Clustering as an efficient tool for assessing fluid content and movability by resistivity logs

Pedram Masoudi, Babak Nadjar Araabi, Tahar Aïfa, Hossein Memarian

4th International Mine & Mining Industries Congress & Expo & 6th Iranian Mining Engineering Conference
October 2016 - Tehran

Clustering as an efficient tool for assessing fluid content and movability by resistivity logs

Pedram Masoudi[1,3], Babak Nadjar Araabi[2], Tahar Aïfa[1*], Hossein Memarian[3*]

1* : Géosciences-Rennes, CNRS UMR6118, Université de Rennes 1, Bat.15, Campus de Beaulieu, 35042 Rennes cedex, France, tahar.aifa@univ-rennes1.fr and pedram.masoudi@univ-rennes1.fr

2 : School of Electrical and Computer Eng., Univ. of Tehran, Tehran, Iran, araabi@ut.ac.ir

3 : School of Mining Eng., Univ. of Tehran, Tehran, Iran, memarian@ut.ac.ir and masoudip@ut.ac.ir

Abstract

Neither core measurements nor well tests provide precise measurement of fluid contents; so, at the moment there is no validated saturation measurement in the oil industry. However, resistivity logs contain some valuable information about reservoir fluids, and classes of saturation based on the behaviour of resistivity data. In order to extract saturation value, clustering algorithms are proposed and tested here. Clustering is an unsupervised categorization method, which relies on natural groupings of real data, instead of predefined labels. Distance or dissimilarity plays an important role in the formation of clusters in an algorithm. The application of three clustering algorithms on prediction of water saturation is discussed here. It is shown that Fuzzy C-Means clustering divides data only according to saturation property; while the Gustafson-Kessel algorithm considers not only saturation but also permeability. The reason is that Gustafson-Kessel can detect linear patterns. This algorithm is introduced as the most appropriate clustering method for predicting permeability, and for understanding the reservoir quality of the formation under examination. Gath-Geva clustering did not provide as much information as Gustafson-Kessel. In addition to these three clustering algorithms, two other methods were likewise checked, but they did not provide acceptable interpretations, so not reported. Another achievement of this study is the introduction of a cluster label instead of a single value which is very unreliable for expressing saturation and permeability simultaneously. Predictions of these two petrophysical properties are provided merely by two conventional resistivity well-logs: deep and shallow. The output of cluster analysis is much closer to well-scale reservoir properties as compared with that of core-scale properties, since clustering algorithms are applied to logs that are volumetric recordings. The applicability and efficiency of the proposed methods are examined and verified through the use of well-logs of Sarvak Formation in an anticlinal oil field in the Abadan Plain, Iran.

Keywords: water saturation in carbonate reservoirs; clustering permeable zones; clustering saturated zone; application of unsupervised learning; Gustafson-Kessel clustering

1. INTRODUCTION

The importance of evaluating water saturation in petroleum exploration is taken for granted by petroleum geoscientists. The widest usage of water saturation is in net pay detection. Petrophysicists apply cut-off values to water saturation logs, to distinguish oil-bearing intervals in both horizontally [1] and vertically drilled wells [2, 3, 4 and 5]. Due to the effect of water saturation on the plasto-elastic properties of rocks, accurate determination of this property is essential in geomechanics [6]. In addition, Tokhmechi et al. (2009) have used water saturation log in the process of fracture detection within wells [7]. Production planning is another criterion which water saturation is essential for [8].

Still, empirical relations for estimating water saturation are the most widely used methodologies. Among them, Archie relations, developed by early Archie works [9, 10, 11 and 12] are the best-known. There are so many other empirical relations which have also been developed, such as Carman [13] and Timur [14]. For the comprehensive study of empirical methods in estimating water saturation, please refer to [15]. Worthington (2004) offered a wise fit-for-purpose usage of Archie and non-Archie relations due to rock type and our purpose of estimating water saturation [16].

There are some other methodologies, based on supervised methods and trainable machines, which have recently been developed. To use these supervised methods, a reference criterion is required to train a machine. For example, Kadkhodaie-Ilkhchi et al. (2009) proposed a Committee Machine with Training Algorithm (CMTA), utilizing a combination of Levenberg-Marquardt, Bayesian regularization, gradient descent, one step secant and resilient back-propagation to train a neural network for predicting Normalized Oil Content (NOC) [17].

The Bayesian classifier can be applied, while using Gassmann equation as labels [18]. Also, Radial Basis Function Neural Networks [19] and support vector

* Corresponding author

regression machine [20] could be trained according to core measurements on saturation as reference criteria. It seems that these trainable (supervised) methodologies are more reliable as we have evidence for measuring precision of output.

Barros and Andrade (2013) predicted water saturation by angular competitive neural network. In this methodology there is no training, and the algorithm only searches for patterns of cross-plots to predict a saturation parameter. This algorithm is really applicable and useful in cases in which we lack labelled data as validation [21]. We consider the current manuscript as complementary to [21], since here, the proposed approach searches for patterns without considering labelled data for machine learning. In this study, a review is addressed on previous works regarding water saturation. Then, some clustering-methodologies are introduced as possible solutions to the problem of saturation prediction. Finally, outputs of different cluster analysis methods in studying water saturation are discussed.

As mentioned above, the industrial importance of this work is its applicability in cases where there is no evidence for validating outputs. Still, there is no reliable continuous saturation measurement within drilled wells. It is noteworthy that the experience and wisdom of petrophysicists is essential for a successful interpretation, and cluster-based methodology is only a powerful tool in distinguishing similar data due to different criteria of similarity.

2. HYPOTHESES: WHY CLUSTER ANALYSIS FOR SATURATION PREDICTION?

As discussed previously, there are several saturation prediction methods. All of them provide an indirect procedure for predicting water content in each horizon, using well-log data. Nearly all saturation prediction methods on well-log data assign a single value (not a range) of water content to each horizon through wells. We believe that such a fixed saturation value is not a credible way of introducing this variable in oil wells due to three issues:

(i) A fixed and single value is not an appropriate representative of the volume of investigation regarding recorded measurements within the well. In other words, the certainty and accuracy of a single value for expressing saturation of a volume is subject to doubt. Experts are keen to have a single value as the value of saturation; but, here a range of values, instead of a single value for saturation is recommended as to render the predictions more reliable. When we have vagueness in the parameter under study, we have to use a vague language too. In other words, at the moment, science and technology do not let us express water saturation by a fixed value. It is just an approximation without any error measurement!

(ii) If we can train a machine in a specific environment, in this case a well which is drilled in a reservoir, it is more accurate to use the trained machine in evaluating water saturation instead of empirical universal equations. So, the use of empirical relations such as Archie is not recommended when we are able to use trainable and locally available machines. In addition, empirical relations have only been developed in specific environments, assumptions and considerations. For example, the Archie equation was primarily developed in sandy reservoirs, but if it is addressed in shale-contaminated reservoirs, some corrections have to be considered. Therefore, the Archie relation in carbonate reservoirs is only a rough approximation of water saturation.

(iii) Using supervised methods in saturation prediction is not accurate enough at the moment, because we do not have any reliable evidence or label for training the machine. Evidences for training have to be either core measurements or well-tests. Training a machine based on the label of water saturation is not an accurate job due to the following points:

(a) Whereas each core plug could be tested several times at several intervals (with a vertical resolution of about centimeter) for porosity measurements; it could only be used once for measuring water saturation. Therefore it is impossible through cores to get real, accurate data for saturation with high vertical resolution. Also, measuring water saturation in laboratory conditions results in different outputs, compared to the in-situ situation. Water saturation is too dependent on the place of analysis (because of environmental causes such as temperature, pressure, etc.); i.e. temperature and pressure of reservoir conditions differ greatly from laboratory conditions. As we recover the core from the reservoir, its pressure would be dropped, and the balance between fluid contents would be violated. Even in measuring porosity, in-situ and laboratory measurements differ somewhat due to stress relief, but this variation boosts in measuring saturation property, since dynamic processes take part.

(b) Surprisingly, well-tests cannot provide a practical label of saturation. Two issues are under doubt in well-test results as labels of supervised classification of water saturation: (b1) the very first issue is that wells are tested at intervals, not horizons. Therefore, when we do not have more than one test in a specific reservoir interval through a well, we would not have more than an average value for that specific interval, which is not enough for training a machine. In order to be able to train a machine in the highly heterogeneous conditions of the reservoirs, we require much more than 100 observations, while by well testing; we always obtain less than 10 observations, i.e. well test intervals. (b2) Besides, the nature of well-test results differs from the characters of water saturation, i.e. a well-test result is a combination of water saturation and reservoir permeability. The effect of permeability has to be removed from the testing results if we want to use the tests as a measurement of saturation. This incompatibility between well-test output and saturation value is shown in Figure 1. As it is obvious, well-test class numbers 2 and 3 (i.e. oil production lower and higher than 1500 barrel per day) show the reverse order of the expected pattern, compared to water saturation value [9].

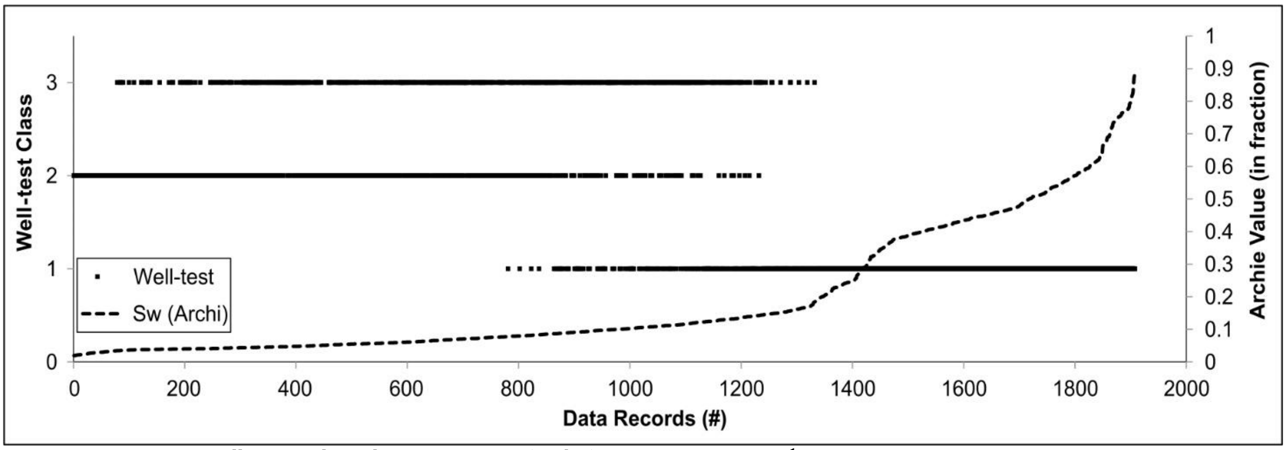


Figure 1. Comparing well-test results and water saturation (Archie) value in five wells of Sarvak Formation. Horizontal axis is a simple data numbering, which makes saturation values in ascending order. Well-test class value means water production (1), oil production with a rate of less than (2) or of higher than (3) 1500 barrel per day.

Therefore, in predicting water saturation, we face a problem: there is not enough credible measurement for saturation. To cope with this situation, we believe that the assessment has to be done by an unsupervised method able to analyze the behavior of datasets according to the target (here, saturation). So, cluster analysis methods are recommended and investigated here as unsupervised clustering methods to evaluate the water saturation of reservoir rocks.

3. DATASETS

Datasets of this work belong to the Sarvak Formation in an anticlinal (north-south trend) oil field in the Abadan Plain, Iran. For the sake of confidentiality, the name of the field is not disclosed. There are six drilled wells for exploratory purposes, and well-log data is available through all the wells. But well-test and core data are not available at all. The summary of the available information is presented in Table 1.

From the standpoint of sequence stratigraphy, the Sarvak Formation is a shallowing upward carbonate reservoir, deposited on a shallow Upper Albian to Upper Turonian platform. It is conformably laid on the Kazhdumi Formation, and the Laffan Formation is overlaid on the Sarvak Formation above this sharp erosional unconformity [22].

4. METHODOLOGY

A. Archie Equation

The famous empirical Archie equation was first developed in sandy reservoirs; however, it is used in carbonate rocks too. Two general forms of this relation are [9]:

The famous empirical Archie equation was first developed in sandy reservoirs; however, it is used in carbonate rocks too. Two general forms of this relation are [9]:

$$R_t = R_o \times S_w^{-n} \quad (1)$$

$$S_w = \left(\frac{R_o}{R_t}\right)^{\frac{1}{n}} \quad (2)$$

where, R_t is the recorded resistivity of rock, filled with fluids at reservoir conditions. R_o is the resistivity of sand when all the pores are filled with brine; S_w is the water saturation, in fraction form; and n is the saturation exponent. The saturation exponent for consolidated sands and clean unconsolidated sand appears to be close to 2. Hence, the equation for clean sand could be simply rewritten as:

$$S_w = \sqrt{\frac{R_o}{R_t}} \quad (3)$$

R_t is the resistivity measured by the logging instrument at each depth. Therefore, by having R_o, we can easily estimate water saturation by relation 3. Here, R_o is considered as the minimum value of R_t through the well. The logic behind choosing the minimum of R_t as the value of R_o could be stated through two facts: (i) Resistivity logs are mostly affected by reservoir fluids; therefore, they are measures of reservoir fluids. (ii) The resistivity of water is the lowest amongst other reservoir fluids.

A. Fuzzy Clustering

Finding available structures (clusters) within datasets is the main goal of clustering algorithms. The problem of finding several cluster centers that are appropriately representative of relevant classes of X, where X is a finite set of data [23]. The word "relevant classes" in the above sentence shows the impreciseness in defining clusters. In other words, the sentence mentioned is a general definition, and should be specified by each clustering algorithm. Hartigan and Wong, (1979) have defined k-means clustering as: "The aim of the k-means algorithm is to divide M points in N dimensions into K clusters so that the within-cluster sum of squares is minimized" [24].

In this study, three fuzzy clustering algorithms are applied to datasets to compare differences between the outputs of these algorithms. Also to find the most suitable one for the purpose of studying water saturation properties among Fuzzy C-Means (FCM), Gustafson-Kessel (GK) and Gath-Geva (GG) algorithms. The application of two other cluster analysis methods, Affinity Propagation (AP) and Spectral clustering, is also checked. The results of neither of the latter algorithms are included since their outputs were not well-interpretable.

TABLE I. DATASET AVAILABLE FROM TWO WELLS OF THE OIL FIELD. "NPV" STANDS FOR "NET PAY VALUE". NPV=1 MEANS THAT THE INTERVAL OF TESTING IS NOT OIL PRODUCING, NPV=2 AND 3 MEAN THAT THE TESTED INTERVAL PRODUCES OIL LESS AND MORE THAN 1500 BARREL PER DAY (IN SI UNITE SYSTEM) RESPECTIVELY. AVAILABLE AND UNAVAILABLE INFORMATION ARE SHOWN BY SYMBOLS (✓) AND (×), RESPECTIVELY. FURTHER INVESTIGATION IS LIMITED TO AVAILABLE LOGS ONLY.

		Well 1	**Well 2**	**Well 3**	**Well 4**	**Well 5**	**Well 6**
No. of Well Test Intervals	npv=1	3	3	4	1	0	1
	npv=2	3	0	1	1	0	0
	npv=3	0	1	1	1	0	1
Petrophysical Well Logs	Caliper (CALI)	✓	✓	×	✓	✓	✓
	Gamma Ray (GR)	✓	×	✓	✓	✓	✓
	Corrected Gamma Ray (CGR)	✓	✓	✓	✓	✓	✓
	Sonic Log (ΔT)	✓	✓	✓	✓	✓	✓
	Neutron Porosity (NPHI)	✓	✓	✓	✓	✓	✓
	Bulk Density (RHOB)	✓	✓	✓	✓	✓	✓
	Density Correction (DRHO)	✓	×	✓	✓	×	×
	Deep Laterolog Resistivity (LLD)	✓	✓	✓	✓	✓	✓
	Shallow Laterolog Resistivity (LLS)	✓	✓	✓	✓	✓	✓
	Microspherically Focused Log (MSFL)	✓	✓	✓	✓	✓	✓
	Photoelectric Effect Log (PEF)	✓	×	×	✓	✓	×
Core Tests	Porosity	✓	✓	✓	✓	✓	×
	Permeability	✓	✓	✓	✓	✓	×
	Water Saturation	Not Reliable	Not Reliable	Not Reliable	Not Reliable	Not Reliable	×

Fuzzy C-Means (FCM) clustering is an extension of hard C-means (or k-means) clustering. The main difference between FCM and hard C-means is in a fuzzy weight, powered by a weighting exponent, which is multiplied by Euclidian distance values in cost function. FCM is not an optimum clustering algorithm for oriented or multi-size clusters. The FCM algorithm is introduced in four main steps in the literature [23, 25]:

(1) Define fixed number of clusters, C, fuzziness, m, and norm type (i.e. distance or similarity function). Also randomly create similarity matrix, U, for all cluster centers.

(2) Compute the average value of similarities of all C clusters using the following equation:

$$d(\underline{x}^i, w_j) = \frac{\sum_{k=1}^{n} u_{kj}^{m} \|\underline{x}^i, w_j\|^2}{\sum_{k=1}^{n} u_{kj}^{m}};\ j \in [1, c]; i \in [1, n] \qquad (4)$$

(3) Compute an updated membership matrix by:

$$u_{kj} = \frac{1}{\sum_{j=1}^{c} \left[\frac{d(\underline{x}^i, w_l)}{d(\underline{x}^i, w_j)}\right]^{\frac{1}{m-1}}};\ l \in [1, c]; i \in [1, n] \qquad (5)$$

(4) Comparing Uk with Uk+1. If $\|U^k, U^{k+1}\| \leq \varepsilon$, stop; otherwise set $U^k = U^{k+1}$, and go back to stage (2).

The core difference between GK and FCM is in the use of the Mahalanobis distance, instead of the Euclidian one in cost function (norm in relation 4), which results in the recognition of oriented clusters [26, 27]. Therefore, GK tends to recognize oriented (linear at extreme mode) clusters. As in FCM, the identified clusters of GK are of the same size in terms of number of data, i.e. the number of data in each cluster is about the same.

The limitation of multi-size clusters is solved in the GG algorithm. The definition of cluster means and covariance matrix are derived from FCM and GK algorithms, respectively [28, 29]. The novelty of GG that results in solving the limitation of multi-size clusters is in defining distance based on decision surface. For example, a small blue cluster, a medium-sized green cluster and a large black cluster are detected by the GG algorithm in a single run (Figure 5c,d). The algorithm is much the same as FCM; and for detailed study, please refer to [29].

5. INPUT SELECTION FOR CLUSTERING

Clustering algorithms do not consider any class label for data and only search the optimum clusters within input datasets. Therefore, selected features/datasets are very important in a clustering project. A logical interrelation between petrophysical parameters, inspired from [11] is depicted in Figure 2.

Five features are related directly to saturation property: capillary pressure, permeability, fluid relief, fluid-solid interaction and electrical response. Four of the features are mostly categorized as dynamic properties, and usually we do not have any accurate estimation of these parameters, especially in exploratory stages. Only resistivity has a reliable and continuous recording throughout the reservoir. In the datasets of the current work, we have three resistivity log recordings: Deep Laterolog Resistivity (LLD), Shallow Laterolog Resistivity (LLS) and Microspherically Focused Log (MSFL). By trial and error, we found that removing MSFL provides better clustering results; therefore, two resistivity logs (LLS and LLD) are selected as input features of clustering algorithms. Cross plots of LLD to LLS show upper and lower bounds of data (Figure 3a), and the area of concentrated data (Figure 3b).

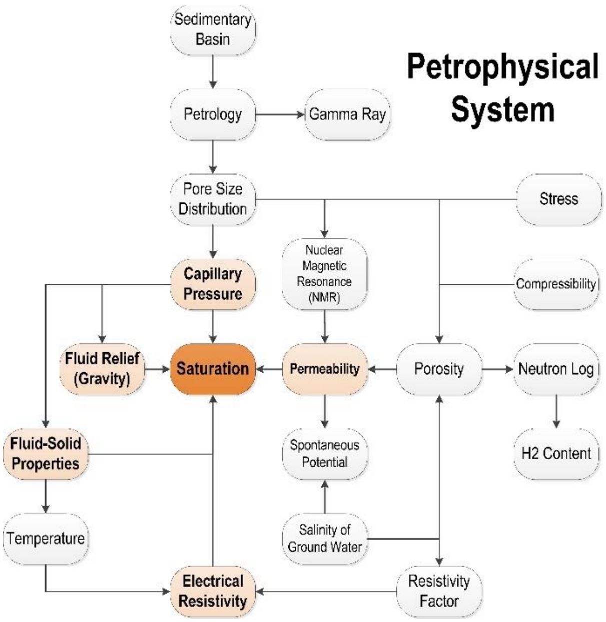


Figure 2. Interrelations between petrophysical features. Water saturation and directly related features are highlighted by color and bold fonts (modified after [11]).

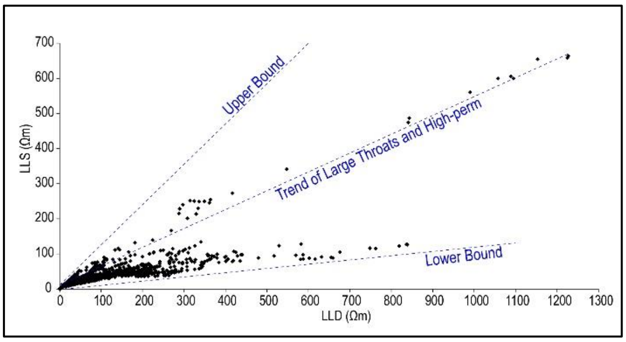
6. LLD TO LLS, A MEASURE OF PERMEABILITY

In Figure 3b, we have interpreted two arrows: (i) the direction of increasing water saturation, which is downward, and is taken for granted by researchers; (ii) the direction of increasing permeability, which is the main questionable issue of this figure. The interpretation (ii) is inspired from a previous publication, for which the ratio of LLD to LLS is used to relate invasion to permeability [30].

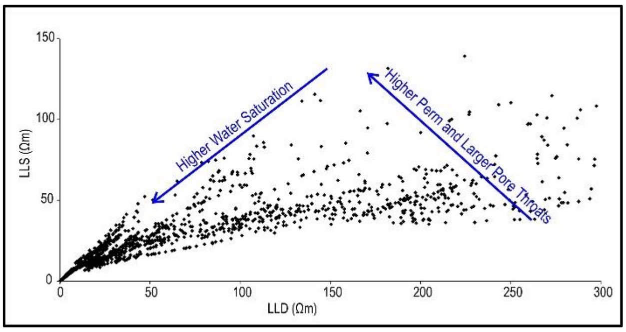
It is not far from our intuition of the reservoir conditions that the more permeability around the well, the more uniform the fluid distribution outward from the well-axis. Here, LLS is representative of near-well conditions (including infiltration), compared to LLD which represents rather a larger zone around the wellbore. Therefore, in an extreme situation that recorded value of LLD is equal to LLS (i.e. ratio of LLD to LLS is one), the high permeability of the reservoir resulted in a mixture of reservoir fluids and drilling mud effectively around the well. In the other extreme case (i.e. LLD>>LLS), the mud cake and infiltration zone will be easily constructed due to the insufficiency of mixing drilling mud particles with reservoir fluids.

Besides all analytical descriptions, we have checked the correctness of this hypothesis (i.e. correlation between permeability and the ratio of LLD to LLS) in wells number one to five. This is fundamental to find the applicability of our analytics in real conditions and datasets. So, we plotted a logarithmic scale of permeability vs. the ratio of LLD to LLS in the five above-mentioned wells to investigate correlations (Figure 4). It is noteworthy that permeability is a lognormal property in most reservoir conditions.

In the visual assessment of Figure 4, the correlation between permeability and LLD:LLS, on a logarithmic scale could be easily found in wells two, three and five. There is no correlation in wells number one and four; however it does not mean that LLD:LLS is not a good measure of permeability in these two wells. This is because core permeability only shows permeability of intact rock on a small scale, without considering fractures and joints, whereas well-logs are volumetric responses and could be better representatives of real heterogeneous reservoir conditions.



(a)



(b)

Figure 3. Cross plots of input features. Trends show interpretation of petrophysical interrelations. (a) Whole the data, and (b) focused on dense area. Arrows show interpretation of permeability, pore throat size and saturation on cross plot.

It is worth mentioning that from a structural viewpoint, well number four is drilled on a deviation of the axis of the anticline from its normal trend, i.e. the trend of the anticline changes from normal, which is NS, to another trend, which is NW-SE, at the emplacement of well number four. Therefore, lots of fractures and joint sets are probably to be found within well number four, thus increasing the permeability of the reservoir, while the permeability of the intact rock (cores) is unchanged. Intact rock properties are not controlled by structural geology. These properties are controlled by sedimentary processes and lithological properties. All in all, there is a correlation, however not that strong between LLD:LLS and permeability on the reservoir scale.

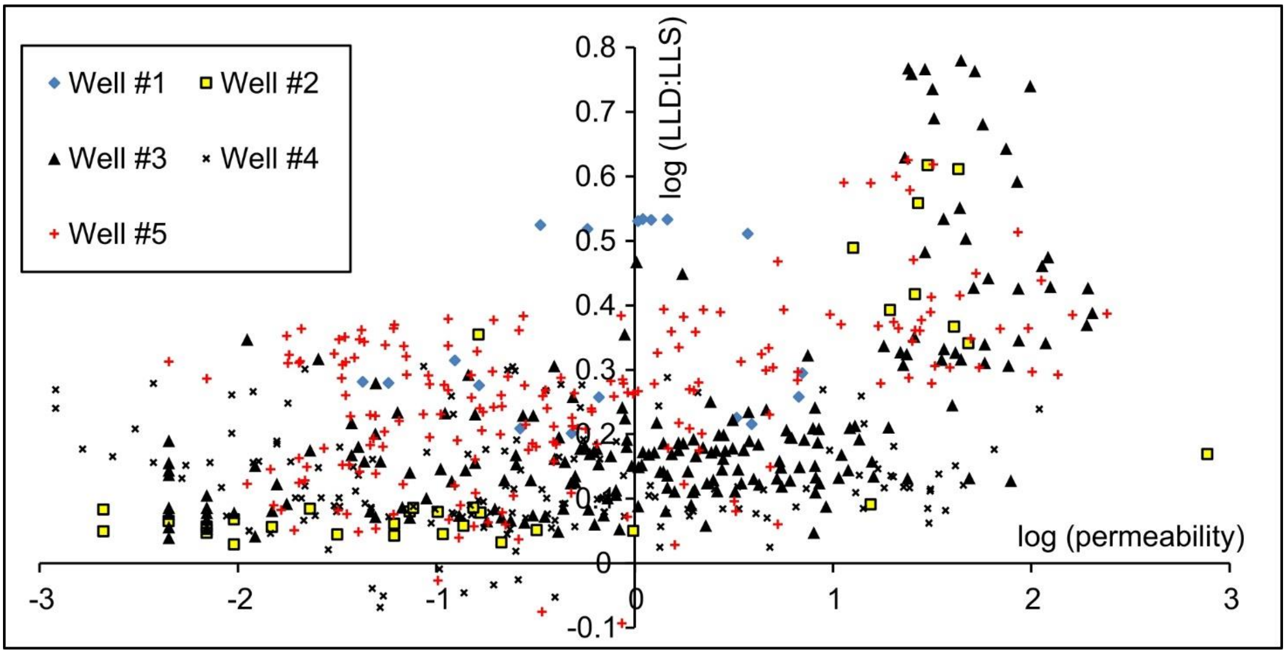


Figure 4. Checking correlation of permeability (mD) and LLD (Ω.m) to LLS (Ω.m) in five wells, logarithmic scale.

7. Results: Output of Clustering Methods

Clustering algorithms, FCM, GK and GG were run on the datasets (Figure 3). The number of clusters is set to be 3, 5, 10 and 20 to find the optimum number of clusters. The optimum number of clusters is found to be 5 for all mentioned algorithms. The outputs of all the fuzzy clustering methods are given in Figure 5.

Clusters of FCM are relatively of the same size, with no elongation. Clustering is done along the direction of "higher water saturation" (Figure 3b). Therefore, the blue cluster represents the lowest water saturation, (Figure 5a) while the red cluster shows the highest water content data. But as this algorithm is not able to detect lineation, it is unable to differentiate along the arrow of "Higher Perm and Larger Pore Throats" on Figure 3b. The green cluster shows two different trends: One trend belongs to high permeability values, the other represents the less permeable or shale contaminated part.

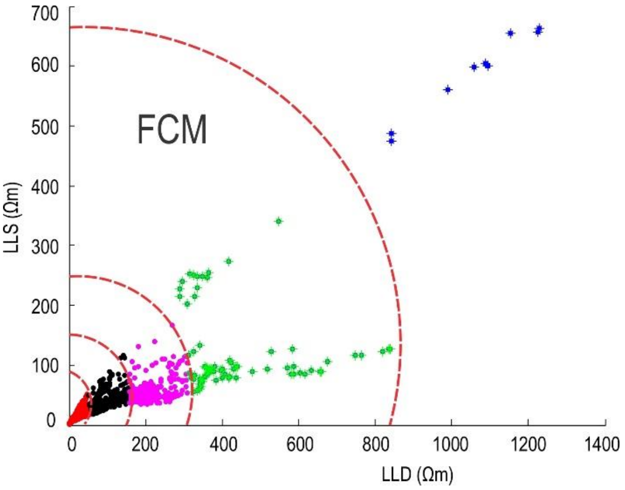
This phenomenon is detected by GK clustering as this algorithm is able to distinguish elongated clusters. The blue cluster represents low water saturation and high permeability (Figure 5b), which is ideal to be known as net pay zone, i.e. a hydrocarbon production zone. The red data shows below oil water contact but still permeable, i.e. a water-producing zone. The area of green data is the least permeable part of the Sarvak Formation, either oil-bearing or water-bearing. The black and magenta parts have less permeability than red and blue clusters; and magenta contains less water, compared to black data. GK provided the best clustering due to petrophysical interpretations.

While GK clusters are relatively similar in size (number of data in each cluster), the GG clusters differ considerably in size (Figure 5c,d) with a net pay zone represented by the magenta colour (Figure 5c). Red, black, green and blue clusters have lower reservoir quality, respectively. Contrary to FCM and GK, clusters of GG do not show a specific pattern; therefore they cannot be as easily interpreted as FCM and GK clusters.

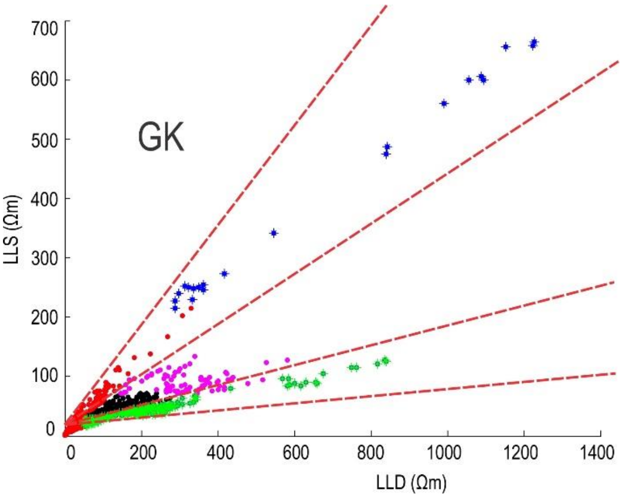
8. Discussion

In the previous section, the outputs of three clustering methods on a two-dimensional dataset were tested. FCM was the most compatible tool for saturation prediction, while GK is the best for permeability study and GG could be used as a rough measure of productivity. The reason why the functionality of these clustering methods differ from each other could be seen at the bottom of Table 2.

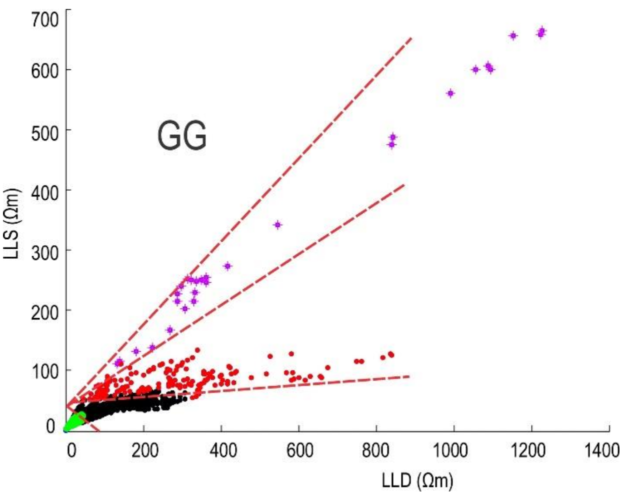
FCM provides simple clusters of the same size, while GG provides elongated unequal clusters, which is difficult to interpret due to the complexity of the structure of clusters. On the other hand, GK provides elongated clusters (same size), but the structure of the clusters is not complex for interpretation. It is not difficult to imagine that when we have a change in the specifications of clusters, their physical meaning (characterization target) would be changed too. So, three issues have to be kept in mind for successful clustering: (i) our goal of study or characterization target; (ii) the relation or correlation of input dataset with target, which was discussed above; and (iii) the specifications of produced clusters.



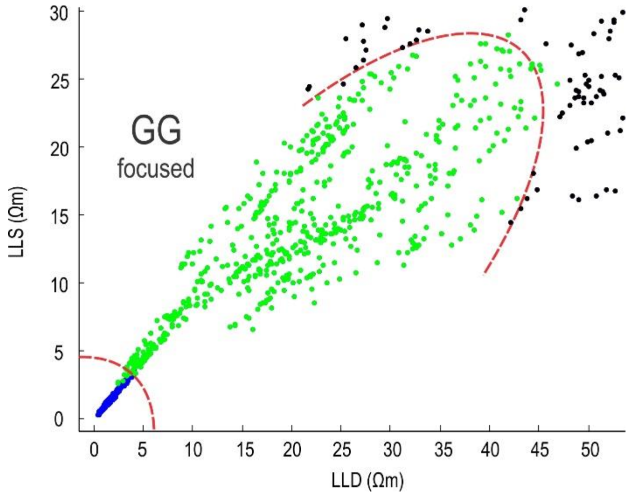
(a)



(b)



(c)



(d)

Figure 5. Outputs of fuzzy clustering algorithms. (a) Fuzzy C-Means clustering with 5 clusters. (b) Gustafson-Kessel with 5 clusters. (c) Gath-Geva with 5 clusters. (d) Gath-Geva with 5 clusters. Focused to show blue cluster.

Among the three issues mentioned, first we define the purpose of our study; i.e. “What is to be measured or predicted?” According to this goal, the dataset is chosen; e.g. here we used resistivity logs for the sake of their correlation with saturation and reservoir fluids. It is noteworthy that sometimes the dataset is not optional, and we have to use some fixed parameters. But we are always able to choose among various clustering algorithms. To choose the most appropriate one, we can either go in deep through the mathematical basis of the algorithm, analytically; or we can simply check the specifications of their clusters as is provided in Table 2.

9. CONCLUSIONS

This paper shows the strength of clustering algorithms in interpreting resistivity well-logs to extract various inferences regarding water saturation, and other reservoir parameters just by means of two well-known resistivity well-logs. The number of publications on saturation prediction is much less than other petrophysical properties such as porosity and permeability, mostly due to a lack of real saturation data for the validation procedure. However, its importance is not less than other reservoir properties. Furthermore, we proposed clustering algorithms for water saturation prediction, because clustering methods are unsupervised methods that do not require real labeled data for training the machine.

Due to petrophysical interpretations, the FCM method provided separation of the space only by considering saturation property. However, GK clustering divided the datasets according to two properties: saturation and permeability (or pore throat size). GG could divide the space according to overall reservoir quality. The output of the Spectral algorithm was much the same as GG but more noise-contaminated. The affinity propagation clustering method did not provide well-interpretable outputs though we are not disappointed with this algorithm, and believe that further investigations have to be carried out on the application of clustering algorithms, on petrophysical datasets.

It is noteworthy that by means of unsupervised methods, we do not train the machine, considering cores, i.e. the minute scale. Therefore, the results would be free from the falsification resulting from the use of idealistic small scale measurements of cores, when interpreting highly heterogeneous reservoir conditions based on core analysis. The validation procedure is done qualitatively by checking the behaviour of clusters against the behaviour of core permeability. The ratio of LLD:LLS has established a link between the reservoir permeability and core permeability. Quantitative comparison is not carried out here because: (i) there are two different languages in expressing permeability values: core permeability, in milliDarcy, is a continuous value, while inferred well-log permeability is in discrete labels (i.e. clusters); (ii) There is no guarantee for compatibility of core-scale permeability with reservoir-scale permeability, inferred from well-logs. Hence, we lack a reliable base point for quantitative evaluation of water saturation and permeability in reservoir-scales; however, qualitative reasoning is possible and necessary as is done here.

TABLE II. SUMMARY OF APPLICATION OF CLUSTERING METHODS IN RESERVOIR CHARACTERIZATION. SYMBOLS (✓) AND (✗) MEAN COMPATIBILITY AND INCOMPATIBILITY, RESPECTIVELY; E.G. FCM IS COMPATIBLE FOR PREDICTING WATER SATURATION BUT IT IS NOT COMPATIBLE FOR PERMEABILITY INFERENCE. "SIMPLICITY OF CLUSTERS" SHOWS WHETHER THE CLUSTERING ALGORITHM COULD BE INTERPRETED EASILY OR NOT; E.G. GG COULD NOT BE INTERPRETED SIMPLY; WHILE GK IS EASILY INTERPRETABLE. THE DOUBLE SYMBOL (✓✓) MEANS MORE COMPATIBILITY, COMPARED TO THE SINGLE SYMBOL (✓).

		Fuzzy C-Means	**Gustafson-Kessel**	**Gath-Geva**
characterization target	water saturation	✓	✗	✗
	permeability	✗	✓✓	✓
	productivity	✗	✓	✓✓
properties of clusters	equity of cluster size	✓	✓	✗
	elongation	✗	✓	✓
	interpretability	✓	✓	✗

ACKNOWLEDGMENT

This work has been supported by the Center for International Scientific Studies & Collaboration (CISSC) and French Embassy in Iran through PHC Gundishapur Program. It is our pleasure to acknowledge the Exploration Directorate of National Iranian Oil Company (NIOC) for providing data, and permitting the publication of scientific achievements.

REFERENCES

[1] Mostafazadeh, M.; Mousavi, S. A.; Ghadami, N.; Aghdasinia, H.; 2010; "*The Productivity Estimation of Designed Horizontal Oil and Gas Wells Before a Drilling Operation, Using Seismic and Petrophysical Parameters and Modeling*", Petroleum Science and Technology, 28, 1863-1877.

[2] Worthington, P.F.; Cosentino, L.; 2005; "*The Role of Cut-offs in Integrated Reservoir Studies*", SPE Reservoir Evaluation & Engineering, 8, 276-290.

[3] Jensen, J.L.; Menke, J.Y.; 2006; "*Some Statistical Issues in Selecting Porosity Cutoffs for Estimating Net Pay*", PetroPhysics, 47, 315–320.

[4] Worthington, P.F.; 2010; "*Net Pay-What Is It? What Does It Do? How Do We Quantify It? How Do We Use It?*", SPE Reservoir Evaluation & Engineering, 13, 812-822.

[5] Mahbaz, S.; Sardar, H.; Namjouyan, M.; Mirzaahmadian, Y.; 2011; "*Optimization of reservoir cut-off parameters: a case study in SW Iran*", Petroleum Geoscience, 17, 355-363.

[6] Amalokwu, K.; Best, A.I.; Sothcott, J.; Chapman, M.; Minshull, T.; Li, X.Y.; 2014; "*Water saturation effects on elastic wave attenuation in porous rocks with aligned fractures*", Geophysical Journal International, 197, 943-947.

[7] Tokhmechi, B.; Memarian, H.; Rasouli, V.; Noubari, H. A.; Moshiri, B.; 2009 ; "*Fracture detection from water saturation log data using a Fourier–wavelet approach*", Journal of Petroleum Science and Engineering, 69, 129-138.

[8] Lopes, S.; Lebedev, M.; Müller, T.M.; Clennell, M.B.; Gurevich, B.; 2014; "*Forced imbibition into a limestone: measuring P-wave velocity and water saturation dependence on injection rate*", Geophysical Prospecting, 62(5), 1126-1142.

[9] Archie, G.E; 1942; "*The electrical resistivity log as an aid in determining some reservoir characteristics*", Petroleum Transactions of AIME, 146, 54-62.

[10] Archie, G.E.; 1947; "*Electrical resistivity an aid in core-analysis interpretation*", AAPG Bulletin, 31, 350-366.

[11] Archie, G.E.; 1950; "*Introduction to petrophysics of reservoir rocks*", AAPG Bulletin, 34, 943-961.

[12] Archie, G.E.; 1952; "*Classification of Carbonate Reservoir Rocks and Petrophysical Considerations*", AAPG Bulletin, 36, 278-298.

[13] Carman, P.; 1937; "*Fluid flow through granular beds*", Chemical Engineering Research and Design, 75, 32-48.

[14] Timur, A.; 1968; "*An Investigation Of Permeability, Porosity, and Residual Water Saturation Relationships For Sandstone Reservoirs*", The Log Analyst, IX.

[15] Tiab, D.; Donaldson, E.C.; 2004; "*Petrophysics: theory and practice of measuring reservoir rock and fluid transport properties*", Gulf Professional Publishing.

[16] Worthington, P.F.; 2004; "*Improved Quantification of Fit-for-Purpose Saturation Exponents*", SPE Reservoir Evaluation & Engineering, 7.

[17] Kadkhodaie-Ilkhchi, A.; Rezaee, M. R.; Rahimpour-Bonab, H.; 2009; "*A committee neural network for prediction of normalized oil content from well log data: An example from South Pars Gas Field, Persian Gulf*", Journal of Petroleum Science and Engineering, 65, 23-32.

[18] Mollajan, A.; Mehrgini, B.; Memarian, H.; 2013; "*Zonal classification by pattern recognition methods: An example from Asmari Formation (Mansuri oil field, south of Iran)*", Energy, Exploration & Exploitation, 31, 367-380.

[19] Mollajan, A.; Memarian, H.; 2013; "*Estimation of water saturation from petrophysical logs using radial basis function neural network*", Journal of Tethys, 1, 156-163.

[20] Mollajan, A.; Memarian, H.; Jalali, M.; 2013; "*Prediction of Reservoir Water Saturation Using Support Vector Regression in an Iranian Carbonate Reservoir*", 47th US Rock Mechanics/ Geomechanics Symposium, American Rock Mechanics Association, 137-138.

[21] Barros, C.; Andrade, A.; 2013; "*Determination of water saturation by angular competitive neural network*", Journal of Petroleum Science and Engineering, 102, 47-56.

[22] Ghabeishavi, A.; Vaziri-Moghaddam, H.; Taheri, A.; Taati, F.; 2010; "*Microfacies and depositional environment of the*

Cenomanian of the Bangestan anticline, SW Iran", Journal of Asian Earth Sciences, 37, 275-285.

[23] Klir, G.J.; Yuan, B.; 1995; "*Fuzzy Sets and Fuzzy Logic, Theory and Applications*", Prentice Hall New Jersey.

[24] Hartigan, J.A.; Wong, M.A.; 1979; "*Algorithm AS 136: A k-means clustering algorithm*", Applied statistics, 100-108.

[25] Bezdek, J.C.; Ehrlich, R.; Full, W.; 1984; "*FCM: The fuzzy c-means clustering algorithm*", Computers & Geosciences, 10, 191-203.

[26] Babuka, R.; Van Der Veen, P.; Kaymak, U.; 2002; "*Improved covariance estimation for Gustafson-Kessel clustering*", Fuzzy Systems 2002: FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference, 1081-1085.

[27] Gustafson, D.E.; Kessel, W.C.; 1978; "*Fuzzy clustering with a fuzzy covariance matrix*", Decision and Control including the 17th Symposium on Adaptive Processes, 1978 IEEE Conference, 761-766.

[28] Gath, I.; Geva, A.B.; 1989; "*Unsupervised optimal fuzzy clustering*", Pattern Analysis and Machine Intelligence, IEEE Transactions on, 11, 773-780.

[29] Abonyi, J.; Babuska, R.; Szeifert, F.; 2002; "*Modified Gath-Geva fuzzy clustering for identification of Takagi-Sugeno fuzzy models*", Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 32, 612-621.

[30] Ibrahim Sami, N.; Adel, M.; 2010; "*Permeability Prediction from Wireline Well Logs Using Fuzzy Logic and Discriminant Analysis*", SPE Asia Pacific Oil and Gas Conference and Exhibition. Brisbane, Queensland, Australia: Society of Petroleum Engineers.