Smart learning strategy for predicting viscoelastic surfactant (VES) viscosity in oil well matrix acidizing process using a rigorous mathematical approach

Mehdi Mahdaviara¹ · Alireza Rostami² · Khalil Shahbazi²

Received: 17 October 2020 / Accepted: 8 September 2021

Published online: 27 September 2021

© The Author(s) 2021

Abstract

This piece of study attempts to accurately anticipate the apparent viscosity of the viscoelastic surfactant (VES) based self-diverting acids as a function of VES concentration, temperature, shear rate, and pH value. The focus not only is on generating computer-aided models but also on developing a straightforward and reliable explicit mathematical expression. Towards this end, Gene Expression Programming (GEP) is used to connect the aforementioned features to and the target. The GEP network is trained using a wide dataset adopted from open literature and leads to an empirical correlation for fulfilling the aim of this study. The performance of the proposed model is shown to be fair enough. The accuracy analysis indicates satisfactory Root Mean Square Error and R-squared values of 7.07 and 0.95, respectively. Additionally, the proposed GEP model is compared with literature published correlations and established itself as the superior approach for predicting the viscosity of VES-based acids. Accordingly, the GEP model can be potentially served as an efficient alternative to experimental measurements. Its obvious advantages are saving time, lowering the expenses, avoiding sophisticated experimental procedures, and accelerating the diverter design in stimulation operations.

Article Highlights

- (1) The Gene Expression Programming evolutionary algorithm is proposed for modeling the viscosity of Viscoelastic Surfactant-based self-diverting acids.
- (2) The viscoelastic surfactant viscosity correlation presents high accuracy which is demonstrated through multiple analyses.
- (3) The Gene Expression Programming algorithm is a reliable tool expediting the diverter design phase of each stimulation operation.

Keywords Viscoelastic surfactant · Artificial intelligence · Gene expression programming · Matrix acidizing · Carbonate rock

✉ Alireza Rostami, alireza.rostami.put2014@gmail.com; alireza.rostami@afp.put.ac.ir | ¹Department of Petroleum Engineering, Amirkabir University of Technology (AUT), Tehran, Iran. ²Department of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran.

1 Introduction

In the last decades, matrix acidizing has been established as a prosper stimulation approach for alleviating the formation damage from various drilling/production operations. Although this technique constitutes an extensive part of the carbonate stimulation projects all over the world, it suffers from irregular and uneven distribution of the acids into the rock due to the heterogeneous nature of the carbonate reservoirs. The treatment fluid intrinsically tends to enter the high permeability zones where there is lower resistance to flow. As a consequence, the major portion of the acid is spent in a particular part of the formation, leaving the main targets untreated [1].

Several efforts have been made to deal with the foregoing challenge by developing different mechanical and chemical diverter systems [2, 3]. The diverting procedure can be mechanically conducted by setting the packers and/or bridge plugs to entirely isolate the zone of interest from the rest of the formation [4, 5] or using the ball sealers for effectively sealing the high permeability zones and perforations based on the differential pressure across the pathways. The prosperity of these approaches has been widely confirmed through various field applications [3]. However, they are mainly applicable in cased holes. Furthermore, they suffer from different shortcomings including time, expenses, and complexity of the operation. Accordingly, several chemical materials have been introduced to equalize the acids flow over producing zones.

Smith et al. [5] have introduced foams as the appropriate diverting agents for both open hole and cased hole completions. Crowe [6] has assessed the diverting capability of diverse chemical agents including sulfamic acids, solid organic acids, oil-soluble acids, swellable synthetic polymers or gums, deformable solids, viscous liquids (hydroxyethylcellulose and guar gum gels), and acid in oil emulsions. The application of the finely ground friable and pliable oil-soluble resins was deemed more advantageous in comparison to other alternatives. Following the work of Crowe [6], Cooper and Bolland [7] introduced a special benzoic acid improved by incorporating the benzoate salt as a proper diverter in the course of acidizing in the water injection wells. Crowe et al. [8] apprised different polymer/thickening materials including xanthan polymers (XP), guar gum (GC), carboxymethyl hydroxyethyl cellulose (CMHEC), and hydroxyethyl cellulose (HEC) as gelling agents for enhancing the viscosity of the hydrochloric acids (HCl). They found the XP as the superior agent from

the efficiency, stability, and the spent acid condition points of view.

Introduced by Chang and Frenier [9], the viscoelastic surfactant (VES) self-diverting acids are a recent generation of chemical diverters that have grabbed the attention of the petroleum community in the last decade [10, 11]. The VES diverter deflects the treatment fluids into interested zones by automatically heightening the acid viscosity while reacting with the carbonate rock. In the surface condition, the live VES-based fluid benefits from low apparent viscosities due to its micellar-like structure which facilitates the pumping operation. The acid viscosity increases by increasing the temperature from the surface to the deeper areas of the well. The foregoing increasing trend, however, ceases in temperatures higher than a critical value depending on the type and concentration of the acid, surfactant, and other additives [12]. Additionally, the shear rate is discussed in the literature as one of the significant parameters that inversely affects the VES fluid viscosity [9, 12]. The major viscosity increment occurs by spending the acid in the zone of interest, where the pH value enhances (in a range from 0 to 2) and the reaction product (CaCl_2) increment causes the acid molecules to alter to the rode-like structure [13, 14]. Subsequently, the spent acid viscosity slightly decreases at the higher pH values facilitating the fluids flow back to the surface.

The wealth of the VES-based stimulation fluids has been asserted through several experimental and field investigations [13, 15]. Notwithstanding, success is not ensured without recognizing the rheological behavior of the treatment fluid. The laboratory measurements could be expensive and time-consuming. Moreover, they require complex procedures. This provides an impetus to make use of the alternative empirical correlations. However, it should be mentioned that correlating the viscosity of foregoing fluids is not a simple task due to its robust dependency on numerous of factors including temperature, shear rate, pH, and Ca^{2+} concentration. The aforesaid factors, in turn, are remarkably influenced by the type/concentration of acid and other additives such as corrosion inhibitors, iron control, H_2S scavenger, de-emulsifier, and mutual solvent agents [12]. This captures the essence of utilizing vigorous computational approaches to establish models predicting this parameter with minimum input variables, yet possessing maximum accuracy.

The soft computing methodologies are the trailblazing techniques being extensively utilized for alleviating

the sophisticated challenges pertinent to petroleum and chemical engineering. Several authors have attempted to model the static properties of the hydrocarbon-bearing rocks such as absolute permeability, porosity, and irreducible water saturation by employing different artificial intelligence (AI) approaches [16, 17]. Many researchers have employed the foregoing methodologies to anticipate the dynamic properties such as relative permeability of conventional and unconventional hydrocarbon deposits [18–20]. Furthermore, there is a large body of literature working with prognostication of rheological properties of different fluids such as gas/liquid hydrocarbons as well as nanofluids [21–23].

The aim of this piece of study is to develop an AI-based model describing the viscosity of VES-based self-diverting acids as a function of VES concentration, pH value, shear rate, and temperature. It has been particularly focused on the model's output to be user-accessible while benefiting from high accuracy and precision. To do this, an evolutionary algorithm named Gene Expression Programming (GEP) technique was employed to appropriately link the aforesaid input variables to the output. The correctness of the suggested model was evaluated through several measures containing graphical descriptions and statistical parameters. Furthermore, a secondary analysis was carried out to provide a comparison between the recently developed GEP-based model and the published correlations from the accuracy point of view. As significant progress in estimating the viscosity of VES-self diverting acids, the contribution of this work is to expedite the diverter designing procedure which is the prerequisite for any stimulation operation. Because the developed model herein can be tuned with a smaller number of experimentations, it can easily be used in the petroleum industry. This is the outline of the subsequent section of this paper. In the second section, the GEP algorithm employed in this study is described in detail. The third section addresses the model development and the validation analysis. The conclusions are presented in the fourth section.

2 Model architecture

Inspired by the natural biological evolution theorem, the Evolutionary Algorithms (EAs) are novel optimization methodologies that have emerged as promising and versatile candidates for working with complex problems in the past decades. Commenced with the Genetic Algorithm (GA), several EA algorithms such as Genetic Programming (GP), Modified Shuffled Frog Leaping Algorithm (SFLA),

the Particle Swarm Optimization (PSO), the Memetic Algorithm (MA), the Ant Colony Optimization (ACO), etc. were introduced for modeling and optimizing purposes. The idea behind all of them is unique. The evolutionary mechanisms such as natural selection, mutation, and cross-over are applied to a given population of individuals, and the fittest individuals are survived and constitute the next generation of the population.

Introduced by Ferreira [24], Gene Expression Programming (GEP) is a generation of evolutionary algorithms that has come into widespread approval due to its substantial success in dealing with a plethora of sophisticated problems in various research communications [25, 26]. This algorithm is a successor of the GA and GP optimizers with the difference of the individual's type. Chromosomes (or so-called genotypes), symbolic strings of fixed length, comprise the individuals of the GA method. In the case of the GP algorithm, the individuals are expressed in the form of phrase trees representing the code data in a tree-like structure, in which the leaves and nodes involve the operands and operators, respectively. The GEP algorithm embodies both chromosome and expression tree individuals. In other words, the GEP individuals are expressed as expression trees following encoding as chromosomes [24].

The GEP commences with the establishment of a random population of chromosomes; namely linear strings of fixed length which embody multiple genes. The genes, a combination of variables, constants, and mathematical operators, are linked by employing a pre-assigned linking operator such as addition, subtraction, division, and multiplication [27, 28]. Then, all chromosomes are expressed as expression trees and the tree-shaped structures. The performance of the chromosomes is assessed according to a special pre-defined fitness function in order to recognize the superior expressions [27]. The GEP algorithm ceases if the program meets the termination threshold. Otherwise, an improved generation of the chromosome population is produced by manipulating various nature-inspired operators including transportation, mutation, and recombination. The algorithm iterates until fulfilling the aforementioned termination criteria. An example of the GEP structure is presented in the schematic of Fig. 1. Furthermore, a flowchart describing the utilized algorithm in this study is depicted in Fig. 2. For an overview of the GEP technique, the interested reader is referred to useful works available in relevant literature [28–30].

In this study, achieving the maximum fitness was selected to be the termination criteria of the GEP algorithm. Root mean square of error (RMSE) was considered as

Fig. 1 A schematic representation of the GEP structure [30]

the fitness function to determine the consistency between predictions and the target values. The number of chromosomes, head sizes, and the number of genes were set to be 30, 8, and 3, respectively. The addition operator was considered as the linking function of the GEP model. It should be mentioned that the large values for the number of chromosomes, head sizes, and genes will increase the complexity of the model which, in turn, leads to an overfitting (high variance) problem. In contrast, low values of the aforementioned parameters give rise to a weak model suffering from underfitting (high bias) issues. Accordingly, these hyper-parameters were tuned after several attempts

and iterations. Following the optimization procedure proposed in the work of Creton et al. [31], the optimal number of chromosomes, head sizes, and genes were determined as 30, 8, and 3, respectively.

3 Results and discussion

3.1 Training the model

In this study, the GEP methodology was utilized to accurately predict the viscosity of the VES-based self-diverting

Fig. 2 The flowchart of the GEP algorithm

acids as a function of more straightforward-to-measure variables. For this purpose, an experimental databank representing the viscosity changes in response to concentration of VES (C_{VES} , %), pH, temperature (T , °K), and shear rate (SR , s^{-1}) input variables were gathered from the literature [32]. The employed VES self-diverting acid embodies the hydrochloric acid (5%), VES (5–7.5%), corrosion inhibitor (0.5%), and methanol (1%) components. Table 1 represents a detailed description of the collected

data [32]. Furthermore, the contour plots of Fig. 3 demonstrate the changes of the VES viscosity with different input parameters.

The algorithm was implemented after separating the data based on the hold out approach, based on which 80% and 20% of the points were randomly specified to the training and test sub-data, respectively. Table 2 summarizes the setting parameters of the predescribed GEP model. Additionally, Fig. 4 depicts the expression trees associated with the established model. The model development led to the following simplified and rearranged user-accessible correlation:

$$\mu_{VES} = A + B + C \quad (1)$$

$$A = a \frac{[\ln(PH * C_{VES})]^5}{T^4} \quad (2)$$

$$B = PH + \frac{(b + C_{VES})^3}{2SR + c^{PH}} \quad (3)$$

$$C = d \frac{PH * C_{VES}}{PH + SR} \quad (4)$$

where a , b , c , and d are constant parameters which are equal to 271,736.7, 7.21, 11.62, and 91.63, respectively.

3.1.1 Accuracy assessment

In an aim to prove the capability of the suggested GEP-derived model for predicting the μ_{VES} , several significant accuracy indices including Standard Deviation (SD), Average Relative Deviation (ARD), Absolute Average Relative Deviation (AARD), Mean Square of Error (MSE), Root Mean Square of Error (RMSE), and Determination Coefficient (R^2) were utilized. The mathematical expressions of the statistical error functions are represented in Table 3. Table 4 lists the values of the error parameters associated with

Table 1 Details of the experimental data sets gathered from the literature [32]

Parameter	Min.	Avg.	Max.	SD
VES-based diverting acid composition				
HCl (5%) + VES (5–7.5%) + Corrosion Inhibitor (0.5%) + Methanol (1%)				
VES concentration (C_{VES}), %	5	6.25	7.5	1.2520
T , °K	298.15	315.65	333.15	272.90
pH	1	4	7	1.8738
Shear rate (SR), 1/s	25	104.1666	200	62.2432
VES-based acid viscosity (μ_{VES}), mPa.s	9.6	41.5157	135.9	30.9261

Min. refers to minimum value; Avg. refers to average value; Max. refers to a maximum value; SD refers to standard deviation value

Fig. 3 The changes of the VES viscosity with different input parameters of VES concentration (C_{VES}), pH, shear rate (SR), and temperature

Table 2 The setting parameters of the proposed GEP based model

Setting	Value
Independent variables	4
Training records	251
Validation records	61
Program size	42
Literals	14
Used variables	VES (weight: 3), T (weight: 1), pH (weight: 7), SR (weight: 3)
Chromosomes	30
Genes	3
Head Sizes	8
Tail Sizes	9
Linking function	Addition
Mutation	0.002
IS transposition	0.005
RIS transposition	0.005
Gene recombination	0.002
Function set	$\pm, \times, /, \sqrt{}, \text{EXP}, X^Y, 10^X, \text{Ln}, X^2, X^3, X^4, X^5, X^{1/3}, X^{1/4}, X^{1/5}$

the developed GEP model for the training, test, and total subsets. As it can be seen, the proposed correlation offers satisfactory predictions with the AARD values of 12.0721, 14.8839, and 12.6218 percent, the RMSE values of 6.3284, 9.5323, and 7.0699, and the R-squares of 0.9552, 0.9254, and 0.9477 for the train, test, and all the data points, respectively.

In addition to the previously mentioned quantitative analysis, a set of graphical representations were employed to qualitatively evaluate the correctness of the extended model. Figure 5 presents the cross plot and the relative deviation plot of the developed model to probe the consistency between the calculated and measured values. As it is evident in the cross plot (Fig. 5a), the majority of training and test points are allocated nearby the unit slop line and only a few deviations can be observed. This is an affirmation regarding the truthfulness of the extended GEP model. Furthermore, Fig. 5b depicts the distribution of the relative deviation for the GEP model over various data points including training and test subsets. It can be seen that the cloud of data points is predominantly situated nearby the zero line. This is further evidence showing the correctness of the suggested model.

Fig. 4 A schematic description of the expression trees and k-expressions of the developed GEP model

The distribution of absolute relative deviation percent (ARD%) versus pH and shear rate is portrayed in the 2D contour plots of Fig. 6. As it is shown, the main portion of this figure is covered by ARD less than 15%. This map shows the reliability of the GEP based model over a wide range of operational parameters.

Another error assessment was conducted by splitting the data points into different ranges of input variables to indicate that the success of the GEP model is not limited to a particular range of inputs and outputs. The results are graphically illustrated in the bar plots of Fig. 7. As shown, the highest discrepancies between the predictions and

Table 3 The mathematical expressions of the statistical error functions used in this study

Statistical measure	Equation
Standard deviation	$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(\frac{\mu_i^{exp} - \mu_i^{pred}}{\mu_i^{exp}} \right)^2}$
Mean square of error	$MSE = \frac{1}{N} \sum_{i=1}^N \left(\mu_i^{exp} - \mu_i^{pred} \right)^2$
Root mean square of error	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\mu_i^{exp} - \mu_i^{pred} \right)^2}$
Average relative deviation	$ARD\% = \frac{100}{N} \sum_{i=1}^N \left(\frac{\mu_i^{exp} - \mu_i^{pred}}{\mu_i^{exp}} \right)$
Absolute average relative deviation	$AARD\% = \frac{100}{N} \sum_{i=1}^N \left \frac{\mu_i^{exp} - \mu_i^{pred}}{\mu_i^{exp}} \right $
Determination coefficient	$R^2 = 1 - \frac{\sum_{i=1}^N \left(\mu_i^{pred} - \mu_i^{exp} \right)^2}{\sum_{i=1}^N \left(\mu_i^{pred} - \overline{\mu_i^{exp}} \right)^2}$

Table 4 The values of the statistical parameters for the training, test, and all data sets

Statistical parameter	Train data	Test data	All data
SD	0.1565	0.1778	0.1608
ARD (%)	-3.9630	-5.2377	-4.2122
ARD ^{min} (%)	0.0734	0.2844	0.0734
AARD (%)	12.0721	14.8839	12.6218
ARD ^{max} (%)	52.4571	38.1910	52.4571
MSE	40.0491	90.8649	49.9842
RMSE	6.3284	9.5323	7.0699
R ²	0.9552	0.9254	0.9477

ARD^{min} refers to minimum absolute relative deviation; ARD^{max} refers to maximum absolute relative deviation

experimental points occur in the pH and SR (s⁻¹) ranges of 2.5–3.5 and 75–120, respectively. This signifies that in modeling the viscosity of VES in a well stimulation process, the predictions pertinent to these ranges are most likely prone to errors. Nevertheless, it should be mentioned that the errors are not remarkable with the AARDs lower than about 18% and 14% in all ranges of pH and SR, respectively. In addition, the model is associated with maximum errors for estimating the low μ_{VES} values ranging from 9.6 to 40.25 mPa.s with the AARD of about 13.20%. The superior performance of the model is observed at the pH, SR,

and μ_{VES} ranges of 4–4.5, 25–45 s⁻¹, and 71.5–98.25 mPa.s, respectively. Moreover, their corresponding AARD value is approximately 9.29, 10.23, and 10.11 percent, respectively. It is worth mentioning that there are only marginal differences between the AARDs associated with various ranges and the model represents overall high accuracy for viscosity prediction.

The relationship between the independent variables and the predicted VES viscosity was assessed by employing the contour plots of Fig. 8. As can be seen, the trend of viscosity changes with input parameters is in good agreement with previously mentioned input-target relationships. In addition, the contour plots of Fig. 8 are in line with the plots of Fig. 3, and there is little to distinguish between the predicted viscosities (Fig. 8) and the experimental values (Fig. 3).

The outliers are the unlikely points highly deviated from the rest of the data, which can be raised from errors in the laboratory measurements. The foregoing abnormalities can lead to remarkable problems in the training procedure of an algorithm. Accordingly, the real performance of a model may not be achieved without an accurate outlier removal step. Further information in this regard is represented in the literature [33, 34]. Herein, a couple of the Leverage methodology and the Williams' plot is utilized to detect the outliers. The results of the analysis are depicted in Fig. 9. As shown, the plot is separated into various areas on the basis of the Leverage values (Hat indices) and the standardized

Fig. 5 **a** The cross plot and **b** the relative deviation plot of the established GEP-based model

residuals (R). The squared area of $-3 \leq R \leq +3$ and $H < \hat{H}$ and the range of $-3 \leq R \leq 3$ and $H > \hat{H}$ are so-called “applicability domain of the model” and “Bad High Leverage points”, respectively. Although the latter is not the desirable area, its embedded points are not considered as abnormalities. The outliers are the points situated out of the $-3 \leq R \leq 3$ range in all values of H [35]. According to Fig. 9, vast portion of the data are positioned in the applicability domain of the extended model, and only 7 suspected data (4 outliers and 3 Bad High Leverage points) can be detected. This is an indication of the high

applicability of the GEP model for predicting the viscosity of the acids.

A comparison of the extended GEP model with the literature published correlations was considered as the last step of the accuracy assessment. In this analysis, the GEP and alternative models were utilized for calculating the viscosity of VES-based self-diverting acids as a function of different input parameters such as C_{VES} (%), pH, T ($^{\circ}\text{F}$), and SR (S^{-1}). Subsequently, the results were compared with the experimental targets, and the corresponding errors were indicated through different statistical indexes. Eventually,

Fig. 6 The contour plot of the absolute relative deviation changes as a function of pH and shear rate

the graphical plots were used for visualizing the performance of each model and introducing the best one from the standpoints of accuracy and precision.

The calculations were carried out based on the datasets employed in this study [32]. To the best of the authors' knowledge, only a few literature studies have been undertaken to correlate the viscosity of the VES-based self-diverting acids as a function of the previously mentioned input variables [36]. Therefore, the comparison analysis was implemented by utilizing a single correlation proposed by Ratnakar et al. [36], in which the viscosity of in-situ cross-linked acids (ICA) is connected to the pH, temperature, and shear rate parameters as follows:

$$\mu_{ICA} = \mu_o SR^{\alpha-1} \left(1 - \alpha \frac{T - T_0}{T_0} \right) \left[1 + (\mu_m - 1) \exp \left\{ \frac{-\beta (PH - PH_m)^2}{PH(7 - PH)} \right\} \right] \quad (5)$$

where μ_o stands for the viscosity corresponding to pH values of 0 or 7, μ_m denotes the maximum viscosity of the treatment fluid after gel formation and PH_m is the value of pH at μ_m . This correlation encompasses three constants of n , α , and β . A curve fitting analysis was conducted using the Levenberg–Marquardt and Gauss–Newton algorithms to appropriately tune the foregoing constant parameters

to the experimental data point utilized in this study. The optimum constant parameters were found to be 2, 31, and 1 for n , α , and β entities, respectively.

The comparison was carried out through the aforementioned statistical parameters and plots. The result of the analysis indicated the RMSE, AARD, SD, and R^2 values of about 47.58, 74.42%, 0.83, and 0.7% for the correlation of Eq. (5). Hence, the errors associated with the Ratnakar et al. [36] are extreme to be of any use in the prediction of viscosity of VES-based acids. Additionally, three graphical plots of contour, relative deviation, and cumulative frequency were exploited to visualize the results of the comparison. Figure 10 shows the comparison of the extended

GEP model here with the correlation proposed by Ratnakar et al. [36] by using the cross plot and relative deviation diagrams. According to Fig. 10a, as opposed to the GEP model, the predictions pertinent to Eq. (5) highly deviate from the unit slop line. Besides, the distribution of the relative deviation exhibited in Fig. 10b demonstrates the wide errors associated with Eq. (5), in which the main portion of

Fig. 7 The average absolute relative deviations (%) associated with the proposed GEP model at different ranges of **a** pH, **b** shear rate, and **c** the VES-based acid viscosity

points falling within the ARD range of 50 to 100%. On the whole, both plots of Fig. 10 indicate that only a few calculations of Eq. (5) are close-to-reality predictions, the rest of which are associated with remarkable under-estimation as well as over-estimation. Figure 11 compares the cumulative frequency of absolute relative error percent for the GEP model and Eq. (5). The greater proximity of the curve to the top left side of the figure corresponds to the better performance of the model. As it is evident in Fig. 11, the GEP model curve shows a steeper ascending trend indicating the superiority of this model over the correlation of Eq. (5). For shedding light on this fact, it is suggested the Absolute Relative Error (%) of 27% to be considered. As it can be seen, more than 90% of estimations (Cumulative Frequency equal to 90%) by the proposed GEP model correspond to errors lower than this value. In the case of Eq. (5), however, only about 15% of predictions (Cumulative Frequency equal to 15%) have errors equal to or less than 27%, the rest of which (about 85%) suffer from higher errors. On the whole, the model developed here is seen easily outperforms the literature correlation of Eq. (5).

4 Conclusions

The current study was aimed at developing a reliable model for anticipating the viscosity of VES-based self-diverting acids. In this regard, the GEP was implemented by utilizing widespread experimental data points reported in the literature. The GEP approach led to a straightforward mathematical expression linking the VES viscosity to various input variables including VES concentration, temperature, pH, and shear rate. This AI-based model benefits from simplicity and the low number of input variables and tuning parameters. The prosperousness of the extended model was revealed using several statistical evaluations. The AARD, RMSE, and R^2 parameters were found to be 12.62%, 7.07, and 0.95, respectively, which are desirable values considering the complex nature of the VES-based acids. Different graphical representations were provided

Fig. 8 The changes of the predicted VES viscosity with different input parameters of VES concentration (C_{VES}), pH, shear rate (SR), and temperature

Fig. 9 The William's plot of the developed GEP model

Fig. 10 A comparison between the suggested GEP model and a traditional published correlation [36] using **a** cross plot and **b** the relative deviation plot

to visualize the foregoing results. Additionally, the correctness of the model over various ranges of input variables was demonstrated through statistical graphs. Furthermore, outlier detection revealed the wide applicability domain of the established GEP algorithm. Finally, in a comparison between the proposed GEP model and a traditional literature published model, the GEP established itself as the most precise approach for estimating the viscosity of the VES-based self-diverting acids. Even though the

established tool here is accurate to estimate VES viscosity, its application beyond the applicability range may lead to some deviations from real viscosity. However, for in-range data, this correlation could give a suitable estimate of the viscosity. As a further extension of this study, it would be useful to assess the capability of the GEP algorithm for working with a wider range of data sets containing more extensive acid systems.

Fig. 11 A comparison between the cumulative frequency of errors associated with the GEP model and the Ratnakar et al. [36] correlation

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Liu N, Liu M (2016) Simulation and analysis of wormhole propagation by VES acid in carbonate acidizing. *J Petrol Sci Eng* 138:57–65
- Liu P, Xue H, Zhao LQ, Fu Y, Luo Z, Qu Z (2015) Analysis and simulation of rheological behavior and diverting mechanism of In Situ Self-Diverting acid. *J Petrol Sci Eng* 132:39–52
- Nozaki M, Zhu D, Hill AD (2013) Experimental and field data analyses of ball-sealer diversion. *SPE Production Operations* 28:286–295
- Dill W (1969) Effect of bridging agents and carrier fluids on diverting efficiency. *J Petrol Technol* 21:1347–341352.
- Smith CL, Anderson JL, Roberts P (1969) New diverting techniques for acidizing and fracturing. In: *SPE California regional meeting, society of petroleum engineers*
- Crowe C (1971) Evaluation of oil soluble resin mixtures as diverting agents for matrix acidizing. In: *Fall meeting of the society of petroleum engineers of AIME. Society of Petroleum Engineers*
- Cooper R, Bolland J (1984) Effective diversion during matrix acidization of water injection wells. In: *Offshore technology conference, offshore technology conference*
- Crowe C, Martin R, Michaelis A (1981) Evaluation of acid-gelling agents for use in well stimulation. *Soc Petrol Eng J* 21:415–424
- Chang F, Qu Q, Frenier W (2001) A novel self-diverting-acid developed for matrix stimulation of carbonate reservoirs. In: *SPE international symposium on oilfield chemistry. Society of Petroleum Engineers*
- Nasr-El-Din HA, Samuel MM (2007) Lessons learned from using viscoelastic surfactants in well stimulation. *SPE Prod Oper* 22:112–120
- Zeiler C, Alleman D, Qi Q (2004) Use of viscoelastic surfactant-based diverting agents for acid stimulation: case histories in GOM. In: *SPE annual technical conference and exhibition. Society of Petroleum Engineers*
- Wang G, Nasr-El-Din HA, Zhou J, Holt S (2012) A new viscoelastic surfactant for high temperature carbonate acidizing. In: *SPE Saudi Arabia section technical symposium and exhibition. Society of Petroleum Engineers*
- Nasr-El-Din HA, Hill AD, Chang FF, Sultan AS (2007) Chemical diversion techniques used for carbonate matrix acidizing: an overview and case histories. In: *International symposium on oilfield chemistry. Society of Petroleum Engineers*
- Liu M, Zhang S, Mou J, Zhou F, Shi Y (2013) Diverting mechanism of viscoelastic surfactant-based self-diverting acid and its simulation. *J Petrol Sci Eng* 105:91–99
- Alleman D, Qi Q, Keck R (2003) The development and successful field use of viscoelastic surfactant-based diverting agents for acid stimulation. In: *International symposium on oilfield chemistry. Society of Petroleum Engineers*
- Ahmadi MA, Chen Z (2018) Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. *Petroleum*
- Cvetković M, Velić J, Malvić T (2009) Application of neural networks in petroleum reservoir lithology and saturation prediction. *Geol Croat* 62:115–121
- Mahdaviara M, Menad NA, Ghazanfari MH, Hemmati-Sarapardeh A (2020) Modeling relative permeability of gas condensate reservoirs: Advanced computational frameworks. *J. Petrol Sci Eng*, pp 106929
- Guler B, Ertekin T, Grader A (2003) An artificial neural network based relative permeability predictor. *J Can Petrol Technol* 4
- Arigbe OD, Oyeneyin M, Arana I, Ghazi M (2019) Real-time relative permeability prediction using deep learning. *J Petrol Explor Prod Technol* 9:1271–1284
- Rostami A, Hemmati-Sarapardeh A, Shamshirband SJF (2018) Rigorous prognostication of natural gas viscosity. *Smart Model Comparat Study* 222:766–778
- Dargahi-Zarandi A, Hemmati-Sarapardeh A, Hajirezaie S, Dabir B, Atashrouz S (2017) Modeling gas/vapor viscosity of hydrocarbon fluids using a hybrid GMDH-type neural network system. *J Mol Liq* 236:162–171
- Rostami A, Kalantari-Meybodi M, Karimi M, Tatar A, Mohammadi AH (2018) Efficient estimation of hydrolyzed polyacrylamide (HPAM) solution viscosity for enhanced oil recovery process by

- polymer flooding. *Oil & Gas Sciences and Technology–Revue d'IFP Energies nouvelles* 73:22
24. Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. *arXiv preprint cs/0102027*
 25. Zorn CR, Shamseldin AY (2015) Peak flood estimation using gene expression programming. *J Hydrol* 531:1122–1128
 26. Rostami A, Arabloo M, Kamari A, Mohammadi AH (2017) Modeling of CO₂ solubility in crude oil during carbon dioxide enhanced oil recovery using gene expression programming. *Fuel* 210:768–782
 27. Hong T, Jeong K, Koo C (2018) An optimized gene expression programming model for forecasting the national CO₂ emissions in 2030 using the metaheuristic algorithms. *Appl Energy* 228:808–820
 28. Zhong J, Feng L, Ong Y (2017) gene expression programming: a survey [review article]. *IEEE Comput Intell Mag* 12:54–72
 29. Ferreira C (2006) *Gene expression programming: mathematical modeling by an artificial intelligence*. Springer, Berlin
 30. Mahdaviara M, Rostami A, Shahbazi K (2020) State-of-the-art modeling permeability of the heterogeneous carbonate oil reservoirs using robust computational approaches. *Fuel* 268:117389
 31. Creton B, Lévêque I, Oukhemanou F (2019) Equivalent alkane carbon number of crude oils: a predictive model based on machine learning. *Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles* 74:30
 32. Mehri Ghahfarrokhi A, Kazemzadeh E, Shokrollahzadeh Behbahani H, Safian GA (2019) An experimental study of acid diversion by using gelled acid systems based on viscoelastic surfactants: a case study on one of Iran Southwest Oilfields. *Iran J Oil Gas Sci Technol* 8:32–46
 33. Rousseeuw PJ, Leroy AM (2005) *Robust regression and outlier detection*. John Wiley & sons, Hoboken
 34. Gramatica P (2007) Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 26:694–701
 35. Hemmati-Sarapardeh A, Varamesh A, Husein MM, Karan KJR, Reviews SE (2018) On the evaluation of the viscosity of nano-fluid systems: Modeling and data assessment 81:313–329
 36. Ratnakar RR, Kalia N, Balakotaiah V (2013) Modeling, analysis and simulation of wormhole formation in carbonate rocks with in situ cross-linked acids. *J Chem Eng Sci* 90:179–199
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.