REGULAR ARTICLE

Application of a new approach for modeling the oil field formation damage due to mineral scaling

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

[1] Department of Petroleum Engineering, Petroleum University of Technology (PUT), P.O. Box 6198144471, Ahwaz, Iran
[2] Department of Chemical and Petroleum Engineering, Sharif University of Technology (SUT), P.O. Box 113659465, Tehran, Iran

Abstract. Mineral scaling has been considered a great concern for developing the oil production from the underground petroleum reservoirs. One of the main causes of this phenomenon is known as the chemical incompatibility of injected brine, frequently sea water, with the reservoir brine leading to the deposition of various supersaturated salts such as calcium carbonate, calcium sulfate and barium sulfate. In present communication, an evolutionary approach namely, Gene Expression Programming (GEP), was employed for rigorous modeling of formation damage by mineral scaling of mixed sulfate salt deposition. At first, a large databank of damaged permeability datapoints as a function of injected volume, injection flowrate, temperature, differential pressure and ionic concentrations of the existing chemical species in the porous media was employed. In this regard, a user-friendly correlation was extended for the first time by the aforementioned technique in the literature. Professional evaluation of the suggested GEP-based model was implemented by different statistical parameters and appealing visualization tools. Having proposed the GEP-based correlation, statistical parameters of the Average Absolute Relative Deviation Percent (AARD%) of 0.640% and determination coefficient (R^2) of 0.984 was calculated. Accordingly, it is demonstrated that the proposed model has a superior performance and great potential for efficient prediction of damaged permeability due to the mixed sulfate salt scaling. Moreover, the implemented outlier diagnosis technique verified the validity of the databank used for modeling, as well as the high robustness of the suggested model was confirmed. In conclusion, the developed correlation in this work can be of enormous practical value for skillful engineers and scientists in any academic study and industrial applications dealing with mixed salt deposition.

1 Introduction

In water injection scheme, deposition of mineral scales has been considered a great challenge for production development of subterranean petroleum reservoirs [1]. Success of a water injection project can be jeopardized by such scale, and even worse it can be terminated in the worst operational conditions. Therefore, the efficiency of this process, depending upon the degree of pressure maintenance and the oil production level, can be reduced. The amount of formation damage caused by the scaling is chiefly quantified by the well-known terms of permeability and porosity reduction [2, 3].

The mechanism of permeability reduction leading to the formation damage in the porous media by precipitated minerals is the deposition of the supersaturated minerals on the pore walls due to the attraction forces between the pore surface area and scale solid particles. In such condition,

a number of bridges made of the scale particles across the pore throats will be created, as well as pore throats blockage by a single particle will strongly happen. The degree of formation damage is affected by the features of minerals precipitate. The morphology and quantity of the growing crystals on the pore walls are monitored by means of several factors including existence of impurities, mixing rate, temperature change and supersaturation owing to the variation in physical conditions (*e.g.*, chemical incompatibility) [4–7].

Chemical incompatibility is classified as the main reason of the severe oilfield scaling when incompatible waters are mixed. While water is injected into the reservoir, a reaction between the formation minerals and formation water with the injected fluid will take place resulting in the establishment of scale formation through the porous media. In other words, based on the so-called phenomenon of the chemical incompatibility, the reservoir brine and injection water will undergo the strong chemical interactions together leading

---

* Corresponding authors: alireza.rostami.put2014@gmail.com; alireza.rostami@afp.put.ac.ir; shahbazi@put.ac.ir

to the precipitation of the supersaturated minerals. Precipitation of inorganic scale is dependent upon the ionic composition of fluid and the nature of scale due to their influence on the molecular or atomic interaction amongst the involved particles of commingling brines [8–11].

Precipitation instigates as long as the nucleation process have turned on. Nucleation is defined as the joining of atoms attracted for establishing submicron nuclei. The presence of impurities in fluid makes lower energy needed for construction of such nuclei than the required energy for a pure fluid. Because the impure components perform as proper sites for nucleation, in which it is termed as heterogeneous nucleation. The process of heterogeneous nucleation extensively takes place where suspended inorganic particles exist in the formation water coming into contact with the injection water [4, 12–16].

Generally, sea water is used as the injected fluid in both secondary and tertiary oil recovery processes, which is rich in sulfate and bicarbonate anionic chemical species. Despite sea water as the injection fluid, high contents of cationic species like calcium, barium and strontium exist in the formation water which can form diverse types of sulfate scales (*i.e.*, $CaSO_4$, $BaSO_4$, $SrSO_4$) and carbonate scales (*i.e.*, $CaCO_3$) in the oilfield. Hence sulfate and carbonate scales are recognized as the two key classes of scaling minerals in the oilfields [17, 18]. Notwithstanding carbonate scaling minerals in which they are extremely reliant on pressure alterations and pH changes, sulfate scales commonly happen on account of incompatible waters mixing and temperature ups and downs [19–21]. In addition to the sea water, disposal water, which is produced in association with oilfield production, can also be recycled and re-injected into the reservoir, even though the problem of chemical incompatibility is still present in this case. It is noteworthy that injection of compatible water resources with the connate and formation water is something impossible to implement [22, 23].

Relatively good solubility of carbonate scaling minerals makes the easier inhibition of them by conventional techniques such as soaking with suitable acid dissolver than the sulfate scales. Thus, the deposition of such scales downstream of wellhead equipment can be resolved through nonstop injection of an appropriate inhibitor into the transportation lines. Nevertheless, sulfate scales (*e.g.*, barium sulfate) have very poor solubility, low tendency for reacting with most acids, great hardness and exposures of very little surface area at the time of deposition. In consequence, restricted number of removal methods is existing in order to preclude their deposition in sensitive regions. For this reason, squeeze treatments of scale inhibitor are usually applied so that sulfate mineral scaling will be inhibited in downstream or upstream of first completion and any mixing location [18, 20, 24, 25].

Aiming for evaluating the impact of different parameters on scale deposition such as incompatibility of mixing fluids, temperature, pressure, concentration of chemical species present in commingling fluids, innumerable investigations have been implemented [17, 18, 26, 27]. The impact of incompatible waters on the formation of carbonate and calcium sulfate scales in the synthetic porous media was extensively studied by Moghadasi *et al.* [2]. In continuum study, more experimental and theoretical investigations were conducted on the degree of permeability reduction as a measure of formation damage caused by calcium carbonate and calcium sulfate scaling through mixing carbonate/sulfate rich and calcium rich solutions by Moghadasi *et al.* [28].

Along with the experimental studies, a number of theoretical studies have been focused on the prediction of mineral scale deposition in porous media [29–32]. In view of the hydrodynamic and kinetics of gypsum deposition, Jamialahmadi *et al.* [33] have instituted a model with outstanding performance for mathematical modeling of deposition and removal of gypsum scaling integrating the impact of salt supersaturation, injection flowrate and temperature. This model may work ineffectively in order to predict the value of permeability reduction when mixed salt precipitation happens in solution. In more recent years, Safari and Jamialahmadi [19] initiated a highly nonlinear simulator on the basis of hydrodynamic, kinetics and thermodynamic laws for modeling deposition of both single salt (*i.e.*, barium sulfate), and mixed salt (*i.e.*, strontium sulfate in connection with barium sulfate) during fluid flow through porous media. Then, the authors optimized the kinetic coefficients by means of a hybrid approach namely, Pattern Search (PS) algorithm in cooperation with Particle Swarm Optimization (PSO) technique. Finally, they concluded that their model has an acceptable agreement with the experimental data with deviations less than 10% [19]. Even though the authors carried a great job out, their model is highly complex with lots of coefficients to be tuned. So, there is a vital requisite for constructing a universal model in order to have a rapid and precise estimation of permeability impairment for mixed sulfate salts scaling during water flooding process in the porous media.

Genetic based calculations have been commonly used in petroleum industry as a promising approach in estimating several parameters [34–36]. In recent years, Gene Expression Programming (GEP) [37] as an evolutionary algorithm, has been increasingly applied in different disciplines of petroleum and chemical engineering. The application of GEP [37] algorithm resulted in development of accurate models in wide varieties of petroleum industry which generally gives more precise estimates than the pre-existing models. In accordance with GEP [37] mathematical strategy, the optimum correlation format will be initiated without any assumption about the form of equation. Successful examples of GEP scheme applied in the literature can be found in the work of several researchers in the open literature [38, 39].

In current study, potential application of GEP algorithm as a powerful technique is presented so as to prepare a possible solution to the all disadvantages earlier mentioned in the field of mixed sulfate salt scaling in the porous media. For this reason, a large database was adopted from the open literature [40–43]. Afterwards, the database is divided into two sets of training (about 341 datapoints) and testing (about 85 datapoints). According to the

Table 1. Statistical specifications of the database utilized for developing the correlation.

Parameter	Unit	Minimum	Average	Maximum	SD[a]
ΔP	psi	100.00	150.12	200.00	40.85
T	°C	50.00	66.71	80.00	12.48
Q	cc/min	8.55	17.78	31.33	5.37
K_i	md	12.30	12.99	13.87	0.52
V_{inj}	PV	1.47	23.72	83.80	15.55
$C_{Ca^{2+}}$	ppm	780	9592.76	30 000	11 999.84
$C_{Ba^{2+}}$	ppm	10	618.91	2200	919.49
$C_{Sr^{2+}}$	ppm	370	583.81	1100	301.36
$C_{SO_4^{2-}}$	ppm	2750	2854.76	2960	105
K_d	md	9.81	12.05	13.81	0.77

[a] SD refers to the standard deviation which is calculated as follows:

$$SD = \left(\frac{1}{N-1} \sum_{i=1}^{N} \left(S_i - \bar{S} \right)^2 \right)^{\frac{1}{2}}.$$

training dataset, the GEP-based empirically derived equation is developed. Throughout various statistical parameters and visualization tools the performance of the proposed model is exhibited. To the best of authors knowledge, there is no report on modeling permeability impairment as a representative of formation damage caused by mixed sulfate salt scaling in the open literature. To this end, the validity of the databank used for modeling is assessed by means of outlier analysis.

2 Data gathering

Based on the previous modeling studies in the field of soft computation, it has been demonstrated that development of a comprehensive model is crucially in the need of a large database. A database with the feature of all-inclusiveness makes the constructed model to be applied for a wider ranges of operational conditions. In other words, the proposed model is not limited to a specific condition and can be employed for mathematical description of the interest phenomenon at different conditions [35, 44–62]. In present research, 431 datapoints of permeability reduction values as a function of temperature, differential pressure, volume of injected water, initial permeability, flowrate, and ionic concentrations of cationic species (*i.e.*, strontium, calcium and barium) and anionic species (*i.e.*, sulfate) is taken from the open literature (see Supplementary Material) [40–43]. This database is used for developing and testing the capability of the proposed model. For model development and examining its capability, nearly 80% and 20% of the entire database are employed, respectively. This data division is carried out by a random computational process defined in GEP modeling. Table 1 shows the specification of the employed database for GEP modeling.

3 Gene expression programming

In an attempt to develop the genetic-based calculation, the most recent version of genetic computational models, namely GEP, in which the shortcomings of the preceding genetic models like Genetic Algorithm (GA) and Genetic Programming (GP) were modified in GEP computation strategy [63]. Unlike GP approach working with one element of Expression Trees (ETs), GEP scheme deals with two components, including ETs and chromosomes. Symbolic ETs are defined as the population individuals, and the chromosomes are responsible for encoding and translating the candidate solution into a real candidate solution as ETs [64]. In this regard, a typical chromosome is categorized into functions and variable/constant terminals. The constants are determined by the model program, however, the variables and functions are set as the inputs of the model. For each gene, the inputs and terminals are corresponded to, respectively, gene's head and gene's tail which are related as follows [65]:

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

where, the symbols n, h and t denote the largest function arity, the magnitude of gene's head and the length of gene's tail, respectively. Setting parameters of the used GEP strategy for modeling damaged permeability in this study are reported in Table 2. The similar translation procedure is observed in biological genes encoded in DNAs which are constantly transformed into proteins. Owing to the structural features of the chromosomes and reproduction processes accomplished to this technique, unlimited modifications of programs are obtained leading to effective solution to the problem [65]. It is confirmed that the convergence speed of the GEP mathematical strategy is two to four orders of magnitude larger than that of the

Table 2. Setting parameters of the used GEP strategy for modeling damaged permeability in this study.

GEP algorithm parameters	Value
No. of chromosomes	30
No. of genes	3
Head size	7
Linking function	$+$
Generations without change	2000
Fitness function	Root Mean Square Error
Inversion	0.00546
Mutation	0.00138
IS transposition	0.00546
RIS transposition	0.00546
One-point recombination	0.00277
Two-point recombination	0.00277
Gene transposition	0.00277
Gene recombination	0.00277
Permutation	0.00546
Constants per gene	10
Random chromosomes	0.0026
Type of data	Floating point
Random cloning	0.00102
Operators used	$+, -, \times, /, \sqrt{}$, EXP, INV, LN, LOG, X^2, POW

Expression Tree:

Tanh(l×m)+√(l-m)

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

GP scheme [66]. Figure 1 shows a typical two-gene chromosome made of three terminals l, m, n, and four functions "$\times, \pm, \sqrt{}, \tanh$" with its decoded ET and corresponding algebraic expression (correlation) with the Karva language illustration. The authors used a well-known and optimized programming code for GEP modeling to simulate the interested parameter in this study.

Developing the correlation

Based on the existing literature concentrated on the mineral scale formation in porous media, it is fully understood that the amount of permeability reduction as a measure of formation damage is under the influence of several independent variables. These variables include ionic concentrations of sulfate anion and divalent cations (*i.e.*, calcium, strontium, barium), differential pressure, temperature, injected volume and flowrate [19, 67–70]. Therefore, the proposed GEP-based model is extended, as follows:

$$K_\mathrm{d} = f\Big(\Delta P, \ T, \ Q, \ V_\mathrm{inj}, \ K_\mathrm{i}, \ C_\mathrm{Ca^{2+}}, \ C_\mathrm{Ba^{2+}}, \ C_\mathrm{Sr^{2+}}, \ C_\mathrm{SO_4^{2-}}\Big), \tag{2}$$

where, the symbols K_d, K_i, ΔP, T, Q, V_inj, $C_\mathrm{Ca^{2+}}$, $C_\mathrm{Sr^{2+}}$, $C_\mathrm{Ba^{2+}}$ and $C_\mathrm{SO_4^{2-}}$ indicate the damaged permeability, initial permeability, differential pressure, temperature, flowrate, injected volume, concentration of calcium ion, concentration of strontium ion, concentration of barium
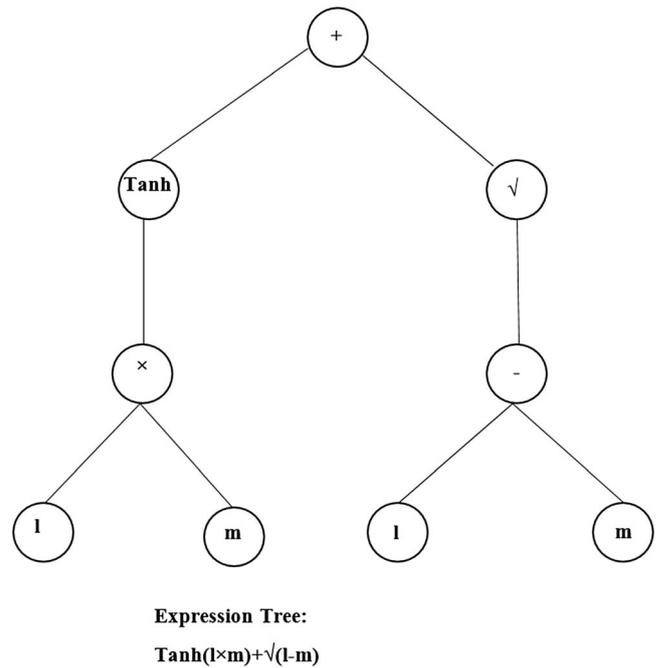
ion and concentration of sulfate ion, respectively. When the decision variables are defined, the subsequent mathematical strategy will be applied to find the optimal equation format described as below:

1. *Population preparation.* Randomly selected individual chromosomal structures throughout checking various algebraic operators (*e.g.*, $\times$, $/$, $\sqrt{}$, $\pm$), and setting terminals as functions of output and input data [71].

2. *Predicting fitness value.* For each individual, the Objective Function (OF) is predicted by the subsequent formulation:

$$\mathrm{OF}(i) = \frac{100}{N} \sum_i^N \frac{\left| K_\mathrm{di}^\mathrm{exp} - K_\mathrm{di}^\mathrm{pred} \right|}{K_\mathrm{di}^\mathrm{exp}}. \tag{3}$$

   In equation (3), N denotes the number of datapoints, and the superscripts "exp and pred" are, in turn, representatives for experimental and predicted values of permeability reduction [71].

3. *Individuals selection.* For replacement goals, the OF value gives an indication to select the appropriate individuals indicating suitable parents. For this reason, the so-called approach of tournament is utilized to prepare the adequate variety of dataset during each generation process [64, 72].

4. *Genetic operations.* Several operators including replication, mutation and inversion are applied for the goals of genes modification and reproduction. In replication stage, the chosen chromosomes used in step 3, are accurately duplicated [71]. Moreover, through

the application of the mutation operator, the effective adaptation of individuals' population will be resulted by selecting randomly engaged nodes, and replacing saved information with the random primitive from the similar arity. The mutation will be applied via alteration in the magnitudes of gene's head and gene's tail. For this, the mutation operation can be occurred everywhere in chromosomal structure by introducing the known quantity of mutation rate (p_{m}). By dint of modifying the randomly chosen gene's head, new individuals are developed in inversion process of GEP modeling. Having defined the well-known term of inversion rate (p_{i}), the efficiency of inversion operation can be evaluated simply [71].

5. *Insertion and transposition sequence components.* These transposable components can be shifted easily from one location to another one in a chromosomal structure [71]. Ferreira [37] introduced three types of such elements in his work, as follows: short fragments with first position function that move to the gene' root (RIS components); short fragments with either first position terminal or first position function moving to the gene's head; and whole genes that move to start of chromosomes.

6. *Recombination.* In this process, three different types of recombination, including one-point, two-point and gene, randomly select two chromosomes for exchanging certain materials together resulting in the extension of two new chromosomes. As a result, a new generation will be established. Considering a user-defined stopping condition, the aforementioned procedure will be repeated until the interested with satisfactory precision will be achieved. For more information, the interested readers are suggested to refer the extensive instances explained in the work of Ferreira [66].

4 Results and discussions

4.1 Benchmarks for evaluation of the proposed GEP model

In this section, a GEP-based empirically derived equation will be presented subsequently as a result of applying GEP mathematical strategy for the first time in this field of study. For better assessment of the proposed GEP-based correlation, various statistical quality measures are utilized including the Average Absolute Relative Deviation Percent (AARD%), Root Mean Square Error (RMSE), determination coefficient (R^2) and Average Relative Deviation Percent (ARD%). One of the most important statistical criterion applied in a wide variety of mathematical and numerical modeling in chemical and petroleum engineering is calculation of the AARD% value. The AARD% which is defined as the degree of model precision, directly indicates the total magnitude of the estimation error relative to the target experimental data. The higher value of AARD% shows the lower model accuracy. The quantity of RMSE illustrates the amount of inaccuracy happened in the process of modeling. In other words, it shows the magnitude

of deviation between the real and simulated data. The other important and widely used statistical parameter is R^2 which shows the goodness of fit, or it displays how well the model estimates are matched with the actual data. When the value of R^2 approaches to unity, the most satisfactory agreement will be achieved. The final benchmark is the ARD% value, which shows the quality of deviation distribution in the vicinity of zero deviation. The lower ARD% value near to zero confirms the more compacted concentration of error distribution around the zero deviation. In addition to the parametric evaluation of the proposed model, diverse graphical illustrations are utilized to confirm the superiority and large capability of the GEP-based empirically derived correlation. The most significant of all applied diagrams are crossplot, index plot, and error distribution plot which will be represented subsequently in this study.

4.2 Assessment of the proposed GEP model

Based on the GEP processing, a user-friendly equation is developed which can be used for fast estimation of damaged permeability by this novel approach for the first time in the bulk of research have been paying attention to the formation damage caused by mineral scaling. The proposed empirically-derived GEP equation is given, as below:

$$\log{(K_{\mathrm{d}})} = A_1 + A_2 + A_3, \tag{4}$$

$A_1 =$

$$\frac{(T - V_{\mathrm{inj}}) \times (Q - K_{\mathrm{i}})}{Q \times T^2 \times \left\{ \log\left(C_{\mathrm{Ba}^{2+}} + C_{\mathrm{Sr}^{2+}} + C_{\mathrm{Ca}^{2+}} + C_{\mathrm{SO}_4^{2-}} \right) - 2.77550702075275 \right\}}, \tag{5}$$

$A_2 =$

$$\frac{40.7928649580713 \times Q^2}{(\Delta P - 2.43561403505515)^2 \times \{0.0656999407983034 \times T - 1.46554080471316\}}, \tag{6}$$

$$A_3 = 0.0266417196782303 \times \left(\ln\left(302.487101618176 - \frac{Q}{V_{\mathrm{inj}}} \right) \right)^2, \tag{7}$$

where, the units of K_{d}, K_{i}, $\Delta\mathrm{P}$, T, Q, V_{inj}, $C_{\mathrm{Ca}^{2+}}$, $C_{\mathrm{Sr}^{2+}}$, $C_{\mathrm{Ba}^{2+}}$ and $C_{\mathrm{SO}_4^{2-}}$ in the above equations are md, md, psi, °C, cc/min, PV, ppm, ppm, ppm and ppm, respectively. The statistical details of the proposed GEP-derived correlation for the individual sets of training, test and total database are shown in Table 3. Based on this table, the values of R^2, ARD%, AARD% and RMSE for the total database are 0.9843, 0.0355%, 0.6409% and 0.0967, respectively. It means that the AARD% < 1% and R^2 > 0.98 which show the satisfactory performance of the proposed correlation. Moreover, this table confirms the successful testing of the proposed model because of the better performance of the test set than the training set.

The results of GEP predictions/calculations against the experimental damaged permeability are indicated in

Table 3. The statistical parameters of the developed GEP model for prediction of damaged permeability.

Parameter	Value
Training set	
$R^{2,a}$	0.9841
Average relative deviation, %[b]	0.0629
Average absolute relative deviation, %[c]	0.6565
Root mean square error[d]	0.0990
Number of data samples	345
Test set	
R^2	0.9858
Average relative deviation, %	−0.0746
Average absolute relative deviation, %	0.5782
Root mean square error	0.0872
Number of data samples	86
Total	
R^2	0.9843
Average relative deviation, %	0.0355
Average absolute relative deviation, %	0.6409
Root mean square error	0.0967
Number of data samples	431

[a] Determination coefficient:

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{N} \left((K_\mathrm{d})_i^{\mathrm{exp}} - (K_\mathrm{d})_i^{\mathrm{pred}} \right)^2}{\sum\limits_{i}^{N} \left((K_\mathrm{d})_i^{\mathrm{exp}} - (\bar{K}_\mathrm{d}) \right)^2}.$$

[b] Average relative deviation percent (ARD):

$$\mathrm{ARD}\% = \frac{100}{N} \sum_{i=1}^{N} \left(\frac{(K_\mathrm{d})_i^{\mathrm{exp}} - (K_\mathrm{d})_i^{\mathrm{pred}}}{(K_\mathrm{d})_i^{\mathrm{exp}}} \right).$$

[c] Average absolute relative deviation percent (AARD):

$$\mathrm{AARD}\% = \frac{100}{N} \sum_{i=1}^{N} \left(\left| \frac{(K_\mathrm{d})_i^{\mathrm{exp}} - (K_\mathrm{d})_i^{\mathrm{pred}}}{(K_\mathrm{d})_i^{\mathrm{exp}}} \right| \right).$$

[d] Root mean square error (RMSE):

$$\mathrm{RMSE} = \left(\frac{\sum_{i=1}^{N} \left((K_\mathrm{d})_i^{\mathrm{exp}} - (K_\mathrm{d})_i^{\mathrm{pred}} \right)^2}{N} \right)^{\frac{1}{2}}.$$

Figure 2. As can be seen in this crossplot, a compressed cloud of datapoints can be observed around the unit slope line or 45° line which shows the better agreement of the GEP model estimates in comparison with the corresponding measured data of damaged permeability. Because the most ideal performance of a model will be achieved when the predictions and actual data are the same leading to settling of the datapoints on unit slope line. For GEP model the values of AARD%, RMSE and ARD% are acceptably nearby the zero error value, and the R^2 value is close to unity. The other characteristic diagram for justified judgment of the GEP model is represented in Figure 3 exhibiting the distribution of GEP predictions error against the actual data of damaged permeability. In Figure 3, the relative deviation percent mainly changes in an acceptable range of −2 to 2%. Furthermore, the total value of ARD% which is equal to 0.0355%, demonstrates that the central focus of relative error distribution is sufficiently near to zero value. Figure 4 shows the schematic diagram for the error distribution of the proposed GEP-based equation *versus* the data
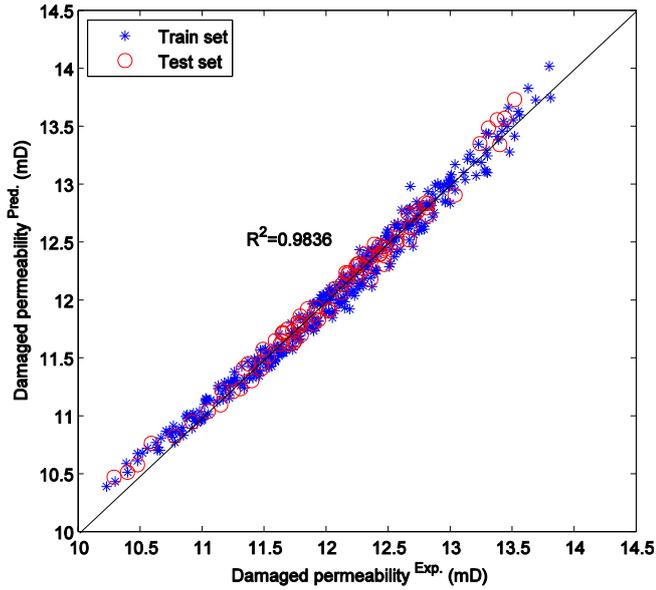
Fig. 2. Comparison between experimental damaged permeability and GEP predictions/calculations.
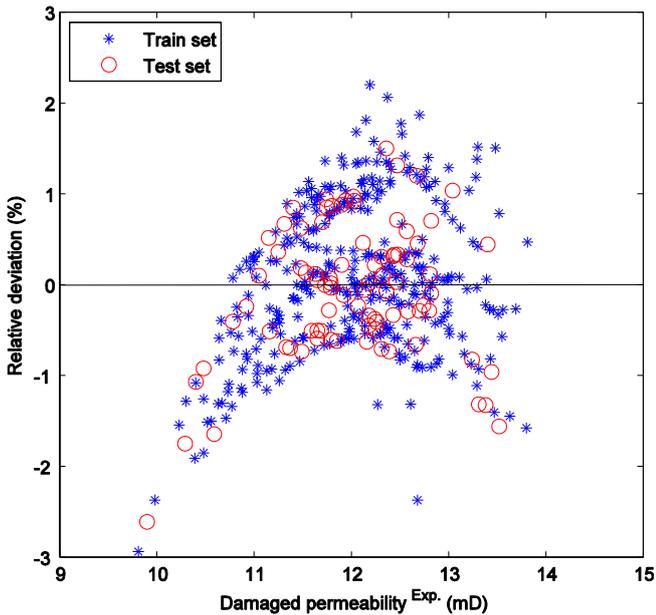

Fig. 3. Illustration of relative error distribution *versus* the damaged permeability.

Fig. 4. Distribution of relative deviation in the experimental dataset including train set, and test set.

frequency. Having focused on this figure, a normal distribution can be easily observed for both training and test subsets indicating the symmetry in the outcomes from current paper.

4.3 Outlier diagnosis

It has been known that the reliability of a utilized database directly affects the accuracy and validity of the constructed model. Nonetheless, proper data measurement is often not practicable, and diverse types of undesirable experimental
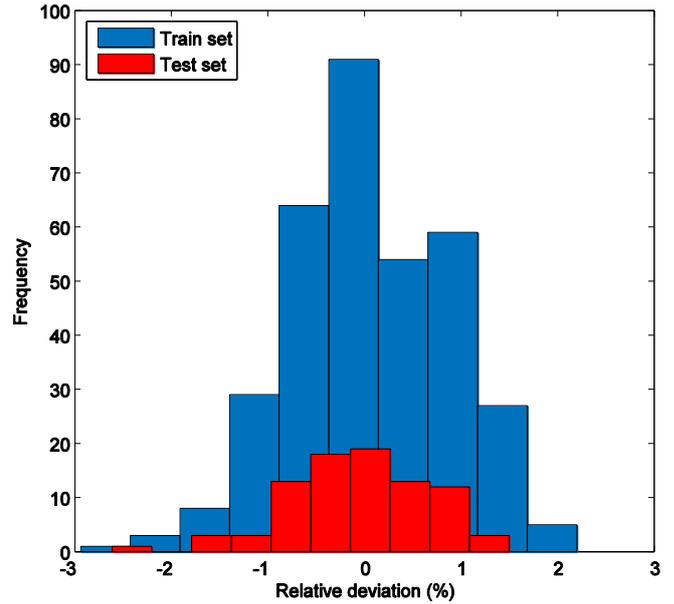
errors originated from human and equipment mistakes may diffuse into the measurements. Such unwanted deviations in experimental work are accounted as a menace to the success of modeling. Hence, detection of these suspected measurements from data is vital for any modeling study [67].

In an attempt to distinguish the invalid data, the so-called technique of Leverage Value Statistics (LVS) was conducted in current study. The statistical technique of LVS is commonly applied for describing the outlier data detecting the existing arrangement between the independent and dependent parameters. As it was previously explained, the suggested GEP model has a robust capability for predicting impaired permeability as a result of mineral scaling. In LVS processing, mathematical strategies contain the computation of Hat matrix and residual values for whole database. For calculating the residuals, the deviation between the target and the corresponding GEP predicted value have to be calculated for each datapoint. Besides, Hat matrix includes the GEP model estimates and the target experimental data, as below [73–76]:

$$H = X(X^t X)^{-1} X^t. \tag{8}$$

In equation (8), X is defined as the two-dimensional Hat matrix in which the possible area of the problem lay on the diagonal of this matrix, and the symbol t denotes the transpose operator. The total numbers of the model parameters and used data determine the number of the columns (n) and rows (m) of the Hat matrix, respectively.

The well-known diagram of William is broadly applied to recognize the suspected data according to the calculated residuals and the values of Hat matrix estimated by the equation (8). The relationship between the standardized residuals (R) and H indices are shown in William's plot. A cautionary leverage limit (H^*) is calculated as 3 $(n + 1)/m$. The criterion for measurement acceptability is
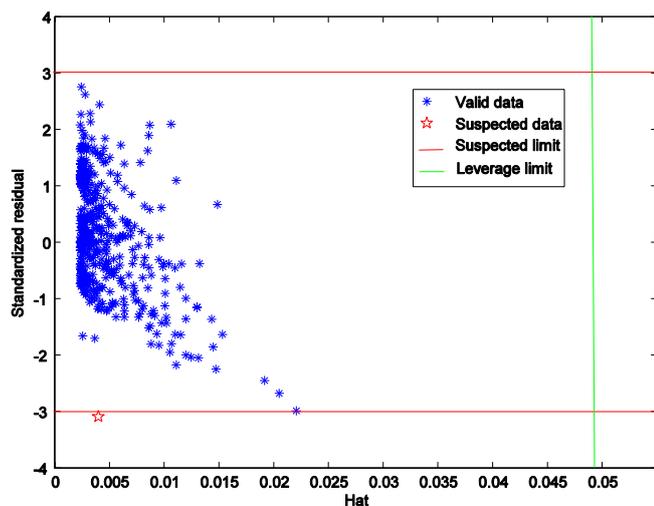
Fig. 5. Detection of probable outliers and applicability domain of the GEP model.

the presence of data in the cut-off limit of residuals that is equal to ± 3 (indicated by two horizontal lines in William's plot). In Figure 5, the result of outlier analysis is represented in which the most of data exist in the range of $-3 \leq R \leq 3$ and $0 \leq H \leq H^*$ verifying the high truthfulness and robustness of the GEP-based model in this study. With respect to the H and R ranges, three classes of outliers can be suggested including Regression, Bad High Leverage and Good High Leverage outliers [73–76].

When the conditions $-3 \leq R \leq 3$ and $H^* \leq H$ are fulfilled, the outlier is called "Good High Leverage". In this outlier, the measurements do not adequately affect the determination coefficient and accommodate nearly on the regression line which passes through the measured data, even though they have large values of leverage. The measurements with R values larger than 3 or less than -3 are under the class of "Bad High Leverage" outlier which is a serious intimidation for strong modeling. The intercept and the slope of the regression line are extremely affected by this outlier. The final type of outlier is named as "Regression" not disturbing the valid range and it has no impact on the regression line in spite of having large values of residual [73–76].

In this study, for recognizing the more likely invalid data, the Hat values were calculated via equation (8), then the William's plot was drawn in Figure 5. Accordingly, it is clear that just 1 datapoint of 427 datapoints is recognized as the off-range (outlier) data. As a consequence, the suggested GEP-derived model in this study is reliable and efficiently precise owing to the fact that the datapoints are widely held in the interior region of $-3 < R < 3$ and $0 < H < 0.0493$.

5 Conclusion

In current study, GEP as an evolutionary mathematical strategy was applied in order to estimate the damaged permeability as a result of mixed salt scaling. For this purpose, an extensive database was taken from the open

literature to develop a user-friendly and empirically-derived equation by this novel approach. The processed database was grouped into two subsets of training and test. The training group includes 80% of the entire database used for model development, as well as testing group includes 20% of the whole database applied for examining the model. Moreover, various schematic diagrams including crossplot, index plot and relative distribution plot were employed for better model evaluation. Calculation of different statistical quality measures for the proposed model results in the AARD% of 0.640%, the R^2 of 0.984, RMSE of 0.097 and the ARD% of 0.036%. Therefore, the GEP-based model developed in this study has proven to have an excellent precision and superior performance in estimating the impaired permeability in comparison with the experimentally measured datapoints. Finally, the proposed tool in this study is of tremendous practical value for quick and cheap prediction of impaired permeability in water flooding schemes during co-precipitation of mixed sulfate salts.

Supplementary Material

Supplementary Material is available at https://ogst.ifpenergiesnouvelles.fr/10.2516/ogst/2019032/olm.

References

1 Zabihi R., Schaffie M., Nezamabadi-pour H., Ranjbar M. (2011) Artificial neural network for permeability damage prediction due to sulfate scaling, *J. Pet. Sci. Eng.* **78**, 575–581.

2 Moghadasi J., Jamialahmadi M., Müller-Steinhagen H., Sharif A., Ghalambor A., Izadpanah M.R., Motaie E. (2003) Scale formation in Iranian oil reservoir and production equipment during water injection, in: *International Symposium on Oilfield Scale*, Society of Petroleum Engineers, Aberdeen, United Kingdom.

3 Lindlof J.C., Stoffer K.G. (1983) *A case study of seawater injection incompatibility* **35**, 1256–1262.

4 Aliaga D.A., Wu G., Sharma M.M., Lake L.W. (1992) Barium and calcium sulfate precipitation and migration inside sandpacks, SPE-19765-PA, *SPE Formation Evaluation*, **7**, 01, 79–86.

5 Boon J.A., Hamilton T., Holloway L., Wiwchar B. (1983) Reaction between rock matrix and injected fluids in cold lake oil sand – spotential for formation damage, *J. Can. Pet. Technol.* **22**, 04, 55–66.

6 Cusack F., Brown D.R., Costerton J.W., Clementz D.M. (1987) Field and laboratory studies of microbial/fines plugging of water injection wells: Mechanism, diagnosis and removal, *J. Pet. Sci. Eng.* **1**, 39–50.

7 El-Hattab M.I. (1985) Scale deposition in surface and subsurface production equipment in the Gulf of Suez, *J. Pet. Technol.* **37**, 09, 1640–1652.

8 Bayona H.J. (1993) A review of well injectivity performance in Saudi Arabia's Ghawar field seawater injection program, in: *Middle East Oil Show*, Society of Petroleum Engineers, Bahrain.

9 Stalker R., Collins I.R., Graham G.M. (2003) The impact of chemical incompatibilities in commingled fluids on the efficiency of a produced water reinjection system: A North

Sea example, in: *International Symposium on Oilfield Chemistry*, Society of Petroleum Engineers, Houston, Texas.

10 Bedrikovetsky P., Marchesin D., Shecaira F., Serra A.L., Marchesin A., Rezende E., Hime G. (2001) Well impairment during sea/produced water flooding: Treatment of laboratory data, in: *SPE Latin American and Caribbean Petroleum Engineering Conference*, Society of Petroleum Engineers, Buenos Aires, Argentina.

11 Ahmed S.J. (2004) *Laboratory study on precipitation of calcium sulphate in berea sandstone cores*, Doctoral dissertation, King Fahd University of Petroleum & Minerals, Saudi Arabia.

12 Gunn D.J., Murthy M.S. (1972) Kinetics and mechanisms of precipitations, *Chem. Eng. Sci.* **27**, 1293–1313.

13 Liu S.-T., Nancollas G.H. (1975) The crystal growth and dissolution of barium sulfate in the presence of additives, *J. Coll. Interf. Sci.* **52**, 582–592.

14 Walton A.G., Füredi H., Elving P.J., Kolthoff I.M. (1967) *The formation and properties of precipitates*, Vol. **23**, Interscience Publishers, New York, pp. 36–38.

15 Nancollas G.H., Eralp A.E., Gill J.S. (1978) Calcium sulfate scale formation: A kinetic approach, *Soc. Pet. Eng. J.* **18**, 133–138.

16 Nancollas G.H., Purdie N. (1963) Crystallization of barium sulphate in aqueous solution, *Trans. Faraday Soc.* **59**, 735–740.

17 Mitchell R.W., Grist D.M., Boyle M.J. (1980) Chemical treatments associated with North Sea projects, *J. Pet. Technol.* **32**, 904–912.

18 Yuan M. (1989) *Prediction of sulphate scaling tendency and investigation of barium and strontium sulphate solid solution scale formation*, Doctoral dissertation, Heriot-Watt University, Edinburgh.

19 Safari H., Jamialahmadi M. (2014) Thermodynamics, kinetics, and hydrodynamics of mixed salt precipitation in porous media: Model development and parameter estimation, *Transp. Porous Media* **101**, 477–505.

20 Safari H., Jamialahmadi M. (2014) Estimating the kinetic parameters regarding barium sulfate deposition in porous media: A genetic algorithm approach, *Asia-Pacific J. Chem. Eng.* **9**, 256–264.

21 Vitthal S., Sharma M.M. (1992) A Stokesian dynamics model for particle deposition and bridging in granular media, *J. Coll. Interf. Sci.* **153**, 314–336.

22 Andersen K.I., Halvorsen E., Sælensminde T., Østbye N.O. (2000) Water management in a closed loop – Problems and solutions at brage field, in: *SPE European Petroleum Conference*, Society of Petroleum Engineers, Paris, France.

23 Paulo J., Mackay E.J., Menzies N., Poynton N. (2001) Implications of brine mixing in the reservoir for scale management in the Alba field, in: *International Symposium on Oilfield Scale*, Society of Petroleum Engineers, Aberdeen, United Kingdom.

24 Mackay E. (2003) Predicting in situ sulphate scale deposition and the impact on produced ion concentrations, *Chem. Eng. Res. Des.* **81**, 326–332.

25 McElhiney J.E., Sydansk R.D., Lintelmann K.A., Benzel W.M., Davidson K.B. (2001) Determination of in-situ precipitation of barium sulphate during coreflooding, in: *International Symposium on Oilfield Scale*, Society of Petroleum Engineers, Aberdeen, United Kingdom.

26 Weintritt D.J., Cowan J.C. (1967) Unique characteristics of barium sulfate scale deposition, *J. Pet. Technol.* **19**, 1381–1394.

27 Read P.A., Ringen J.K. (1982) The use of laboratory tests to evaluate scaling problems during water injection, in: *SPE Oilfield and Geothermal Chemistry Symposium*, Society of Petroleum Engineers, Dallas, Texas.

28 Moghadasi J., Jamialahmadi M., Müller-Steinhagen H., Sharif A. (2004) Formation damage due to scale formation in porous media resulting from water injection, in: *SPE International Symposium and Exhibition on Formation Damage Control*, Society of Petroleum Engineers, Lafayette, Louisiana.

29 Chang F., Civan F. (1991) Modeling of formation damage due to physical and chemical interactions between fluids and reservoir rocks, in: *SPE Annual Technical Conference and Exhibition*, Society of Petroleum Engineers, Dallas, Texas.

30 Yeboah Y.D., Somuah S.K., Saeed M.R. (1993) A new and reliable model for predicting oilfield scale formation, in: *SPE International Symposium on Oilfield Chemistry*, Society of Petroleum Engineers, New Orleans, Louisiana.

31 Bertero L., Chierici G.L., Gottardi G., Mesini E., Mormino G. (1988) Chemical equilibrium models: Their use in simulating the injection of incompatible waters, SPE-14126-PA, *SPE Reservoir Engineering*, **3**, 01, 288–294.

32 Thomas L.G., Albertsen M., Perdeger A., Knoke H.H.K., Horstmann B.W., Schenk D. (1995) Chemical characterization of fluids and their modelling with respect to their damage potential in injection on production processes using an expert system, in: *SPE International Symposium on Oilfield Chemistry*, Society of Petroleum Engineers, San Antonio, Texas.

33 Jamialahmadi M., Muller-Steinhagen H. (2008) Mechanisms of scale deposition and scale removal in porous media, *Int. J. Oil Gas Coal Technol.* **1**, 81–108.

34 Creton B., Lévêque I., Oukhemanou F. (2019) Equivalent alkane carbon number of crude oils: A predictive model based on machine learning, *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **74**, 30.

35 Rostami A., Shokrollahi A., Ghazanfari M.H. (2018) New method for predicting n-tetradecane/bitumen mixture density: Correlation development, *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **73**, 35.

36 Sales L.d.P.A, Pitombeira-Neto A.R., de Athayde Prata B. (2018) A genetic algorithm integrated with Monte Carlo simulation for the field layout design problem, *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **73**, 24.

37 Ferreira C. (2006) Designing neural networks using gene expression programming. In *Applied soft computing technologies: The challenge of complexity*, Springer, Berlin, Heidelberg, pp. 517–535.

38 Gharagheizi F., Ilani-Kashkouli P., Farahani N., Mohammadi A.H. (2012) Gene expression programming strategy for estimation of flash point temperature of non-electrolyte organic compounds, *Fluid Phase Equilib.* **329**, 71–77.

39 Gharagheizi F., Eslamimanesh A., Sattari M., Mohammadi A.H., Richon D. (2013) Development of corresponding states model for estimation of the surface tension of chemical compounds, *AIChE J.* **59**, 613–621.

40 Merdhah A. (2007) The study of scale formation in oil reservoir during water injection at high-barium and high-salinity formation water, in: *Chemical and Natural Resources Engineering*, Universiti Teknologi, Malaysia.

41 Merdhah A.B., Yassin M., Azam A. (2008) Study of scale formation due to incompatible water, *Jurnal Teknologi* **49**, 9–26.

42 Merdhah A., Yassin A. (2009) Scale formation due to water injection in Berea sandstone cores, *J. Appl. Sci.* **9**, 3298–3307.

43 Merdhah A.B., Yassin A.A.M., Muherei M.A. (2010) Laboratory and prediction of barium sulfate scaling at high-barium formation water, *J. Pet. Sci. Eng.* **70**, 79–88.

44 Rostami A., Kamari A., Joonaki E., Ghanaatian S. (2018) Accurate estimation of minimum miscibility pressure during nitrogen injection into hydrocarbon reservoirs, in *80th EAGE Conference and Exhibition 2018*, Copenhagen, Denmark.

45 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.

46 Moghadasi R., Rostami A., Hemmati-Sarapardeh A., Motie M. (2019) Application of Nanosilica for inhibition of fines migration during low salinity water injection: Experimental study, mechanistic understanding, and model development, *Fuel* **242**, 846–862.

47 Rostami A., Arabloo M., Lee M., Bahadori A. (2018) Applying SVM framework for modeling of CO_2 solubility in oil during CO_2 flooding, *Fuel* **214**, 73–87.

48 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.

49 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**, 1–15.

50 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.

51 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.

52 Rostami A., Hemmati-Sarapardeh A., Karkevandi-Talkhooncheh A., Husein M.M., Shamshirband S., Rabczuk T. (2019) Modeling heat capacity of ionic liquids using group method of data handling: A hybrid and structure-based approach, *Int. J. Heat Mass Trans.* **129**, 7–17.

53 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**, 270–282.

54 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.

55 Rostami A., Kamari A., Panacharoensawad E., Hashemi A. (2018) New empirical correlations for determination of Minimum Miscibility Pressure (MMP) during N_2-contaminated lean gas flooding, *J. Taiwan Ins. Chem. Eng.* **91**, 369–382.

56 Rostami A., Arabloo M., Esmaeilzadeh S., Mohammadi A.H. (2018) On modeling of bitumen/n-tetradecane mixture viscosity: Application in solvent-assisted recovery method, *Asia-Pacific J. Chem. Eng.* **13**, e2152.

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

58 Rostami A., Baghban A., Mohammadi A.H., Hemmati-Sarapardeh A., Habibzadeh S. (2019) Rigorous prognostication of permeability of heterogeneous carbonate oil reservoirs: Smart modeling and correlation development, *Fuel* **236**, 110–123.

59 Rostami A., Shokrollahi A., Esmaeili-Jaghdan Z., Ghazanfari M.H. (2019) Rigorous silica solubility estimation in super-heated steam: Smart modeling and comparative study, *Environ. Prog. Sustain. Energy*, doi: [10.1002/ep.13089](10.1002/ep.13089), in press.

60 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.

61 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.

62 Rostami A., Ebadi H., Mohammadi A.H., Baghban A. (2018) Viscosity estimation of Athabasca bitumen in solvent injection process using genetic programming strategy, *Energy Sources Part A Recovery Utilization Env. Eff.* **40**, 922–928.

63 Ferreira C. (2001) Gene expression programming: A new adaptive algorithm for solving problems, *Compl. Syst.* **13**, 87–129.

64 Koza J.R. (1992) *Genetic programming: On the programming of computers by means of natural selection*, MIT Press, Cambridge, Massachusetts, USA.

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

66 Ferreira C. (2006) *Gene expression programming: Mathematical modeling by an artificial intelligence*, 2nd edn., Springer, Berlin, Heidelberg.

67 Shokrollahi A., Safari H., Esmaeili-Jaghdan Z., Ghazanfari M.H., Mohammadi A.H. (2015) Rigorous modeling of permeability impairment due to inorganic scale deposition in porous media, *J. Pet. Sci. Eng.* **130**, 26–36.

68 Moghadasi J., Müller-Steinhagen H., Jamialahmadi M., Sharif A. (2004) Model study on the kinetics of oil field formation damage due to salt precipitation from injection, *J. Pet. Sci. Eng.* **43**, 201–217.

69 Yassin M.R., Arabloo M., Shokrollahi A., Mohammadi A.H. (2014) Prediction of surfactant retention in porous media: A robust modeling approach, *J. Dispers. Sci. Technol.* **35**, 1407–1418.

70 BinMerdhah A.B., Yassin A.A.M., Muherei M.A. (2010) Laboratory and prediction of barium sulfate scaling at high-barium formation water, *J. Pet. Sci. Eng.* **70**, 79–88.

71 Kamari A., Arabloo M., Shokrollahi A., Gharagheizi F., Mohammadi A.H. (2015) Rapid method to estimate the minimum miscibility pressure (MMP) in live reservoir oil systems during CO_2 flooding, *Fuel* **153**, 310–319.

72 Ferreira C. (2002) Gene expression programming in problem solving. In *Soft computing and industry*, Springer, London, pp. 635–653.

73 Goodall C.R. (1993) Computation using the QR decomposition, in: *Handbook of Statistics*, Elsevier, Amsterdam, North Holland, 467–508.

74 Eslamimanesh A., Gharagheizi F., Mohammadi A.H., Richon D. (2013) Assessment test of sulfur content of gases, *Fuel Process. Technol.* **110**, 133–140.

75 Gramatica P. (2007) Principles of QSAR models validation: Internal and external, *QSAR Comb. Sci.* **26**, 694–701.

76 Fayazi A., Arabloo M., Shokrollahi A., Zargari M.H., Ghazanfari M.H. (2014) State-of-the-art least square support vector machine application for accurate determination of natural gas viscosity, *Ind. Eng. Chem. Res.* **53**, 945–958.