Petroleum 6 (2020) 423–437

Contents lists available at ScienceDirect

Petroleum

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

Application of adaptive neuro-fuzzy inference system and data mining approach to predict lost circulation using DOE technique (case study: Maroon oilfield)

Farough Agin[a,b,*], Rasool Khosravanian[a], Mohsen Karimifard[c], Amirhosein Jahanshahi[d]

[a] Department of Petroleum Engineering, Amirkabir University of Technology (Tehran Polytechnic), 424 Hafez Avenue, Tehran, 15875-4413, Iran
[b] Research Institute of Petroleum Industry (RIPI), West Blvd., Near Azadi Sports Complex, Tehran, 14857-33111, Iran
[c] Department of Petroleum Engineering, Petroleum University of Technology, Ahwaz, Iran
[d] School of Mining Engineering, College of Engineering, University of Tehran, 14174-66191, Tehran, Iran

ARTICLE INFO

Keywords:
Lost circulation
Maroon oilfield
ANFIS
RMSE
Expert system

ABSTRACT

Lost circulation is the most common problem encountered while drilling oil wells. Occurrence of such a problem can cause a lot of time and cost wastes. In order to drill oil wells, a fast and profitable way is necessary to predict and solve lost circulation problem. Expert system is a method used lately for problems that deal with uncertainty. In this paper, three approaches are carried out for prediction of lost circulation problem. These approaches include design of experiments (DOE), data mining, and adaptive neuro-fuzzy inference system (ANFIS). Data of 61 wells of Maroon oilfield are selected and sorted as the feed of the systems. Seventeen variables are used as inputs of the approaches and one variable is used as the output. First, DOE is conducted to observe the effects of variables. Plackett-Burman method is used to determine the effects of variables on lost circulation. After that, data mining is conducted to predict the amount of lost circulation. The class of regression is used to determine a function to model the data and the error of the model. Then, ANFIS is applied to predict the amount of lost circulation. The chosen data are used in order to train, test, and control the ANFIS. Furthermore, subtractive clustering is used to train the fuzzy inference system (FIS) of the model. The performance of the ANFIS model is assessed through the root mean squared error (RMSE). The results suggest that ANFIS method can be successfully applied to establish lost circulation prediction model. In addition, results of ANFIS and data mining are investigated through their prediction performances. The comparison of both methods reveals that ANFIS error is much lower than data mining.

1. Introduction

Lost circulation problem is usually the most important problem in drilling industry and causes the most drilling non-productive time. Therefore, the industry spend about US$800 million per year to cure problem [1,2]. Losses also consist over 90% of the lost returns expenditures of the industry [3]. However, the first step in the lost circulation subject is to prevent lost circulation. This can be conducted by identifying the problematic zones during drilling [4,5]. The planning of lost circulation includes both prevention and remediation methods. Stopping losses while occurring is as important as preventing it to occur because it can save time and money [6]. Lost circulation problems occur at any depth. There are four types of formations which lost circulation occur in: a) Natural or induced fractured formations; b) Vugular or cavernous formation; c) Highly permeable formations; and d) Unconsolidated formations. In formations such as naturally fractured, cavernous, vugular and unconsolidated formations lost circulation cannot be avoided [7]. There are a plethora of factors that affect the lost circulation. These factors are formation pressure gradient, formation fracture pressure gradient, petrophysical properties of the formation, fluid properties, existence of fractures and caves in the formation, some drilling parameters such as weight on bit (WOB), pump rate, and lots of other known and unknown parameters which might affect severity of lost circulation [8]. The main concern in the history of lost circulation was to propose best ways to cure the problem instead of predicting it [1–8]. Thereby, in oil industry, the need of innovating new

Peer review under responsibility of Southwest Petroleum University.

* Corresponding author. Department of Petroleum Engineering, Amirkabir University of Technology (Tehran Polytechnic), 424 Hafez Avenue, Tehran, 15875-4413, Iran.

E-mail addresses: Aginf@ripi.ir, Farough.agin@gmail.com (F. Agin).

https://doi.org/10.1016/j.petlm.2018.07.005
Received 15 April 2018; Received in revised form 26 June 2018; Accepted 27 July 2018
Available online 31 July 2018

approaches to forecast the problem has grown within recent years. Intelligent modeling of problems has been improved by the development of new approaches such as soft computing in all fields of science and technology [9–21]. One of the approaches of soft computing is ANFIS. For the first time, Zadeh introduced fuzzy logic [22]. In this method, instead of dealing with crisp premises, linguistic terms are considered [22]. Afterwards, Jang proposed a new model containing both fuzzy logic and neural network [10]. This approach can be applied for situations which deal with uncertainty and large data set [10]. ANFIS application has been widely used in different fields of science and technology [15–21,23–30]. In oil industry, ANFIS functions have been utilized in various fields of petroleum like enhancement of drilling parameters, reservoir characterization, exploration, etc. [14,28–33]. As it is mentioned above, one of the major problems in drilling operations is lost circulation. Shermetov et al. proposed an approach for solving lost circulation problem based on fuzzy expert system. They used fuzzy expert system to determine the best way to cure the lost circulation [33]. In addition, Moazeni et al. used artificial neural network to predict the amount of lost circulation in one of the Iranian oilfields [8].

The main focus relevant to lost circulation in past years was how to cure the loss while it happens. This is due to the lack of knowledge about the mechanisms of lost circulation and variables, which may affect the lost circulation. The major problem in this case is uncertainty about how lost circulation occurs. A new model is presented here to predict the lost circulation. Data of one of the Iranian oilfields are used to feed the approaches. Seventeen variables are used as inputs and one variable is used as the output of the systems. By collecting drilling and geomechanical data of Maroon oilfield, proper results about predicting the amount of lost circulation are gained. In this paper, three approaches are carried out for prediction of lost circulation problem. These approaches include DOE, data mining, and ANFIS. First, DOE is conducted to observe the effects of each variable. DOE is an approach applied at the stage of collecting data in order to ensure the reliability and validity of data. Furthermore, by applying this approach the effect of each variable can be determined. Plackett-Burman method is used to determine the effects of variables on the amount of lost circulation [34–36]. The results of DOE reveal that the highest effect of variables relates to shear stress at shear rate 300, and the lowest effect relates to drilling time. Then, one of the classes of data mining is used to predict the amount of lost circulation. The class of regression is used to determine a function to model the data and the error of the model [37]. In the last section, lost circulation is predicted using ANFIS. Subtractive clustering is utilized to train the FIS. Backpropagation method is used as the optimization method. In addition, the structure of ANFIS, surfaces obtained from ANFIS and RMSE of the system are illustrated. Finally, results of ANFIS and data mining are compared. The comparison of both methods shows that ANFIS error is much lower than data mining.

2. Design of experiments

DOE is a method by which sensitivity of variables and their independencies are examined. The purposes of DOE are as below [8,35,36]:

(1) Comparison: this is used when there are more than one option to design. For example, there are several materials and it is decided which material or materials are the best.
(2) Variable screening: this is used when there is a large data set and amount of variables are too high. By conducting this work, those variables which are more important and their effects are higher than some others, can be selected. Therefore, the number of variables decreases and better results can be obtained with low variables.
(3) Transfer function exploration: after identifying the important variables, effects of these variables and their responses to the system can be tested. Transfer function is the relationship between the input variables and response of the output of the system.
(4) System optimization: the purpose of designing a system is a highly improved system in some aspects such as performance, reliability, quality, and efficiency. After identifying the transfer function between variables and outputs, optimization can be conducted in order to improve the performance.
(5) System robustness: after conducting optimization, system has to be robust against noise.

3. Data mining

Data mining is the process of extracting nontrivial and potentially useful information, or knowledge, from the enormous data sets available in experimental sciences (historical records, reanalysis, general circulation method (GCM) simulations, etc.), providing explicit information that has a readable form and can be used to solve diagnosis, classification or forecasting problems. Traditionally, these problems are solved by direct hands-on data analysis using standard statistical methods, but the increasing volume of data motivates the study of automatic data analysis using more complex and sophisticated tools which can operate directly from data. Thus, data mining identifies trends within data that go beyond simple analysis. Modern data mining techniques (association rules, decision trees, Gaussian mixture models, regression algorithms, neural networks, support vector machines, Bayesian networks, etc.) are used in many domains to solve association, classification, segmentation, diagnosis, and prediction problems [31,37].

There are six task classes of data mining: 1) anomaly detection: identification of unusual data records, 2) association rule learning: this class tries to find the relationship between variables, 3) clustering: without knowing the structures in the data, this class identifies groups and structures in the data, 4) classification: this class applies generalized known structure to new data, 5) regression: considering the minimum error, this class tries to find a function to model the data, 6) summarization: represents the data set, including visualization and report generation [32]. A statistical correlation coefficient (R^2), expressed between 0 and 1, between the model and measured data demonstrates the high performance of the model in representing the field measurements. R^2 is calculated with the following relationships [34,37]:

$$R^2 = 1 - \frac{SSE}{SSy} \tag{1}$$

$$SSE = \sum_{i=1}^{n} (x_i - \hat{x})^2 \tag{2}$$

$$SSy = \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{3}$$

Where; x = measured value, $\hat{x}$ = the predicted value, and $\bar{x}$ = the

mean of the measured value.

The closer R^2 is to 1, the more confidence there is in the ability of the model to predict the output variable based on input measures provided for the field in which the input data is compiled.

4. Adaptive neuro-fuzzy inference system

ANFIS is one type of artificial neural network, which is based on the Takagi and Sugeno FIS [15]. ANFIS comprises these both methods, which means that it has the benefits of both methods in one framework [21]. FIS of the system has the ability of learning by which nonlinear functions can be approximated [21,24]. In the following, the mathematical concept of ANFIS is explained.

We consider for simplicity that the FIS has two inputs x and y, and one output f (x, y). Therefore, the rule base consists of two fuzzy if-then rules of Takagi and Sugeno's type [10].

Rule1: if x is A1 and y is B1, then $f_1 = p_1x + q_1y + r_1$ (4)

Rule2: if x is A2 and y is B2, then $f_2 = p_2x + q_2y + r_2$ (5)

Fig. 1 shows the structure of the ANFIS. The node functions in the same layer are of the same function family as presented in the following. The ANFIS model contains five layers illustrated in Fig. 1.

4.1. Layer 1

Every node i, in this layer is a square node with a node function [10].

$$o_i^1 = u_{A_i}(x) \tag{6}$$

Where, x is the input to node i and A_i, is the linguistic label. In other words, O_i^1 is the membership function of A_i, and shows the degree to which the given x satisfies the quantifier A_i. $u_{A_i}(x)$ is often chosen to be bell-shaped with maximum equal to 1 and minimum equal to 0, as below [10];

$$u_{A_i}(x) = \frac{1}{1 + |\frac{x - c_i}{a_i}|2b_i} \tag{7}$$

Where, $\{a_i, b_i, c_i\}$ is the set of parameter. If these parameters change, the bell-shaped functions change too. Therefore, various forms of membership functions on linguistic label A_i can be obtained. As a matter of fact, any continuous and piecewise differentiable functions, such as triangular-shaped membership functions, are also qualified candidates for node functions in this layer. Parameters in this layer are referred to as premise parameters.

4.2. Layer 2

Every node in this layer is a circle node, which multiplies the incoming signals and sends the product out. For example [10];

$$w_i = u_{A_i}(x) \times u_{B_i}(y);\ i = 1,2 \tag{8}$$

Each node output represents the firing strength of a rule.

4.3. Layer 3

Every node in this layer is a fixed node "Norm". The i_{th} node calculates the ratio of the ith rule's firing strength to the sum of all rules' firing strength [10].

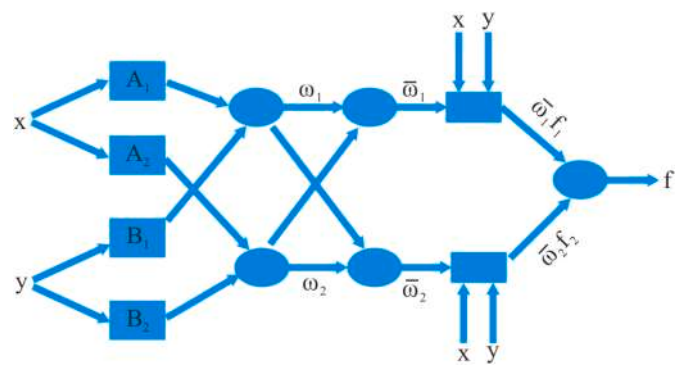

Fig. 1. Structure of ANFIS with two-input Takagi-Sugeno model [9].

$$\overline{w_i} = \frac{w_i}{w_1 + w_2};\ i = 1,2 \tag{9}$$

Outputs are called normalized firing strengths.

4.4. Layer 4

Every node "i" in this layer is a square node with a node function [10]:

$$O_i^4 = \overline{w_i}f = \overline{w_i}(p_2x + q_2y + r_2) \tag{10}$$

Where, $\overline{w_i}$ is the normalized firing strength from layer 3 (layer 3 output) and $\{p_i, q_i, r_i\}$ is the parameter set of this node. These are referred to as consequent parameters.

4.5. Layer 5

The single node in this layer is a fixed node labeled "sum" which computes the overall outputs as the summation of all incoming signals: overall output [10]:

$$O_i^5 = f = \sum_i \overline{w_i}f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{11}$$

The premise parameters and the consequent parameters are calculated during the training process [13]:

5. Maroon oilfiled

Maroon oilfield is one of the greatest oilfields in the western south of Iran. It is 67 km long and about 7 km wide. This oilfield is an asymmetric anticline. Oil is produced from Asmari formation in this field. Because of the magnitude of this field, it is divided into eight sections [8]. The location of this oilfield is illustrated in Fig. 2.

6. Data collecting

The data of 61 wells of Maroon oilfield are used to predict the lost circulation by ANFIS. These data include a large quantity of data that comprises the zones whether lost circulation is occurred. Because of the variety of scales of variables, all data are normalized using the below equation in order to get better results on ANFIS [8].

Fig. 2. Location of Maroon oilfield in Iran [38].

Table 1
Variables used as inputs and output.

S.N.	Variable	Unit	Minimum	Maximum
1	Drilling Meterage	ft	1.05	2132.546
2	Hole Size	in	4.125	26
3	WOB	tons	0.5	70
4	Pump Rate	GPM	80	1000
5	Pump Pressure	psi	50	2950
6	Viscosity	cp	27	100
7	θ_{600}	lb/(100ft)^2	3	293
8	θ_{300}	lb/(100ft)^2	2	163
9	Gel Strength	lb/(100ft)^2	1	49
10	Drilling Time	hr	0.1	24
11	Mud Velocity	ft/sec	45	644
12	Solid Percent	percent	0	61
13	Rotation of Bit	RPM	50	250
14	Formation Type	number	1	18
15	Pore Pressure	psi	12.07513	11121.56
16	Mud Pressure	psi	12.78161	16170.85
17	Fracture Pressure	psi	18.96326	14599.97
18	Loss Severity	bbl/hr	0	999

$$\widetilde{\chi}_i = \frac{\chi_i - \chi_{i_{\min}}}{\chi_{i_{\max}} - \chi_{i_{\min}}} \tag{12}$$

These data are shown in Table 1. The minimum and maximum of each variable in the table are used to normalize the variable into the interval [0 1].

Inputs of ANFIS must be numbers. Therefore, types of formations are assigned to numbers. 18 different formations are considered in the Maroon oilfield. Therefore, 18 numbers are allocated to these formations. Finally, these numbers are also normalized in the interval 0 to 1, so that these data will be in similar scale of the others. Each formation according to its number is shown in Table 4 (Appendix B). The sample of the used data in the ANFIS is shown in Tables 5 and 6 (Appendix B).

7. Results and discussions

The practical importance of this project resides in the prediction of lost circulation rather than solving the problem when it occurs. This work shows that in encountering problems such as lost circulation which deals with uncertainty, traditional approaches cannot predict the loss because of the complexity of the problem. This work offers a new approach for future work to prevent occurrence of lost circulation. It can be concluded that using smart systems such as ANFIS and data mining which handle a large data set, instead of curing problems, would save time and money. As a matter of fact, instead of solving problems, new approaches must be implemented to predict the problem. Because of the limitation of providing industrial data, we could collect and sort eighteen variables from Maroon oilfield. As ANFIS and data mining can handle a larger set of data, increasing number of inputs would possibly increase the accuracy of the model and decrease the error. However, it shows that other variables affect the occurrence of loss or even decrease the amount of it. Therefore, in planning new wells of the same field, these variables must be considered and optimized in order to decrease the possibility of lost circulation occurrence.

7.1. Determination the effect of each variable on lost circulation by DOE

According to previous discussion, DOE is used to analyze the sensitivity of data. By conducting DOE, effects of each parameter can be determined. Fig. 3 shows the effect of each variable on lost circulation. According to this figure, the highest effect of variables is related to shear stress at shear rate 300 (θ_{300}) and the lowest effect is drilling time. It should be noted that, θ_{600} and θ_{300} have the greatest effect on loss severity according to DOE. This means that proper designing of rheological properties of drilling fluids such as apparent viscosity, plastic viscosity, yield point, and gel strength could prevent the lost circulation.

7.2. Predicting lost circulation using data mining (regression)

Before using ANFIS to predict the amount of loss circulation, one class of data mining (regression) is used to find a function to model the data. Data mining is an interdisciplinary subfield in the science of computer. It is a process of finding patterns in large data sets which consists methods in artificial intelligence, machine learning, statistics, and database systems. The aim of using data mining is to extract information from a data set and change it to an analytical structure [31].

In this paper, class of regression is used to determine a function to model the data and the error of the model. The aim of this work is to show that regression only is not suitable for large data sets like lost circulation with seventeen variables. In ANFIS, part of data is used for training, another part is used for testing, and the rest is used for checking the model. In regression, the whole data are used together to determine the loss severity. Fig. 4 shows the results of regression of this data set. As it can be observed from Fig. 4, the error of regression is too high and not in acceptable range. Standard error of this model is 0.123667, root square is 40.71%, root square (adj) is 40.29%, and root square (pred) is 39.40%.

7.3. Predicting lost circulation using ANFIS

As it is mentioned before in this paper, the data of Maroon oilfield are used to feed the ANFIS. Total amount of the data are 42948 which 66% of them are used for training, 17% are used for testing, and 17%

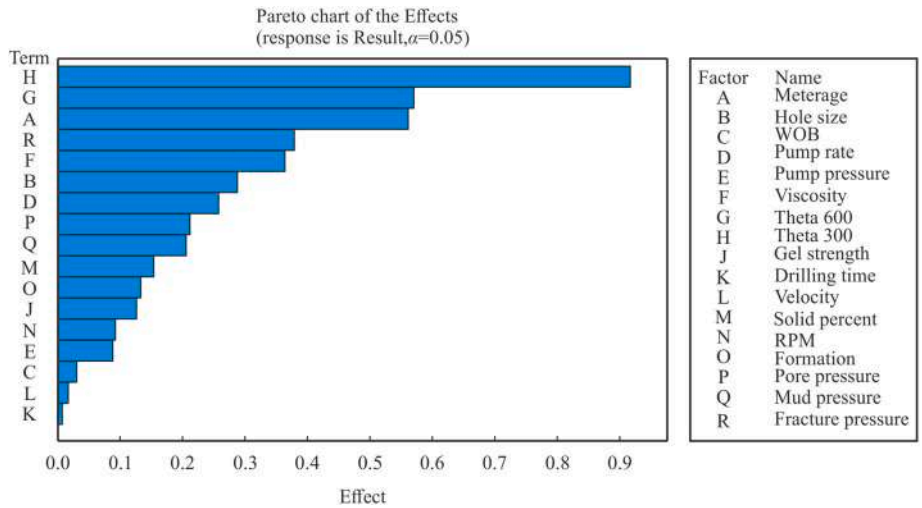

Fig. 3. Analyzing sensitivity of each variable on lost circulation using DOE technique.

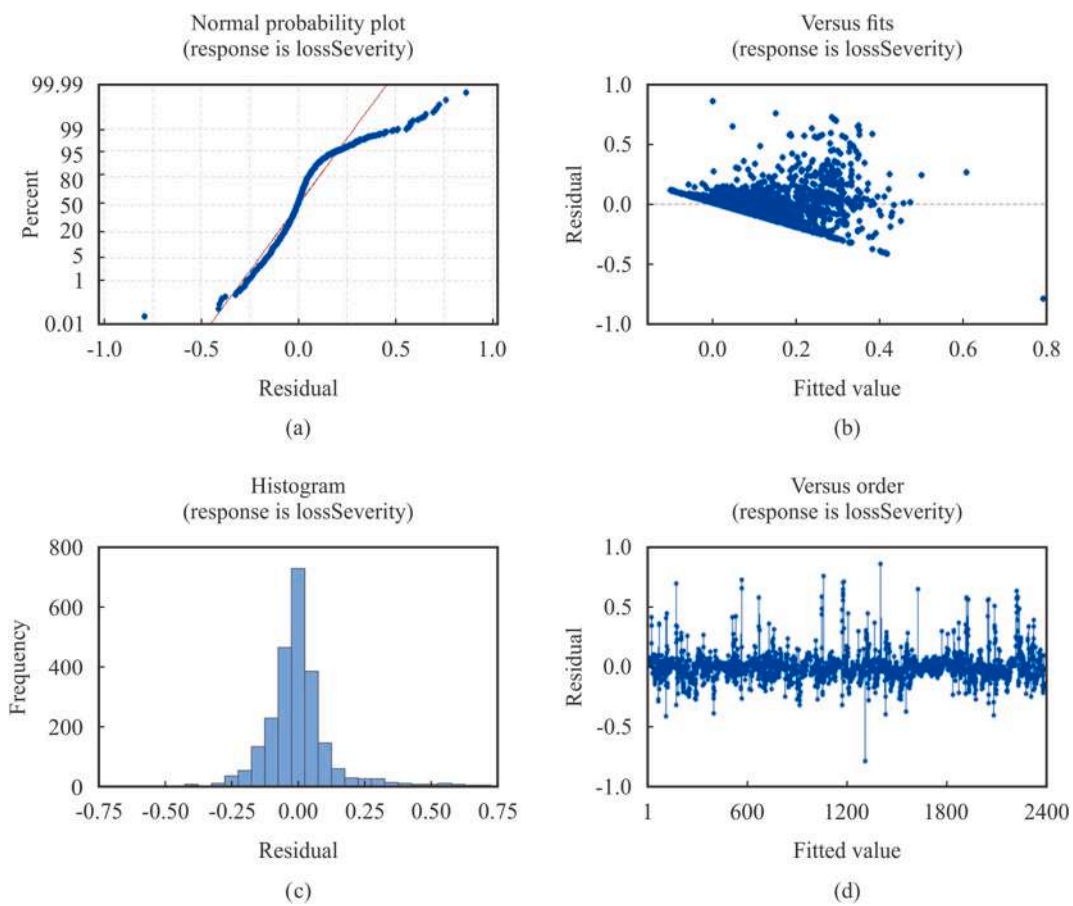

Fig. 4. Results of regression obtained from Data Mining a) Normal probability plot (residuals versus percent), b) Fitted value versus residuals, c) Histogram plot (residuals versus frequency), and d) Observation order versus residual.

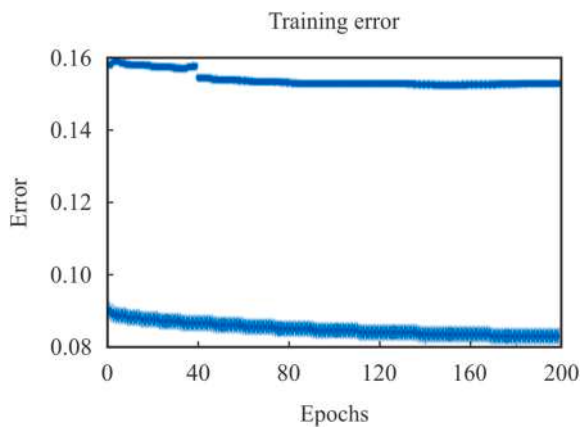

Fig. 5. Error of training and checking data using backpropagation method.

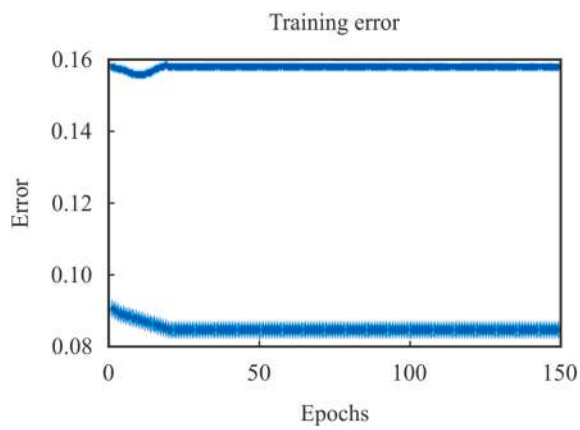

Fig. 6. Error of training and checking data using hybrid method.

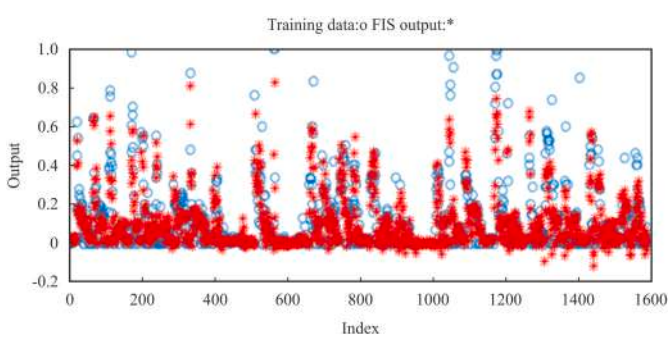

Fig. 7. Comparison of real and ANFIS outputs in training data.

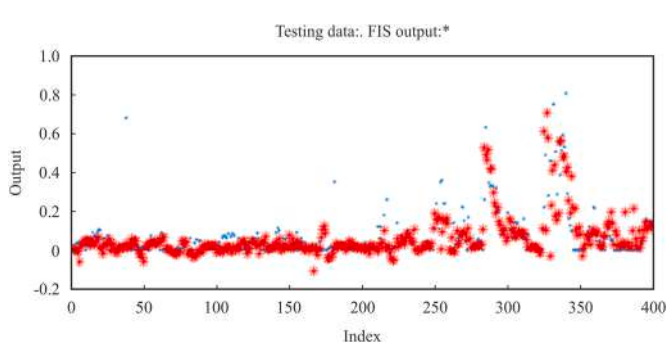

Fig. 8. Comparison of real and ANFIS outputs in testing data.

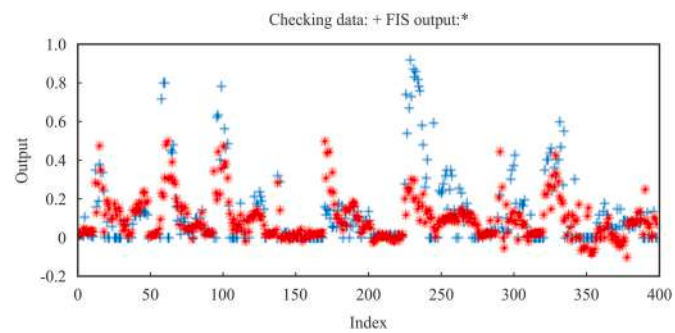

Fig. 9. Comparison of real and ANFIS outputs in checking data.

are used for checking the system. It should be noted that all of these data are combined together and are randomly selected for each section.

In order to generate FIS, subtractive clustering approach is used. Squash factor is considered to be 1.25, range of influence is 0.5, accept ratio is 0.5 and reject ratio is 0.15.

We used not only backpropagation, but also hybrid optimization method to train the FIS of the model. In the backpropagation method the minimum error is obtained at epoch 150. In the hybrid method the minimum error is obtained at epoch 10. The results of these two methods are shown in Figs. 5 and 6. As a result, backpropagation method is selected since it contains minimum error.

In the following Figs. 7–9 the results of model from training, testing, and checking data are illustrated. Blue spots represent the real data, and red spots represent those that model has predicted. As it can be seen, for the losses greater than 600 bbl/hr, the model does not effectively and precisely predict the amount of lost circulation. RMSE of the model according to both methods of backpropagation and hybrid for training, testing, and checking data is shown in Table 2.

Fig. 10 shows the structure of ANFIS for lost circulation. As it is shown, this structure is similar to the structures in ANFIS consists of five layers. First layer consists of inputs and weights of each input. Second layer consists membership functions of the 17 inputs used in the paper (Table 1). Third layer consists of if-then fuzzy rules. Fourth layer consists of the output of each cluster which is obtained using Sugeno's method. Subtractive clustering is used to generate FIS. As it is shown in Fig. 10, the ANFIS divided data into 12 clusters. Each cluster contains specific features relevant to the prediction of lost circulation.

Appendix A consists of 2D and 3D surfaces which are obtained from ANFIS. These surfaces show the effect of each variable on loss severity. Figs. 11–13 show the effect of each variable itself on loss severity. Fig. 14 illustrates the combined effect of some variables on loss severity. These figures show how these variables can affect the amount of lost circulation. It should be noted that in major data, fracture pressure is higher than mud pressure. Therefore, it can be concluded that the difference between these two variables is not the only factor in occurring lost circulation. Some factors such as rheological properties of drilling fluid, type of formations which drilling occur, petrophysical

Table 2
RMSE of ANFIS calculated from backpropagation and hybrid methods.

Method	Training	Testing	Checking
Backpropagation Method	0.083733	0.091039	0.15432
Hybrid Method	0.087428	0.09404	0.15615

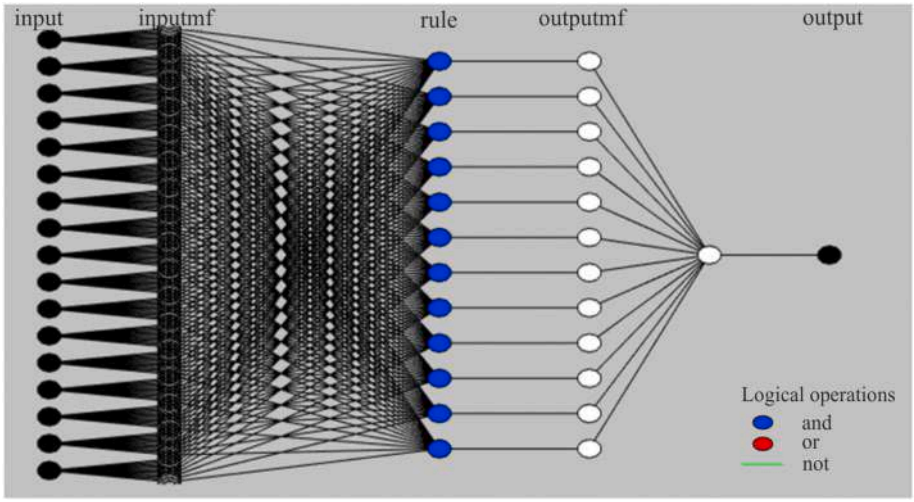

Fig. 10. Structure of ANFIS with two-input Takagi-Sugeno model For lost circulation having seventeen inputs.

Table 3
Real, ANFIS, and data mining Outputs of a chosen well.

#	Formation	Real Output (bbl)	ANFIS Output (bbl)	Data Mining Output (bbl)	ANFIS Error (%)	Data Mining Error (%)
1	Aghajari	540	407	207	25	62
2	Aghajari	180	142	190	21	6
3	Aghajari	250	177	198	29	21
4	Mishan	200	142	120	29	40
5	Mishan	150	100	102	33	32
6	Mishan	120	131	124	9	3
7	Gachsaran	168	123	141	27	16
8	Gachsaran	80	100	47	25	41
9	Gachsaran	210	95	74	55	65
10	Asmari	30	20	78	33	160
11	Asmari	48	38	33	21	31
12	Asmari	18	10	98	44	444
13	Pabdeh	30	27	32	10	7
14	Pabdeh	45	40	44	11	2
15	Pabdeh	70	56	21	20	70
16	Gurpi	40	45	48	13	20
17	Gurpi	21	26	31	24	48
18	Gurpi	22	18	32	18	45
19	Ilam	19	15	0	21	100
20	Ilam	8	3	0	63	100
21	Ilam	10	12	5	20	50
22	Sarvak	12	8	56	33	367
23	Sarvak	36	25	79	31	119
24	Sarvak	24	33	87	38	263
25	Kazhdomi	26	19	8	27	69
26	Kazhdomi	6	10	44	67	633
27	Kazhdomi	0	0	63	0	63
28	Dalian	30	24	19	20	37
29	Dalian	35	40	0	14	100
30	Dalian	16	12	0	25	100
31	Gadvan	18	13	27	28	50
32	Gadvan	8	5	10	38	25
33	Gadvan	130	110	0	15	100
34	Fahlian	90	75	0	17	100
35	Fahlian	0	0	7	0	0
36	Fahlian	44	37	20	16	55

characteristics, drilling parameters, and other factors can determine the amount and type of lost circulation. Since a large number of wells (61 wells) are investigated in this project, the results and effects of these variables can be extended to the whole oilfield.

7.4. Comparing results of ANFIS with data mining

From previous sections, errors of both ANFIS and data mining are calculated. In this paper, all variables such as inputs and output (loss severity) are given to the system and the goal is to find a relationship between these variables. Both ANFIS and data mining are conducted to find the desired relationship.

According to Fig. 2, standard error of this model is 0.123667, root square is 40.71%, root square (adj) is 40.29%, and root square (pred) is 39.40% which these results are not in acceptable range. The performance of the ANFIS models is assessed via RMSE. RMSE of this model using backpropagation method for training, testing, and training are 0.083733, 0.091039, and 0.15432, respectively.

By comparing these errors obtained from both approaches it can be concluded that ANFIS results are much better than of data mining. In addition, data of one well are chosen to compare the prediction ability of ANFIS and data mining methods. Results are shown in Table 3. These predictions are conducted according to each formation from top to the bottom of the chosen well in Maroon oilfiled. Lost circulation outputs are expressed in bbl/hr. Errors of each formation in which lost circulation is occurred are calculated by subtracting output of ANFIS and data mining from the real output and the output is divided by the real output. Errors of both methods are shown in percentage. From Table 3 it can be concluded that ANFIS method offers better results comparing to the data mining. Therefore, as a result, for problems such as lost circulation which contain large data set and input numbers and uncertainty is one of the major concerns, data mining cannot be conducted. However, ANFIS is a practical choice to be used in such problems.

8. Conclusions

In this paper, three approaches (DOE, data mining, and ANFIS) were carried out to predict the lost circulation in Maroon oilfield of Iran. 42948 data of 61 wells of Maroon oilfield were used to predict the lost circulation. DOE was conducted to observe the effect of each variable on the amount of lost circulation. Plackett-Burman method was used to determine the effects of variables on the amount of lost circulation. According to the results of this approach the highest effect of variables relates to shear stress at shear rate 300 (θ_{300}), and the lowest effect relates to drilling time. Then, one of the classes of data mining was used to predict the amount of lost circulation. Standard error of this model was 0.123667, root square was 40.71%, root square (adj) was 40.29%, and root square (pred) was 39.40%. Then, one of the classes of data mining was conducted to observe the results and errors of this approach. Results of this approach were not in acceptable range and errors were too high. Afterwards, lost circulation was predicted using ANFIS. In this approach, 66%, 17% and 17% of the data were used for training, testing and checking the system, respectively. Seventeen variables were used as inputs of the system and one variable (loss severity) was used as the output of the system. Subtractive clustering was used to generate the FIS. The optimization method for minimizing the error of the system was backpropagation method. RMSE of the model for training, testing and checking data were 0.083733, 0.091039, and 0.15432, respectively. Surfaces of the ANFIS showed the effect of each variable or combined effect of two variables on lost circulation. Finally, the comparison between ANFIS and data mining (regression) was carried out. It was concluded that ANFIS can predict lost circulation more precisely than data mining.

Nomenclature

ANFIS	adaptive neuro-fuzzy inference system
DOE	design of experiments
FIS	fuzzy inference system
MF	membership function
WOB	weight on bit
R	statistical correlation coefficient
RMSE	root mean square error
RPM	revolution per minute
θ_{600}	shear stress at shear rate 600 1/sec
θ_{300}	shear stress at shear rate 300 1/sec
O_i	output of ANFIS layers
p, q, r	sugeno's constants
u_{Ai}	membership function
w_i	firing strength

Appendix A

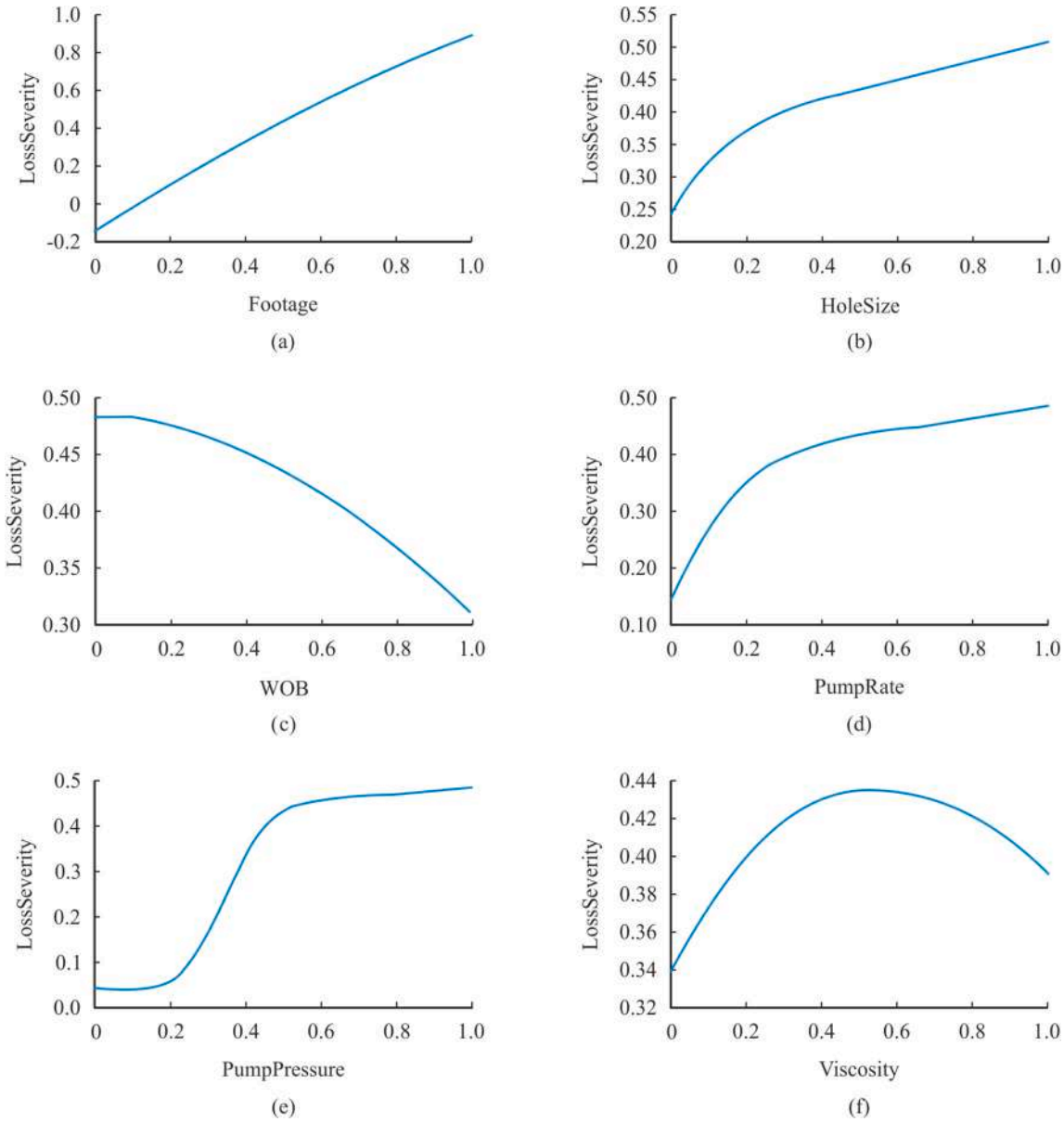

Fig. 11. Effect of each variable on lost circulation in 2-D surfaces: a) Effect of drilling Meterage on loss severity, b) Effect of hole size on loss severity c) Effect of WOB on loss severity, d) Effect of pump rate on loss severity, e) Effect of pump pressure on loss severity, and f) Effect of mud viscosity on loss severity.

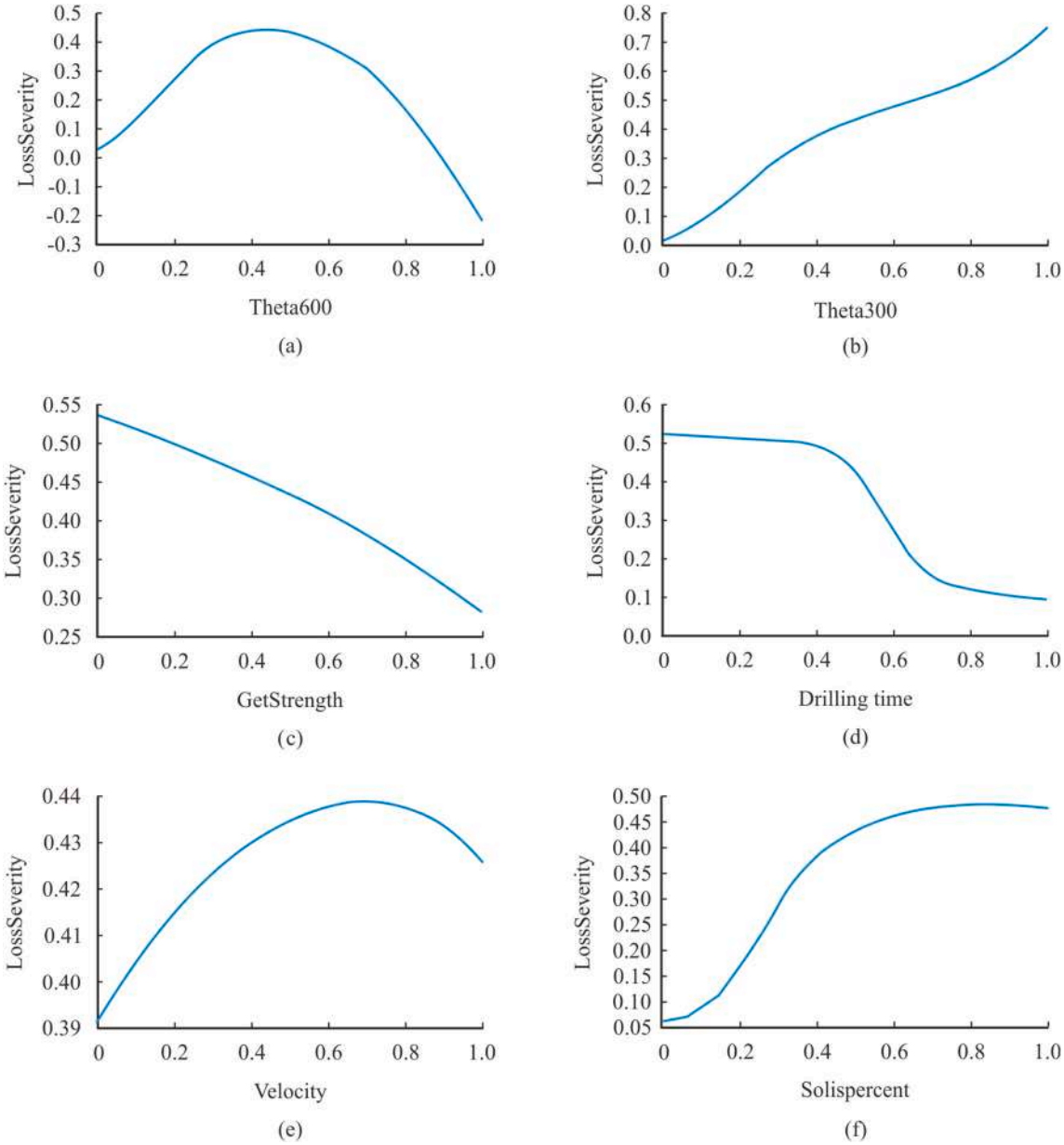

Fig. 12. Effect of each variable on lost circulation in 2-D surfaces: a) Effect of θ_{600} on loss severity, b) Effect of θ_{300} on loss severity, c) Effect of gel strength on loss severity, d) Effect of drilling time on loss severity, e) Effect of mud velocity on loss severity, and f) Effect of solid percent on loss severity.

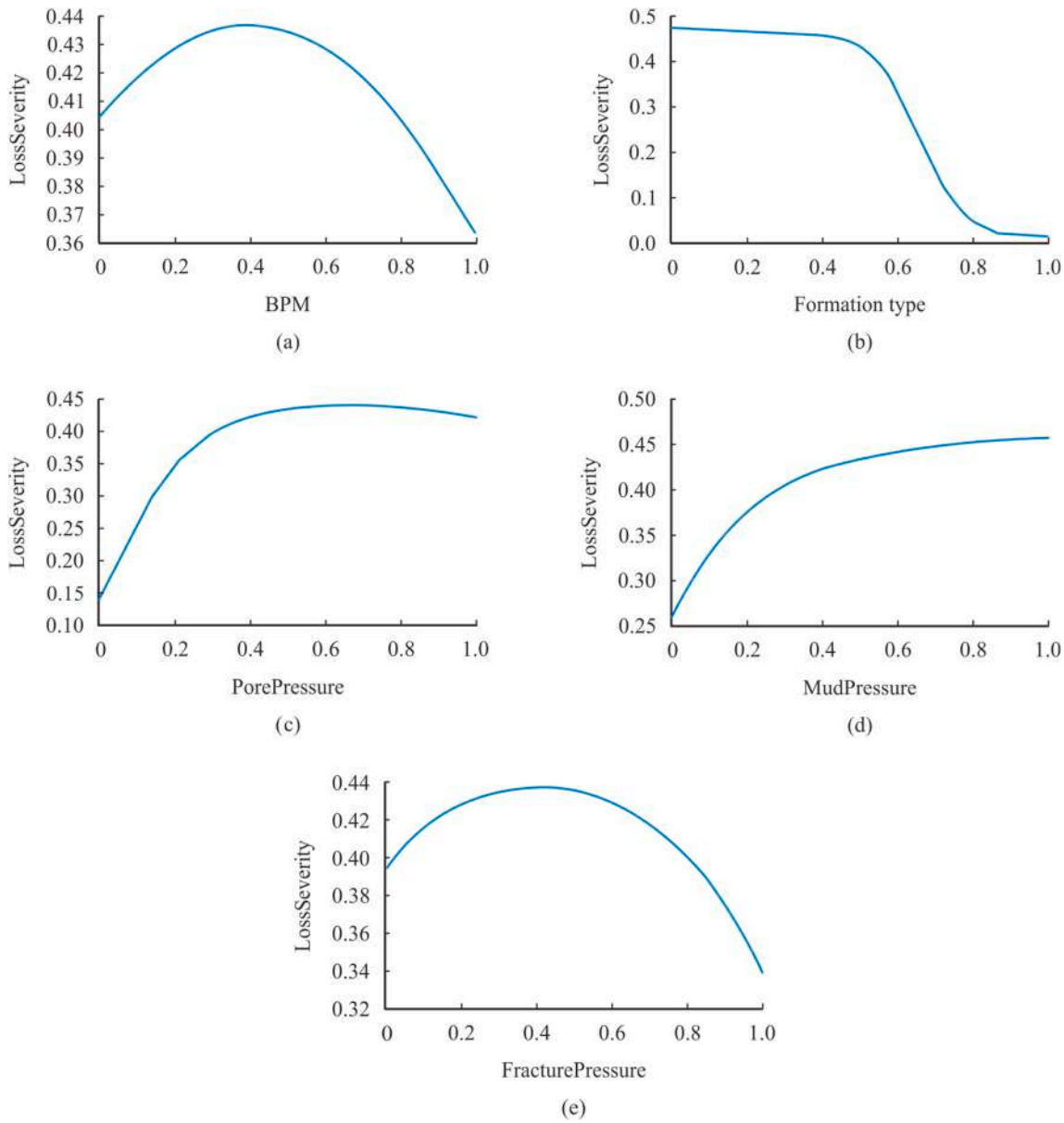

Fig. 13. Effect of each variable on lost circulation in 2-D surfaces: a) Effect of RPM on loss severity, b) Effect of formation type on loss severity, c) Effect of pore pressure on loss severity, d) Effect of mud pressure on loss severity, and e) Effect of fracture pressure on loss severity.

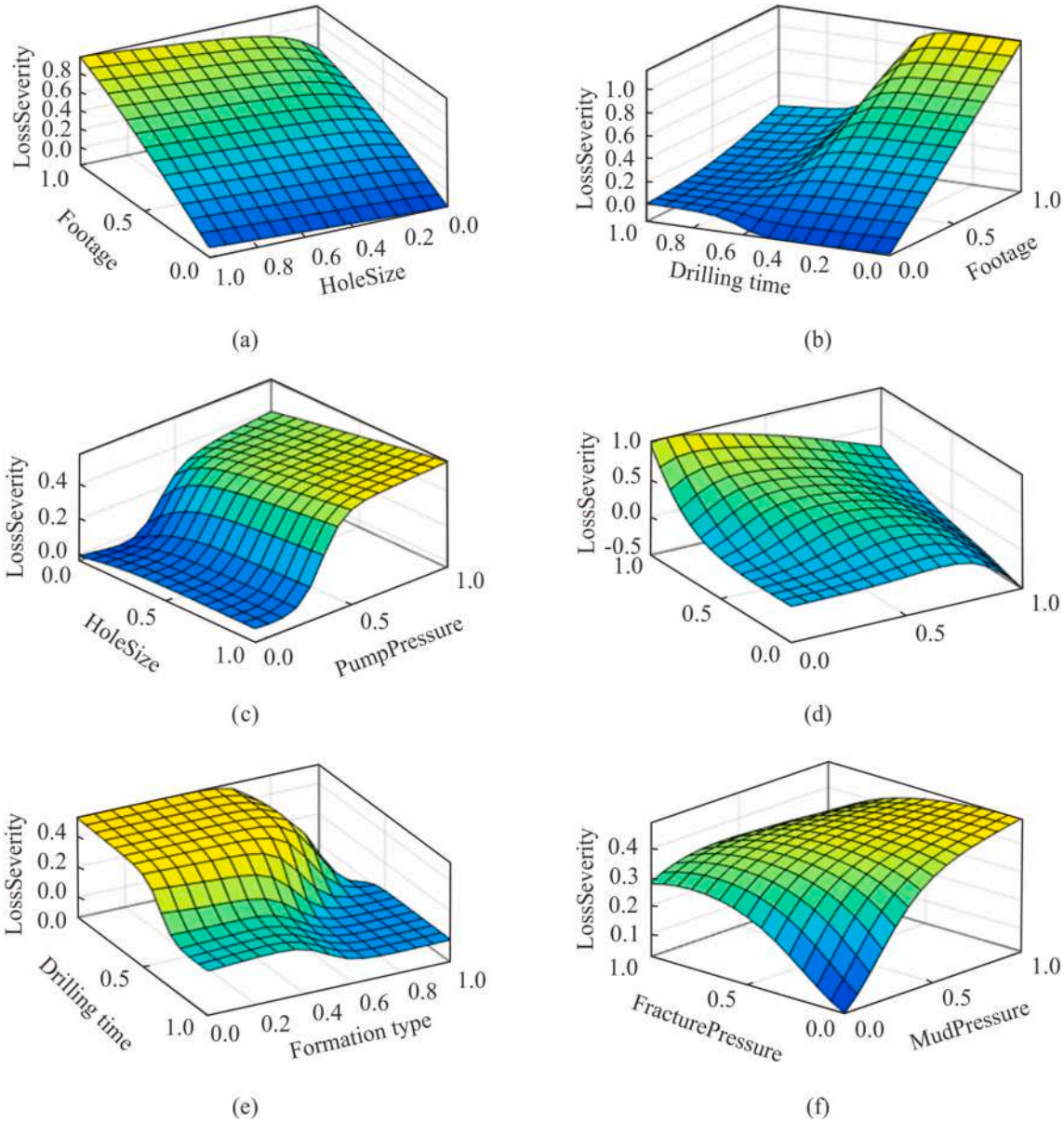

Fig. 14. Effects of two variables on lost circulation in 3-D surfaces: a) Combined effect of drilling meterage and hole size on loss severity, b) Combined effect of drilling Meterage and drilling time on loss severity, c) Combined effect of hole size and pump pressure on loss severity, d) Combined effect of θ600 and θ300 on loss severity, e) Combined effect of drilling time and formation type on loss severity, f) Combined effect of mud pressure and fracture pressure on loss severity.

By analyzing surfaces obtained from ANFIS, the below results are concluded.

(1) Fig. 11(a) shows that increasing drilling footage increases loss severity, which has a wide effect on this problem. Increasing of drilling footage causes amount of lost circulation to change and increase between zero to 850 bbl/hr.
(2) Fig. 11(b) shows that increasing hole size (bit diameter) increases loss severity. This figure shows that changing the hole size has increased the amount of lost circulation between 250 and 500 bbl/hr.
(3) Fig. 11(c) shows that increasing WOB decreases loss severity. By analyzing this figure it can be obtained that increasing WOB, has decreased the amount of lost circulation from 480 bbl/hr to 310 bbl/hr.
(4) Fig. 11(d) shows that increasing pump rate, increases loss severity. This figure shows that increasing pump rate has caused loss severity to

increase from 150 bbl/hr to 500 bbl/hr.

(5) Fig. 11(e) shows that increasing pump pressure, increases loss severity. In pump pressures between 50 and 485 psi, the amount of lost circulation is about 50 bbl/hr. in pump pressures between 485 and 1500 psi, the amount of lost circulation has increased from 50 to 450 bbl/hr. finally, in pressures between 1500 and 2950 psi, the amount of is approximately constant and in its highest amount.

(6) Fig. 11(f) shows that increasing viscosity, first increases loss severity, after that decreases it. In amount of viscosity between 27 and 67 cp, lost circulation increases from 340 to 440 bbl/hr but in amount of viscosity between 67 and 100 cp, the amount of lost circulation has decreased from 440 to 390 bbl/hr.

(7) Fig. 12(a) shows that increasing shear stress at shear rate 600, first increases loss severity, then decreases it. This figure shows that increasing shear stress at shear rate 600 from 3 to 133 has increased the amount of lost circulation from 30 to 450 bbl/hr but in the amount of shear stress between 133 and 293, the amount of lost circulation has reached to zero.

(8) Fig. 12(b) shows that increasing shear stress at shear rate 300, increases loss severity. This figure shows that increasing shear stress at shear rate 300 has caused the amount of lost circulation to increase from zero to 750 bbl/hr.

(9) Fig. 12(c) shows that increasing gel strength of the mud, decreases loss severity. This figure shows that increasing gel strength has caused the amount of lost circulation to decrease from 550 to 280 bbl/hr.

(10) Fig. 12(d) shows that increasing drilling time, decreases loss severity. In low drilling time, between 6 min and 10 h, the highest amount of lost circulation is about 530 bbl/hr. between 10 and 19 h the amount of lost circulation decreases from 530 to 100 bbl/hr and in time between 19 and 24 h, the amount of lost circulation has decreased slowly, which in time of 24 h has reached below 100 bbl/hr.

(11) Fig. 12(e) shows that increasing mud velocity, first increases loss severity, after that decreases it. In velocities between 45 and 464 ft/min the amount of lost circulation has increased from 390 to 440 bbl/hr and between 464 and 644 ft/min its amount has decreased from 440 to 425 bbl/hr. this result shows that mud velocity has low effect on loss severity.

(12) Fig. 12(f) shows that increasing solid percent obtained from retort test, increases loss severity. This figure shows that increasing solid percent from zero to 36% has increased amount of lost circulation between 50 and 450 bbl/hr. increasing solid percent from 36% to 48% has increased amount of lost circulation from 450 to 480 bbl/hr. Finally, the amount of lost circulation is constant and in its highest amount till the solid percent reaches to 61.

(13) Fig. 13(a) shows that RPM has low effect on loss severity.

(14) Fig. 13(b) shows that the most lost circulation problem has occurred in formations such as Aghajari, Mishan, Gachsaran 7 to 1, Asmari and Pabdeh.

(15) Fig. 13(c) shows that increasing pore pressure, increases loss severity. This figure shows that increasing pore pressure from 12.07 to 6678 psi, increases the amount of lost circulation from 140 to 440 bbl/hr and after that the amount of lost circulation is constant and it is in highest amount.

(16) Fig. 13(d) shows that increasing mud pressure, increases loss severity. This figure shows that increasing mud pressure has increased the amount of lost circulation from 260 to 460 bbl/hr.

(17) Fig. 13(e) shows that increasing fracture pressure, first increases loss severity, then causes it to decrease. This figure shows that increasing fracture pressure from 18.96 to 5852 psi, has increased the amount of lost circulation from 395 to 440 bbl/hr and after that by increasing the fracture pressure, the amount of lost circulation decreases from 440 to 340 bbl/hr.

(18) Fig. 14(a) shows that if drilling footage and hole size increase, loss severity will increase. In a constant drilling footage, loss severity will increase if the hole size increases.

(19) Fig. 14(b) in fact shows the ROP. It shows that increasing ROP, increases loss severity.

(20) Fig. 14(c) shows that the effect of pump pressure on loss severity is very larger than the effect of hole size.

(21) Fig. 14(d) shows that, in low amount of θ300, increasing θ600 firs increases loss severity, then causes it to decrease. But, in high amount of θ300, increasing θ600 decreases loss severity.

(22) Fig. 14(e) shows that effect of drilling time on lost circulation in formations like Gurpi, Ilam, Sarvak, Kazhdomi, Dalian, Gadvan and Fahlian is very low but in the other formations decreasing drilling time, increases loss severity.

(23) Fig. 14(f) shows that increasing mud pressure and fracture pressure, increase loss severity.

Appendix B

Table 4
Allocated numbers to each formation.

#	Formation	Number	Normalized Number
1	Aghajari	1	0
2	Mishan	2	0.06

(*continued on next page*)

Table 4 (*continued*)

#	Formation	Number	Normalized Number
3	Gachsaran 7	3	0.12
4	Gachsaran 6	4	0.18
5	Gachsaran 5	5	0.24
6	Gachsaran 4	6	0.29
7	Gachsaran 3	7	0.35
8	Gachsaran 2	8	0.41
9	Gachsaran 1	9	0.47
10	Asmari	10	0.53
11	Pabdeh	11	0.59
12	Gurpi	12	0.65
13	Ilam	13	0.71
14	Sarvak	14	0.76
15	Kazhdomi	15	0.82
16	Dalian	16	0.88
17	Gadvan	17	0.94
18	Fahlian	18	1

Table 5
Inputs of ANFIS.

#	Drilling Footage	Hole Size	WOB	Pump Rate	Pump Pressure	Viscosity	θ_{600}	θ_{300}	Gel Strength
1	160.7612	8.375	17.5	450	1250	39	33	20	2
2	13.12336	8.375	17.5	450	1250	40	36	22	4
3	16.4042	8.375	17.5	300	910	40	28	18	4
4	45.93176	8.375	25	300	865	40	27	17	4
5	78.74016	8.375	25	300	670	39	29	18	4
6	82.021	8.375	25	300	670	39	29	18	4
7	88.58268	17.5	27.5	880	2775	43	44	29	7
8	91.86352	17.5	27.5	800	2650	44	48	32	8
9	52.49344	17.5	27.5	780	2675	45	46	31	8
10	95.14436	17.5	27.5	780	2725	45	45	30	8
11	32.8084	17.5	27.5	780	2725	45	47	31	8
12	39.37008	17.5	27.5	750	2875	45	49	32	7
13	6.56168	17.5	27.5	750	2875	45	51	33	7
14	72.17848	17.5	22.5	760	2875	46	56	36	8
15	259.1864	12.25	30	580	2800	55	139	79	3
16	232.9396	12.25	30	560	2800	57	144	82	3
17	131.2336	12.25	30	560	2825	57	143	82	4
18	269.0289	12.25	30	550	2825	58	149	85	4
19	157.4803	26	37.5	980	2290	33	25	20	6
20	121.3911	26	35	980	2390	34	28	22	7
21	236.2205	17.5	25	970	1875	35	20	15	9
22	219.8163	17.5	25	970	1875	35	21	14	8
23	209.738	17.5	25	970	1975	37	22	14	9
24	85.30184	17.5	25	970	1975	38	25	16	10
25	9.84252	17.5	25	970	2175	39	24	15	7

Table 6
Inputs and output of ANFIS.

#	Drilling Time	Mud Velocity	Solid Percent	RPM	Formation Type	Pore Pressure	Mud Pressure	Fracture Pressure	Loss Severity
1	24	244	6	120	10	4764.28	5348.031	9659.088	8
2	8	244	6	120	10	4796.014	5383.653	9723.426	8
3	6	163	6	60	10	4801.403	5389.702	9734.351	105
4	13	163	6	70	10	4812.779	5402.473	9757.415	26
5	24	163	6	70	10	4835.532	5428.013	9803.544	16
6	24	163	6	70	10	4864.871	5460.947	9863.025	14
7	23	76	23	200	2	3491.314	4849.047	525.742	0
8	24	69	23	200	3	3533.268	5012.85	592.772	0
9	19	67	26	200	3	3566.831	5167.004	5983.071	0
10	24	67	27	180	3	3601.157	5270.51	6040.65	12
11	8	67	27	180	3	3630.906	5314.049	6090.551	45
12	21	65	27	150	3	3647.687	5338.61	6118.701	120
13	8	65	27	180	3	3658.366	5354.24	6136.614	60
14	18	0	28	180	3	3676.673	5545.759	6167.323	75
15	23	113	40	180	4	7625.492	7943.221	9320.046	0
16	24	110	40	180	5	8074.311	8410.741	9868.603	90
17	14	110	40	180	5	8238.189	8581.447	10068.9	98

(*continued on next page*)

Table 6 (*continued*)

#	Drilling Time	Mud Velocity	Solid Percent	RPM	Formation Type	Pore Pressure	Mud Pressure	Fracture Pressure	Loss Severity
18	24	107	41	180	5	8418.307	8834.026	10289.04	120
19	18	37	5	120	1	1508.681	1621.145	2369.291	340
20	7	37	5	120	1	1569.057	1686.021	2464.108	130
21	24	87	0	180	1	3211.985	3966.558	5044.226	360
22	24	87	14	180	1	3310.717	4088.485	5199.278	200
23	24	87	16	180	1	3403.767	4421.752	5345.407	60
24	24	87	17	180	1	3467.694	4504.798	5445.801	150
25	3	87	19	180	1	3488.292	4755.338	5478.15	0

References

[1] W.J. Murchison, Lost Circulation for the Man on the Rig, Murchison Drilling Schools, Inc, Albuquerque, NM, 2006.
[2] C. Ivan, J. Bruton, B. Bloys, How can we best manage lost circulation? AADE National Technology Conference, American Association of Drilling Engineers, 2003.
[3] F.E. Dupriest, Fracture closure stress (FCS) and lost returns practices, SPE/IADC Drilling Conference, Society of Petroleum Engineers, 2005.
[4] S. Savari, D.L. Whitfill, Lost circulation management in naturally fractured formations: efficient operational strategies and novel solutions, IADC/SPE Drilling Conference and Exhibition, Society of Petroleum Engineers, 2016.
[5] A. Nasiri, A. Ghaffarkhah, M.K. Moraveji, A. Gharbanian, M. Valizadeh, Experimental and field test analysis of different loss control materials for combating lost circulation in bentonite mud, J. Nat. Gas Sci. Eng. 31 (2017) 1–8.
[6] D.L. Whitfill, D.E. Jamison, M. Wang, A. Angove-Rogers, Preventing lost circulation requires planning ahead, SPE International Oil Conference and Exhibition in Mexico, Society of Petroleum Engineers, 2007.
[7] E. Fidan, T. Babadagli, E. Kuru, Use of cement as lost circulation material-field case studies, IADC/SPE Asia Pacific Drilling Technology Conference and Exhibition, Society of Petroleum Engineers, 2004.
[8] A.R. Moazzeni, M. Nabaei, S. Ghadami jegarluei, Prediction of lost circulation using virtual intelligence in one of the Iranian oilfileds, SPE Nigeria Annual International Conference and Exhibition, Society of Petroleum Engineers, 2010.
[9] P. Chen, M.M. Rahman, H.K. Sarma, Interaction between hydraulic fracture and natural fracture- a new prediction model by adaptive neuro-fuzzy inference system (ANFIS), SPE Abu Dhabi International Petroleum Exhibition and Conference, Society of Petroleum Engineers, 2014.
[10] J.S. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE. Trans. Syst. Man. Cybern. Syst. 23 (1993) 665–685.
[11] A.H. Dehaghani, M.H. Badizad, A soft computing approach for prediction of P-ρ-T behavior of natural gas using adaptive neuro-fuzzy inference system, Petroleum 4 (2017) 447–453.
[12] G.R. SHI, Superiorities of support vector machine in fracture prediction and gassiness evaluation, Pet. Exp. Develop. 35 (2008) 588–594.
[13] L. Xiongyan, Z. Jinyu, L. Hongqi, S. ZHANG, C.H. Yihan, Computational intelligent methods for predicting complex ithologies and multiphase fluids, Pet. Exp. Develop. 39 (2) (2012) 261–267.
[14] A. Kamari, A.H. Mohammadi, M. Lee, A. Bahadori, Decline curve based models for predicting natural gas well performance, Petroleum 3 (2) (2017) 242–248.
[15] D. Petković, Ž. Ćojbašič, V. Nikolić, Adaptive neuro-fuzzy approach for wind turbine power coefficient estimation, Renew. Sustain. Energy Rev. 28 (2013) 191–195.
[16] D. Petković, Ž. Ćojbašič, V. Nikolić, S. Shamshirband, M.L. Kiah, N.B. Anuar, A.W. Wahab, Adaptive neuro-fuzzy maximal power extraction of wind turbine with continuously variable transmission, Energy 64 (2014) 868–874.
[17] D. Petković, S.H. Ab Hamid, Ž. Ćojbašić, N.T. Pavlović, Adapting project management method and ANFIS strategy for variables selection and analyzing wind turbine wake effect, Nat. Hazards 74 (2) (2014) 463–475.
[18] D. Petković, V. Nikolić, V.V. Mitić, L. Kocić, Estimation of fractal representation of wind speed fluctuation by artificial neural network with different training algorothms, Flow Meas. Instrum. 54 (2017) 172–176.
[19] V. Nikolić, V.V. Mitić, L. Kocić, D. Petković, Wind speed parameters sensitivity analysis based on fractals and neuro-fuzzy selection technique, Knowl. Inf. Syst. 52 (1) (2017) 255–265.
[20] V. Nikolić, S. Sajjadi, D. Petković, S. Shamshirband, Ž. Ćojbašić, L.Y. Por, Design and state of art of innovative wind turbine systems, Renew. Sustain. Energy Rev. 61 (2016) 258–265.
[21] D. Petković, N.T. Pavlović, Ž. Ćojbašić, Wind farm efficiency by adaptive neuro-fuzzy strategy, Int. J. Elec. Power 81 (2016) 215–221.
[22] L.A. Zadeh, Fuzzy sets, Inf. Control. 8 (3) (1965) 338–353.
[23] V. Uraikul, C.W. Chan, P. Tontiwachwuthikul, Artificial intelligence for monitoring and supervisory control of process systems, Eng. Appl. Artif. Intell. 20 (2) (2007) 115–131.
[24] Q. Zhou, Y. Wu, C.W. Chan, P. Tontiwachwuthikul, From neural network to neuro-fuzzy modeling: applications to the carbon dioxide capture process, Energy. Procedia. 4 (2011) 2066–2073.
[25] Q. Zhou, C.W. Chan, P. Tontiwachwuthikul, R. Idem, D. Gelowitz, Application of neuro-fuzzy modeling technique for operational problem solving in a CO2 capture process system, Int. J. Greenh. Gas. Con. 15 (2013) 32–41.
[26] Q. Zhou, C.W. Chan, P. Tontiwachwuthikul, An application of neuro-fuzzy technology for analysis of the CO2 capture process, Fuzzy Set Syst. 161 (19) (2010) 2597–2611.
[27] A. Chakma, G.H. Huang, Z. Chen, P. Tontiwachwuthikul, A fuzzy set approach for evaluating environmental risks at petroleum storage sites, Technical Meeting/Petroleum Conference of the South Saskatchewan Section, Petroleum Society of Canada, 1997.
[28] S.S. Gill, J. Singh, An Adaptive Neuro-Fuzzy Inference System modeling for material removal rate in stationary ultrasonic drilling of sillimanite ceramic, Expert Syst. Appl. 37 (8) (2010) 5590–5598.
[29] M.J. Shojaei, E. Bahrami, P. Barati, S. Riahi, Adaptive neuro-fuzzy approach for reservoir oil bubble point pressure estimation, J. Nat. Gas Sci. Eng. 20 (2014) 214–220.
[30] A. Ja'fari, R.H. Moghadam, Integration of ANFIS, NN and GA to determine core porosity and permeability from conventional well log data, J. Geophys. Eng. 9 (5) (2012) 473.
[31] K. Singh, S.C. Guntuku, A. Thakur, C. Hota, Big data analytics framework for peer-to-peer botnet detection using random forests, Inf. Sci. 278 (2014) 488–497.
[32] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, From data mining to knowledge discovery in database, AI Mag. 17 (1996) 37.
[33] L. Sheremetov, I. Batyrshin, D. Filatov, J. Martinez, H. Rodriguez, Fuzzy expert system for solving lost circulation problem, Appl. Soft Comput. 8 (2008) 14–29.
[34] D.C. Montgomery, Design and Analysis of Experiments, Jhon Wiley and Sons Inc, New York, 2001.
[35] C.F.J. Wu, M.S. Hamada, Experiments Planning Analysis and Parameter Design Optimization, Jhon Wiley and Sons Inc, Singapore, 2000.
[36] J.G. Carton, A.G. Olabi, Design of experiment study of the parameters that affect performance of three flow plate configurations of a proton exchange membrane fuel cell, Energy 35 (7) (2010) 2796–2806.
[37] A.K. Pujari, Data Mining Techniques, Universities press, 2001.
[38] E. Aghanabati, Geology of Iran, Publication of the Organization of Geology and Mineral Exploration, 2004.