Petroleum 6 (2020) 415–422

Contents lists available at ScienceDirect

Petroleum

journal homepage: http://www.keaipublishing.com/en/journals/petroleum

Application of hybrid support vector regression artificial bee colony for prediction of MMP in CO_2-EOR process

Menad Nait Amar*, Noureddine Zeraibi

Laboratoire Génie Physique des Hydrocarbures, Faculty of Hydrocarbons and Chemistry, University M'hamed Bougara of Boumerdes, Avenue de l'Indépendance, 35000, Boumerdes, Algeria

ARTICLE INFO

Keywords:
CO_2- EOR process
CO_2-Crude oil minimum miscibility pressure
Support vector regression (SVR)
Artificial bee colony (ABC)

ABSTRACT

Minimum miscibility pressure (MMP) is a key parameter in the successful design of miscible gases injection such as CO_2 flooding for enhanced oil recovery process (EOR). MMP is generally determined through experimental tests such as slim tube and rising bubble apparatus (RBA). As these tests are time-consuming and their cost is very expensive, several correlations have been developed. However, and although the simplicity of these correlations, they suffer from inaccuracies and bad generalization due to the limitation of their ranges of application. This paper aims to establish a global model to predict MMP in both pure and impure CO_2-crude oil in EOR process by combining support vector regression (SVR) with artificial bee colony (ABC). ABC is used to find best SVR hyper-parameters. 201 data collected from authenticated published literature and covering a wide range of variables are considered to develop SVR-ABC pure/impure CO_2-crude oil MMP model with following inputs: reservoir temperature (T_R), critical temperature of the injection gas (T_c), molecular weight of pentane plus fraction of crude oil (MW_{C5+}) and the ratio of volatile components to intermediate components in crude oil (x_{vol}/x_{int}). Statistical indicators and graphical error analyses show that SVR-ABC MMP model yields excellent results with a low mean absolute percentage error (3.24%) and root mean square error (0.79) and a high coefficient of determination (0.9868). Furthermore, the results reveal that SVR-ABC outperforms either ordinary SVR with trial and error approach or all existing methods considered in this work in the prediction of pure and impure CO_2-crude oil MMP. Finally, the Leverage approach (Williams plot) is done to investigate the realm of prediction capability of the new model and to detect any probable erroneous data points.

1. Introduction

Inside mature oil reservoirs remains a considerable amount of residual oil even after applying the primary and secondary oil recovery mechanisms. Nowadays, several methods have been applied with the aim to decrease the residual oil and improve the oil recoverability. These techniques are categorized under on so-called "enhanced oil recovery (EOR)". Among EOR techniques, miscible CO_2 injection is one of the most effective and attractive [1], since it has a great potential to increase oil recovery by enhancing microscopic sweep efficiency on one hand; and its lower cost in comparison to other methods in addition to environmental benefits [2], on the other. The right design of miscible CO_2 injection projects depends greatly on a key factor that must be taken into account before the execution phase: minimum miscible pressure (MMP). MMP is defined as the lowest pressure at which the flood changes from immiscible (multiple phase flow) to miscible (single phase flow) [3–5]. Hence, accurate and fast calculation/estimation of the MMP is of great importance. Three main categories of methods can be distinguished in the aim to determine MMP for either pure CO_2-crude oil or impure CO_2-crude oil [6]: experimental tests, empirical/analytical correlations and computational methods (artificial intelligence).

The main experimental tests are slim tube [7], rising bubble apparatus (RBA) [8], multi-contact mixing-cell experiment [5] or the vanishing interfacial tension (VIT) technique [9]. Although the accuracy of these tests, all of them are very expensive and time consuming.

To overcome the limitations of experimental test, several empirical/analytical correlations have been developed for determining the MMP of pure CO_2-crude oil, impure CO_2-crude oil or both pure and impure CO_2-crude oil. A quick estimation of the MMP can be done by applying these correlations. However, as all these latter have been developed under specified experimental data, they have certain constraints and

Peer review under responsibility of Southwest Petroleum University.

* Corresponding author.
E-mail addresses: manad1753@gmail.com, m.naitamar@univ-boumerdes.dz (M. Nait Amar).

https://doi.org/10.1016/j.petlm.2018.08.001
Received 9 June 2018; Received in revised form 8 August 2018; Accepted 27 August 2018
Available online 29 August 2018

conditions of application. Among these correlations, we can cite Cronquist correlation [10], Yellig and Metcalfe correlation [11], Orr & Jensen correlation [12], Emera-Sarma correlation [13], Shokir correlation [14]. Other correlations such as Alston et al. [15], Sebastian et al. [16], Eakin and Mitch [17], Lee [18], Dong [19], and Yuan et al. [20] have focused directly on specific oil reservoirs, and this cannot satisfy the level of comprehensiveness and generalization required by various other oil reservoirs where the characteristics are different.

A good number of studies have been carried out on the application of various artificial intelligence (AI) techniques such as artificial neural network (ANN), gene expression programming (GEP) and support vector regression (SVR) in many fields of science and engineering [21–30]. In petroleum and reservoir engineering, these techniques are becoming more attractive in resolving many problems of practical importance such as well-test analysis [31], PVT parameters estimation [32–34], history matching [35], WAG optimization [36], bottom hole pressure estimation [37], reservoir characterization [38–40] and others. MMP estimation of CO_2-crude oil system can also be done by applying AI methodologies. Many articles shed light the success of these methodologies in establishing reliable models to estimate MMP of CO_2-crude oil systems. Huang et al. [41] applied multilayer perceptron to predict MMP for both pure and impure CO2 injection cases. Ahmadi et al. [42] combined least square support vector machine (LS-SVM) and evolutionary algorithms to the same aim. Sayyad et al. [43] proposed an ANN model based on PSO (particle swarm optimization) to estimate MMP. Ahmadi et al. [44] established a MMP model based on fuzzy logic. Three different CO_2-crude oil MMP correlations using GEP have been developed by Kamari et al. [45], Fathinasab-Ayatollahi [46] and Ahmadi et al. [47], respectively. Although the accuracy showed by the aforementioned CO_2-crude oil MMP based AI, it is necessary to underline that some of them were developed under a limited number of data (such as in Refs. [42] and [44] with only 68 and 58 data respectively).

Among computational intelligence tools, support vector regression (SVR) has gained considerable popularity and importance thanks to its high generalization capability that is assured by its exploitation of both structural risk minimization (SRM) and empirical risk minimization (ERM) principles [48]. The prediction accuracy of an SVR model relies heavily on appropriate determination of model parameters. Therefore, to improve this selection and find the optimal values of these parameters, hybridizations of SVR with metaheuristic algorithms, which are known for their global solution to such problems, have been proposed in many studies [39,49].

In this study, we propose a new robust, fast, cheap and global model to predict MMP of both pure and impure CO_2-crude oil system. This model consists in hybridization SVR with artificial bee colony (ABC). This latter is used to optimize SVR-hyper parameters. 201 measured MMP data collected from published literature (105 of 201 data are pure CO_2-oil while the remain 96 are impure CO_2-oil) are used for SVR-ABC development. Four inputs are considered: reservoir temperature (T_R), critical temperature of the injection gas (T_c), molecular weight of pentane plus fraction of crude oil (MW_{C5+}) and the ratio of volatile components to intermediate components in crude oil (x_{vol}/x_{int}). Subsequently, the results obtained for the proposed SVR-ABC are compared against literature reported data, SVR-TE (support vector with trial and error method) and previously published correlations. In addition, several statistical parameters and graphical error analysis are considered. In assessing the validity of the SVR-ABC, the Leverage approach (Williams plot) is utilized to determine its prediction capability and also to identify probable erroneous data points. There are two main differences between this work and other research works in this field (such in Ref. [50]): (1) the number of data used in the developed model and their intervals that cover a big variety, this gives an excellent generalization of the established model, and (2) besides the application of SVR as an effective tool with excellent predictive and generalization ability for estimating MMP of CO_2-crude oil system, this study shows the efficiency and the robustness of the ABC algorithm in the optimization of SVR hyper parameters to better estimate MMP of CO_2-crude oil system.

2. Methodology

2.1. Support vector regression

Support Vector Regression (SVR) is a supervised learning technique based on statistical learning theory introduced by Vapnik [51,52]. SVR trains to identify a function which computes the relationship dependency between targets $T = \{t_1, t_2, ..., t_k\}$ defined on R, and inputs $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_k\}$ that $\boldsymbol{x}_i \in R^n$ and k is the data size. This is done by mapping the original problem into a high dimensional space by introducing a kernel function to make the original problem be able to be fitted by a linear regression function and to conveniently solve nonlinear regression problems. This can be formulated as follows:

$$f(x) = w.\ \varphi(x) + b \tag{1}$$

where $\varphi(x)$ is a high dimensional feature space that mapped the input space vector x, w is a weight vector and b is a bias.

To estimate w and b, first, the regression problem is transformed into a strict minimization problem of the regularized risk function that defines both the model complexity and the empirical error under ε-insensitive loss. This can be expressed as follows:

$$\begin{gathered} minimize\ \frac{1}{2}w^2 + C\sum_{i=1}^{k}(\xi_i^- + \xi_i^+) \\ subject\ to \begin{cases} t_i - (w.\ \varphi(x_i) + b) \le \varepsilon + \xi_i^+ \\ (w.\ \varphi(x_i) + b) - t_i \le \varepsilon + \xi_i^- \\ \xi_i^-,\ \xi_i^+ \ge 0,\ \ i = 1,2,...,k \end{cases} \end{gathered} \tag{2}$$

where w^2 is the measure of function flatness, $C\sum_{i=1}^{k}(\xi_i^- + \xi_i^+)$ reflects the empirical error, the constant $C > 0$ is the penalty parameter that computes the trade-off between the empirical error and the model complexity (regularization factor), $\xi_i^-\ and\ \xi_i^+$ are positive slack variables which represent lower and upper excess deviation, respectively and ε is the error tolerance.

Then, the constrained optimization problem (2) is resolved after its transformation into dual space using Lagrange multipliers. To keep the work concise, the procedure is not described here. The obtained solution is shown below [53]:

$$f(x) = \sum_{i=1}^{k}(\alpha_i - \alpha_i^*)K(x_i, x_j) + b \tag{3}$$

where $\alpha_i\ and\ \alpha_i^*$ are Lagrange multipliers and are subjected to the constraints $0 \le \alpha_i,\ \ \alpha_i^* \le C$, and the term $K(x_i, x_j)$ is the kernel function. Among kernel functions proposed in the literature [53], radial basis function (RBF) and Gaussian function are the popular ones. In this paper, RBF is used as the kernel function. RBF is defined as shown below:

$$K(x_i, x_j) = \exp(-\gamma x_i, x_j) \tag{4}$$

where γ is the kernel parameter.

To achieve a high performance and high accuracy of SVR, the combination of C, ε and the kernel function parameter γ should be selected properly. Hence, optimizing these parameters using robust metaheuristic algorithms able to select automatically the optimum combination is a favorite alternative to overcome the use of the traditional trial and error method (TE).

Meanwhile, it should be noted that one of the main advantages of SVR is that its computational complexity does not depend on the dimensionality of the input space [54]. In addition, sequential minimal optimization (SMO) [55], one of quadratic programming solvers for

SVRs, has become a quasi-standard in SVR training as it reduces the computational complexity of SVR to approximately $O(N^2)$, where N is the number of training data [56].

2.2. Artificial bee colony (ABC)

Artificial Bee Colony (ABC) algorithm was proposed by Karaboga in 2005 [57] for optimizing numerical problems. ABC is inspired from honey bee swarms, which represent the intelligent foraging behavior of bee swarms. This algorithm provides solution in organized form by dividing the bee objects into 3 different tasks/categories of bee: employed bees, onlooker bees, and scout bees. These latter determine the objects of problems by sharing information to others bees. Each food source discovered so far is tracked by an employed bee. Hence, the number of food sources is equal to employed bees. The food source is the solution candidate and is coded at time (iteration) t with a D dimensional vector as:

$$x_{i,t} = \{x_{i1,t}, x_{i2,t}, \dots, x_{iD,t}\} \tag{5}$$

The best food source is the most qualified optimized solution for the problem. An employed bee computes a modified position from her memorized food location depending on the local information and tests the quality of food on the new source. If the new location is a better food source compared to the previous one, the bee memorizes the new position and forgets the old one. After all employed bees complete their exploration; they share new food source information and their position information with the onlooker bees on the dance area. Onlooker bees select food sources based on the dance performed by the employed bees: they watch the dance of hive bees and select the best food source according to the probability proportional to the quality of that food source [57,58] (for this purpose, a roulette wheel selection technique can be used). An employed bee is transformed into a scout bee if its food source is abandoned. Scout bees carry out random search in the model space.

Four phases are followed in ABC algorithm: initialization, employed bees phase, onlooker bees phase and scout bee phase, each of which is explained below:

(1) *Initialization phase*: Initialize the population of solutions x_i where $i \in \{1,2,\dots,NP\text{: Population size}\}$.

After producing food sources and assigning them to the employed bees, the objective function specific for the optimization problem is operated, its value is obtained, and all the fitness values of the food sources are calculated by using:

$$fit_{i,t} = \begin{cases} \frac{1}{1+f(x_{i,t})}, & \text{if } f(x_{i,t}) \geq 0 \\ 1 + abs(f(x_{i,t})) & \text{otherwise} \end{cases} \tag{6}$$

where $fit_{i,t}$ is the fitness of the source i at the iteration t (t = 1 for the initialization phase), and $f(\,.)$ is the objective function (the optimization problem is to minimize this objective function).

(2) *Employed bees phase:* during this phase, each bee modifies its current solution according to the information of individual experiences and the fitness value of the new solution. If the fitness value of the new food source is higher than that of the old, the bee updates its position with the new one and discards the old one. The position is updated by the equation below:

$$v_{iD,t+1} = x_{iD,t} + \lambda_i (x_{iD,t} - x_{\chi D,t}) \tag{7}$$

where χ is a random from $\in \{1,2,\dots,NP\text{: Population size}\}$ which must differ to i, and λ_i is a random from [0; 1].

(3) *Onlookers phase:* Onlooker bee phase is started after finishing the employed bee phase, this phase operates on sharing the fitness information that has been collected by the employed bee phase, this information includes updated solutions and their position information. Onlooker bees analyze the available information and select a solution with a probability related to its fitness:

$$P_i = \frac{fit_i}{\sum_{i=1}^{NP} fit_i} \tag{8}$$

where fit_i is the fitness of ith solution. As in the case of the employed bee, an onlooker bee produces a modification in the position in its memory with the same principle to employed bees.

(4) *Scout Bees Phase:* In this phase, the employed bee become scout bee if the employed bee is associated with an abandoned food source or after a certain defined number of iteration, also, the food source is replaced by randomly choosing another food source from the search space.

In this paper, ABC is employed to tune SVR hyper-parameters (C, ε and γ). The manner of implementation of ABC into SVR is summarized in flowchart shown in Fig. 1.

3. Model development

3.1. Data acquisition and pre-processing

The number and the variety of initial input data sets affect the accuracy and generalization of the proposed SVR-ABC model for MMP estimation of pure/impure CO_2-crude oil system. In this study, 201 MMP of pure/impure CO_2-crrude oil data were collected from the published literature [10–20,59-66]. Among these 201 data, 105 and 96 refer to pure and impure CO_2 respectively.

A review on recent researches shows that [14–17,20,66]:

(1) The main factors affecting CO_2–oil MMP are: the reservoir temperature, the purity of injected gas, and the components and properties of oil.
(2) As the temperature increases, the MMP increases as well and vice versa.
(3) MMP decreases when the crude oil is characterized with high content of $C_2 \sim C_6$ and low molecular weight. On the contrary, themore the heavy components in the crude oil are, the less favorable it will be for miscibility.
(4) The presence of methane and nitrogen (CH_4 and N_2) impurities in the injected gas substantially increase the MMP, but intermediate hydrocarbons (C_2—C_5) or H_2S impurities decrease the MMP.

Accordingly, four factors that have great effects on MMP were selected as the input variables: reservoir temperature (T_R), oil composition which is generally summarized by the ratio of volatile components (C_1 and N_2) to intermediate components (C_2—C_4, H_2S and CO_2), and the critical temperature of injection gas. Before proceeding to the model development, it is very useful to get an overview about the statistical analysis of the employed datasets. Statistical criteria including standard deviation (SD), which measures the extent of variation or dispersion within the data points of each variable, relative standard deviation (% RSD) which shows the extent of variability in relation to the mean of the data points of each variable, mean, maximum and minimum value which allow a better comprehension of the content of the dataset, are considered for this purpose. A full statistical description of the system input/output is reported in Table 1. MMP for this dataset is ranged from 6.50 to 38.52 MPa, while the inputs used varied as follows: Reservoir temperature from 307.5 to 410 K, the molecular weight of the C5+ fraction from 136.26 to 391 g/mol, the ratio of volatile components to intermediate components in crude oil from 0 to 13.07 and the critical temperature of the injection gas from 281.45 to 338.77 K. To improve

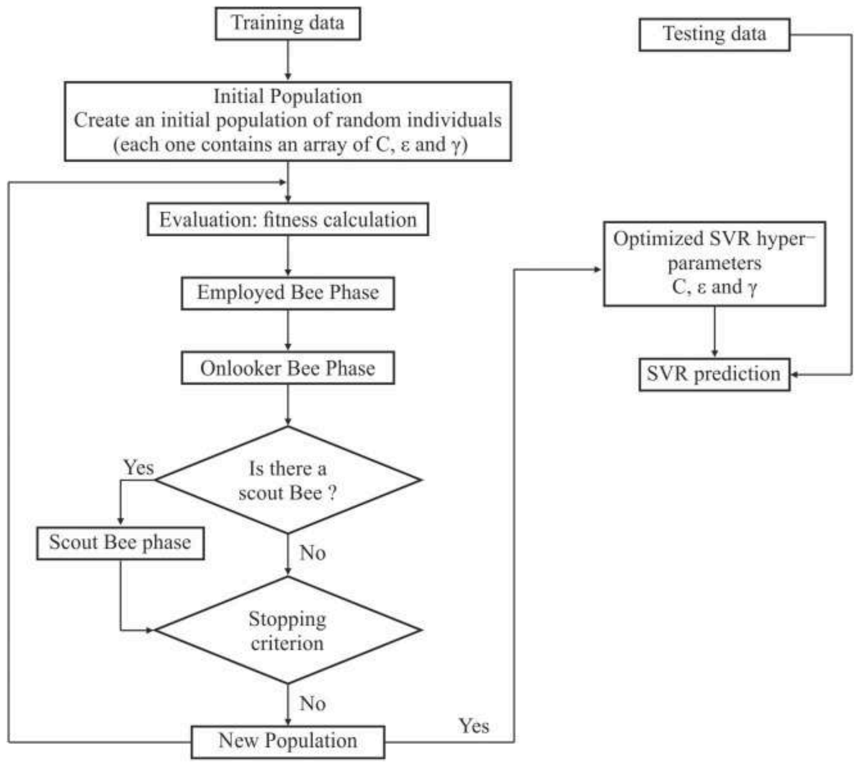

Fig. 1. SVR-ABC flowchart.

Table 1
Statistical description of the input/output data used for developing SVR-ABC.

	Variables	Max	Min	Mean	SD	%RSD
Output	MMP (MPa)	38.52	6.50	16.02	6.12	38.20
Inputs	MW_{C5+} (g/mol)	391	136.26	194.63	40.10	20.60
	T_R (°K)	410	307.5	345.45	24.31	7.04
	x_{vol}/x_{int}	13.07	0	1.60	2.09	130.62
	T_c (°K)	338.77	281.45	302.72	8.31	2.75

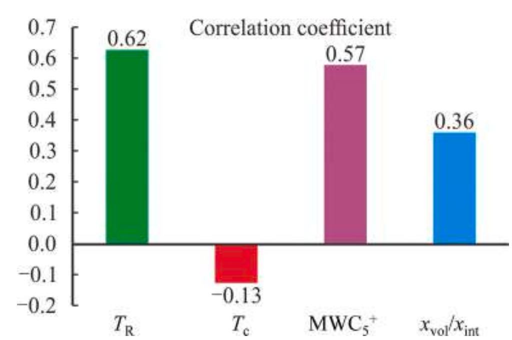

Fig. 2. Correlation coefficient of the independent variables.

the convergence conditions of SVR-ABC model, the used data are standardized.

To demonstrate the correlation between MMP and the used independent variables the correlation matrix is implemented. This matrix aims to distinguish linear relationship between two different variables in multi-variables system [67,68].

Its values are between [−1, 1]. Two variables are said to be positively linearly related if their correlation coefficient is close to 1, and negatively linearly if it is close to −1. For values nearby zero, it would indicate a weak linear relationship between the variables.

The obtained results are shown in the bar graph of Fig. 2 and in Table 2. According to this figure, it can be seen that reservoir temperature (T_R) has the highest linearly relation with MMP (0.62). Then molecular weight (MW) of the $C_5{}^+$ and ratio of volatile components to intermediate components in crude oil (x_{vol}/x_{int}) by 0.57 and 0.37, respectively. Furthermore, it can be deduced that critical temperature of the injected gas (T_c) is in negative linearly relation with MMP which means that the MMP is high, if T_c is low. It is obvious from the aforementioned results that in multivariable phenomena, the relationship of the input-output variables has a weak linear relationship and cannot be handled using a linear mode, hence the need of SVR model to identify this relationship.

Table 2
Correlation coefficients between the descriptors and the target.

	MW_{C5+} (g/mol)	T_R (°K)	x_{vol}/x_{int}	T_c (°K)
Correlation coefficient	0.62	−0.13	0.57	0.36

3.2. Computational procedure

The used database is divided into two parts: the first one contains 175 data (from the 201) and it is used in the training of the model, and the other part (with the remain 26 data) is devoted to test its accuracy and robustness to predict MMP of pure/impure CO_2-EOR with blind data.

All the programming tasks developed in this work are carried out using MATLAB® 2016-a computing environment [69]. The corresponding parameters of ABC algorithm used in the optimization process of SVR hyper-parameters (C, ε and γ) are selected using parameter-tuning method [70], and their corresponding values are as follows: the total number of bees is 40 (whose 20 are employed and the same goes for onlookers); scout bees' max iteration (i.e. after how many iterations an employed bee becomes scout bee): 5; and the maximum number of iteration is 50. Then, the steps of the flowchart of Fig. 1 which describes the process of hybridization SVR with ABC, are implemented. As a comparison of SVR-ABC performances against SVR with the classical trial and error method (SVR-TE) is done, this last is developed by manually searching a specified subset of the hyper-parameter (C, ε and γ). For this purpose, a control loop is added to the main program of the SVR, which varied the values of these constants and examined the accuracy of the model in each situation in finding the more appropriate conditions.

3.3. Performance assessment

To evaluate how well the model fits the data was used, a group of statistical indicators, including the average absolute percent error (AARD%), standard deviation (SD), the coefficient of determination (R^2) and the root mean square error (RMSE) are considered. These statistical indexes are calculated by the following equations:

$$AARD\% = \frac{1}{k}\sum_{i=1}^{k}\left|\frac{MMP_i^{exp} - MMP_i^{cal}}{MMP_i^{exp}}\right| \times 100 \tag{9}$$

$$SD = \sqrt{1/(k-1)\sum_{i=1}^{k}\left(\frac{MMP_i^{exp} - MMP_i^{cal}}{MMP_i^{exp}}\right)^2} \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{k}(MMP_i^{exp} - MMP_i^{cal})^2}{\sum_{i=1}^{k}(MMP_i^{cal} - \overline{MMP})^2} \tag{11}$$

$$RMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(MMP_i^{exp} - MMP_i^{cal})^2} \tag{12}$$

where k represents the number of the measured information, MMP_i^{exp} is the experimental minimum miscibility pressure values, while MMP_i^{cal} is the calculated MMP values which are predicted by the developed model. Average value of the MMP data is shown by $\overline{MMP}$.

To better show the model ability to estimate MMP, it is compared against existing correlations and approaches. Furthermore, as the leverage technique (detection of the outlier data points) plays a significant role to identify the applicability domain of a model and detect suspect data [71–73], this technique through Williams plot is used to check whether the SVR-ABC model is statistically valid. A detailed description of computational procedure and equations for the leverage technique can be found elsewhere [71–73].

4. Results and discussions

The above-described computational steps of SVR-ABC and SVR-TE are followed to achieve best SVR hyper-parameters. Moreover, as previously mentioned, for assessing the performance capabilities, statistical error analysis, in which AARD, R^2, SD and RMSE, as well as graphical error analysis, in which cross plot and APRE error distribution plot are sketched, have been implemented. As a consequence, the obtained best SVR hyper-parameters (C, ε and γ), which yield the most precise results for both trial and error method and ABC algorithm are shown in Table 3. According to this table, high values of the regularization factor (C) and small Epsilon values are achieved for both SVR-ABC and SVR-TE, while medium values of the kernel parameter (γ) are expected.

Table 3
Obtained SVR hyper-parameters for developed models.

	C	ε	γ
SVR-ABC	3253.70	0.2690	0.9258
SVR-TE	3065	0.40	1.35

Cross plots between output and target values for training and test data of SVR-ABC and SVR-TE models are illustrated in Fig. 3. For each model, all the predicted values are sketched against the experimental values and therefore across plot is created and compared against a unit slope line that shows the perfect model line: the closer the plotted data to this line, the higher is the reliability of the model. As clear be seen, the prediction capability of the SVR-ABC is documented. For a deep comparison between the calculated MMP by SVR-ABC and SVR-TE and the experimental data, the results of the aforementioned statistical indexes are reported in Table 4. As can be observed in Table 4, the SVR model optimized by artificial bee colony (SVR-ABC) shows the higher accuracy in comparison with SVR-TE by total values of 3.16%, 0.9879, 0.78 and 0.052 for AARD, R^2, RMSE and SD, respectively. In order to further evaluate the SVR-ABC accuracy, a performance distinction of this model in predicting the MMP of the cases pure and impure CO_2-crude oil is presented separately in Table 5. This table highlights that the proposed SVR-ABC model predictions are in excellent agreement with experimental data with AARD of 2.75% for pure CO_2 case and 4.33% for impure CO_2 case are noticed, respectively. Fig. 4 depicts the error distribution of the SVR-ABC model for estimating MMP CO_2-crude oil. In this figure, the error is the percent relative error (PRE%) which measures the relative deviation of predicted data from the experimental data. The figure confirms that SVR-ABC provides favorable results so that it has a low scatter around the zero error line, and most of its predictions have a relative error between −15% and 15%. This indicates the potential of SVR-ABC model for predicting CO_2-crude oil MMP with small expected errors.

The performance of the model has been compared with that of some of the most widely utilized correlations. These include seven correlations for pure CO_2-crude oil cases, namely those of: Cronquist [10], Lee [18], Yellig & Metcalfe [11], Orr & Jensen [12], Emera-Sarma [13] and Shokir [14]; and two correlations for impure CO_2-crude oil cases: Kamari et al. correlation [45] and Fathinasab-Ayatollahi correlation [46]. Table 6 reports the corresponding results. As it is shown in Table 6, SVR-ABC model leads to very satisfactory performances for determination of MMP values either for pure or impure CO_2-crude oil cases, and outperforms largely the published correlations in terms of accuracy of the predictions.

Finally, in order to check whether the SVR-ABC model is statistically acceptable, the Williams plot has been illustrated in Fig. 5. The existence of the majority of data points in the ranges $0 \le H \le 0.0746$ and $-3 \le$ Standardized Residuals ≤ 3 confirms that the SVR-ABC model developed for the calculation of CO_2-crude oil MMP is statistically accurate and reliable. Consequently, good high leverage data points are located in the domain of $H < 0.0746$ for SVR-ABC: the entire data except seven in the data set are located within the applicability domain. As a result, there are only seven points in the datasets that are not within this domain, and accordingly, these data can be stated as probable doubtful datum.

5. Conclusions

In this work, a new model based on optimized support vector regression (SVR) by artificial bee colony algorithm (ABC) namely SVR-

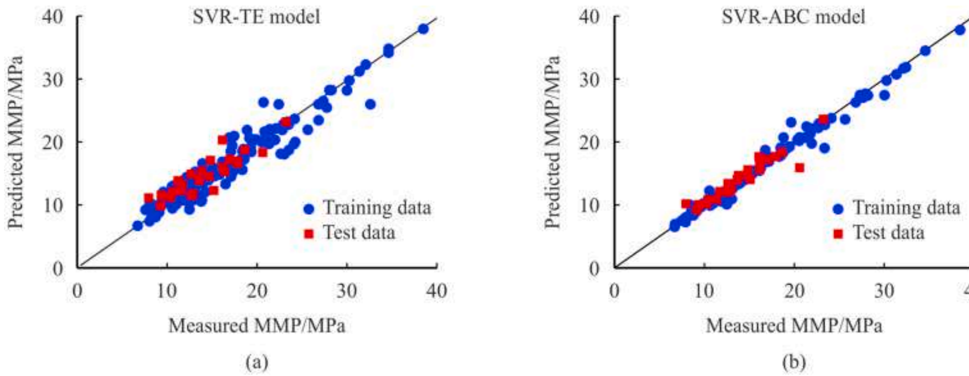

Fig. 3. Cross-Plots of the Results: (a) MMP measured vs MMP SVR-TE (training + test); (b) MMP measured vs SVR-ABC (training + test).

Table 4
Statistical analysis of performance of SVR-ABC and SVR-TE.

		R^2	AARD (%)	RMSE	SD
SVR-ABC	Training (175 data)	0.9936	2.71	0.72	0.046
	Test (26 data)	0.9497	6.15	1.20	0.094
	All (201 data)	0.9879	3.16	0.78	0.052
SVR-TE	Training (175 data)	0.9669	7.03	1.61	0.101
	Test (26 data)	0.8912	13.05	1.92	0.169
	All (201 data)	0.9571	7.81	1.65	0.11

Table 5
Statistical analysis of performance of SVR-ABC for pure CO_2-crude oil and impure CO_2-crude oil MMP estimation.

		R^2	AARD (%)	RMSE	SD
SVR-ABC	Pure CO_2 (105 data)	0.9948	2.75	0.73	0.047
	Impure CO_2 (96 data)	0.9708	4.33	1.17	0.077

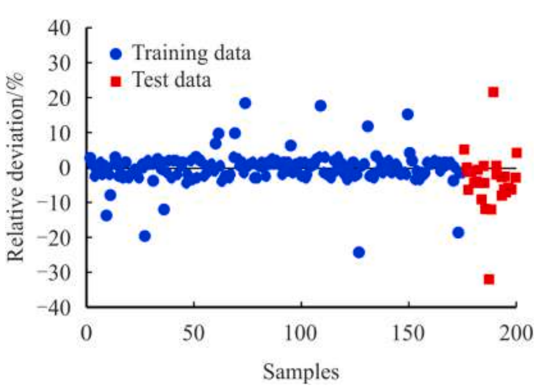

Fig. 4. Relative deviations of the MMP values obtained by SVR-ABC model from data base values.

Table 6
Comparison of SVR-ABC performances against available correlations for both pure/impure CO_2-crude oil.

	Model/ correlation	R^2	AARD (%)	RMSE	SD	Max AARD (%)
Pure CO_2-crude oil	SVR-ABC	0.9948	2.75	0.73	0.047	23.83
	Cronquist	0.9043	16.42	4.23	0.2085	69.13
	Lee	0.6393	19.73	6.05	0.2782	98.68
	Yelling Metcalfe	0.6771	18.01	6.06	0.2341	66.09
	Orr-Jensen	0.6331	20.17	6.71	0.2548	78.71
	Alston et al.	0.8546	19.15	6.05	0.2640	91.54
	Emera-Sarma	0.8868	14.02	3.92	0.1867	56.82
	Shokir	0.8574	12.58	3.45	0.1674	57.94
Impure CO_2-crude oil	SVR-ABC	0.9708	4.33	1.17	0.077	33.72
	Kamari et al.	0.9329	8.65	1.71	0.109	28.54
	Fathinasab-Ayatollahi	0.8243	23.94	5.79	0.255	98.73

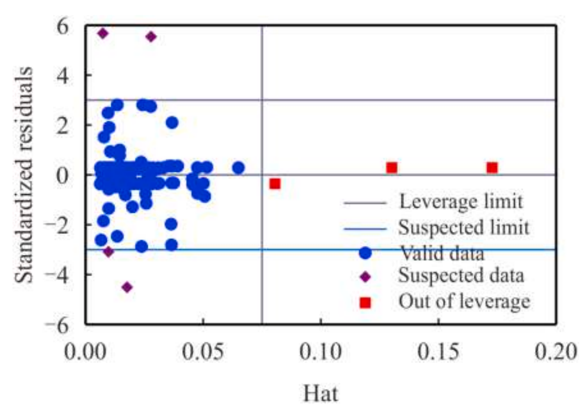

Fig. 5. Detection of the probable doubtful data and the applicability domain of the developed SVR-ABC model.

ABC was proposed and applied to predict the MMP of pure/impure CO_2-crude oil systems. In this model, four factors (i.e. reservoir temperature, critical temperature of the injection gas, molecular weight of C_5^+ fraction of crude oil, and the ration of volatile components to intermediate components in crude oil) representing the most comprehensive and robust set were selected as the input variables, while MMP was considered as the output variable. A large number of data points collected from open literature and covering a wide range of independent variables was considered for developing the model. Statistical and graphical error analysis have been carried out to establish the adequacy and accuracy of the model as well as to compare it with classical SVR based on trial and error method (TE), and with existing correlations. The results show that SVR-ABC model provides predictions in satisfactory agreement with experimental data and outperforms all the other models either SVR-TE or correlations. The precision analysis of the developed model shows its high ability to predict MMP with a low mean absolute percentage error (3.24%) and root mean square error (0.79) and a high coefficient of determination (0.9868). Furthermore, leverage approach reveals that this model is statistically correct and valid with only seven probably doubtful data points in the whole experimental data set.

Nomenclatures

EOR	Enhanced Oil Recovery
ABC	artificial bee colony
SVR	support vector regression
Int.	intermediate components
MMP	Minimum Miscibility Pressure
Mw	Molecular Weight
T_R	reservoir temperature
T_c	critical temperature
R^2	coefficient of determination
Vol.	volatile components
AARD	average absolute relative deviation
RMSE	root mean Squared Error
SD	standard deviation

References

[1] M.F. Al-Ajmi, O.A. Alomair, A.M. Elsharkawy, Planning miscibility tests and gas injection projects for four major Kuwaiti reservoirs, Kuwait Int. Pet. Conf. Exhib, Society of Petroleum Engineers, 2009, , https://doi.org/10.2118/127537-MS.
[2] P. Zanganeh, S. Ayatollahi, A. Alamdari, A. Zolghadr, H. Dashti, S. Kord, Asphaltene deposition during CO_2 injection and pressure depletion: a visual study, Energy Fuels 26 (2012) 1412–1419, https://doi.org/10.1021/ef2012744.
[3] E.H. Benmekki, G.A. Mansoori, Minimum miscibility pressure prediction with equations of state, SPE (Society Pet. Eng. Reserv. Eng.; (United States)). 3 (1988) 2.
[4] A. Fazlali, M. Nikookar, A. Agha-Aminiha, A.H. Mohammadi, Prediction of minimum miscibility pressure in oil reservoirs using a modified SAFT equation of state, Fuel 108 (2013) 675–681, https://doi.org/10.1016/j.fuel.2012.12.091.
[5] J.-N. Jaubert, L. Wolff, E. Neau, L. Avaullee, A very simple multiple mixing cell calculation to compute the minimum miscibility pressure whatever the displacement mechanism, Ind. Eng. Chem. Res. 37 (1998) 4854–4859.
[6] K. Ahmadi, Advances in Calculation of Minimum Miscibility Pressure, PhD Thesis (2011), p. 267.
[7] R.S. Wu, J.P. Batycky, Evaluation of miscibility from slim tube tests, J. Can. Pet. Technol. 29 (1990), https://doi.org/10.2118/90-06-06.
[8] R.L. Christiansen, H.K. Haines, Rapid measurement of minimum miscibility pressure with the rising-bubble apparatus, SPE Reservoir Eng. 2 (1987) 523–527.
[9] F.M. Orr, K. Jessen, An analysis of the vanishing interfacial tension technique for determination of minimum miscibility pressure, Fluid Phase Equil. 255 (2007) 99–109, https://doi.org/10.1016/J.FLUID.2007.04.002.
[10] C. Cronquist, Carbon dioxide dynamic miscibility with light reservoir oils, Proc. 4th Annu. U.S. DOE Symp, 1978, pp. 28–30.
[11] W.F. Yellig, R.S. Metcalfe, Determination and prediction of CO2 minimum miscibility pressures (includes associated paper 8876), J. Petrol. Technol. 32 (1980) 160–168, https://doi.org/10.2118/7477-PA.
[12] F.M. Orr, C.M. Jensen, Interpretation of pressure-composition phase diagrams for CO2/crude-oil systems, Soc. Petrol. Eng. J. 24 (1984) 485–497, https://doi.org/10.2118/11125-PA.
[13] M.K. Emera, H.K. Sarma, Use of genetic algorithm to estimate CO2–oil minimum miscibility pressure—a key parameter in design of CO2 miscible flood, J. Petrol. Sci. Eng. 46 (2005) 37–52, https://doi.org/10.1016/J.PETROL.2004.10.001.
[14] E.M.E.-M. Shokir, CO2–oil minimum miscibility pressure model for impure and pure CO2 streams, J. Petrol. Sci. Eng. 58 (2007) 173–185.
[15] R.B. Alston, G.P. Kokolis, C.F. James, CO2 minimum miscibility pressure: a correlation for impure CO2 streams and live oil systems, Soc. Petrol. Eng. J. 25 (1985) 268–274, https://doi.org/10.2118/11959-PA.
[16] H.M. Sebastian, R.S. Wenger, T.A. Renner, Correlation of minimum miscibility pressure for impure CO2 streams, J. Petrol. Technol. 37 (1985) 2076–2082.
[17] B.E. Eakin, F.J. Mitch, Measurement and correlation of miscibility pressures of reservoir oils, SPE Annu. Tech. Conf. Exhib, Society of Petroleum Engineers, 1988, , https://doi.org/10.2118/18065-MS.
[18] J.I. Lee, Effectiveness of Carbon Dioxide Displacement under Miscible and Immiscible Conditions, Rep. RR-40, Pet. Recover. Inst., Calgary, 1979.
[19] M. Dong, Task 3-minimum Miscibility Pressure (MMP) Studies, Pet. Res. Center, Saskatchewan Res. Counc. Saskatchewan, Canada, 1999.
[20] H. Yuan, R.T. Johns, A.M. Egwuenu, B. Dindoruk, Improved MMP correlation for CO2 floods using analytical theory, SPE Reservoir Eval. Eng. 8 (2005) 418–425.
[21] T.O. Owolabi, M.A. Gondal, Development of hybrid extreme learning machine based chemo-metrics for precise quantitative analysis of LIBS spectra using internal reference pre-processing method, Anal. Chim. Acta (2018) 33–41.
[22] A. Garg, J.S.L. Lam, B.N. Panda, A hybrid computational intelligence framework in modelling of coal-oil agglomeration phenomenon, Appl. Soft Comput. 55 (2017) 402–412.
[23] T.O. Owolabi, M. Gondal, Novel techniques for enhancing the performance of support vector regression chemo-metric in quantitative analysis of LIBS spectra, J. Anal. At. Spectrom. (2017), https://doi.org/10.1039/C7JA00229G.
[24] M. Raos, D. Petković, M. Protić, M. Jovanović, D. Marković, Selection of the most influential flow and thermal parameters for predicting the efficiency of activated carbon filters using neuro-fuzzy technique, Build. Environ. 104 (2016) 68–75.
[25] M. Gocic, S. Shamshirband, Z. Razak, D. Petković, S. Ch, S. Trajkovic, Long-term precipitation analysis and estimation of precipitation concentration index using three support vector machine methods, Adv. Meteorol. 2016 (2016).
[26] S. Shamshirband, M. Tabatabaei, M. Aghbashlo, Support vector machine-based exergetic modelling of a DI diesel engine running on biodiesel – diesel blends containing expanded polystyrene, Appl. Therm. Eng. 94 (2016) 727–747.
[27] S. Jović, N. Arsić, J. Vilimonović, D. Petković, Thermal sensation prediction by soft computing methodology, J. Therm. Biol. 62 (2016) 106–108.
[28] T.O. Owolabi, K.O. Akande, S.O. Olatunji, Application of computational intelligence technique for estimating superconducting transition temperature of YBCO superconductors, Appl. Soft Comput. 43 (2016) 143–149.
[29] T.O. Owolabi, M. Faiz, S.O. Olatunji, I.K. Popoola, Computational intelligence method of determining the energy band gap of doped ZnO semiconductor, Mater. Des. 101 (2016) 277–284.
[30] T.O. Owolabi, M. Faiz, S.O. Olatunji, IdrisK. Popoola, Computational intelligence method of determining the energy band gap of doped ZnO semiconductor, Mater. Des. 101 (2016) 277–284, https://doi.org/10.1016/J.MATDES.2016.03.116.
[31] M. Adibifard, S.A.R. Tabatabaei-Nejad, E. Khodapanah, Artificial Neural Network (ANN) to estimate reservoir parameters in Naturally Fractured Reservoirs using well test data, J. Petrol. Sci. Eng. 122 (2014) 585–594, https://doi.org/10.1016/J.PETROL.2014.08.007.
[32] S. Rafiee-Taghanaki, M. Arabloo, A. Chamkalani, M. Amani, M.H. Zargari, M.R. Adelzadeh, Implementation of SVM framework to estimate PVT properties of reservoir oil, Fluid Phase Equil. 346 (2013) 25–32, https://doi.org/10.1016/J.FLUID.2013.02.012.
[33] E.A. El-Sebakhy, Forecasting PVT properties of crude oil systems based on support vector machines modeling scheme, J. Petrol. Sci. Eng. 64 (2009) 25–34.
[34] S. Nowroozi, M. Ranjbar, H. Hashemipour, M. Schaffie, Development of a neural fuzzy system for advanced prediction of dew point pressure in gas condensate reservoirs, Fuel Process. Technol. 90 (2009) 452–457.
[35] A. Shahkarami, S.D. Mohaghegh, V. Gholami, S.A. Haghighat, Artificial intelligence (AI) assisted history matching, SPE West. North Am. Rocky Mt. Jt. Meet, Society of Petroleum Engineers, 2014, , https://doi.org/10.2118/169507-MS.
[36] M. Nait Amar, N. Zeraibi, K. Redouane, Optimization of WAG process using dynamic proxy, genetic algorithm and ant colony optimization, Arabian J. Sci. Eng. (2018) 1–14, https://doi.org/10.1007/s13369-018-3173-7.
[37] M. Nait Amar, N. Zeraibi, K. Redouane, Bottom hole pressure estimation using hybridization neural networks and grey wolves optimization, Petroleum (2018), https://doi.org/10.1016/j.petlm.2018.03.013.
[38] F. Anifowose, A. Abdulraheem, Fuzzy logic-driven and SVM-driven hybrid computational intelligence models applied to oil and gas reservoir characterization, J. Nat. Gas Sci. Eng. 3 (2011) 505–517.
[39] K.O. Akande, T.O. Owolabi, S.O. Olatunji, A. AbdulRaheem, A hybrid particle swarm optimization and support vector regression model for modelling permeability prediction of hydrocarbon reservoir, J. Petrol. Sci. Eng. 150 (2017) 43–53.
[40] F. Anifowose, S. Adeniye, A. Abdulraheem, A. Al-Shuhail, Integrating seismic and log data for improved petroleum reservoir properties estimation using non-linear feature-selection based hybrid computational intelligence models, J. Petrol. Sci. Eng. 145 (2016) 230–237.
[41] Y.F. Huang, G.H. Huang, M.Z. Dong, G.M. Feng, Development of an artificial neural network model for predicting minimum miscibility pressure in CO2 flooding, J. Petrol. Sci. Eng. 37 (2003) 83–95, https://doi.org/10.1016/S0920-4105(02)00312-1.
[42] M.A. Ahmadi, M. Zahedzadeh, S.R. Shadizadeh, R. Abbassi, Connectionist model for predicting minimum gas miscibility pressure: application to gas injection process, Fuel 148 (2015) 202–211, https://doi.org/10.1016/J.FUEL.2015.01.044.
[43] H. Sayyad, A.K. Manshad, H. Rostami, Application of hybrid neural particle swarm optimization algorithm for prediction of MMP, Fuel 116 (2014) 625–633.
[44] M.-A. Ahmadi, M. Ebadi, Fuzzy modeling and experimental investigation of minimum miscible pressure in gas injection process, Fluid Phase Equil. 378 (2014) 1–12, https://doi.org/10.1016/J.FLUID.2014.06.022.
[45] A. Kamari, M. Arabloo, A. Shokrollahi, F. Gharagheizi, A.H. Mohammadi, Rapid method to estimate the minimum miscibility pressure (MMP) in live reservoir oil systems during CO2 flooding, Fuel 153 (2015) 310–319.
[46] M. Fathinasab, S. Ayatollahi, On the determination of CO2–crude oil minimum miscibility pressure using genetic programming combined with constrained multivariable search methods, Fuel 173 (2016) 180–188.
[47] M.A. Ahmadi, S. Zendehboudi, L.A. James, A reliable strategy to calculate minimum miscibility pressure of CO2-oil system in miscible gas flooding processes, Fuel 208 (2017) 117–126, https://doi.org/10.1016/J.FUEL.2017.06.135.
[48] S.R. Na'imi, S.R. Shadizadeh, M.A. Riahi, M. Mirzakhanian, Estimation of reservoir porosity and water saturation based on seismic attributes using support vector regression approach, J. Appl. Geophys. 107 (2014) 93–101.
[49] E.T. Al-Shammari, A. Keivani, S. Shamshirband, A. Mostafaeipour, P.L. Yee, D. Petković, S. Ch, Prediction of heat load in district heating systems by Support Vector Machine with Firefly searching algorithm, Energy 95 (2016) 266–273.
[50] Z. Zhong, T.R. Carr, Application of mixed kernels function (MKF) based support vector regression model (SVR) for CO2– Reservoir oil minimum miscibility pressure prediction, Fuel 184 (2016) 590–603, https://doi.org/10.1016/j.fuel.2016.07.030.
[51] V.N. Vapnik, M. Jordan, S.L. Lauritzen, J.L. Lawless, V. Nair (Eds.), The Nature of Statistical Learning Theory, 1995.
[52] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (1998) 121–167, https://doi.org/10.1023/A:1009715923555.
[53] J. Shawe-Taylor, N. Cristianini, Kernel Methods for Pattern Analysis, Cambridge University Press, 2004 10.2277.

[54] M. Awad, R. Khanna, Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, Apress, 2015.
[55] J.C. Platt, 12 fast training of support vector machines using sequential minimal optimization, Adv. Kernel Methods. (1999) 185–208.
[56] Y. Kaneda, H. Mineno, Sliding window-based support vector regression for predicting micrometeorological data, Expert Syst. Appl. 59 (2016) 217–225.
[57] D. Karaboga, An Idea Based on Honey Bee Swarm for Numerical Optimization, Tech. Rep. TR06, Erciyes Univ., 2005, p. 10 doi:citeulike-article-id:6592152.
[58] D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm, J. Global Optim. 39 (2007) 459–471, https://doi.org/10.1007/s10898-007-9149-x.
[59] H. Zhang, D. Hou, K. Li, An improved CO2-crude oil minimum miscibility pressure correlation, J. Chem. 2015 (2015), https://doi.org/10.1155/2015/175940.
[60] M. Dong, S. Huang, S.B. Dyer, F.M. Mourits, A comparison of CO2 minimum miscibility pressure determinations for Weyburn crude oil, J. Petrol. Sci. Eng. 31 (2001) 13–22, https://doi.org/10.1016/S0920-4105(01)00135-8.
[61] A.M. Elsharkawy, F.H. Poettmann, R.L. Christiansen, Measuring minimum miscibility pressure: slim-tube or rising-bubble method? SPE/DOE Enhanc. Oil Recover. Symp, Society of Petroleum Engineers, 1992, , https://doi.org/10.2118/24114-MS.
[62] W.N. Adyani, N.I. Kechut, Advanced technology for rapid minimum miscibility pressure determination (Part 1), Asia Pacific Oil Gas Conf. Exhib, Society of Petroleum Engineers, 2007, , https://doi.org/10.2118/110265-MS.
[63] O. Adekunle, Experimental Approach to Investigate Minimum Miscibility Pressures in the Bakken, (2014).
[64] R.M. Dicharry, T.L. Perryman, J.D. Ronquille, Evaluation and design of a CO2 miscible flood project-SACROC unit, Kelly-snyder field, J. Petrol. Technol. 25 (1973) 1309–1318, https://doi.org/10.2118/4083-PA.
[65] R. Winzinger, J.L. Brink, K.S. Patel, C.B. Davenport, Y.R. Patel, G.C. Thakur, Design of a major CO2 flood, north ward estes field, ward county, Texas, SPE Reservoir Eng. 6 (1991) 11–16, https://doi.org/10.2118/19654-PA.
[66] Y. Zuo, J. Chu, S. Ke, T. Guo, A study on the minimum miscibility pressure for miscible flooding systems, J. Petrol. Sci. Eng. 8 (1993) 315–328.
[67] N. Samani, M. Gohari-Moghadam, A.A. Safavi, A simple neural network model for the determination of aquifer parameters, J. Hydrol. 340 (2007) 1–11.
[68] B. Vaferi, F. Samimi, E. Pakgohar, D. Mowla, Artificial neural network approach for prediction of thermal behavior of nanofluids flowing through circular tubes, Powder Technol. 267 (2014) 1–10, https://doi.org/10.1016/J.POWTEC.2014.06.062.
[69] MATLAB - MathWorks 2016.
[70] A.E. Eiben, R. Hinterding, Z. Michalewicz, Parameter control in evolutionary algorithms, IEEE Trans. Evol. Comput. 3 (1999) 124–141, https://doi.org/10.1109/4235.771166.
[71] P. Rousseeuw, A. Leroy, Robust Regression and Outlier Detection, Wiley, 2003.
[72] C.R. Goodall, 13 Computation using the QR decomposition, Handb. Stat. 9 (1993) 467–508, https://doi.org/10.1016/S0169-7161(05)80137-3.
[73] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (2007) 694–701, https://doi.org/10.1002/qsar.200610151.