\times10^{-4}\times P\times\mu_{ob} + 10.926\,\mu_{ob}}{0.001\,P_b + 10.712} \qquad (8)$$

where: μ_o represents reservoir oil viscosity (cP); P denotes pressure (psi); P_b stands for pressure at bubble point (psi) and finally μ_{ob} is saturated oil viscosity (cP).

In Table 7 (bottom section), the statistical error factors calculated for the model associated with under-saturated oil viscosities, are reported. The AARD %, RMSE, and R^2 values calculated are 1.47, 0.09, and 0.99, respectively. From these values of deviations it can be concluded that the model is reliable for the calculation of under-saturated reservoir oil viscosities.

Fig. 8 provides a crossplot, and a relative deviation distribution plot, for the estimated under-saturated oil viscosities against the experiemntal data. As can be seen, the data points

are almost on the line of Y = X, illustrating that there is acceptable agreement between the GEP-based model calculations and the experimental data for under-saturated reservoir oil viscosities. Moreover, a low distribution of data points is observed in the bottom section of Fig. 8.

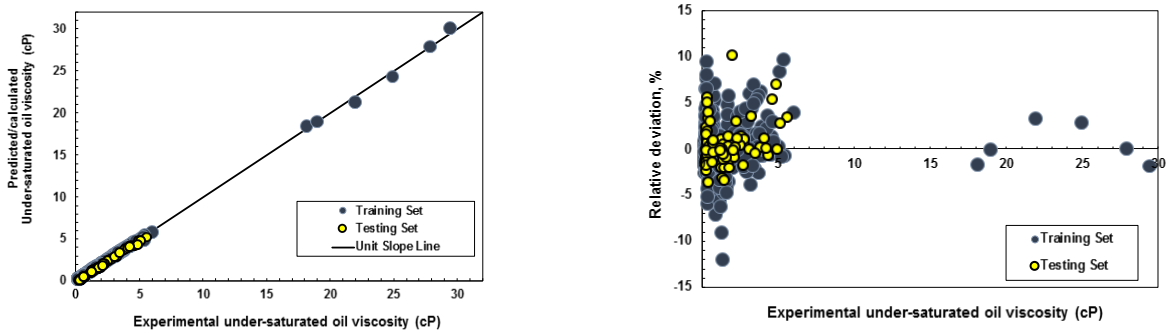
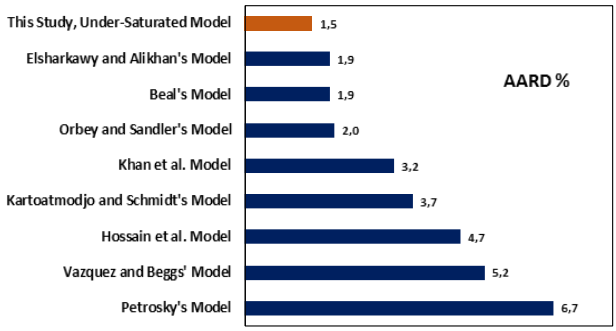


Fig. 8. Parity plot and relative deviation distribution plot for the developed under-saturated oil viscosity model



Additionally, comparison of the values estimated by Eq. (8) with the previously published models [2, 15, 22, 27, 29, 31, 34, 36, 37] for the determination of under-saturated oil viscosities clearly reveals the accuracy of the model compared to the other methods (Table 10 and Fig. 9).

Fig. 9. Graphical comparison between the AARD % values obtained by the model developed in this work for the determination of under-saturated oil viscosity as well as the studied comparative methods; the obtained AARD % for Labedi [2] method is 1078.4 %

Table 10. Summarized statistical error factors including AARD, RMSE and R^2 for the model developed for under-saturated oil viscosity as well as the studied empirical correlations results from the actual data

Method	AARD %	R^2	RMSE
Beal [22]	1.8	0.9978	0.133
Vazquez and Beggs [36]	5.2	0.9713	0.575
Khan *et al.* [34]	3.2	0.9881	0.298
Petrosky [27]	6.7	0.8301	0.940
Labedi [2]	1078.4	0.0969	43.557
Orbey and Sandler [37]	1.9	0.9624	0.495
Kartoatmodjo and Schmidt [29]	3.6	0.9986	0.107
Elsharkawy and Alikhan [15]	1.8	0.9372	0.618
Hossain *et al.* [31]	4.7	0.9948	0.225
This study	1.4	0.9990	0.09

Finally, Fig. 10 demonstrates the smooth behaviour of the GEP-based model, with changes of under-saturated oil viscosity in relation to decreasing and/or increasing saturated oil viscosity and pressure, ultimately to the pressure at bubble point. The model is therefore shown to be able to estimate under-saturated oil viscosity in terms of a trend analysis.

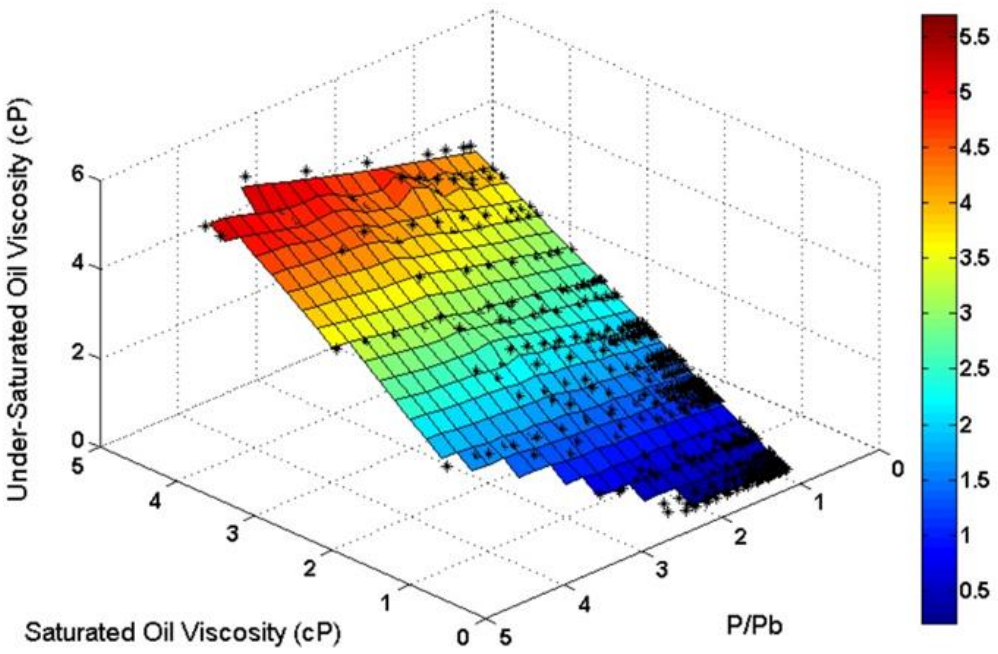


Fig. 10. 3D plot of change of under-saturated oil viscosity versus change in pressure ultimately to the pressure at bubble point and saturated oil viscosity

4.4. Identification of outlier data points

Normally, there are some outlier data points in the databases used for developing mathematical models. Therefore, it is important to identify these data points in order to improve the accuracy and efficiency of the models. A determination of the out of range, or outlier, data points, helps in finding the data that may differ from the bulk of the data [42-45]. Therefore, the databanks associated with dead, saturated and under-saturated reservoir oil viscosities were assessed to find outlier data points, in order to avoid any uncertainties that may lead to a high deviation in estimating reservoir oil viscosities.

The Leverage value statistics technique was employed in this study for detecting outlier data points in the reservoir oil viscosities databanks [45-46]. As a consequence, the Williams plot, in the Leverage analysis, is used to show outlier data points, on the basis of the H values calculated [42-43]. 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 [42-43].

Figs. 11-13 illustrate the Williams plots, for the calculated values of dead, saturated, and under-saturated reservoir oil viscosities, with the application of the GEP-based models. It is evident that the majority of data points are in the range of $0 \leq H \leq 0.08035$ and $-3 \leq R \leq 3$ for the dead oil viscosity model, $0 \leq H \leq 0.02233$ and $-3 \leq R \leq 3$ for the saturated oil viscosity, and $0 \leq H \leq 0.02298$ and $-3 \leq R \leq 3$ for the under-saturated oil viscosity. These results confirm that the equations are statistically valid and accurate in calculating the respective viscosities. In fact, Figs. 11-13 reveal that there are only three points for the developed dead oil viscosity model, five points for the developed saturated oil viscosity, and finally, eight data points for the developed under-saturated oil viscosity, in comparison with their corresponding experimental values, which are outside of the applicability domain of the GEP-based models, and are probably accounted as outliers, whose values may be doubtful.

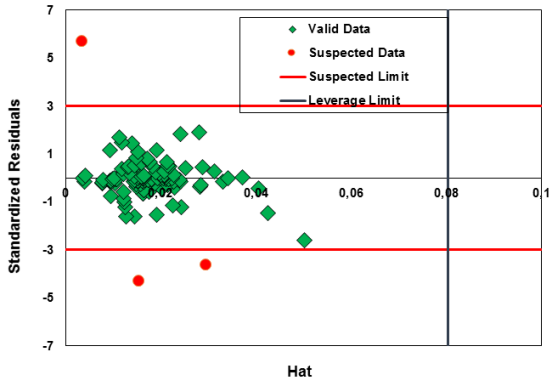


Fig. 11. Detection of outlier data points existing in the dead oil viscosity dataset during development of the model using Leverage approach

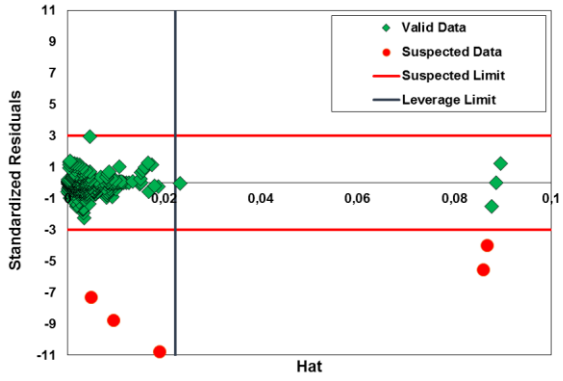


Fig. 12. Detection of outlier data points existing in the saturated oil viscosity dataset during development of the model using Leverage approach

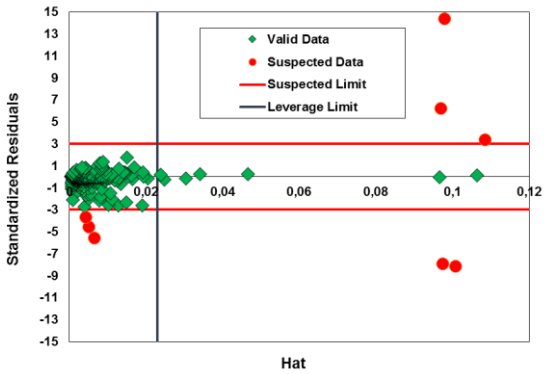


Fig. 13. Detection of outlier data points existing in the under-saturated oil viscosity dataset during development of the model using Leverage approach

5. Conclusions

In this study, a methodology called gene expression programing was utilized to develop consistent, accurate and reliable models for the determination of dead, saturated, and under-saturated reservoir oil viscosities. To this end, more than 1000 sets of experimental data drawn from Iranian oil fields, comprising the gravity of oil, reservoir temperature, solution gas oil ratio, and bubble point pressure, were used in the model development.

The results reveal that GEP-based models are able to calculate the dead, saturated and under-saturated reservoir oil viscosities, within a wide range of reservoir properties of Iranian crude oils.

Additionally, a comprehensive comparative study was performed, to evaluate the performance of the models, by using the most reliable empirically-derived correlations existing in the open literature. It was found that the models in this study, for dead, saturated, and under-saturated reservoir oil viscosities, are more accurate than the comparative methods, in terms of all error factors investigated. The calculated AARD % for the dead, saturated and under-saturated reservoir oil viscosity models are 17.29, 13.55, and 1.47, respectively.

From the results obtained in this study it can be concluded that the proposed models are reliable for the estimation of dead, saturated and under-saturated reservoir oil viscosities.

Acknowledgements

This work is based upon research supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation. Furthermore, the authors are grateful to IOR Research Institute, National Iranian Oil Company (NIOC), R&T for providing us with the required data.

References

[1] Elsharkwy A, Gharbi R. Comparing classical and neural regression techniques in modeling crude oil viscosity. Advances in Engineering Software, 2000; 32: 215-224.

[2] Labedi R. Improved correlations for predicting the viscosity of light crudes. Journal of Petroleum Science and Engineering, 1992; 8: 221-234.

[3] Alizadeh N, Mighani S, Hashemi Kiasari H, Hemmati-Sarapardeh A. Kamari A. Application of Fast-SAGD in Naturally Fractured Heavy Oil Reservoirs: A Case Study. in: 18th Middle East Oil & Gas Show and Conference (MEOS), 2013.

[4] Kamari A, Nikookar M, Sahranavard L, Mohammadi AH. Efficient screening of enhanced oil recovery methods and predictive economic analysis. Neural Computing and Applications, 2014; 25: 815-824.

[5] Hemmati-Sarapardeh A, Majidi SMJ, Mahmoudi B, Ramazani SA, Mohammadi AH. Experimental measurement and modeling of saturated reservoir oil viscosity. Korean Journal of Chemical Engineering, 2014; 31(7): 1253-1262.

[6] Al-Marhoun MA. Evaluation of empirically derived PVT properties for Middle East crude oils. Journal of Petroleum Science and Engineering, 2004; 42: 209-221.

[7] Naseri A, Nikazar M, Mousavi Dehghani S A correlation approach for prediction of crude oil viscosities. Journal of Petroleum Science and Engineering, 2005; 47: 163-174.

[8] Obanijesu E, Omidiora E. Artificial neural network's prediction of wax deposition potential of Nigerian crude oil for pipeline Safety. Petroleum Science and Technology, 2008; 26: 1977-1991.

[9] Xu Z, Zhang J, Feng Z, Fang W, Wang F. Characteristics of remaining oil viscosity in water- and polymer-flooding reservoirs in Daqing Oilfield. Science in China Series D: Earth Sciences, 2010; 53: 72-83.

[10] Torabi F, Abedini A, Abedini R. The development of an artificial neural network model for prediction of crude oil viscosities. Petroleum Science and Technology, 2011; 29: 804-816.

[11] Riazi MR, Al-Sahhaf TA. Physical properties of heavy petroleum fractions and crude oils. Fluid Phase Equilibria, 1996; 117: 217-224.

[12] Teja A, Rice P. Generalized corresponding states method for the viscosities of liquid mixtures. Industrial & Engineering Chemistry Fundamentals, 1981; 20: 77-81.

[13] Johnson SE, Svrcek WY, Mehrotra AK. Viscosity prediction of Athabasca bitumen using the extended principle of corresponding states. Industrial & engineering chemistry research, 1987; 26: 2290-2298.

[14] Johnson S, Svrcek WY. J. Can. Pet. Technol., 26 (1991) 60.

[15] Elsharkawy A, Alikhan A. Models for predicting the viscosity of Middle East crude oils. Fuel, 1999; 78: 891-903.

[16] Obanijesu E, Omidiora E. The artificial neural network's prediction of crude oil viscosity for pipeline safety. Petroleum Science and Technology, 2009; 27: 412-426.

[17] Hemmati-Sarapardeh A, Shokrollahi A, Tatar A, Gharagheizi F, Mohammadi AH, Naseri A. Reservoir oil viscosity determination using a rigorous approach. Fuel, 2014; 116: 39-48.

[18] Al-Anazi A, Gates I. Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study. Computers & Geosciences, 2010; 36: 1494-1503.

[19] Al-Anazi A, Gates I. Support vector regression to predict porosity and permeability: Effect of sample size. Computers & Geosciences, 201; 39: 64-76.

[20] Parhizgar H, Dehghani MR, Eftekhari A. Modeling of vaporization enthalpies of petroleum fractions and pure hydrocarbons using genetic programming. Journal of Petroleum Science and Engineering, 2013; 112: 97-104.

[21] Ferreira C. Gene expression programming: a new adaptive algorithm for solving problems. Complex Systems, 2001; 13: 87-129.

[22] Beal C. The Viscosity of Air Water Natural Gas Crude Oil and Its Associated Gases at Oil Field Temperatures and Pressures. Transactions of the AIME, 1946; 165: 94-115.

[23] Beggs HD, Robinson J. Estimating the viscosity of crude oil systems. Journal of Petroleum technology, 1975; 1140-1141.

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

[25] Kaye S. Offshore California viscosity correlations. COFRC, TS85000940, (Aug. 1985).

[26] Al-Khafaji A, Abdul-Majeed G, Hassoon S. Viscosity correlation for dead, live and undersaturated crude oils. J. Pet. Res., 1987; 6: 1-16.

[27] Petrosky GE. PVT correlations for gulf of mexico crude oils. in, University of Southwestern Louisiana, 1990.

[28] Egbogah EO, Ng JT. An improved temperature-viscosity correlation for crude oil systems. Journal of Petroleum Science and Engineering, 1990; 4: 197-200.
[29] Kartoatmodjo T, Schmidt Z. Large data bank improves crude physical property correlations. Oil and Gas Journal, 1994; 92: 51-55.
[30] Bennison T. Prediction of heavy oil viscosity. in: Presented at the IBC Heavy Oil Field Development Conference, London, 1998.
[31] Hossain MS, Sarica C, Zhang H-Q, Rhyne L, Greenhill K. Assessment and development of heavy oil viscosity correlations. in: SPE International Thermal Operations and Heavy Oil Symposium, Society of Petroleum Engineers, Calgary, Canada, 2005.
[32] Alomair OA, Elsharkawy AM, Alkandari HA. Viscosity prediction of Kuwaiti heavy crudes at elevated temperatures. in: SPE Heavy Oil Conference and Exhibition, Society of Petroleum Engineers, Kuwait City, Kuwait, 2011.
[33] Chew JN, Connally CA Jr. A viscosity correlation for gas-saturated crude oils. Trans. AIME 1959; 216: 23-25.
[34] Khan S, Al-Marhoun M, Duffuaa S, Abu-Khamsin S. Viscosity correlations for Saudi Arabian crude oils. in: Middle East Oil Show, Manama, Bahrain, 1987.
[35] Bergman DF, Sutton RP. An update to viscosity correlations for gas-saturated crude oils. in: SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers, Anaheim, California, USA, 2007.
[36] Vazquez M, Beggs HD. Correlations for fluid physical property prediction. Journal of Petroleum Technology, 1980; 32: 968-970.
[37] Orbey H, Sandler SI. The prediction of the viscosity of liquid hydrocarbons and their mixtures as a function of temperature and pressure. Canadian Journal of Chemical Engineering, 1993; 71: 437-446.
[38] Hashmi MZ, Shamseldin AY, Melville BW. Statistical downscaling of watershed precipitation using Gene Expression Programming (GEP). Environmental Modelling & Software, 2011; 26: 1639-1646.
[39] Huo L-M, Fan X-Q, Huang L-H, Liu W-N, Zhu Y-l. Short-term Load Forecasting Based on Gene Expression Programming With Error Cycling Compensation. Proceedings of the CSEE, 2008; 28: 103-107.
[40] Cranganu C, Bautu E. Using Gene Expression Programming to estimate sonic log distributions based on the natural gamma ray and deep resistivity logs: A case study from the Anadarko Basin, Oklahoma. Journal of Petroleum Science and Engineering, 2010; 70: 243-255.
[41] Ahmed T. Reservoir engineering handbook. Access Online via Elsevier, 2006.
[42] Mohammadi AH, Eslamimanesh A, Gharagheizi F, Richon D. A novel method for evaluation of asphaltene precipitation titration data. Chemical Engineering Science, 2012; 78: 181-185.
[43] Mohammadi AH, Gharagheizi F, Eslamimanesh A, Richon D. Evaluation of experimental data for wax and diamondoids solubility in gaseous systems. Chemical Engineering Science, 2012; 81: 1-7.
[44] Rousseeuw PJ, Leroy AM. Robust regression and outlier detection, John Wiley & Sons, Inc. New York, NY, USA, ISBN: 0-471-85233-3.
[45] Gramatica P. Principles of QSAR models validation: internal and external. QSAR & combinatorial science, 2007; 26: 694-701.
[46] 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, and Professor Deresh Ramjugernath Thermodynamics Research Unit, School of Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa E-mails amir_h_mohammadi@yahoo.com, E-mail ramjuger@ukzn.ac.za