Petroleum and Coal

Article Open Access

DISTINCT METHODOLOGIES TO ASSESS THE CONDITIONS OF PETROLEUM RESERVOIRS WITH RESPECT TO ONSET OF SAND PRODUCTION

Mohammad M. Ghiasi[1], Mohammad F. Ghasemi[2], Vahid Heidaripour[3], Amir H. Mohammadi[1]**

[1] *Discipline of Chemical Engineering, School of Engineering, University of KwaZulu-Natal, Howard College Campus, South Africa*
[2] *Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Bolshaya Gruzinskaya, 10-1 Moscow 123242, Russia*
[3] *Department of Computer Science, Kharazmi University, Tehran, Iran*

Received December 27, 2018; Accepted February 18, 2019

Abstract

Handling such problem as sanding in the petroleum industry leads to safe and more economic access to the required energy for development. With the objective of making strategy in order to deal with sanding the results of assessing the conditions at which sand production takes place in petroleum reservoirs have been reported in this communication. In this work, CART-Decision Tree, Random Forest and Extremely Randomized Tree (Extra Tree) have been employed to conduct the classification for the first time. The dominating variables in sanding including total vertical depth, transmit time, cohesive strength of the formation, water and gas flow rates, bottom hole flowing pressure, drawdown pressure, effective overburden stress, shut per foot, and perforation interval have been introduced into the employed approaches for obtaining the best models. The modeling process was performed on the basis of the gathered field data from the literature. All the developed models classify the sanding conditions with 100% accuracy. The performance of the presented models proves their capability in determining the possible sand production in a real petroleum field. Hence, the presented methodologies will pave the way for effective sand control plan.

***Keywords**: Sand production; classification; modeling; reservoir; decision tree.*

1. Introduction

1.1. Energy security and Sand production phenomenon

In view of growth in population and advancement in standards of living as well, the energy consumption is increased. According to the U.S. Information Administration (EIA) [1], the increase in the world energy consumption is fully expected to happen for the upcoming decades. Hence, a lot of importance is attached to energy security and conservations.

The growth in the world economy can be attributed to the advances in technologies since the 1750s. It is clear that development of new technologies/ methodologies, leading to efficient utilization of energy resources, is crucial for sustainable development. Nowadays, large share of energy demand of the world is provided by petroleum fluids. Indeed, fossil fuels, especially crude oil and natural gas, play a significant role in supplying the energy demand [2]. Hence, solving/ handling the petroleum production, processing, and transportation problems provides safe and more economic access to the required energy for development.

The sand production is a costly phenomenon that brings substantial damages and problems to the petroleum industry. In the petroleum industry, sand production is known to be the coexistence of solid particles with the produced reservoir fluid. It is believed that about 70% of the hydrocarbons exist in oil and gas reserves are in not well-consolidated reservoirs [3-4]. Because of initiation of drilling and starting hydrocarbon production from a well, leading to redistributing the pore pressure and stresses around the production cavity, the hydro-mechanical equilibrium of the sandstone formation is disrupted. Consequently, the sand particles travel from the reservoir into the production well [5-7]. Based on the field observations, the sand volumetric concentration in systems of oil pipe ranges between 1% and 40% [8-9].

Various problems are associated with the sand production. Further to the damages of sanding on production facilities, this phenomenon is responsible for the instability of the wellbore and production cavities [10-12]. As a result, production losses increase. Additional costs may also be imposed owing to depositional of waste sand [3,13]. This can be the reason for the increase in costs of maintenance and operation. Equipment erosion indicates another problem caused by sanding [14]. It should be noted that the limited sand production might noticeably increase the productivity of the well in conventional reservoirs [11]. In view of the aforementioned facts, implementation of strategies for sand control and management is highly crucial. To achieve this, the potential of the defined reservoir to sand production is extremely beneficial.

1.2. Literature review

Great efforts have been made in the study of sand production by many researchers. Since the sand arches have been observed in the field around each hole in the casing of well, several researchers employed the phenomenon of sand arching for identifying the possibility of production of sand [15-17]. In 2001, the sand arches behavior, stability, and morphology were evaluated by Bianco and Halleck [3]. In 1989, Morita *et al.* [18] studied the effects of different variables using an analytical approach. Based on the reported results, stress and pressure distribution of wellbore around well, drag forces induced by fluid flow, formation rock strength, perforation geometry and shot density, and history of cyclic loading are the main parameters that affect sanding. In 1991, analyses of five common sand problems in the field were provided by Morita and Boyd [19]. More works on the subject of sand production is reviewed by Ranjith *et al.* [7].

In 2014, the permeability evolution law within process of sand production of weak sandstone was investigated by Nie *et al.* [20]. In another work, a coupled numerical approach was presented on the basis of Lattice Boltzmann Method and Discrete Element Method by Ghassemi and Pak [21]. The proposed model was then employed for simulating the sand production. More recently, Jiang *et al.* [22] studied the ureolytic activities of purified urease enzyme and Bacillus megaterium in both oxic and anoxic conditions for their promising use in control of subsea floor production.

Several models including numerical, empirical, and theoretical are available in the literature for sand prediction [5,17,23-24]. In 2000, Doan *et al.* [25] presented a numerical model for sand gravitational deposition in a horizontal well in heavy oil reservoirs. Similar to other fields of petroleum and natural gas engineering, the algorithms developed based on the artificial intelligence (AI) have been employed in investigating the sand production. In 1999, a neural-based method was proposed by Kanj and Abousleiman [26] for estimating the sanding onset for Northern Adriatic Basin gas wells. Further to the above, Azad *et al.* [27] presented another neural network model for predicting the critical bottom hole flowing pressure inhibiting sanding. Recently, Khamehchi *et al.* [28] presented two models including back-propagation neural network and particle swarm optimization neural network to estimate the critical total drawdown as an indicator of sanding in gas and oil wells.

2. Objective of the study

In 2016, Gharagheizi *et al.* [29] developed a model based on the least square version of support vector machine (LS-SVM) classification method for predicting the sand production onset in reservoirs. It was shown that the LS-SVM classification approach can successfully be employed for prediction of conditions under which sand production occurs. To the best of our knowledge, there is no other published work in this area.

The high performance of the SVM methodology in classification problems has been approved [30]. However, direct understanding of the rules obtained by SVM approach is hard.

Furthermore, they are costly in computation [31]. In view of this, the aim of this study was applying CART-decision trees, Random Forest, and Extra Trees for the application of interest and exploring their performances in classification. Decision trees present several benefits. For example, the background is easy to understand and interpret; they are convertible to set of if-then rules, and there is no need to know about the nature of the data [32].

In this work, three classification methods including CART-Decision Tree, Random Forest and Extremely Randomized Trees (Extra Trees) are utilized to identify the sanding conditions. These algorithms are used for the first time for the application of interest.

The used data in this study for classification include a databank of 31 wells of Northern Adriatic Basin [14]. Amongst the investigated wells, 23 wells are reported as problematic wells with respect to the production of sand; the other wells are considered to be sand free [14]. According to the work of Moricca *et al.* [14] the following parameters are main variables that affect the production of sand: effective overburden stress (EOVS), bottom hole flowing pressure (BHFP), total vertical depth (TVD), transmit time (TT), drawdown pressure (DD), cohesive strength of the formation (COH), shut per foot (SPF), water and gas flow rates (Qw & Qg), perforation interval (Hperf). Before introducing the collected data points into the aforementioned classification algorithms, the data points were tested for incomplete data. Consequently, two sets of gathered databank were removed. Table 1 gives the finalized data points used for the modeling process. In the last column of Table 1, entitled field data, 1 means sand production is observed and 0 means sand production is not observed.

To achieve the goal, the rest of the work is pursued as follows: first, the background of Decision Tree, Random Forest and Extremely Randomized Trees (Extra Trees) are introduced. Second, the development of models for predicting the possibility of the sand production is presented. Finally, the developed models and their results are investigated and discussed.

3. Theory

3.1. Decision tree

As one of the most well-known algorithms for classification, decision trees can be utilized for extraction of classification rules from the data [33]. Decision trees have highly flexible hypothesis space and their theory is easy to understand. Commonly, the obtained results by decision trees are comparable or higher than the outputs of the available methods of classification [34]. In the development of decision trees, there is no need to tune a large number of variables [35].

Decision trees are categorized as non-parametric methods of supervised learning that employ the strategy of divide and conquer. There is no need of assumptions regarding the distributions of the input data. A decision tree is consisting of a root node, internal nodes and terminal (leaf) nodes. Classification in decision tree starts at the root node and each non-leaf node (non-terminal node) asks a question about some features and has N children, where N is the number of possible answers. According to the answer, this process continues to the subtree of one child, while a leaf node is met.

For example, if in a decision tree all features have binary values (in other words all question have "yes" or "no" answer) the algorithm of prediction is like this:

```
Algorithm DecisionTreePredict(tree, test point)
if tree is of the form Leaf(guess) then
    return guess
else if tree is of the form Node(f, left, right) then
    if f = yes in test point then
        return DecisionTreePredict(left, test point)
    else
        return DecisionTreePredict(right, test point)
    end if
end if
```

Table 1 Gathered field data for investigating the sand production from petroleum reservoir

No.	TVD	TT	COH	Q_g	Q_w	BHFP	DD	EOVS	SPF	H_{perf}	Field observed
1	319	105	22	42.3	5672	133.2	27.8	651	4	14	1
2	3182	105	21.9	51.2	68	140.4	16.6	642	4	16	1
3	3366	100	24.7	66.9	157	156.2	18.9	601	4	6	1
4	3647	100	29.6	80.6	85	153.8	57.8	670	8	20	1
5	4548	85	53.2	48	886	209.1	58.9	823	4	18	1
6	4088	85	39.5	72.7	116	147	44	781	2	17	1
7	2100	115	10.8	28.5	724	160.1	8.9	300	4	15.5	1
8	1930	132	9.7	27.5	695	175.5	11.2	245	4	11.5	1
9	2139	112	11.1	36.8	280	185.5	6.1	283	4	10.5	1
10	2380	110	13	23	42	113	47.4	413	6	11	1
11	1122	150	5.7	108	0	107	8	115	12	10.5	1
12	1340	130	6.6	51	52	126.6	14.4	140	12	6.5	1
13	1070	170	5.5	82	70	103.8	0.7	111	4	9	1
14	1920	130	9.6	111	0	248	82	153	4	9	1
15	2530	100	14.3	58	68	302.2	97.8	242	4	4.5	1
16	1640	145	8	94	1260	189	46.8	150	12	10	1
17	2130	120	11	86	112	268.3	31.7	179	4	3.5	1
18	3655	100	19.8	69.8	1780	287.6	9.1	553	4	21	1
19	3668	100	30	75.8	150	272.3	9.2	571	4	21	1
20	1503	125	7.3	139.5	35	152.3	2.2	177	4	11.5	1
21	3170	100	21.7	48	2823	222.1	6.4	485	4	16	1
22	3197	95	22.1	73	273	184.6	48.6	535	2	12	0
23	3230	105	22.6	117	68	210	10	517	27	4	0
24	3684	95	30.3	108.7	36	266.6	59.7	581	1	12	0
25	3005	93	19.5	55	91	67	1	615	33	4	0
26	3790	85	32.5	93.4	77	217.2	124.4	654	8.5	12	0
27	2750	98	16.5	125.8	75	251.8	3	372	8.5	4	0
28	2983	98	19.2	48	28	102	6.1	581	12	4	0
29	3175	100	21.8	30.3	1.698	216.1	17.1	492	20	4	0

The traditional decision trees can be classified into several methodologies as follows:

- Decision trees: such algorithms as Iterative Dichotomiser 3 (ID3) [33] and C4.5 [36] are the examples. ID3 creates a multi-way tree, finding for each node (i.e. in a greedy manner) the categorical feature that will yield the largest information gain for categorical targets. Trees are grown to their maximum size and then a pruning step is usually applied to improve the ability of the tree to generalize to unseen data. C4.5 is very similar to ID3 but features do not need to be categorical. It dynamically defines discrete intervals of continues values to achieve this. C4.5 converts the trained trees into sets of rules in the format of if-then. This accuracy of each rule is then evaluated to determine the order in which they should be applied. Pruning is done by removing a rule's precondition if the accuracy of the rule improves without it [36].
- Fuzzy decision trees: like fuzzy ID3. This algorithm is on the basis of a fuzzy implementation of the aforesaid ID3 method [37-38].
- Oblique decision trees: Classification and Regression Trees (CART) [39] is an example of oblique decision trees. It is similar to C4.5. CART uses a generalization of the binomial variance called the Gini index while C4.5 uses entropy for its impurity function. CART constructs binary trees using the feature and threshold for continue data.

CART algorithm selects the most important and significant parameters and eliminates non-significant parameters and is impervious to transformations that means if some parameters change to their logarithm or square root the structure of tree will not change. CART isolates the outliers in a separate node and finally it can easily handle noisy data. With this advantages and scalable complexity and also ease of implementation, CART method seems a reasonable to choose. Hence, in this work, among various algorithms for decision tree, the CART method was employed for the application of interest.

3.2. Ensemble methodologies

Ensemble methods of classification use multiple different classifiers and aggregate their result by voting instead of using a single classifier. The main principle behind ensemble methods is that a set of weak classifiers can form a strong classifier if they come together in a right way. Every error is possible to make by a minority of classifiers and aggregate theme can achieve the optimal classification. Random Forest and Extra trees (Extremely Randomized Trees) are both ensemble methods that are used in this work. The following sections introduce these types of ensemble approaches.

3.2.1. Random forest

Random Forest is an ensemble of trees grown using some form of randomization and have a structure like the decision tree. In test and prediction phase, an instance is classified by sending it down every tree and aggregating the results.

In growing trees, training data are divided into some random subsets and each tree is grown using a different subset. Also, randomness can be injected in the selection of features at each node to determine the split. A well-known procedure namely Forest-RC is introduced for growing random forest [40]. There is also other algorithms for this purpose in the literature [39,41]. In Forest-RC, the split at each node is based on linear combinations of features rather than one single feature. This allows dealing with cases with only a few inputs supplied and a small set of train data, which is our main motivation to use it. And another advantage is that a Random forest does not over-fit [42] and we can run as many trees as we want.

General training procedure of L trees is as follows:

1. N Sample cases at random with replacement to create a subset of the data. The subset size is 2/3 of all samples.
2. At each node:
   - m predictor variables are selected at random from all the predictor variables.
   - The predictor variable that provides the best split, according to some objective function, is used to do a binary split on that node.

Different m values result in different systems for Random forest: m << number of predictor variables. More details about the theoretical background of random forest classifiers are documented elsewhere [42].

3.2.2. Extra trees

In 2006, Geurts *et al.* [43] introduced the extremely randomized trees. This method is similar to the random forests algorithm in the sense that it is based on selecting at each node a random subset of K features to decide on the split and trees are built using complete samples without partitioning [44]. In other words, the same input training set is used to train all trees. In fact, the difference between random forest and an extra tree is where the randomness injected.

In this method, each tree grows as follows:

- At each node K of random splits including random choice of variable xi, and random choice of threshold t, will be selected and kept among these the one which maximizes the score to grow a tree. More details of score measure are available elsewhere [45].
- Growing of trees continues until all subsamples at all leaves are pure in terms of outputs or there are less than nmin learning samples in them.
- When a tree is grown, each leaf Lj is labeled with a prediction y ˆ_j defined as the local sample average of the output variable, given by (|Lj| is the number of learning cases that reach leaf Lj).

- $$\widehat{y}_j = \frac{1}{|L_j|} \sum_{(x^i, y^i): x^i} y_i \qquad (1)$$

Extra Tree also handles the probable outlier effect on the models induced and its computational complexity is N log N; where N is the size of training data set. There is a more formal description of the algorithm and a detailed discussion of its main features in the literature [43].

4. Development of models

Every time we used decision tree in this work, training was done with CART algorithm described above. Here is CART pseudo-code for GUIDE classification tree construction [39]:

1. Start at the root node.
2. For each ordered variable X, convert it to an unordered variable X by grouping its values in the node into a small number of intervals. If X is unordered, set X = X.
3. Perform a chi-squared test of independence of each X variable versus Y on the data in the node and compute its significance probability.
4. Choose the variable X∗ associated with the X that has the smallest significance probability.
5. Find the split set {X∗ ∈ S∗} that minimizes the sum of Gini indices and uses it to split the node into two child nodes.
6. If a stopping criterion is reached, exit. Otherwise, apply steps 2–5 to each child node.
7. Prune the tree with the CART method.

In Ensemble making, injection of randomness is in the classifier construction and the prediction of the ensemble is given as the averaged prediction of the individual classifiers. This is applied for both Random Forest and Extra Tree classifiers.

For developing the models including CART-decision tree, random forest, and extra tree models, the gathered databank was randomly separated into two sub data sets: training dataset (about 80%) and test dataset (about 20%). The allocated data points for training were employed in the development process. On the other hands, the assigned data points for the test were used for evaluation of the capability of the constructed model in predicting the unseen data.

5. Results and discussion

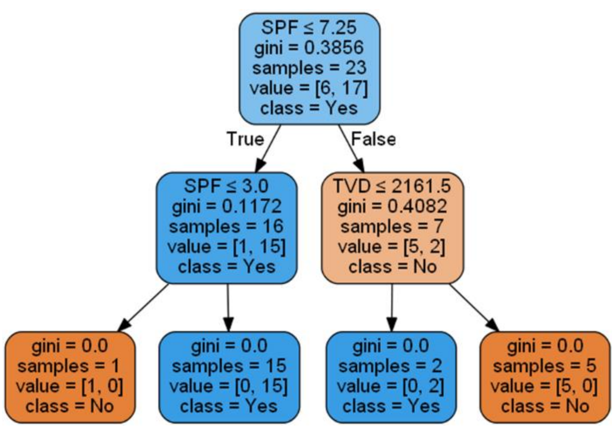


Figure 1. The presented CART model for assessing the conditions of petroleum reservoirs with respect to onset of sand production

As mentioned earlier, 23 data points were used for training and 6 data points were employed as test examples. The obtained model for investigating the sand production status employing CART classifier is shown in Figure 1. Based on the obtained results, the presented CART model reproduces all 23 training samples and 6 test examples same as corresponding field data points. Hence, the accuracy of the proposed CART algorithm is 100%.

The procedure of obtaining the output through the proposed CART model is illustrated using two examples as follows: for the first example, consider the case 22 from Table 1.

According to Figure 1, we start with the root; the SPF of the selected data point is less than 7.25, and this is true. Hence, we go to the left subtree. The SPF is again the criteria in the second stage; SPF of the data is equal to 2 and is less than 3.0 and again this is true. So we go to the left leaf and the class is No. this means there is no sand production. With accordance to the real data, the result is correct.

For the second example, the case 1 from Table 1 is chosen. Start with the root node, the SPF for this case is 4; compare it with the root node, and, in this case, it is less than 7.25. Hence, we go to the left subtree. Again, compare SPF with SPF in this node and is bigger than 3.0; and we go to the right leaf and the class is Yes, i.e. the sand production is observed. According to the real data the predicted result is correct.

As can be seen from Figure 1, among the main parameters including EOVS, BHFP, TVD, TT, DD, COH, SPF, Q_w, Q_g, and H_{perf} only SPF and TVD are considered in the CART model. Indeed, the built CART model is able to forecast the conditions at which sand production is occurred by introducing these two parameters. It may be concluded that these two parameters are the most important variables for investigation of the sand production using CART algorithm.

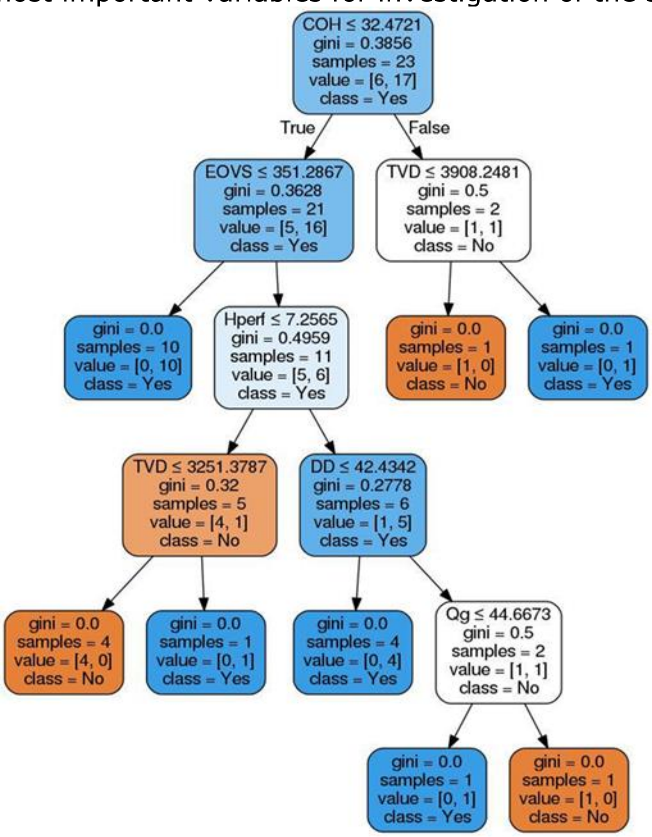


Figure 2. The presented Extra Tree model for assessing the conditions of petroleum reservoirs with respect to onset of sand production

The constructed Extra Tree model for studying the sand production is shown in Figure 2. The proposed Extra Tree model provides the accuracy of 100% for both training and test subdata. As aforesaid before, this type of classifier has extremely randomized behavior. As a result, multiple runs can generate various structures. Similar to the previous model, the procedure of the developed Extra Tree model is clarified using two examples. For the first case, the first data row from Table 1 is selected. The root says that if COH is less than 32.4721, then go to the left subtree; the COH in the first row is 22. Consequently, we are in the left subtree. This stage says if EOVS is less than 351.2867, then go to the left, else go to the right; EOVS is 651. So, we go to the right subtree. In this level, the H_{perf} must be considered. H_{perf} is 14 and is greater than 7.2565; hence, the next node is the root of the right subtree. DD is 27.8 and is less than 42.4342 and we go to the left subtree. Finally, the last leaf is reached and it says that the class is Yes.

As another sample, from Table 1, we chose a case in row 29. COH is 21.8 and is less than 32.4721; so, we go to the left subtree and EOVS is 492 which is greater than 351.2867. Hence, we go to the right subtree; considering the Hperf, we must go to the left subtree. In this stage, we compare the value of TVD; TVD is 3175 and is less than 3251.3787 and is true; and again we go to the left leaf. Based on the model, the class is No.

As it depicted in Figure 1, the presented Extra Tree model considers more parameters as compared to the built CART model. Indeed, in the case of sand production from petroleum reservoirs, decision making via the developed Extra Tree needs the values of EOVS, TVD, DD, COH, Q_g , and H_{perf}.

The third algorithm for studying the sand production is based on Random Forest. The developed Random Forest model consists of 10 weak classifiers. The forest aggregates their votes to predict the final result. All trees of the proposed forest are shown in Figure 3.

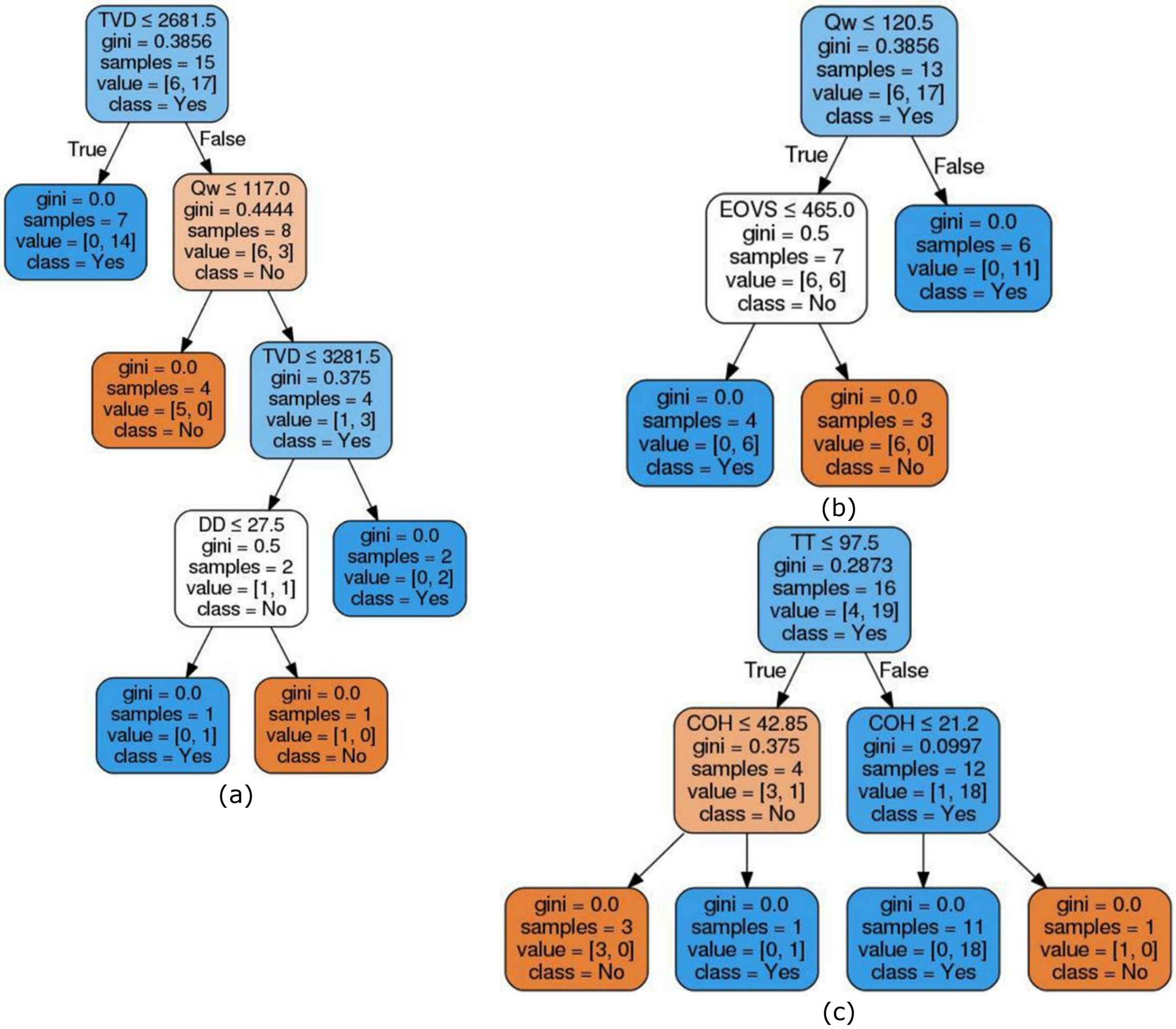


Figure 3. The trees of the presented Random Forest model for assessing the conditions of petroleum reservoirs with respect to onset of sand production (a-c)

For showing the procedure of using the developed Random Forest algorithm, Figure 3 and data in row 8 from Table 1 were employed. The root of Figure 3a says that compare TVD; and the TVD number in this case is 1930 and is less than 2681.5. Hence, we go to left and class is Yes. Based on the next tree of the presented Random Forest model, shown in Figure 3b, if Q_w is less than or equal to 120.5, the left is selected. In the data, Q_w is 724 and is greater than 120.5; this means that the right leaf must be selected. Finally, the class is Yes. Pursuing the procedure, the class obtained by remaining trees of the proposed Random Forest model, shown in Figure 3c to 3j, is Yes. As explained, all the trees in the model present the class of Yes. Hence, the final result of the built Random Forest model for the selected data shows that the sand production is observed.

Demonstrated in this section, all the presented models in this work including CART, Extra Tree, and Random Forest models are capable to predict the right conditions for sand production phenomenon. However, in view of the fact that the available databank for modeling purpose is not extensive and the number of test samples are not adequately enough, the presented models may have low generalization capability on the new and unseen data points.

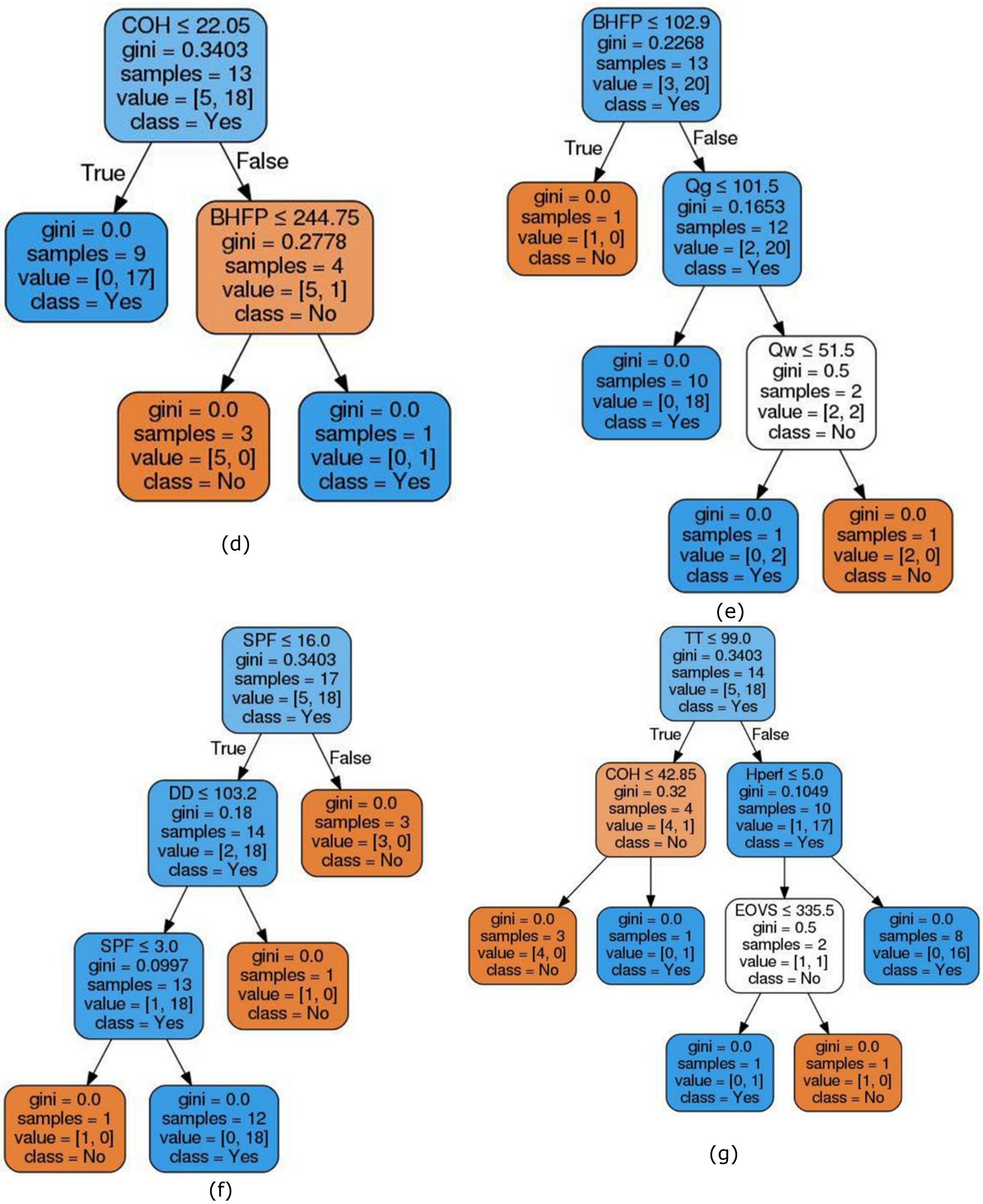
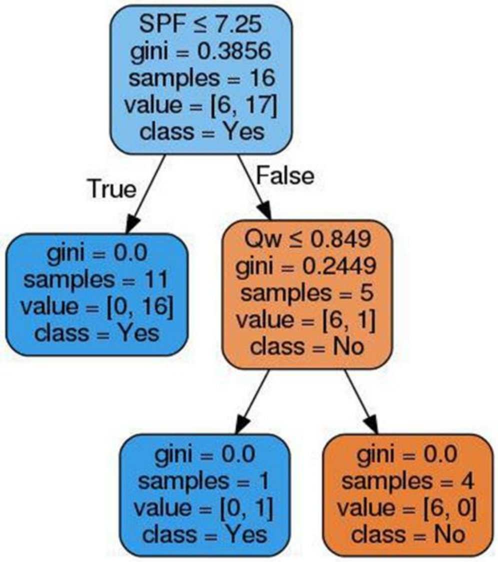
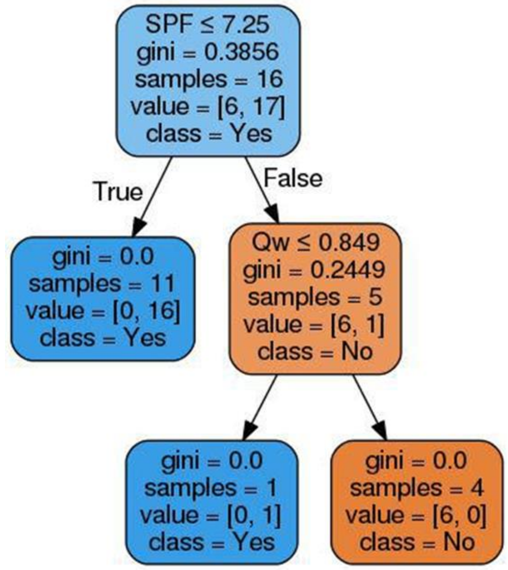


Figure 3. The trees of the presented Random Forest model for assessing the conditions of petroleum reservoirs with respect to onset of sand production (d-g)

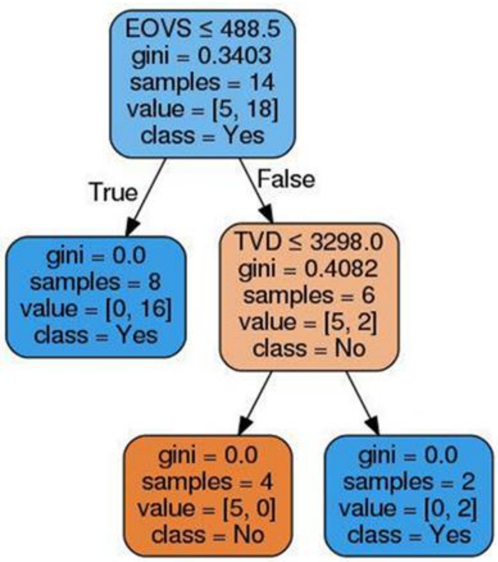
Further to the above, since the CART classifier considers only two parameters, it may have the minimum generalization ability amongst the presented models. It should also be noted that introducing a databank with more reliable data points to the aforementioned algorithms may result in obtaining different structures for the models.



(h)



(i)



(j)

Figure 3 The trees of the presented Random Forest model for assessing the conditions of petroleum reservoirs with respect to onset of sand production (h-j)

From a computational complexity point of view, the CART classifier is the simplest model and the Random forest algorithm is the most complex model. However, in view of the fact that the Random Forest approach is not been over fitted, as explained before, it is suggested to use this model for the application of interest. For predicting the sanding through the proposed Random Forest model, all the investigated parameters are required.

6. Conclusion

Presented in this communication, CART-Decision Tree, Extra Tree, and Random Forest algorithms were utilized for developing models capable to determine the possible sand production in a real petroleum field. To achieve this goal, the models were developed employing a total number of 23 field data points. The performance of the constructed models in classifying the unseen data was evaluated using 6 field data.

Based on the obtained results, all the proposed models have excellent classification power in identifying the petroleum reservoir conditions with respect to sanding. Indeed, the built models provide 100% accuracy for the application of interest. It should be noted that more robust models can be developed if more reliable data points are available. The proposed methodologies in this study can be implemented for facilitating the strategies for sand control plan. However, as it described in the previous section of the work, it is recommended to employ the Random Forest algorithm for investigating the sand production in petroleum reservoirs.

The proposed techniques in this work can play a vital role in the investigation of sand production in various oil and natural gas industrial applications. Indeed, utilizing the presented algorithms will pave the way for sand-control decision making. Consequently, secure energy production can be achieved.

Acknowledgement

The first author acknowledges the help of Mr. Farid D. Maleki from Kharazmi University, Tehran, Iran and Mr. Mohammad Mahmoudi from TU Wien, Wien, Austria who have made a contribution to improve the paper.

References

[1] International energy outlook 2013 Washington, DC 20585, U.S. Energy Information Administration.
[2] Ghiasi MM, Arabloo M, Mohammadi AH, and Barghi T. Application of ANFIS soft computing technique in modeling the CO_2 capture with MEA, DEA, and TEA aqueous solutions. Int J Greenhouse Gas Control, 2016; 49: 47-54.
[3] Bianco LCB, and Halleck PM. Mechanisms of Arch Instability and Sand Production in Two-Phase Saturated Poorly Consolidated Sandstones. Society of Petroleum Engineers 2001.
[4] Nouri A, Vaziri H, Belhaj H, and Islam R. Comprehensive Transient Modeling of Sand Production in Horizontal Wellbores. Society of Petroleum Engineers 2003.
[5] Nouri A, Vaziri H, Kuru E and Islam R. A comparison of two sanding criteria in physical and numerical modeling of sand production. J Petr Sci Eng., 2006; 50(1): 55-70.
[6] Ranjith PG, Perera MSA, Perera WKG, Wu B, and Choi SK. Effective parameters for sand production in unconsolidated formations: An experimental study. J Petr Sci Eng., 2013; 105: 34-42.
[7] Ranjith PG, Perera MSA, Perera WKG, Choi SK, and Yasar E. Sand production during the extrusion of hydrocarbons from geological formations: A review. J Petr Sci Eng., 2014; 124: 72-82.
[8] Almedeij JH, and Algharaib MK. Influence of sand production on pressure drawdown in horizontal wells: Theoretical evidence. J Petr Sci Eng., 2005; 47(3–4): 137-145.
[9] Talaghat MR, Esmaeilzadeh F.and Mowla D. Sand production control by chemical consolidation. J Petr Sci Eng., 2009; 67(1–2): 34-40.
[10] Willson SM, Moschovidis ZA, Cameron JR and Palmer ID. New Model for Predicting the Rate of Sand Production. Society of Petroleum Engineers 2002.
[11] Wang J, Settari AT, Walters D and Wan R. An integrated modular approach to modeling sand production and cavity growth with emphasis on the multiphase and 3D effects. American Rock Mechanics Association 2006.
[12] Fattahpour V, Moosavi M and Mehranpour M. An experimental investigation on the effect of rock strength and perforation size on sand production. J Petroleum Sci Eng., 2012; 86–87: 172-189.
[13] Sanfilippo F, Ripa G, Brignoli M and Santarelli FJ. Economical Management of Sand Production by a Methodology Validated on an Extensive Database of Field Data. Society of Petroleum Engineers 1995.
[14] Moricca G, Ripa G, Sanfilippo F and Santarelli FJ. Basin scale rock mechanics: Field observations of sand production. Society of Petroleum Engineers 1994.
[15] Hall CD Jr., and Harrisberger WH. Stability of Sand Arches: A Key to Sand Control. J. Pet. Technol., 1970; 22(7):
[16] Tippie DB, and Kohlhaas CA. Effect of Flow Rate on Stability of Unconsolidated Producing Sands. Society of Petroleum Engineers 1973.
[17] Bratli RK, and Risnes R. Stability and Failure of Sand Arches. Society of Petroleum Engineers Journal, 1981; 21(2):236-248 1981.
[18] Morita N, Whitfill DL, Fedde OP and Levik TH. Parametric Study of Sand-Production Prediction: Analytical Approach. 1989; SPE Production Engineering 4 (01), 25-33.
[19] Morita N, and Boyd PA. Typical Sand Production Problems Case Studies and Strategies for Sand Control. Society of Petroleum Engineers 1991.
[20] Nie X, Yang S, Ding J, Cao L, Zhou F, Ma Q and Qiu Z. Experimental investigation on permeability evolution law during sand production process of weak sandstone. J Nat Gas Sci Eng., 2014; 21: 248-254.

[21] Ghassemi A, and Pak A. Numerical simulation of sand production experiment using a coupled Lattice Boltzmann–Discrete Element Method. J Petr Sci Eng., 2015; 135: 218-231.
[22] Jiang N-J, Yoshioka H, Yamamoto K and Soga K. Ureolytic activities of a urease-producing bacterium and purified urease enzyme in the anoxic condition: Implication for subseafloor sand production control by microbially induced carbonate precipitation (MICP). Ecol Eng., 2016; 90: 96-104.
[23] Coates GR, and Denoo SA. Mechanical Properties Program Using Borehole Analysis And Mohr's Circle, Society of Petrophysicists and Well-Log Analysts. SPWLA 22nd Annual Logging Symposium, 23-26 June 1981, Mexico City, Mexico
[24] Weingarten JS, and Perkins TK. Prediction of Sand Production in Gas Wells: Methods and Gulf of Mexico Case Studies. Journal of Petroleum Technology, 1995; 47(7):596-600.
[25] Doan QT, Doan LT, Ali SMF and Oguztoreli M. Sand Deposition Inside a Horizontal Well-A Simulation Approach. Annual Technical Meeting, June 8 - 10, 2000,Calgary, Alberta
[26] Kanj MY, and Abousleiman Y. Realistic Sanding Predictions: A Neural Approach, Society of Petroleum Engineers 1999.
[27] Azad M, Zargar G, Arabjamaloei R, Hamzei A and Ekramzadeh MA. A New Approach to Sand Production Onset Prediction Using Artificial Neural Networks. Petr Sci Tech., 2011; 29(19): 1975-1983.
[28] Khamehchi E, Rahimzadeh Kivi I and Akbari M. A novel approach to sand production prediction using artificial intelligence. J Petr Sci Eng., 2014; 123: 147-154.
[29] Gharagheizi F, Mohammadi AH, Arabloo M, and Shokrollahi A. Prediction of Sand Production Onset in Petroleum Reservoirs Using a Reliable Classification Approach. Petr., 2017; 3(2): 280-285.
[30] Bishop C. Pattern Recognition and Machine Learning. Singapore. Springer 2006.
[31] El-Telbany ME, and Warda M. (2016) An Empirical Comparison of Tree-Based Learning Algorithms: An Egyptian Rice Diseases Classification Case Study. Int J Adv Res Art Intel 5(1): 22-26.
[32] Zhao Y, and Zhang Y. Comparison of decision tree methods for finding active objects. Adv Space Res., 2008; 41(12): 1955-1959.
[33] Quinlan JR. Induction of decision trees. Mach Learn, 1986; 1(1): 81-106.
[34] Lim T, Loh W and Shih Y. An empirical comparison of decision trees and other classification methods. Technical Report, Department of Statistics 1997, University of Wisconsin.
[35] Gehrke J, Ganti V, Ramakrishnan R.and Loh W-Y. BOAT-optimistic decision tree construction. Proceedings of the 1999 ACM SIGMOD international conference on Management of data. Philadelphia, Pennsylvania, USA, ACM: 169-180.
[36] Salzberg SL. C4.5: Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., Mach Learn, 1993; 16(3): 235-240.
[37] Weber R. Fuzzy ID3: a class of methods for automatic knowledge acquisition. Proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks 1992, Iizuka, Japan.
[38] Umano M, Okamoto H, Hatono I, Tamura H, Kawachi F, Umedz S and Kinoshita J. Fuzzy Decision Trees by Fuzzy ID3 algorithm and its Applications to Diagnosis System. Proceedings of the third IEEE Conference on Fuzzy Systems 1994, Orlando.
[39] Breiman L, Friedman J, Stone CJ, and Olshen RA. Classification and Regression Trees, Taylor & Francis 1984.
[40] Breiman L. Random Forests. Mach Learn, 2001; 45(1): 5-32.
[41] Robnik-Sikonja M. Machine Learning. ECML 2004 Proceedings, Berlin, Springer.
[42] Breiman L. RANDOM FORESTS--RANDOM FEATURES. CA, Statistics Department, University of California Berkeley 1999.
[43] Geurts P, Ernst D and Wehenkel L. Extremely randomized trees. Mach Learn, 2006; 63(1): 3-42.
[44] Geurts P, and Louppe G. Learning to rank with extremely randomized trees, JMLR: Workshop and Conference Proceedings, 2011; 14: 49-61.
[45] Wehenkel L, Ernst D and Geurts P. Ensembles of extremely randomized trees and some generic applications. RTE-VT workshop 2006. Paris.

To whom correspondence should be addressed: Professor Dr. Amir H. Mohammadi, Discipline of Chemical Engineering, School of Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa