Universal Journal of Petroleum Sciences 1 (2013), 30-38
www.papersciences.com

A Committee Machine with Different Adaptive Neuro Fuzzy Inference System for Water Saturation Prediction

Seyed Ali Jafari Kenari[1,a], Hamid Mosalmannejad[2]
[1] Department of Computer and Communication Systems Engineering, Faculty of Engineering, University Putra Malaysia, Malaysia
[a] Training Center National Iranian Oil Company, Mahmoud Abad, Mazandaran, Iran
s.a.jafari@nioc.ir
[2] Iranian Offshore Oil Company (IOOC), Petro physics department, No 12, Tooraj St, Vali-E-Asr Ave,Tehran, Iran
hmosalmannejad@yahoo.com

Abstract
A committee machine (CM) with different adaptive neuro fuzzy inference system (ANFIS) models is proposed to model highly nonlinear functions to estimate an important parameter in petroleum reservoirs, namely, water saturation. The proposed approach was tested on an Iranian oil field to determine an accurate value of water saturation using well log data. Experimental results show that our proposed method can be used to predict any other reservoir parameters using well logs data in a un-cored interval of the same well or a un-cored well of the same field. The results show that the CM-ANFIS method performed better than three individual members alone.

Key Words and Phrases
Neuro Fuzzy, Ensemble, Committee Machine, Water Saturation-Genetic Algorithm.

1. Introduction

In petroleum industry, obtaining an accurate estimation of the hydrocarbon in place before exploration or production stages is the most important objective. Therefore, reliable prediction of reservoir characterization is very helpful for evaluation and designing any development plan for production of the field. Water saturation is one of the significant reservoir characterizations in petroleum industry. It means the fraction of pore space that is filled by water only. There are many traditional petrophysical methods in petroleum reservoir to evaluate water saturation [1]. One of the most drawbacks of these methods is often the unavailability of the input parameters [2, 3]. Moreover, in shaley sands, the prediction of water saturation using wireline log data is difficult. Core analysis is the most accurate method to calculate water saturation. The obtained water saturation with this direct method is expensive, time consuming and only available for a few wells and a few limited interval of a well [4].

Another method to determine water saturation is the height functions (e.g. J-function and Heseldin) that can be obtained by capillary pressure. The disadvantages of these methods are mainly due to their inaccuracy for complex systems. There are some other methods to determine water saturation in shaly reservoir such as volume of shale (Vsh) and Cation Exchange Capacity (CEC) models. However, the related input parameters to utilize these methods have to be obtained in the library which sometimes gives unreliable results and also are time consuming. Numerous studies have been done to introduce the reliable equations based on well log data to estimate water saturation [5]. One of the most commonly used of them is the Archie equation, (1):

$$S_w = \sqrt[n]{\frac{a \cdot R_w}{R_t \cdot \phi^m}} \tag{1}$$

where R_w, R_t $and\,\phi$ are resistivity of formation water, true formation resistivity and porosity fraction respectively.

Three constants a, n and m are empirical constant, saturation exponent and cementation factor respectively. Their experimental values are defined as (1, 2, 1) for carbonate and (0.62, 2, 2.15) for sandstone reservoirs respectively. The disadvantage of this equation is due to its unsuitability for shaly formation, which means it can be used only in a clean formation. Another empirical equation to determine water saturation is Indonesian equation that can be used for shaly reservoir [5, 6]. The core data dependency is one of the major drawbacks of this method.

Logging tools firstly run to measure different geophysical properties of the reservoir formations. Additionally, it is used to determine the suitable type of Enhanced Oil Recovery (EOR) techniques. This type of measurement is taken on the upward direction in the interest intervals. The most related measurements in wireline logging are recorded continuously. There are many types of logging tools in petroleum industry that can be divided to different kinds such as electrical, sonic, mechanical and nuclear tools. The objective of this paper is to use a set of three neural fuzzy techniques to predict water saturation by utilizing the well logs and a limited core data.

Spontaneous Potential (SP) log

This type of logging tool usually recorded along with resistivity curves and reflects an electrical potential difference between two electrodes. The SP curve can be used to indicate some important properties in rock reservoir such as lithology, permeability, shale volume, porosity and formation water salinity (R_w).

Resistivity (R_t) logs

Resistivity logs can be used to measure the ability of rocks to carry out electrical current such that sand filled with salt water has lower resistivity rather than sand filled with oil or gas. The primary aim of this logging method is to determine hydrocarbon saturation but the other usages of it are to determine porosity, permeability and fracture zones.

Sonic log (DT)

This type of logging is also known as acoustic or velocity log that is widely used for porosity estimation to quantitative interpretation of hydrocarbon saturation. It is based on measuring the sound waves' speed in travel through 1 m of subsurface formations. This logging tool can also be used for many other purposes such as delineate fractures and indicating lithology, determining integrated travel time, secondary porosity, mechanical properties, acoustic impedance, quality of cementation behind the formation and detecting over-pressure.

Gamma Ray (GR) log

This logging tool measures radiation omitted by naturally occurring potassium, thorium, and uranium in formation versus depth. Gamma ray log also known as shale log because the radioactive count in shale is higher than clean sand or carbonates. This instrument can be used in both openhole and casedhole operation and is also suitable to determine bed thicknesses, mineral analysis, correlation between wells and etc.

Density (RHOB) log

Density log is another instrument to determine porosity especially in shale that reacts to variation of the specific formation gravity. This log is also known as gamma-gamma density or densilog and can be used only in openhole well logging. This tool emits gamma ray into the rock and measures the back-scattered radiation that is received by the detector in the instrument. Density log can also be used to determine lithology, gas detection, estimating mechanical properties, evaluation of shaly sands and etc.

Neutron (NPHI) log

The third significant tools to determine formation porosity is neutron log. This log has a radioactive source, and measures the reaction energy between emitted and detected neutrons back to the tool. When the emitted neutrons encounter with the hydrogen’s nucleus, the most emitted energy will be lost. Therefore based on the hydrogen index, apparent neutron porosity can be determined. The neutron logs almost in combination with other logs such as density or sonic can be used to determine porosity, volume of shale, matrix type, lithology, formation fluid type and etc.

2. Related Work

Several studies have been done to predict water saturation by utilizing conventional well log data based on artificial intelligent techniques. In [7], a resilient back propagation Neural Network (NN) model is proposed to predict water saturation based on well logs data as input and core Dean-Stark as target data. The authors also applied the same network to determine the cementation factor and saturation exponent coefficient in Archie equation. In [4], the authors proposed a novel method based on functional network technique to estimate porosity and water saturation from conventional well measurements. The proposed base functions to predict water saturation was logarithm function. They have also used gamma ray, neutron, density, resistivity, and photoelectric effect logging data as inputs data. In order to make sure the input variables are independent from measurement units, they were normalized them between 0 and 1. Later, a feed forward neural network model with Levenberg Marquardt algorithm was used to check the validity of functional network results. In [8], the authors applied a developed NN to predict water saturation and the volume of shale. This method was implemented in shaly formation by utilizing wireline logs and core data as training samples. They have investigated several sensitivity analyses by the variation of some parameters in the proposed algorithm such as the number of hidden neurons, learning algorithms, scaling methods and transfer function. Furthermore, they have used multiple linear and nonlinear regression method but did not obtain any reliable results. In [9], the authors proposed an ANN model to predict water saturation from well log data. They have used four different NN based on two different transfer function (e.g. tansig and logsig) in hidden and output layers. Therefore, they show, the predicted value is more reliable than dual water models because the network uses the well logs as inputs data, which are not laboratory data as well as less time and expense to extract shaley model parameters. In the second part, they have converted Archie's equation to two linear equations to determine cementation factor and saturation exponent (e.g. m and n in (1)). In [10], the authors used ANN method to predict water saturation in southern Iranian oil reservoir. Porosity, permeability, and height above the free water level were used as the input data and water saturation as target data. Finally, they compared predicted water saturation by ANN model with eight saturation height methods. In [11], the authors proposed a committee of nine expert to predict water, oil and gas saturation separately. From twenty number of individually trained network, they have selected nine best experts for final committee neural network. The selected inputs data were resistivity, density, neutron and sonic log and three mentioned saturation as targets. Two combining methods to fuse the outputs of individual experts were simple averaging and optimal linear combination that proposed by [12]. Numerous investigations have been done to predict fluid saturation based on regression method and empirical formula such as [13, 14].

3. Methodology

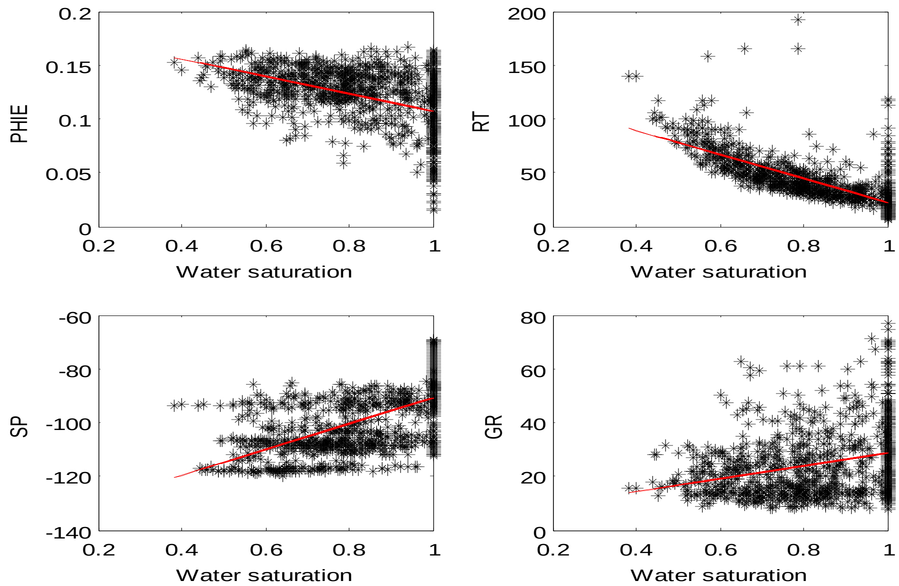
3.1. Data Collection and Input selection

For the purpose of this study, well logs and core data were collected from Iranian oil fields. Selecting the suitable inputs with stronger relationship with the target data plays an important role in model construction. Therefore, the relationship between the available well log data and core data are tested and illustrated in Fig.1. A quick glance at the Figures shows a direct relationship between six independent variables (well logs) and one dependent variable (water saturation). The positive or negative slope means there is a relation between x and y axis. Therefore, effective porosity (PHIE), R_t, SP, GR, NPHI, RHOB are selected as input or independent variable in our study to predict water saturation.

3.2. Neural Fuzzy

Creating fuzzy logic (FL) rule is simple for designers but there is a major drawback for describing the knowledge contained in real and large data sets manually. In contrast, the neural network in large and nonlinear system can provide a solution due to its self-training ability.

Whenever, interpretation and manually modification in certain desired behavior cannot be achieved. The same is in neural network structures, where selecting a suitable training algorithm, learning rate, active function and so on is still difficult task and requires much experience. Hence, both methods have some advantages and limitations and a smart combination of them can improve their ability. Adaptive Neuro-Fuzzy Inference System (ANFIS) is a kind of NN that is based on Takagi-Sugeno FIS that integrates both NNs and FL principles and has the potential to capture the advantages of both systems in a unique framework, which introduced by [15]. Its inference system corresponds to a set of fuzzy rules that have learning capability to approximate nonlinear functions. In this method, the parameters of membership functions (MF) in FIS are adjusted using a hybrid approach which is based on back propagation algorithm and least-squares method. This hybrid approach allows the fuzzy systems to learn from the data that they are modelling. In this paper we have used Sugeno fuzzy inference system with three different clustering methods which were grid partition, subtracting clustering and FCM clustering. Therefore the data sets are divided into three parts which were training, checking and testing data. The checking data set were used for over fitting model validation. These experts are used to predict the mentioned reservoir parameter. Fig. 2 shows an example of ANFIS procedure for testing data set based on grid partition method in Matlab FL toolbox [16]. Fig.3 also illustrated the optimal BPNN obtained by ANFIS based on subtracting clustering method, which has two MF for each input and thirty two rules.



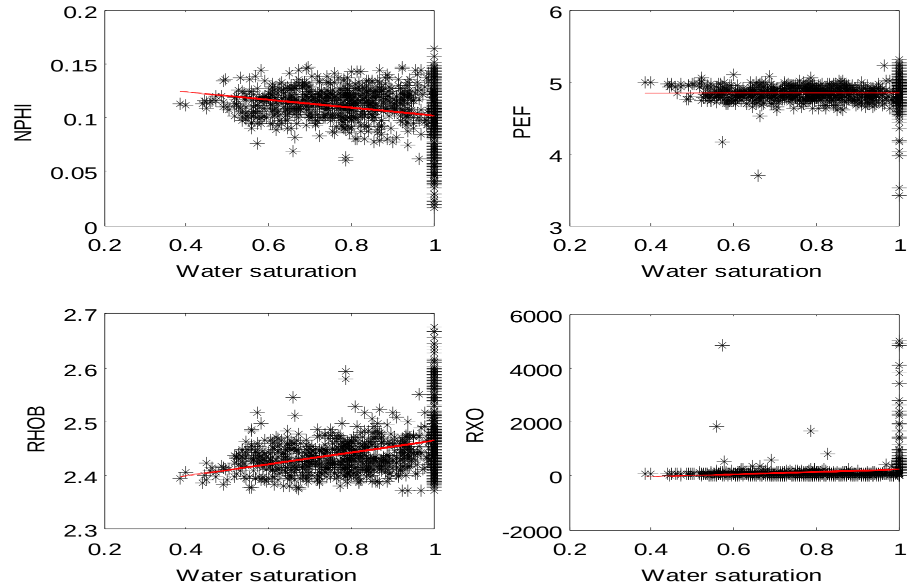


Fig. 1 Cross plots showing the relationship between core water saturation and NPHI, RT, SP, GR, NPHI, PEF, RHOB, RXO logs.

3.3. Why Committee Machine

Although different learning methods applied in various areas of applications, one individual expert is improbable to generalize all possible test pattern perfectly [11, 17]. A simple and easy approach is to train many individual members and then select one expert that gives best result on the test data. However, while one expert reproduces the main patterns, the others may provide the details lost by the first. Therefore, selecting the best expert is similar to loss of information. Then the objective should be to exploit, rather than lose, the information contained in a set of imperfect generalizers [11]. This is the underlying motivation to combine a collection of individual predictors to improve accuracy and increase robustness [18]. The advantages of this technique are summarized as below. First, ensemble methods are very robust when the training data are contaminated with noise [19, 20]. Second, combining several predictors that are prone to over-fitting can be useful to reduce it in numerous practical problems [21, 22]. Finally, the ensemble output gives an improved accuracy compared with the output of any individual member alone [19, 23, 24]. Although, aggregating a collection of the machine learning techniques improves the accuracy of the predictions, its practical implementation can be so hard sometimes. One of the important limitations cited in literature is the number of predictors with a convergence guarantee of the prediction error which are sometime unnecessary large [25, 26]. In addition, the prediction time and memory resource in comparison with a single model increases linearly with the ensemble size that can be critical in online applications [26]. Determine the appropriate ensemble size is also another limitation which is related to the predictive error and strongly depends on the specific problem under investigation [27, 28]. Therefore, a committee machine (CM) or ensemble with large size will result in a waste of resources and also loss of prediction accuracy on small size. In this paper, we have created a CM based on three mentioned experts and genetic algorithm (GA) as fusion method to predict water saturation.

4. Results and Discussion

As mentioned before, three experts have been created based on ANFIS method with different Sugeno model and clustering algorithms. These three clustering methods were grid partition, subtracting clustering and FCM algorithm which are illustrated by ANFIS1, ANFIS2 and ANFIS3. To obtain the best prediction, these methods were retrained by different parameters based on MSE and R^2 as our measuring performance which are widely used techniques. The parameters used in our experiments were set to: the input membership type (trimf); the optimum method (Hybrid); the number of MF (23). As listed in Table 1, the performance results of ANFIS2 and ANFIS3 are more accurate than ANFIS1. It means the MSE value for both of them are less than the MSE value obtained by ANFIS1. Moreover, in the same time, the related R^2 value for ANFIS2 and ANFIS3 are greater than for ANFIS1. Fig. 4 (a-c) also demonstrates a comparison between measured and predicted water saturation for the three mentioned ANFIS methods respectively. Finally, we used genetic algorithm (GA) to obtain an optimal combination of the weights for these three experts for constructing the CM-ANFIS. The GA fitness function and related constrains are shown in Eq.(2).

Therefore, after obtaining the related weight of each individual member, the output of CM-ANFIS is calculated by Eq. (3):

$$Fit\_func = \frac{1}{n}\sum_{i=1}^{n}(w_1 y_{i1} + w_2 y_{i2} + w_3 y_{i3} - T_i)^2$$
$$S.t \quad \begin{cases} \sum_{j=1}^{3} w_j = 1 \\ w_j \le 1 \\ w_j \ge 0 \end{cases} \tag{2}$$

$$OUT_{CM} = \sum_{k=1}^{3} w_k y_k \tag{3}$$

where y_{i1} , y_{i2} and y_{i3} are the outputs of first, second and third expert respectively based on the *i-th* input and T_i is the target value of the *i-th* input, and *n* is the number of data points. The parameters used in GA method were set to: the population size (100); the chromosome length (51); crossover probability (0.7); mutation probability (0.05). Each test function is tested on the GA for 30 times with a maximum of 5000 generations per each run. The final weights of three individual members obtained by (2) are 0.0416, 0.251, and 0.707 respectively. Therefore, by using the (3), the final estimation of water saturation was done through (4):

$$OUT_{CM} = 0.0416 * ANF1 + 0.251 * ANF2 + 0.707 * ANF3 \tag{4}$$

Finally, the proposed committee machine decreases the MSE value up to 0.002 which corresponds to R^2 value of 0.9515. Fig. 4(d) shows a comparison between measured and predicted water saturation based on CM-ANFIS approach.

By comparing the demonstrated results for each expert, it is clear that the performance of the proposed approach is more accurate and reliable than each expert along. Fig. 5 also shows a comparison between core measured and CM-ANFIS predicted water saturation versus depth.

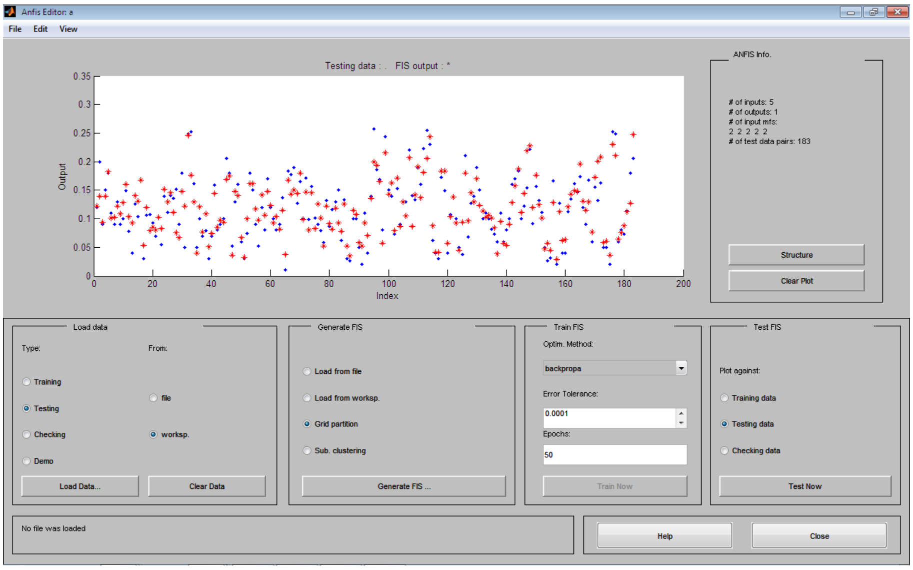


Fig. 2 ANFIS procedure for testing data set.

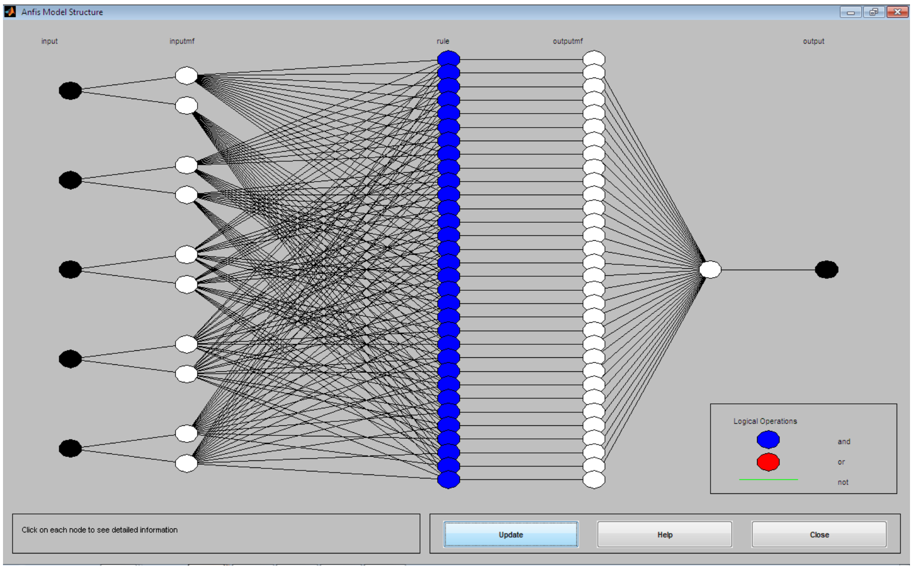


Fig. 3 ANFIS structure for formulating well log data to core reservoir**.**

TABLE 1

A Comparison between NF Predicted and core measured water saturation based on R^2 and MSE values

METHODS	ANFIS1	ANFIS2	ANFIS3
R^2	0.7427	0.9112	0.9447
MSE	0.0112	0.0037	0.0023

5. Conclusion

In this paper, we have created a committee machine with three different experts based on ANFIS methods. These experts were utilized to predict water saturation based on petrophysical data in the Iranian oil filed. Finally, we used GA technique to obtain the optimal weight for each individual to carry out the final results. In the test data, the performance of the proposed models was found to be more reliable than the each expert alone. A comparison between measured and predicted water saturation for the whole experts are illustrated in Fig. 4. As demonstrated in the figures, the performance obtained by CM-ANFIS have been improved to (0.9515, 0.002) in comparison with ANFIS3 with (0.947, 0.0022) for R^2 and MSE. The proposed methods can also be applied to predict any other parameters in oil reservoir using well logs data in an un-cored interval of the same well. These methods also can be applied to predict rock properties from the nearly uncored wells in the same fields.

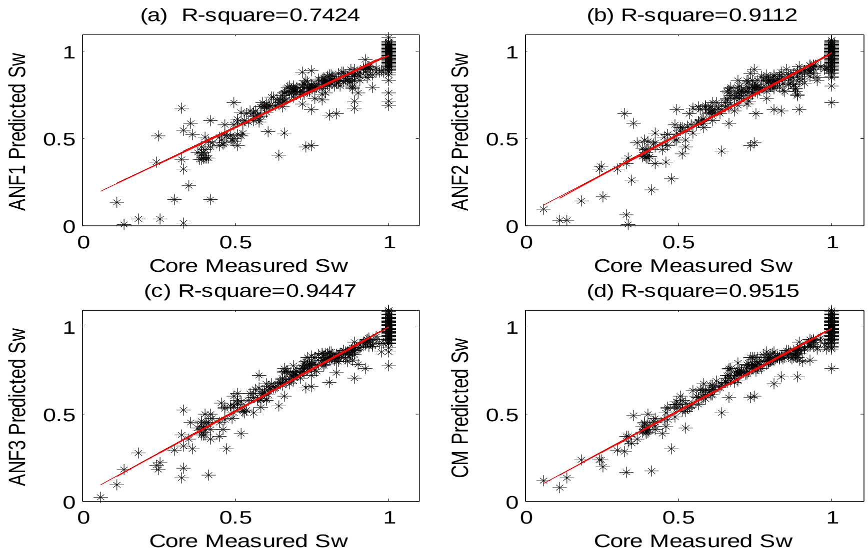


Fig. 4 A scatter plot between core measured and a) ANF1, b) ANF2, c) ANF3, d) CM based on GA combined water saturation.

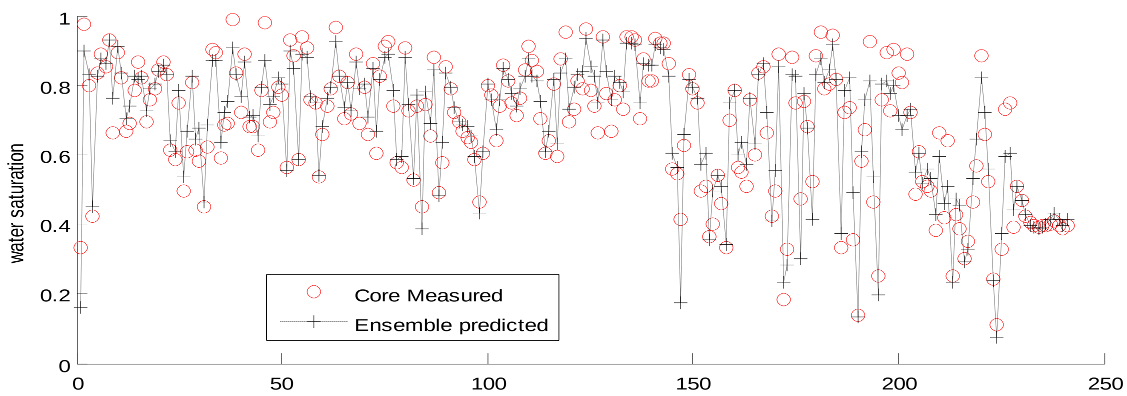


Fig.5 Comparison of CM predicted and core measured water saturation versus depth.

Acknowledgment

The authors wish to express their appreciation to the National Iranian Oil Company (NIOC) for sponsorship, data preparation and their collaboration.

References

1. *Archie.G.E,* 'The electrical resistivity log as an aid in determining some reservoir characteristics', *Transactions of the American Institute of Mining and Metallurgical Engineers,* ***146*** *(1942), 54-62.*
2. *Ipek.G, 'Log-derived cation exchange capacity of shaly sands application to hydrocarbon detection and drilling optimization'*, PhD thesis, Louisiana State University., *2002.*
3. *Rezaee. M.N and Lemon. N.M, 'Petrophysical Evaluation of Kaolinite-bearing Sandstones: I: Water saturation (Sw), an example of the Tirrawarra Sandstone reservoir, Cooper Basin, Australia',* presented at the SPE Asia Pacific Oil and Gas Conference, Adelaide, Australia, *1996.*
4. *Adeniran. A, Elshafei. M, and Hamada. G,* 'Functional Network softsensor for formation porosity and water saturation in oil wells', *Instrumentation and Measurement Technology Conference. IEEE, 2009,*

1138-1143.

5. *Alimoradi. A, Moradzadeh. A and Bakhtiari. M. R,* 'Methods of water saturation estimation: Historical perspective', *Journal of Petroleum and Gas Engineering, **2** (2011), 45-53.*
6. *Poupon. A and Leveaux, J,* 'Evaluation Of Water Saturation In Shaly Formations', *The Log Analyst, 4 , 1971.*
7. *Al-Bulushi. N, King. P. R, Blunt. M. J, and Kraaijveld. M,* 'Development of artificial neural network models for predicting water saturation and fluid distribution', *Journal of Petroleum Science and Engineering, **68** (2009), 197-208.*
8. *Al-Bulushi. N, King. P, Blunt. M, and Kraaijveld. M,* 'Artificial neural networks workflow and its application in the petroleum industry', *Neural Computing and Applications, **21** (2010), 1-13.*
9. *Mardi. M, Nurozi. H, and Edalatkhah. S,* 'A Water Saturation Prediction Using Artificial Neural Networks and an Investigation on Cementation Factors and Saturation Exponent Variations in an Iranian Oil Well', *Petroleum Science and Technology, **30** (2012), 425-434.*
10. *Kamalyar. K, Sheikhi. Y and Jamialahmadi. M,* 'Using an Artificial Neural Network for Predicting Water Saturation in an Iranian Oil Reservoir', *Petroleum Science and Technology, **30** (2012), 35-45.*
11. *Helle. H. B and Bhatt. A,* 'Fluid saturation from well logs using committee neural networks', *Petroleum Geoscience, **8** (2002), 109-118.*
12. *Hashem. S,* 'Optimal Linear Combinations of Neural Networks', *Neural Networks, **10** (1997), 599-614.*
13. *Yan. J,* 'Reservoir parameters estimation from well log and core data: a case study from the North Sea', *Petroleum Geoscience, **8** (2002), 63-69.*
14. *Kamel. M. H and Mabrouk. W. M,* 'An equation for estimating water saturation in clean formations utilizing resistivity and sonic logs: theory and application', *Journal of Petroleum Science and Engineering, **36** (2002) 159-168.*
15. *Jang. J.-S. R.,* 'ANFIS: Adaptive-Network-Based Fuzzy Inference System', *IEEE Transactions on Systems, Man, and Cybernetics, **23** (1993), 665-685.*
16. MATLAB User's Guide, *2011.*
17. *Bhatt. A and Helle. H. B,* 'Committee neural networks for porosity and permeability prediction from well logs', *Geophysical Prospecting, **50** (2002), 645-660.*
18. *Martínez-Muñoz. G, Sánchez-Martínez. A, Hernández-Lobato. D, and Suárez. A,* 'Class-switching neural network ensembles', *Neurocomputing, **71** (2008), 2521-2528.*
19. *Opitz. D and Maclin. R,* 'Popular Ensemble Methods: An Empirical Study', *Journal of Artificial Intelligence Research, **11** (1999), 169 -198.*
20. *Martinez-Muoz. G, Hernandez-Lobato. D, and Suarez. A,* 'An Analysis of Ensemble Pruning Techniques Based on Ordered Aggregation', *IEEE Transactions on Pattern Analysis and Machine Intelligence, **31** (2009), 245-259.*
21. *Dirk. H and Kaspar. A,* 'Modelling Conditional Probabilities with Network Committees: How overfitting can be useful', *Neural Network World, **8**(1998), 417–439.*
22. *Sollich. P and Krogh. A,* 'Learning with ensembles: How overfitting can be useful', *Advances in Neural Information Processing Systems, **8**(1996), 190-196.*
23. *Bühlmann. P. L, 'Bagging, Subagging and Bragging for Improving Some Prediction Algorithms',* Recent Advances and Trends in Nonparametric Statistics, Elsevier, *2003.*
24. *Breiman. L,* 'Bagging predictors', *Machine Learning, **24** (1996), 123-140.*
25. *Banfield. R. E, Hall. L. O, Bowyer. K. W, and Kegelmeyer. W. P,* 'A Comparison of Decision Tree Ensemble Creation Techniques', *IEEE Trans. Pattern Anal. Mach. Intell., **29** (2007), 173-180.*
26. *Margineantu. D and Dietterich. T,* 'Pruning Adaptive Boosting', *in Proceedings of the Fourteenth International Conference on Machine Learning, 1997, 211-218.*
27. *Breiman. L,* 'Random Forests', *Machine Learning, **45** (2001), 5-32.*
28. *Freund. Y and Schapire. R. E,* 'Experiments with a New Boosting Algorithm', *in proceedings of the thirteenth international conference on machine learning, 1996, 148-156.*