Oil & Gas Science and Technology - Rev. IFP Energies nouvelles **73**, 35 (2018)

https://doi.org/10.2516/ogst/2018037

Available online at:
www.ogst.ifpenergiesnouvelles.fr

REGULAR ARTICLE

OPEN ACCESS

New method for predicting *n*-tetradecane/bitumen mixture density: correlation development

Alireza Rostami[1,*], Amin Shokrollahi[2,*], and Mohammad Hossein Ghazanfari[2,*]

[1] Department of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran
[2] Chemical and Petroleum Engineering Department, Sharif University of Technology, Tehran, Iran

Received: 15 January 2018 / Accepted: 13 July 2018

Abstract. Nowadays, incredible growth of the energy consumption has changed the global attention to the production and utilization of the heavy crude oils such as bitumen resources around the globe. Amongst the bitumen properties, density is an important parameter which improves bitumen recovery efficiency and transportation quality. For easy production of bitumen, *n*-alkanes are usually injected into the reservoir to reduce its viscosity and density; however, there are few numbers of models focusing on proper estimation/prediction of diluted bitumen mixture density in literature. In present work, a new method was proposed to accurately prognosticate the bitumen/*n*-tetradecane mixture density as a function of thermodynamic conditions using Gene Expression Programming (GEP) for the first time as a function of solvent composition, pressure and temperature. Consequently, the proposed model here predicts the mixture density with the average Absolute Relative Deviation (AARD%) of 0.3016% and *R*-squared (R^2) of 0.9943. Moreover, it is found out the solvent concentration has the highest impact value on mixture density estimation. In conclusion, results of the present study can be so valuable for field engineers and researchers working on solvent-assisted recovery methods from heavy oil reservoirs.

1 Introduction

In North America, one of the key supplies of energy is heavy oil resources, in which their production mechanism is incredibly challenging [1, 2]. For instance, nearly two million barrels of heavy oils per day are produced in Alberta province individually with viscosity values from thousands to millions of centipoises (*i.e.*, mPa s) at both reservoir and atmospheric conditions leading to the establishment of considerable difficulties during processing, transportation and underground recovery techniques [1, 3]. Therefore, it is crucial to execute a required viscosity reduction by selecting proper methods using the dilution and/or heating mechanisms. For this reason, a number of approaches have been developed such as N-SOLV [4], ES-SAGD [5], SAS [6], LASER [7] and SAP [8] with the simultaneous application of steam and solvent, and VAPEX process with the direct use of *n*-alkane as diluent [9]. To meet preferable specifications for inlet/outlet of refinery systems, heavy oils are mixed with diverse fractions of crude oils [10]. Frequently, heavy oils are diluted by means of naphtha and condensate through the transportation lines [11, 12].

Amongst the all operational factors, density plays a major role in determining the effectiveness of the above processes such as solvent assisted methods. This property is also significant in carrying out volume computations in refineries and pipelines design. Determination of density of blends is not frequently a feasible strategy because a chemical reaction may happen during mixing, in which the properties of the yielded fluids may not comply with the conventional estimating laws [9]. Thereby, it is necessary to present a new method so as to estimate the mixture density with adequate precision and simplicity.

There are a number of researches regarding different aspects of the solvent injection in the open literature [13, 14]. Various techniques have been utilized to predict the mixtures density including empirical models, Equations of State (EoS) [9] and smart computations. In high pressure conditions, reliable determination of volumetric properties and phase behaviors have been failed via EoS analysis by reason of high degree of numerical complexity [15]. Alongside the EoS strategies, a number of empirical correlations have been proposed in the open literature to calculate the mixture density. According to the corresponding states law, two widely applied equations namely COSTALD [16] and Rackett [17] models were developed, in which COSTALD [16] was understood to be more accurate than EoS modeling. For saturated liquid densities, the Rackett [17] equation was

* Corresponding authors: alireza.rostami.put2014@gmail.com; shokrollahi.amin@gmail.com; ghazanfari@sharif.edu

modified in the work of Spencer and Danner [18]. The above-said correlations are in the need of critical properties as the input as well as the acentric factor in some cases. The main inconvenience involved into these methods is that the pseudo-critical parameters which are generally not available must be used for blends of hydrocarbons. Calculating these properties by the use of traditional techniques leads to high inaccuracy; thereby, choosing a suitable mixing rule may resolve this issue. For heavy oils, an extrapolation affair should be conducted to estimate the pseudo-critical properties which introduce extra deviations into the mixture density estimation. Some investigations have been conducted to determine mixture density with regard to density of each constituent which are simpler than the previous methods. For each component, the density must be measured experimentally and/or a proper estimative equation should be used. This is the main drawback of using such approaches [9].

Artificial intelligence is a robust strategy for solving highly nonlinear problem in wide areas of engineering and science [19–26]. The latest strategy developed for density estimation is the application of artificial intelligence which produces reliable estimates heavy oil density [27]. This type of smart computations has been employed for describing intricate problems with considerably nonlinear trends in wide areas of petroleum and chemical engineering [28–30]. Recently, two rigorous versions of smart computations including Radial Basis Function Neural Network (RBFNN) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) have been proposed by Tavakoli *et al.* [31] and Abbasi *et al.* [27] to prognosticate the diluted density of Athabasca bitumen by *n*-tetradecane, respectively. Their results were promising with more accurate precisions than the available mixing rule in literature. Despite the high capabilities of ANFIS and RBFNN strategies, these models are black box with no recognizable governing equation for computation. Amongst the all versions of smart numerical schemes, genetic-based methods have been understood to be one of the most accurate optimization techniques through the literature. The most recent technique in this field of study is Gene Expression Programming (GEP) which is capable of describing any phenomena by developing accurate empirical models [2, 28, 29]. In other words, there is no requirement for determining the correlation format before using GEP strategy. There are a number of forceful investigations in the open literature which applied GEP strategy for accurate prediction of several parameters such as pure hydrocarbon/water Interfacial Tension (IFT) [28], CO_2/brine IFT [32], critical oil flow rate through the wellhead chokes [33], CO_2 solubility in crude oil [2] and thermal conductivity of supercritical CO_2 [29].

In current study, a new symbolic equation with respect to GEP modeling was extended in relation with thermodynamic and independent variables including pressure, solvent concentration and temperature. Statistical quality criteria and illustrative tools were also used to evaluate the effectiveness of the new method proposed here. The impact of each variable was calculated via sensitivity analysis, and the truthfulness of the used database was examined by Williams' plot or outlier detection technique.

2 Data collection

A large database consisting widespread operational parameters is vital before establishing a predictive model [2, 28, 34–37]. So, about 330 datapoints were adopted from the experimental and modeling work of Kariznovi *et al.* [3]. This dataset includes bitumen/*n*-tetradecane mixture density as the output, and pressure, solvent concentration and temperature as the inputs during modeling. By preprocessing of the used database, about 264 data were considered for model training, and for testing goal about 66 data were assigned. This data division was conducted by a random computer selection approach to fulfill the comprehensiveness of both training and testing subsets. Table 1 indicates the description of the adopted database via several parameters such as minimum, average, maximum and median.

3 Gene Expression Programming (GEP)

Originally, the theory of Gene Expression Programming (GEP) was presented in literature through the work of Ferreira [38] with nearly the same computational procedure as Genetic Algorithm (GA). In this computational strategy, a population of potential answers was randomly employed to find the objective function with a certain fitness value to generate the next potential answer to the problem [28, 29, 33, 34]. During this process, GEP applies some operators such as mutation and crossover to properly adjust the chromosomes (population individuals). These calculations are repeated in anticipation of meeting the convergence condition. Frequently, Mean Square Error (MSE) is chosen as cost function during GEP simulation, defined as follows [2, 35, 39]:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\rho_{\mathrm{mi}}^{\mathrm{exp}} - \rho_{\mathrm{mi}}^{\mathrm{est}}\right)^2, \tag{1}$$

in which, the superscripts "exp" and "est", and the symbol N stand for, correspondingly, the target bitumen/*n*-tetradecane mixture density, the calculated values by the GEP model, and the size of the database used for modeling [2, 28, 29, 33, 34]. Translation of a linear string with fixed length which is termed as chromosome, will lead to the creation of a new object known as symbolic Expression Tree (ET), which has a certain size and depth determined by the users. Therefore, GEP strategy has two key constituents of ET and chromosome. Each gene has two main features including functions (arithmetic operators) and terminals (variables). The tail length of each terminal (t) can be formulated as a function of the head length (h) and the largest arity function (n), as follows [2, 28, 29, 33, 34]:

$$t = 1 - h(1 - n). \tag{2}$$

By means of symbolic ETs, an equation will be provided as a consequence of GEP modeling as an advanced symbolic regression technique. Suggestion of correlation and tuning its constants are conducted at the same time in GEP

Table 1. Specification of the databank applied for modeling.

Parameter	Symbol	Unit	Subset	Minimum	Average	Maximum	Median	SD[a]
Temperature	T	K	Train	295.700	314.050	333.200	313.000	13.235
			Test	295.800	313.044	333.200	312.950	12.887
			Total	295.700	313.849	333.200	313.000	13.153
Pressure	P	MPa	Train	0.124	5.069	9.993	4.997	3.089
			Test	0.124	4.771	9.993	4.997	3.382
			Total	0.124	5.009	9.993	4.997	3.147
Solvent concentration	w_s	Mass fraction	Train	0.050	0.257	0.500	0.200	0.161
			Test	0.050	0.265	0.500	0.300	0.155
			Total	0.050	0.258	0.500	0.250	0.259
Mixture density	ρ_m	kg/m^3	Train	841.600	922.930	994.70	923.450	45.467
			Test	854.400	920.965	992.200	918.600	41.811
			Total	841.600	922.537	994.700	922.500	44.705

[a] SD refers to the standard deviation, which can be calculated as follows:

$$SD = \left(\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \right)^{\frac{1}{2}},$$

where, x is a typical data sample and N shows the number of dataset.

modeling, whereas in conventional regression analysis the correlation format is proposed at first step, and then its coefficients are adjusted by fitting the equation to the experimental database [2, 28, 29, 33, 34]. Actually, GEP advanced technology proposes the best equation by itself through checking different algebraic operators randomly among the independent variables, in which these functions are defined by the user, even though traditional regressions have restricted options of equations to fit or the user should define the equation for the program [2, 28, 29, 33, 34]. That is why the GEP strategy can establish a revolution in regression analysis by suggesting powerful empirical models as compared to the classical fitting approaches [2, 28, 29, 33, 34]. A typical chromosome with two genes composed of four operators (*i.e.*, ×, +, /, √) and three terminals a, b, c, is indicated in Figure 1 [35].

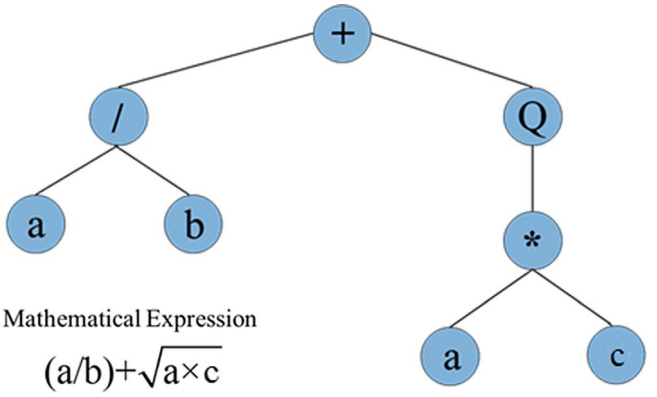

Fig. 1. A typical two-gene chromosome with its corresponding mathematical expression [35].

4 Results and discussions

4.1 GEP model development

In this study, the new method for calculating/estimating the mixture densities of bitumen/*n*-tetradecane was proposed in accordance with the new technology of GEP mathematical strategy for the first time in literature. Despite the existing mixing rules available in literature which consider the component densities, the three independent parameters of temperature, solvent concentration and pressure were regarded as correlating variables in order to predict/estimate the bitumen diluted density. The previous literature studies demonstrated that the pressure has the least effect on the estimation of bitumen density. Therefore, the authors of this study developed a three-parameter model considering the effect of pressure. As a result, the ultimate correlation obtained in this study is as follows:

$$\rho_m = \mathrm{EXP}(A_1 + A_2 + A_3), \tag{3}$$

$$A_1 = \mathrm{EXP}\left[-\left(w_s^2(\ln T + 1.75026656) + 2.23446314\right)\right], \tag{4}$$

$$A_2 = \sqrt[4]{\left[\frac{\sqrt[5]{P} + 6.09188397}{\mathrm{EXP}(w_s + \ln T)}\right]}, \tag{5}$$

Table 2. Parameters of the proposed GEP model.

GEP algorithm parameters	Value
Number of chromosomes	30
Head size	7
Number of genes	3
Linking function	Addition
Generations without change	2000
Number of tries	3
Maximum complexity	3
Fitness function	Mean square error
Mutation	0.00138
Permutation	0.00546
Fixed root mutation	0.00068
Biased mutation	0.00546
Best cloning	0.0026
Random chromosomes	0.0026
Random cloning	0.00102
Inversion	0.00546
Tail inversion and mutation	0.00546
IS and RIS transposition	0.00546
One- and two-point recombination	0.00277
Gene recombination and transposition	0.00277
Data type	Floating-point
Constants per gene	10
Constant fine-tuning	0.00206
Dc mutation	0.00206
Dc IS transposition	0.00546
Constant insertion	0.00123
Constant range finding	0.000085
Stopping condition (maximum fitness)	1000
Best fitness	999.99033
Applied operators	$\pm$, $\times$, /, $\sqrt{}$, EXP, X^2, POW, INV, LN

$$A_3 = 6.41468551 - \frac{0.00397676}{(\ln T - 6.41468551)^4}. \quad (6)$$

In above equations, T stands for temperature (K), P is pressure (MPa), w_s indicates solvent concentration (*mass fraction*), and ρ_m shows mixture density (kg/m^3). For better correlation of the density to the variables, a natural logarithm (i.e., LN) was taken from the temperature and mixture density before running the software. Moreover, it was understood that data normalization has no positive effect on the improvement of modeling. During optimization, a variable impact computation named as sensitivity analysis, was carried out by the simulator to determine the number of decimal points in the GEP-based equation. The parametric description of GEP simulator for developing the new correlation is reported in Table 2. Figure 2 shows the variation of the fitness function and R-squared against the number of generations.

4.2 Evaluation of the new methods

4.2.1 Analytical comparison

For better evaluation, numerous statistical quality measures and visual tools are employed to accurately examine the performance of new method suggested in this study. Table 3 lists the calculated error functions in this study such as Root Mean Square Error (RMSE), Average Deviation (AD), Standard Deviation (*SD*), *R*-squared (R^2) and Average Absolute Relative Deviation (AARD). In addition to these statistical quantities, some useful graphs including relative distribution diagram, crossplot, cumulative frequency diagram and trend analysis were implemented to characterize the behavior and effectiveness of the new method established in this study. The statistical quality parameters of the developed method here are inserted in Table 4 for various subsets of database. The first point implied from this table is that the new GEP based method is a highly precise correlation regarding both training and

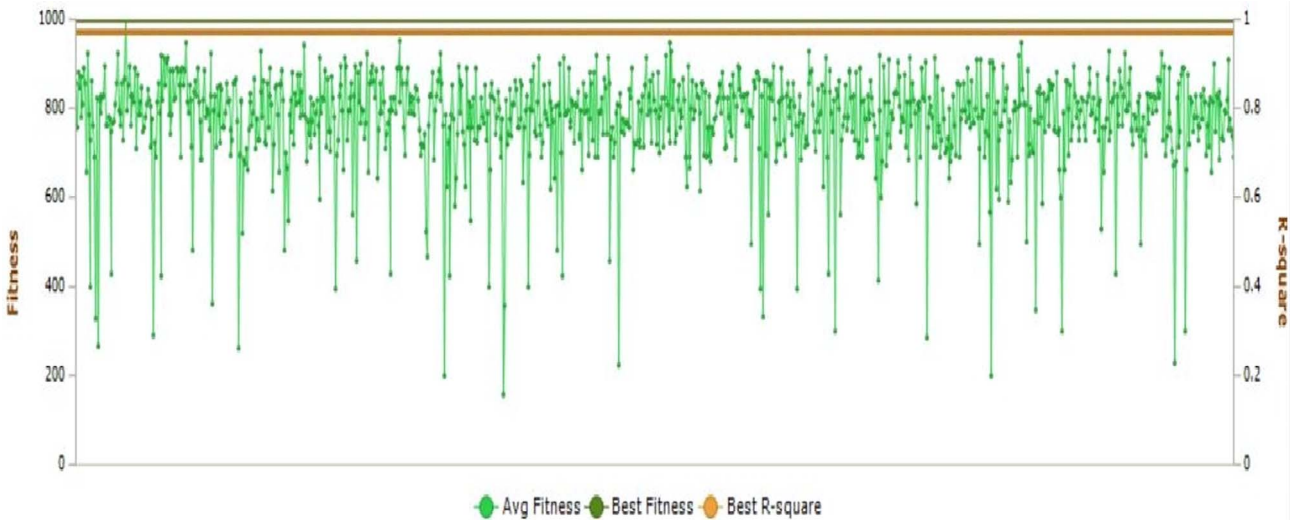

Fig. 2. Variation of the fitness function and R-squared for different generations during GEP modeling.

Table 3. Definitions of statistical quality measures applied for error analysis.

Error	Formula
Minimum deviation	$D^{\min} = \min(\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}})$
Maximum deviation	$D^{\max} = \max(\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}})$
Average deviation	$\text{AD} = \frac{1}{N}\sum_{i=1}^{N}\left(\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}}\right)$
Average absolute deviation	$\text{AAD} = \frac{1}{N}\sum_{i=1}^{N}\left(\left\|\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}}\right\|\right)$
Average relative deviation	$\text{ARD} = \frac{100}{N}\sum_{i=1}^{N}\left(\frac{\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}}}{\rho_{mi}^{\exp}}\right)$
Average absolute relative deviation	$\text{AARD} = \frac{100}{N}\sum_{i=1}^{N}\left(\left\|\frac{\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}}}{\rho_{mi}^{\exp}}\right\|\right)$
Mean square error	$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}}\right)^2$
Root mean square error	$\text{RMSE} = \left(\frac{1}{N}\sum_{i=1}^{N}\left(\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}}\right)^2\right)^{\frac{1}{2}}$
Standard deviation	$SD = \left(\frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}}}{\rho_{mi}^{\exp}}\right)^2\right)^{\frac{1}{2}}$
Coefficient of variation	$CV = \frac{100 \times SD}{\overline{\rho_m}}$
Correlation coefficient	$r = \frac{\sum_{i=1}^{N}\left(\rho_{mi}^{\exp} - \overline{\rho_d}\right) \times \sum_{i=1}^{N}\left(\rho_{mi}^{\text{est}} - \overline{\rho_d}\right)}{\sqrt{\sum_{i=1}^{N}\left(\rho_{mi}^{\exp} - \overline{\rho_d}\right)^2 \times \sum_{i=1}^{N}\left(\rho_{mi}^{\text{est}} - \overline{\rho_d}\right)^2}}$
Determination coefficient	$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(\rho_{mi}^{\exp} - \rho_{mi}^{\text{est}}\right)^2}{\sum_{i=1}^{N}\left(\rho_{mi}^{\exp} - \overline{\rho_m}\right)^2}$

Table 4. Statistical quality measures of the new method for different groups of training, testing and total subsets.

Parameter	New method
Training set	
R^2	0.9946
Average relative deviation, %	−0.1231
Average absolute relative deviation, %	0.3051
Root mean square error,	3.5249
Number of data samples	264
Test set	
R^2	0.9932
Average relative deviation, %	−0.0376
Average absolute relative deviation, %	0.2878
Root mean square error	3.4479
Number of data samples	66
Total set	
R^2	0.9943
Average relative deviation, %	0.0355
Average absolute relative deviation, %	0.3016
Root mean square error	3.5096
Number of data samples	330

Table 5. Comprehensive error investigation of the new method estimating bitumen/*n*-tetradecane mixture density for different values of solvent concentration.

	W_S					
	0.05	0.1	0.2	0.3	0.4	0.5
N	55	55	55	55	55	55
ρ_m^{min}	966.70	951.40	921.50	893.90	866.80	841.60
$\overline{\rho_m}$	981.37	966.57	936.88	909.37	882.93	858.11
ρ_m^{max}	994.70	980.70	950.50	923.30	897.20	872.90
D^{min}	−1.5957	−5.8909	−6.5662	−2.3167	−0.3241	−4.2832
D^{max}	11.3382	5.9291	3.3183	7.1059	8.7189	5.4553
AD	1.9149	−2.8485	−4.1446	−0.2166	1.5040	−2.0721
AAD	2.4655	3.7445	4.6277	1.6815	1.5920	2.6331
ARD (%)	0.1964	−0.2947	−0.4417	−0.0236	0.1702	−0.2426
AARD (%)	0.2521	0.3867	0.4934	0.1850	0.1802	0.3124
MSE	13.4399	16.6813	23.5997	5.0540	6.3763	8.7530
RMSE	3.6660	4.0843	4.8580	2.2481	2.5251	2.9585
SD	0.0038	0.0042	0.0052	0.0025	0.0029	0.0035
CV	0.0004	0.0004	0.0006	0.0003	0.0003	0.0004
r	0.9632	0.9648	0.9684	0.9707	0.9736	0.9742
R^2	0.9278	0.9308	0.9379	0.9423	0.9479	0.9491

testing sets. For this model, the results of training and testing sets are approximately the same; thereby, the overfitting has not occurred through the development of this new empirical model. The results of test set are more accurate than the training set verifying the success of GEP mathematical scheme in developing empirical model by a logical variation of datapoints which are divided randomly for modeling. Table 5 indicates the comprehensive error analysis of the new model here with respect to different solvent concentrations. According to this table, it is clarified

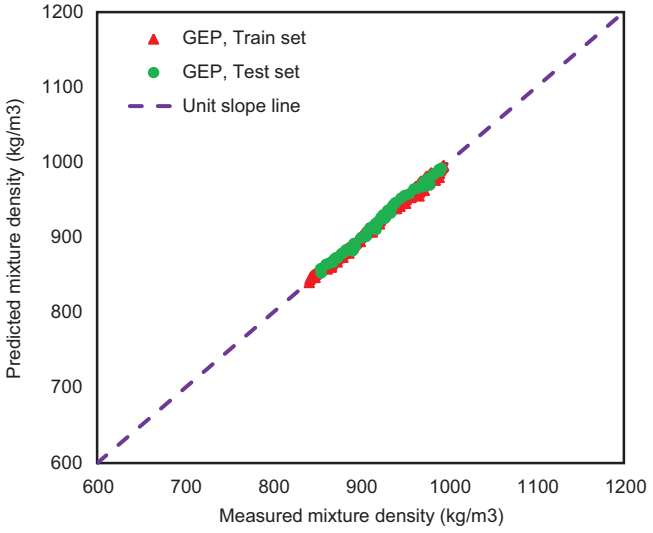

Fig. 3. Comparison of GEP estimates with experimental values of bitumen/*n*-tetradecane mixture density for the new method proposed in this study.

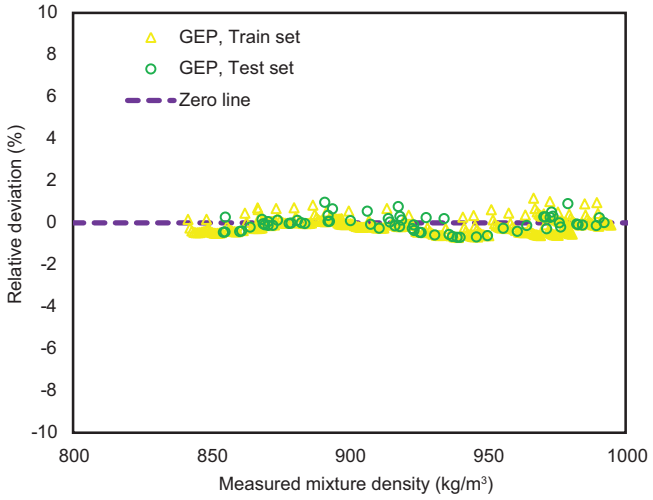

Fig. 5. Distribution of relative deviation for GEP model predictions *versus* the experimental bitumen/*n*-tetradecane mixture density.

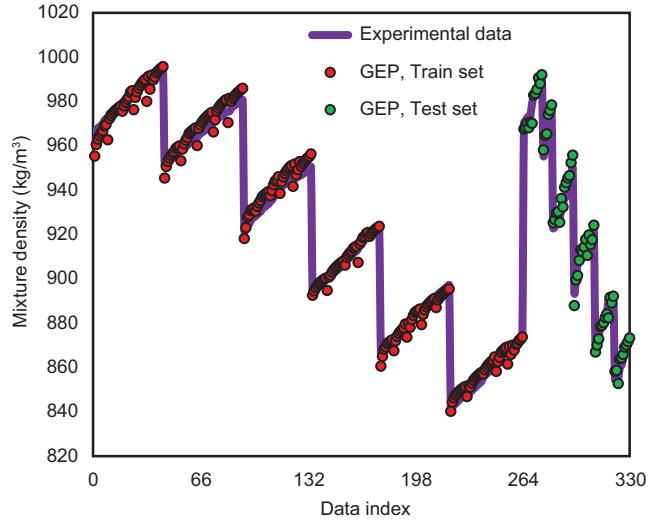

Fig. 4. Experimental and GEP predicted values of bitumen/*n*-tetradecane mixture density *versus* data index for the new method proposed in this study.

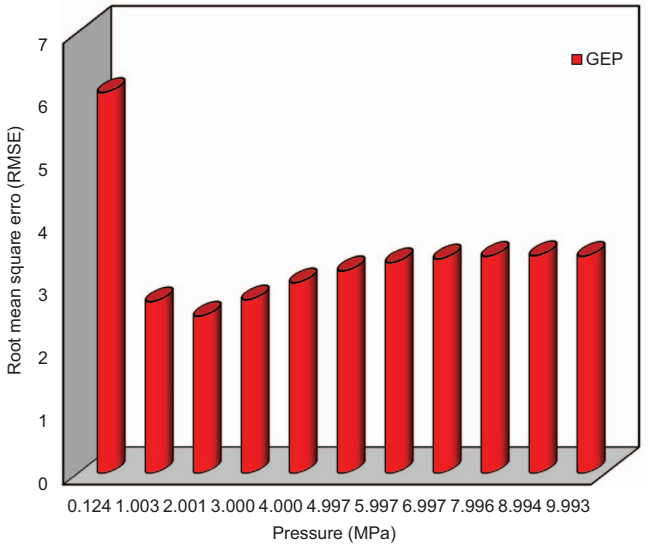

Fig. 6. Root Mean Square Error (RMSE) of the new method here for various pressures.

that the GEP model is an efficient in estimating mixture density owning to the sufficiently low values of AARD%, RMSE, *SD* and AAD parameters.

4.2.2 Graphical comparison

One of the main visual touchstones for analyzing the model performance is known as parity diagram or crossplot. In such diagram, the estimated/predicted data are sketched against the corresponding target experimental values. The strength of model is examined by the degree of data concentration and their neighborhood around the unit slope line or bisection of the first quadrant of coordination system. The parity diagrams for the new method here are illustrated in Figure 3. It is clear that the new method proposed here has a compact cloud of data in the vicinity of 45° line due to the very high R2 value and very small value of total estimation error (e.g., AARD). In Figure 4, simultaneous plots of new method predictions and their corresponding measured values are exhibited against the data index during GEP modeling. The estimates of the GEP model have a satisfactory match with their corresponding target values. Another standard plot is represented in Figure 5 showing the

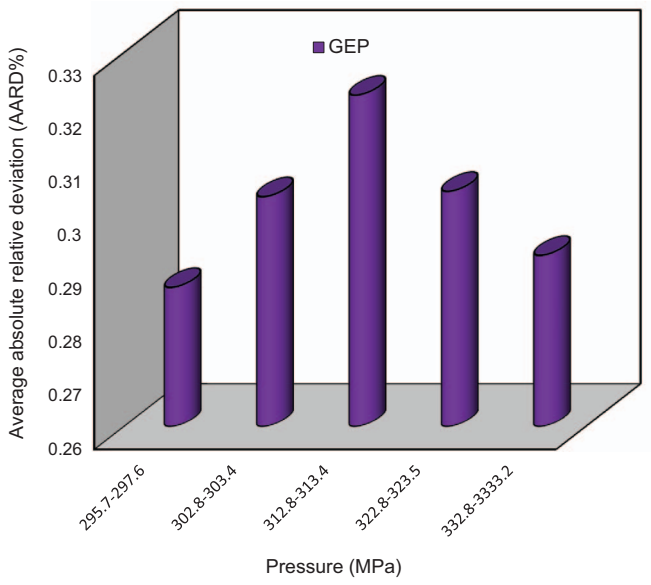

Fig. 7. Average Absolute Relative Deviation (AARD%) of the new method proposed here for various ranges of temperature.

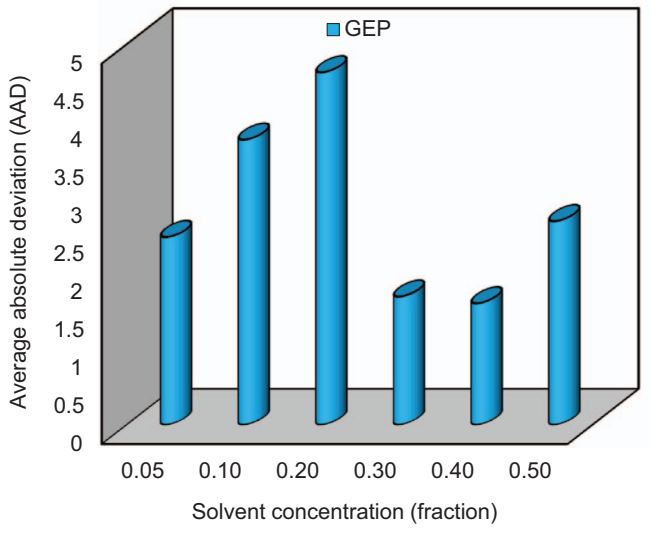

Fig. 8. Comparison of the Average Absolute Deviation (AAD) of the new method here in diverse ranges of solvent concentration.

variation of relative error for GEP model. The variation of error for the new method here shows a compressed range of error varying from −0.5 to +0.5%.

In order to guarantee the outstanding performance of the new method proposed here, the most common statistical parameters are calculated for different ranges of the input parameters. So, a number of bar plot analyses are carried out which are shown in Figures 6–8. Figure 6 declares that at 0.124 MPa, the new method has a high deviation from the reality based on the RMSE parameter. According to Figure 7, the second method has no AARD% value more

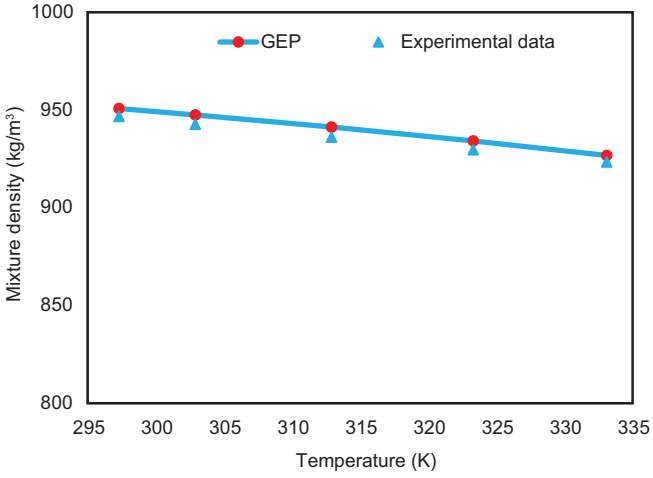

Fig. 9. Comparison of the predictions of the new method with experimental mixture densities *versus* temperature.

than 0.4% at all temperature ranges. The highest AAD value is assigned to the solvent concentration of 0.2 for GEP model which is illustrated in Figure 8. Considering pressure parameter lowers the error value especially for concentrations equal or more than 0.3 mass fraction, which establishes a discontinuity in Figure 8. For the whole ranges of operational conditions, the proposed model almost always gives highly reliable predictions. In other words, considering pressure, temperature and solvent concentration for modeling the mixture density will result in an accurate model with less numbers of experimentations.

Trend analysis of the mixture density *versus* input parameters are indicated through the Figures 9–11 showing the variation of the temperature, pressure and solvent concentration, respectively. Based on these figures, the new method developed here has a good fit with the experimental datapoints. When the pressure varies from 0.124 to 9.994 MPa, temperature from 297.3 to 333.1 K and concentration between 0.05 and 0.50, the magnitude of density difference will be roughly about 6, 23 and 123 kg/m^3, respectively. Therefore, the effect of the independent parameters can be arranged by the following order:

$$\text{Solvent concentration} > \text{Temperature} > \text{Pressure}.$$

For having a better understanding of the estimation potential of the suggested method here, a diagram of cumulative frequency *versus* the absolute relative deviation over the entire database is illustrated in Figure 12. About 100% of the predictions of the GEP method developed here have estimation errors equal or less than 1.2%. This method gives accurate results. Therefore, the new method established in this study exhibits a robust performance in calculating/estimating the bitumen/*n*-tetradecane mixture density. Moreover, the input variables of the new method here are easily available with no need to additional experimentations in comparison with the existing mixing rules in literature. To the best of authors' knowledge, it is the first time in literature that a correlation as a function of thermodynamic conditions for density estimation is proposed. The proposed

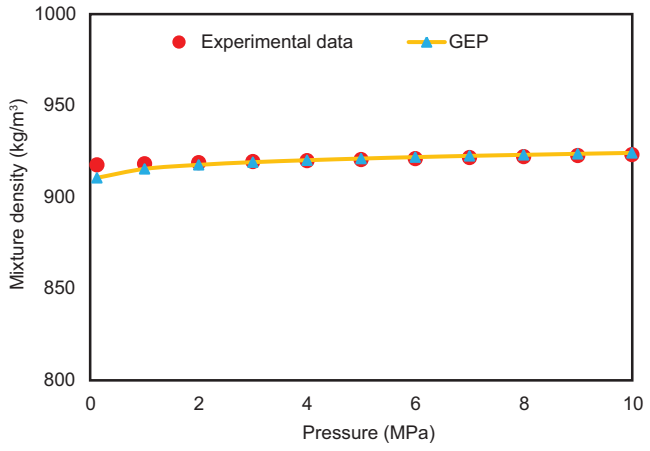

Fig. 10. Comparison of the predictions of the new method with experimental mixture densities *versus* pressure.

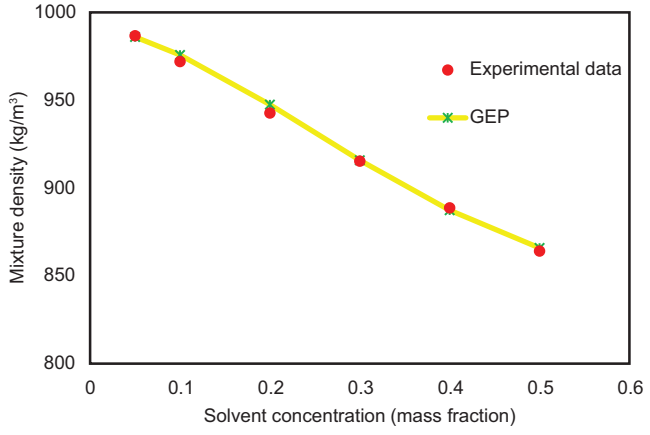

Fig. 11. Comparison of the predictions of the new method here with experimental mixture densities *versus* solvent concentration.

tool in this study can also be applied to density estimation of crude oils and oil fractions because it is a function of thermodynamic condition and solvent composition only; therefore, wide range of applicability can be suggested for this model proposed in this study. However, the existing mixing rules and empirical models which are partly developed on the basis of light to intermediate hydrocarbons, cannot be applied for any type of crude oil and/or mixture of petroleum fractions. This inappropriate application of literature models may lead to large deviations from reality.

4.3 Sensitivity analysis

Quantitative assessment of the effect of each input variable can be calculated by the use of Pearson's technique [40] by assigning an impact value between −1.0 and +1.0 to each independent and input parameter. To obtain this impact value for each input data, the following formula is applied [35]:

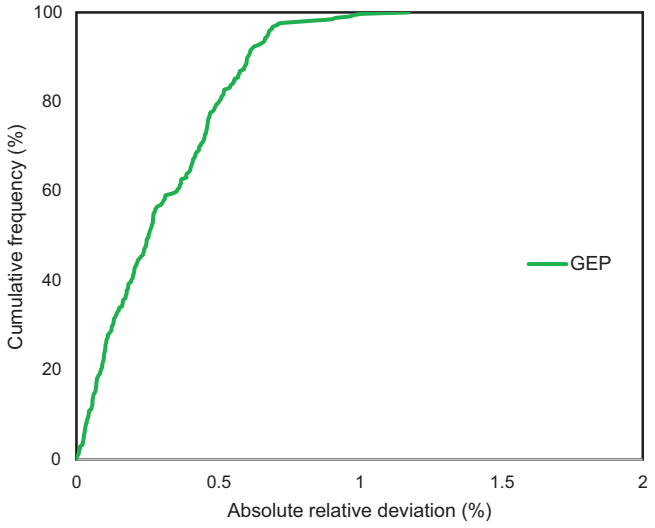

Fig. 12. Cumulative frequency *versus* the absolute relative deviation for the new method proposed in this study for different operational conditions.

$$r(\ln p_k, \mu_g) = \frac{\sum_{i=1}^{n}(\ln p_{k,i} - \overline{\ln p_k})(\mu_i - \overline{\mu})}{\sqrt{\sum_{i=1}^{n}(\ln p_{k,i} - \overline{\ln p_k})^2 \sum_{i=1}^{n}(\mu_i - \overline{\mu})^2}}, \quad (7)$$

where, the symbols r, μ_i, $\overline{\mu}$, $\ln p_{k,i}$ and $\overline{\ln p_k}$ illustrate, respectively, the relative dependency factor, the ith predicted value of mixture density, the mean value of mixture density, the value of ith input variable, and the mean value of the kth input variable, in which k represents the mean values of inputs variables. Based on this approach, a sensitivity analysis is implemented in this study in which its outcomes are shown in Figure 13. As shown, mixture density has a decreasing trend with respect to the temperature and solvent concentration, even though mixture density gradually increases for a rise in pressure.

4.4 Outliers detection

In this section, a useful approach is carried out to detect the suspected data, and separate them from modeling. Because outlier dataset can significantly reduce the reliability of each predictive model during the characterization of physical phenomena. One of the widespread techniques in recognizing such mistrusted data are known as Williams' plot. In this diagram, the standardized residual (R) values are sketched against the Hat parameter. The values of Hat parameter are calculated by the following Hat matrix as follows [41]:

$$H = X(X^tX)^{-1}X^t, \quad (8)$$

where, X is a matrix with m rows and n columns, and superscript t denotes for the matrix transpose operation. In addition, m is the number of datapoints, and n indicates the number of parameters. As a result of this strategy, Figure 14 is plotted showing the outliers data for

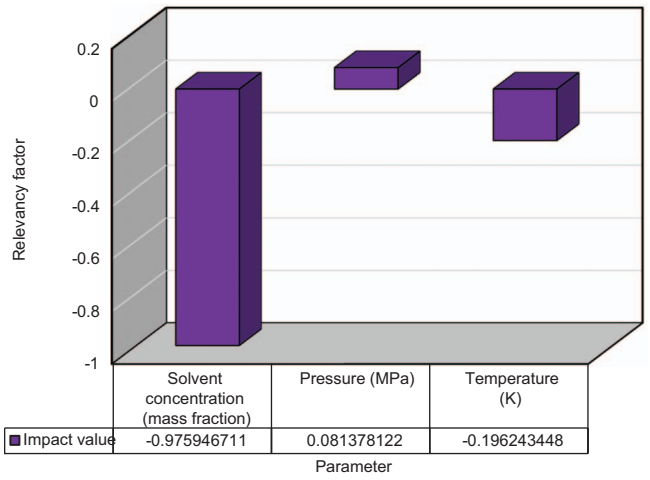

Fig. 13. Sensitivity analysis of the bitumen/*n*-tetradecane mixture density with regard to the pressure, temperature and solvent concentration.

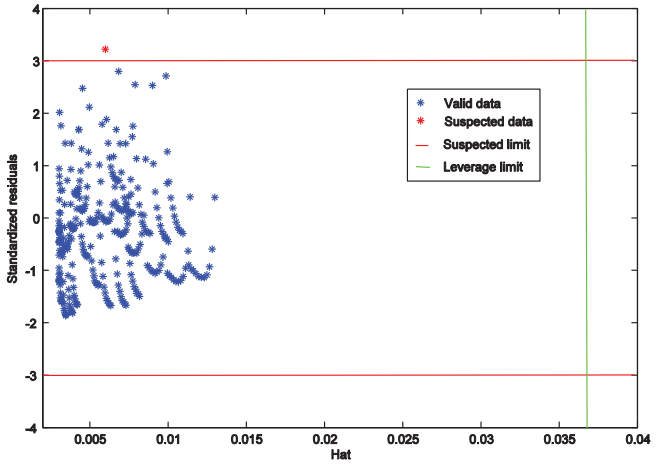

Fig. 14. Outlier detection of the used database for modeling bitumen/*n*-tetradecane mixture density to check the validity of the data.

bitumen/*n*-tetradecane mixture density over the used database. As shown, there is two red horizontal lines named as residual limits, and a green vertical line which shows the leverage limit or critical Hat value (H^*). When the data are located in the interior region of $0 \leq H \leq 0.0364$ and $-3 \leq R \leq +3$, the used data are acceptable (see the blue points in Fig. 14). Obviously, only one point of the density data in this study is unreliable. So, the authors of this study developed two comprehensive models on the basis of a valid database.

5 Conclusion

In current investigation, using Gene Expression Programming (GEP) tool a new method including a three-parameter model was developed to prognosticate the bitumen/*n*-tetradecane mixture density as a function of thermodynamic parameters for the first time in literature. For this purpose, about 80% of the database was utilized for designing the new method, and the remaining data was considered for checking the prediction capability of the trained models. Afterwards, via the most commonly applied statistical functions and graphical methods, the proposed method was assessed. As a result, it is perceived that the GEP method which takes into account the pressure effect is an accurate model with reliable predictions. Moreover, it is demonstrated that the solvent concentration is the most affecting correlative variable for estimating/predicting mixture density. The main benefit of this study over the existing mixing rules is that our newly proposed method requires the pressure, temperature and concentration, which are easily measured through any experimental study; however, the present mixing rules in literature are in the need of each component density which makes their application difficult. Finally, the new methods proposed in this study can be a good nominee for applying in solvent-assisted recovery methods applied on heavy oil reservoirs.

References

1 Teare M., Burrowes A., Baturin-Pollock C., Baturin-Pollock C., Rokosh D., Evans C., Marsh R. (2013) *Alberta's Energy Reserves 2012 and Supply/Demand Outlook 2013–2022. Energy Resources Conservative Board, ST98.*

2 Rostami A., Arabloo M., Kamari A., Mohammadi A.H. (2017) Modeling of CO_2 solubility in crude oil during carbon dioxide enhanced oil recovery using gene expression programming, *Fuel* **210**, 768–782.

3 Kariznovi M., Nourozieh H., Guan J.G.J., Abedi J. (2013) Measurement and modeling of density and viscosity for mixtures of Athabasca bitumen and heavy *n*-alkane, *Fuel* **112**, 83–95.

4 Nenniger J., Nenniger E. (2005) *Method and apparatus for stimulating heavy oil production*, Google Patents.

5 Nasr T.N., Ayodele O.R. (2006) New hybrid steam-solvent processes for the recovery of heavy oil and bitumen, *Abu Dhabi International Petroleum Exhibition and Conference*, Society of Petroleum Engineers.

6 Zhao L. (2004) Steam alternating solvent process, *SPE International Thermal Operations and Heavy Oil Symposium and Western Regional Meeting*, Society of Petroleum Engineers.

7 Leaute R.P. (2002) Liquid addition to steam for enhancing recovery (LASER) of bitumen with CSS: Evolution of technology from research concept to a field pilot at Cold Lake, *SPE International Thermal Operations and Heavy Oil Symposium and International Horizontal Well Technology Conference*, Society of Petroleum Engineers.

8 Gupta S., Gittins S., Picherack P. (2002) Field implementation of solvent aided process, *Canadian International Petroleum Conference*, Petroleum Society of Canada.

9 Sánchez-Lemus M., Okafor J., Ortiz D., Schoeggl F., Taylor S., Van Den Berg F., et al. (2015) Improved density prediction for mixtures of native and refined heavy oil with solvents, *Energy Fuels* **29**, 5, 3052–3063.

10 Gary J.H., Handwerk G.E., Kaiser M.J. (2007) *Petroleum refining: technology and economics*, CRC Press, Florida, USA.

11 Yaghi B.M., Al-Bemani A. (2002) Heavy crude oil viscosity reduction for pipeline transportation, *Energy Sources* **24**, 2, 93–102.

12 Van Den P.J.W.M., Schrijvers FAM (2009) *Process to produce pipeline-transportable crude oil from feed stocks containing heavy hydrocarbons*, Google Patents.

13 Uribe-Vargas V., Carreón-Calderón B., Ramírez-Jaramillo E., Ramírez-de-Santiago M. (2016) Thermodynamic characterization of undefined petroleum fractions of gas condensate using group contribution, *Oil Gas Sci. Technol. - Rev. IFP Energies nouvelles* **71**, 1, 5.

14 Liu H., Cheng L., Xiong H., Huang S. (2017) Effects of solvent properties and injection strategies on solvent-enhanced steam flooding for thin heavy oil reservoirs with semi-analytical approach, *Oil Gas Sci. Technol. - Rev. IFP Energies nouvelles* **72**, 4, 20.

15 Polishuk I. (2011) Hybridizing SAFT and cubic EOS: what can be achieved? *Ind. Eng. Chem. Res.* **50**, 7, 4183–4198.

16 Hankinson R.W., Thomson G.H. (1979) A new correlation for saturated densities of liquids and their mixtures, *AIChE J.* **25**, 4, 653–663.

17 Rackett H.G. (1970) Equation of state for saturated liquids, *J. Chem. Eng. Data* **15**, 4, 514–517.

18 Spencer C.F., Danner R.P. (1973) Prediction of bubble-point density of mixtures, *J. Chem. Eng. Data* **18**, 2, 230–234.

19 Bahari M., Rostami A., Joonaki E., Ali M. (2014) Investigation of a novel technique for decline curve analysis in comparison with the conventional models, *Int. J. Comput. Appl.* **98**, 18, 1–11.

20 Rostami A., Anbaz M.A., Gahrooei H.R.E., Arabloo M., Bahadori A. (2017) Accurate estimation of CO_2 adsorption on activated carbon with multi-layer feed-forward neural network (MLFNN) algorithm, *Egypt J. Petrol.* **27**, 65–73.

21 Rostami A., Arabloo M., Esmaeilzadeh S., Mohammadi A. (2018) On modeling of bitumen/n-tetradecane mixture viscosity: application in solvent-assisted recovery method, *Asia-Pac. J Chem. Eng.* **13**, 1–15.

22 Safari H., Shokrollahi A., Moslemizadeh A., Jamialahmadi M., Ghazanfari M. H. (2014) Predicting the solubility of SrSO4 in Na–Ca–Mg–Sr–Cl–SO4–H2O system at elevated temperatures and pressures, *Fluid Phase Equilibria* **374**, 86–101.

23 Safari H., Shokrollahi A., Moslemizadeh A., Jamialahmadi M., Ghazanfari M. H. (2014) Predicting the solubility of SrSO4 in Na–Ca–Mg–Sr–Cl–SO4–H2O system at elevated temperatures and pressures, *Fluid Phase Equilibria* **374**, 86–101.

24 Rostami A., Hemmati-Sarapardeh A., Shamshirband S. (2018) Rigorous prognostication of natural gas viscosity: Smart modeling and comparative study, *Fuel* **222**, 766–778.

25 Karkevandi-Talkhooncheh A., Rostami A., Hemmati-Sarapardeh A., Ahmadi M., Husein M.M., Dabir B. (2018) Modeling minimum miscibility pressure during pure and impure CO_2 flooding using hybrid of radial basis function neural network and evolutionary techniques, *Fuel* **220**, **1**, 270–282.

26 Rostami A., Kalantari-Meybodi M., Karimi M., Tatar A., Mohammadi A.H. (2018) Efficient estimation of Hydrolyzed Polyacrylamide (HPAM) solution viscosity for enhanced oil recovery process by polymer flooding, *Oil Gas Sci. Technol. - Rev. IFP Energies nouvelles*, **73**, 22.

27 Abbasi P., Madani M., Baghban A., Zargar G. (2017) Evolving ANFIS model to estimate density of bitumen-tetradecane mixtures, *Petrol. Sci. Technol.* **35**, 2, 120–126.

28 Rostami A., Ebadi H., Arabloo M., Meybodi M.K., Bahadori A. (2017) Toward genetic programming (GP) approach for estimation of hydrocarbon/water interfacial tension, *J. Mol. Liq.* **230**, 175–189.

29 Rostami A., Arabloo M., Ebadi H. (2017) Genetic programming (GP) approach for prediction of supercritical CO_2 thermal conductivity, *Chem. Eng. Res. Des.* **122**, 164–175.

30 Bashipour F., Rahimi A., Khorasani S.N., Naderinik A. (2017) Experimental Optimization and Modeling of Sodium Sulfide Production from H2S-Rich Off-Gas via Response Surface Methodology and Artificial Neural Network, *Oil Gas Sci. Technol. - Rev. IFP Energies nouvelles* **72**, 2, 9.

31 Tavakoli H., Khoshkharam A., Sasanipour J., Baghban A. (2017) Modeling of the density of mixtures of Athabasca bitumen and a high boiling *n*-alkane, *Petrol. Sci. Technol.* **35**, **6**, 594–600.

32 Kamari A., Pournik M., Rostami A., Amirlatifi A., Mohammadi A.H. (2017) Characterizing the CO_2-brine interfacial tension (IFT) using robust modeling approaches: A comparative study, *J. Mol. Liq.* **246**, 32–38.

33 Rostami A., Ebadi H. (2017) Toward gene expression programming for accurate prognostication of the critical oil flow rate through the choke: correlation development, *Asia-Pacific J. Chem. Eng.* **12**, 884–893.

34 Rostami A., Shokrollahi A. (2017) Accurate prediction of water dewpoint temperature in natural gas dehydrators using Gene Expression Programming approach, *J. Mol. Liq.* **243**, 196–204.

35 Soroush E., Mesbah M., Shokrollahi A., Bahadori A., Ghazanfari M.H. (2014) Prediction of methane uptake on different adsorbents in adsorbed natural gas technology using a rigorous model, *Energy & Fuels* **28**, **10**, 6299–6314.

36 Kamari A., Gharagheizi F., Shokrollahi A., Arabloo M., Mohammadi A.H. (2016) Integrating a robust model for predicting surfactant–polymer flooding performance, *Journal of Petroleum Science and Engineering* **137**, 87–96.

37 Rostami A., Masoudi M., Ghaderi-Ardakani A., Arabloo M., Amani M. (2016) Effective thermal conductivity modeling of sandstones: SVM framework analysis, *Int. J. Thermophys.* **37**, **6**, 1–15.

38 Ferreira C. (2001) Gene expression programming: a new adaptive algorithm for solving problems, *Complex Syst.* **13**, **2**, 87–129.

39 Teodorescu L., Sherwood D. (2008) High energy physics event selection with gene expression programming, *Comput. Phys. Commun.* **178**, **6**, 409–419.

40 Chok N.S. (2010) Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data *Graduate School of Public Health*, University of Pittsburgh.

41 Meybodi M.K., Naseri S., Shokrollahi A., Daryasafar A. (2015) Prediction of viscosity of water-based Al_2O_3, TiO_2, SiO_2, and CuO nanofluids using a reliable approach, *Chemometr. Intell. Lab. Sys.* **149**, 60–69.