CrossMark

ORIGINAL ARTICLE

Prediction of reservoir permeability from porosity measurements for the upper sandstone member of Zubair Formation in Super-Giant South Rumila oil field, southern Iraq, using M5P decision tress and adaptive neuro-fuzzy inference system (ANFIS): a comparative study

Amna M. Handhal[1]

Abstract This study focus on the determining the reservoir permeability from measured porosity using different techniques namely linear regression, adaptive neuro-fuzzy inference system, and M5 decision trees. The upper sandstone member of Zubair Formation in Rumila oil field in south of Iraq was taken as an example to reveal the efficacy of these techniques for simulating porosity and permeability relationship for this heterogeneous reservoir. The models performances were evaluated using root mean squared error (RMSE) and coefficient of determination (R^2) error statistics. Results demonstrated that ANFIS model with 0.380 (RMSE) and 0.844 (R^2) for the testing phase perform the best among the other models. The results of M5 decision trees model in testing phase were very similar to that of ANFIS with 0.378 (RMSE) and 0.842 (R^2). The complex structure of ANFIS if it compared with a very simple structure and very easy to interpret M5 model make this technique is a best choice for predicting reservoir permeability form core porosity in the study area. The study implies that both M5 and ANFIS were promising tool for identifying porosity–permeability relationship in the study area.

Keywords Reservoir permeability · ANFIS · M5P · Zubair Formation · Rumila oil field · Iraq

Introduction

Permeability, the ability of rock to transmit fluid, is one of the most important parameter in the petroleum reservoir studies and a key parameter in determining the economic value of hydrocarbon accumulation. The accurate determination of spatial distribution of permeability can have a great effect on production and management of petroleum reservoir. Reservoir permeability is hard to detect parameter as it is difficult to determine directly using current subsurface logging tools (Chehrazi and Rezaee 2012). Nowadays, there are two methods to acquire reliable values of permeability, indirect and direct. In indirect methods, three models were used: (1) models based on theoretical or empirical equations these models seek to predict permeability from porosity as a main rock character and other relevant measurable rock properties. The simplicity of these models associated with sound theoretical foundation are the main advantages while the lack of generalization and dependency on the core measurements are the main disadvantages. (2) Models based on artificial intelligence techniques such as artificial neural networks (ANN), fuzzy logic (FL), adaptive neuro-fuzzy inference system (ANFIS). These models generally used well logs data as input and permeability as output. The ability of these models to simulate the non-linear relationship between permeability and well logs data is the main advantage in addition to many other such as ease to construct, understand, and high flexibility. The over fitting problem is regard as the main disadvantage. Lacking of sufficient training instants is another constrain to implement successful simulation models using these models. (3) Models based on porosity and facies which are further subdivided into reservoir layers, elector-facirs, litho-facies, flow zone indicator, rock fabric approach, and petro-facies (Chehrazi

✉ Amna M. Handhal
amhgeo@gmail.com

[1] Geology Department, Sciences College, Basra University, Basra, Iraq

and Rezaee 2012). The main shortcoming of these models is the subjectivity in defining facies boundaries thought the relationship between permeability and facies in cored sections. In direct methods there are also two major approaches: core measurements and well logging. Among the most promising well logging tools for predicting permeability is the Nuclear Magnetic Resonance (NMR) log (Coates et al. 1997). Unfortunately, most of the wells of the upper member of Zubair Formation in Rumaila oil field in southern of Iraq, the subject of this study, have no records for this log. On the other hand, the direct core measurements of this formation were available for most of the old wells. Prediction of permeability from only porosity measurements (core or well logs data) for the upper productive clastic Zubair Formation is the main objective of this study. The direct relationship between porosity and permeability was firstly checked using a linear mathematical model and further enhance using data mining techniques such as ANFIS and 5MP decision tree techniques. ANFIS is a multi-layer adaptive artificial network-based fuzzy inference system which is originally developed by Jang (1993). In the recent years, this technique has been applied in a number of diverse scientific and engineering fields including system modeling, pattern recognition, financial forecasting and water resource management. This technique is known for their efficiency in dealing with complicated problems with only sets of operational data availability (Zoveidavianpoor et al. 2013). However, the application of ANFIS to explore the relationship between porosity and permeability is still limited. M5P is a recreation of Quinlan's M5 algorithm (Quinlan 1992) for inducing trees of regression models. M5P combine a conventional decision tree with the possibility of linear regression at the nodes. For the best knowledge of author, the application of this method for mimic porosity–permeability relationship is not investigated yet. Thus, in this study an attempted was made to explore the abilities of these techniques for estimating reservoir permeability for the upper member of Zubair Formation in Rumaila oil field of Iraq from only porosity measurements. The study also involves a comparison between the efficacies of these techniques and select the best one.

Modeling techniques

ANFIS

ANFIS is a hybrid soft computing technique which has the potential to capture the benefits of artificial neural network technique (ANN) and fuzzy inference system (FIS). It uses the learning ability of ANN to mapping the input–output relationship and construct the fuzzy rules by determining

the input structure (Zoveidavianpoor et al. 2013). The FIS implements a non-linear mapping from its input space to the output space through a specific number of fuzzy IF–THEN rules, each of which describes the local behavior of the mapping (Pramanik and Panda 2009). The fuzzy membership parameters are optimized by well-known back-propagation algorithm which seeks to minimize some measure of error between networks outputs and desired outputs. The ANFIS consists of five layers, (Fig. 1), the basic functions of each layer are the input, fuzzification, rule inference, normalization, and defuzzification. In Fig. 1, a circle refers to a fixed node, whereas a square marks an adaptive node. In layer 1, crisp inputs will be mapped by using membership functions which varies between 0 and 1 to form a fuzzy set (Pramanik and Panda 2009; Cobaner 2013).

$$O_{1,i} = \mu_{A_i}(x) \quad \text{for } i = 1, 2 \tag{1}$$

$$O_{1,i} = \mu_{B_{i-2}}(y) \quad \text{for } i = 3, 4 \tag{2}$$

where x and y are the crisp input to node i, and A_i, B_i are the membership grades of the membership functions μ_{A_i} and μ_{B_i}, respectively. The μ_{A_i} and $\mu_{B_{i-2}}$ can adopt any fuzzy membership function such as bell-shaped, a Gaussian, or any other available membership functions. In this study, a generalized bell-shaped membership function was used. The output from this kind of membership is computed as:

$$O_{1,i} = \mu_{A_i}(x) = \frac{1}{1 + \left(\frac{x - c_i}{a_i}\right)^{2b_i}} \tag{3}$$

where a_i, b_i, and c_i are the parameters of the membership function.

In layer 2, each node calculates the firing strength O_i^2 of each rule using AND fuzzy operator to fuzzify the inputs. The outputs of this layer can by calculated as:

$$O_{2,k} = w_k = \mu_{A_i}(x)\mu_{B_i}(y) \quad i = 1, 2. \tag{4}$$

The main target of the layer 3 is to compute the ration of firing strength of each ith rule to the sum firing strength of all rules (Bacanli et al. 2009). Outputs of this layer is called normalized firing strengths and calculate as:

$$O_{3,k} = \bar{w}_i = \frac{w_i}{w_1 + w_2}. \tag{5}$$

The output of the fourth layer is simply the product of the normalized firing strength and a first-order polynomial (Zoveidavianpoor et al. 2013). The output of this layer is given by the following equation:

$$O_{4,i} = \bar{w}_i f_i = w_i(p_i x + q_i y + r_i) \quad i = 1, 2 \tag{6}$$

where $\bar{w}_i$ is the output of layer 3 and (p_i, q_i, r_i) is the parameter set and referred to as consequent parameters. The final layer, layer 5, is called output nodes. The single

Fig. 1 ANFIS architecture (redraw after Jang et al. (1997))
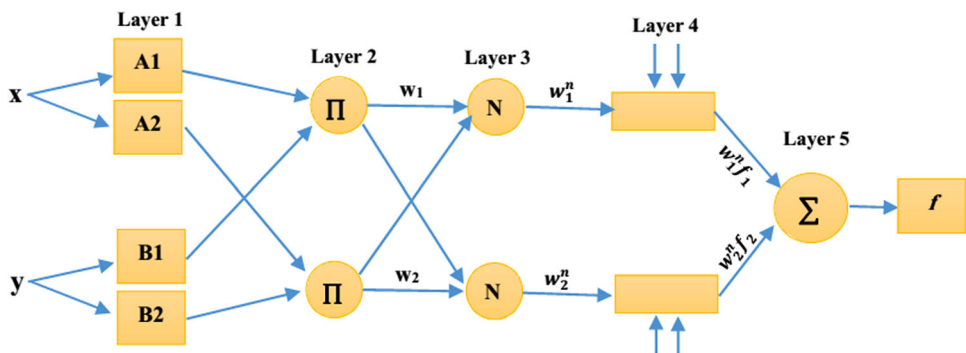
node of this layer computes the overall output by summing all incoming singles:

$$O_{5,l} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}. \tag{7}$$

The ANIFS uses a hybrid learning algorithm (the gradient descent and least square methods) to update all the parameters until acceptable error between actual and desired outputs reach. The detailed mathematical description of these algorithms can be found in Jang et al. (1997) and in Nayak et al. (2004).

M5 decision tree technique

A decision tress is a data mining model that uses an algorithm that identify different ways to split a dataset into branch—like segments. Basically, there are two types of decision tree: classification tree and regression tree. The classification tree is mainly used to predict a symbolic attribute while regression tree is used to predict continuous (numerical) attribute (Witten and Frank 2005). The M5 decision tree algorithm was originally developed by Quinlan (1992). The detail description of this algorithm is out of this article and researchers can be found in Witten and Frank (2005). A short description to illustrate the main idea of this algorithm follows. The M5 algorithm builds a regression trees by recursively splitting the dataset through tests on a single variable that reduce variance on the dependent (target) variable.

The M5 algorithm constructs a regression tress by recursively splitting the instance space using tests on a single attribute that maximally reduce variance in the target variable (Fig. 2). The mathematical formula to compute the standard deviation reduction (SDR) is: (Quinlan 1992).

$$SDR = sd(T) - \sum \left| \frac{T_i}{T} \right| sd(T_i) \tag{8}$$

where is a set of example that reaches the node; T_i is the subset of examples that have the ith outcome of the potential set; and represents the standard deviation.

Once the tree grow, a linear multiple regression is built for every inner node using the data associated with that node and all the attributes that participate for tests in the subtree to that node (Quinlan 1992). To defeat overfitting problem, every subtree is subjected for pruning process. The final step in M5 algorithm is smoothing process. It aims to reward for the sharp discontinuities between adjacent linear models at the leaves of the pruned tree (Witten and Frank 2005).

The study area

The upper part of Zubair Formation in Rumila oil field is the main objective of this study. The Rumaila oil field is a very big oil field located in southern Iraq, approximately (32 km) from the Kuwait border. The field is estimated to contain 17 bbl, which accounts for 12 % for Iraqi's oil reservoir estimated of 143 bbl. The Zubair Formation, the most significant sandstone reservoir in Iraq, is composed of fluvio-deltaic, deltaic and marine sandstones (Aqrawi et al. 2010). The formation is thickest in the type area in south of Iraq, the depocenter is located at the boundary of the Salman and Mesopotamian zones (Jassim and Goff 2006). The Zubair Formation is assumed to represent a prograding delta originating from the Arabian Shield (Ali and Nasser 1989). The sand isolith of the formation in central and S Iraq suggest influx of clastics from the NW in central Iraq and probably from the SW in Iraq. As its type locality in the Zubair field, the formation is divided into five units (Owen and Nasr 1958; Dunnington et al. 1959): These units (from the top): the upper shale (100 m thick), the upper sandstone (the principle reservoir or 'main pay'), and the middle shale, lower sandstone and lower shale (Jamil 1978). The upper sandstone is mainly composed of cross-bedded quartz arenites with thin shale and silt intervals and lignite seams (Aqrawi et al. 2010). The age of formation is earliest Aptian to Hauterivian age (Jamil 1978).

Fig. 2 Example of M5 decision trees technique, 1-6 linear models
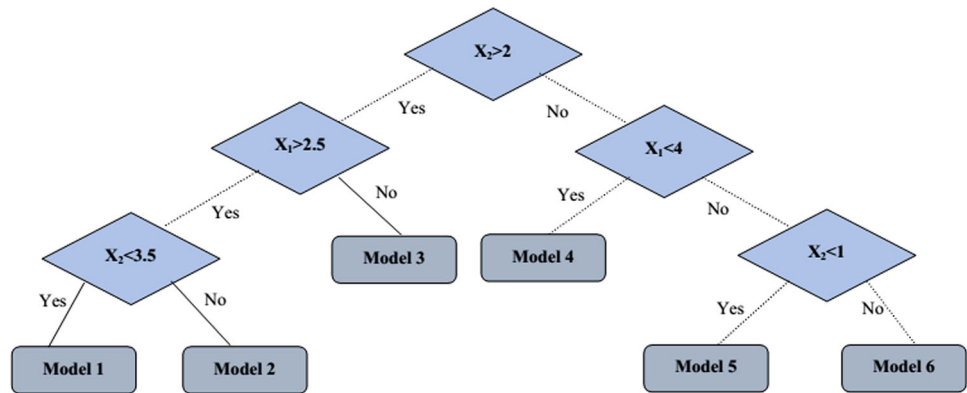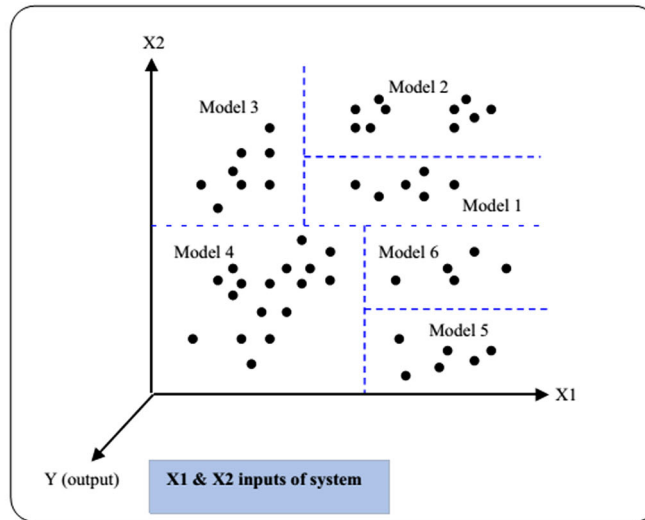
Data used and model performance check error statistics

Data in this study came from the conventional core data from four wells. The studied wells located in suitable distance relative to each other, (Fig. 3). A total number of 149 data points (porosity and permeability measurements) were available, among them 105 data points belongs to the three wells (the training wells) were used to training the ANFIS model, and 44 data points (the testing well) were used for evaluating the performance of the training model for prediction purposes. The statistical summary of the data used were presented in Table 1. It can be seen that the permeability data show higher skewness than porosity data. The minimum and maximum values of permeability in the training phase fall in the range of 0.30–2058 md with an average of 636 md. However, the maximum of the testing permeability data was 3256 md, which may cause difficult in estimating the high permeability values. Also, from Table 1, the variation of permeability data in testing phase

(80 %) is greater than that for training data (77 %). This may cause difficult to estimate permeability values especially for high values. The performance of build models was tested using two error statistics, namely, Root Mean Squared Error (RMSE) and correlation of determination (R^2). These statistics define mathematically as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (K_i - \hat{K}_i)^2}{n}} \tag{9}$$

where K_i is the measured permeability, $\hat{K}_i$ is the predicted permeability, and n number of observed permeability values.

The coefficient of determination is computed as:

$$R^2 = \frac{\sum_{i=1}^{n} (K_i - \hat{Q}_i)^2}{\sum_{i=1}^{n} (K_i - \bar{K}_i)^2} \tag{10}$$

where $\bar{K}_i$ is the mean of the observed permeability.

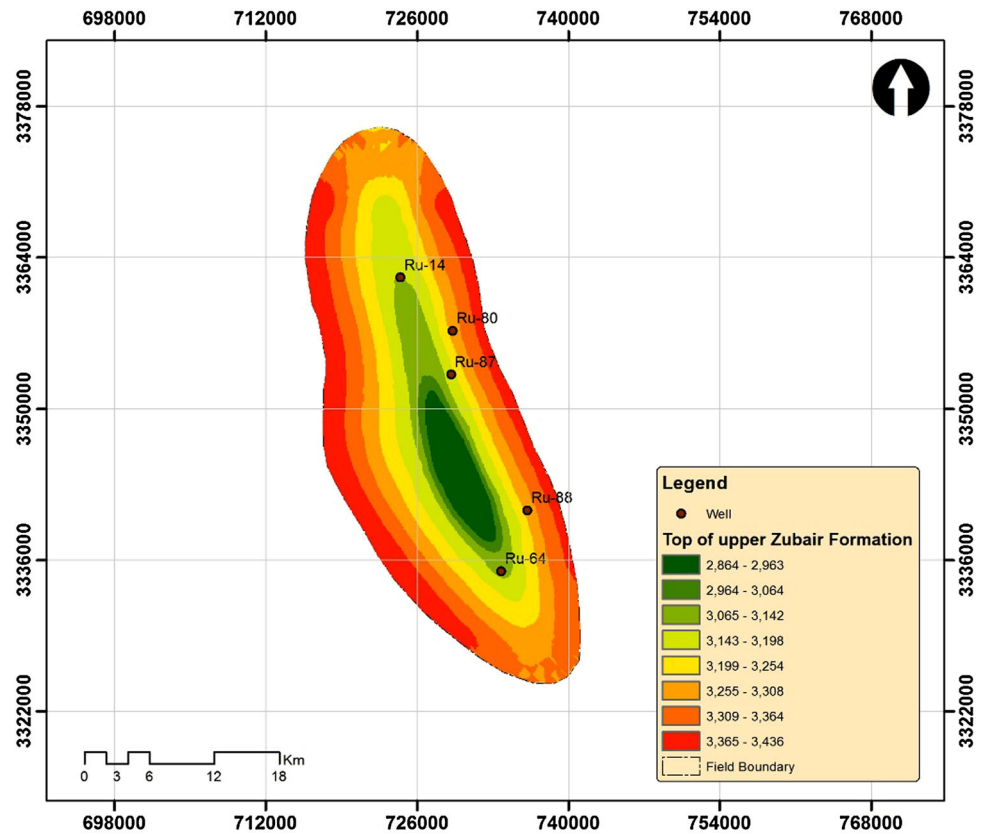**Fig. 3** Top of upper Zubair Formation including location of use wells in this study

Table 1 Statistical summary of variables used in this study

Set	Variable	Statistical parameters					
		Min	Max.	Mean	SD	C_v	Skewness
Training	Porosity	0.096	0.26	0.187	0.035	18.52	−0.65
	Permeability	0.30	2058	635.7	492.3	77.44	0.57
Testing	Porosity	0.096	0.23	0.19	0.031	16.35	−1.15
	Permeability	0	3256	1080	870	80.54	0.69

Min, Max, and Mean are minimum, maximum, and mean of permeability values. SD and Cv are standard deviation and coefficient of variation, respectively. Skewness is skewness coefficient

Results and discussion

The relationship between core porosity as independent variable and permeability as dependent variable was firstly checked using simple linear regression technique. The regression model was build using training data, (Fig. 4), and the resultant equation of linear regression model was used to predict the testing points (blind test), The RMSE and R^2 for the testing phase of this experiment were 0.432 (md) and 0.48, respectively. From this results, it is obvious that linear model was an appropriate to model relationship between porosity and permeability for this heterogeneous reservoir.

In the current study, a simple code along with ANFIS graphical interface of MATLAB 2014a were used to build ANFIS model. For constructing of ANFIS model, the input (porosity) and output (log permeability) for training dataset were initially classified using subtractive clustering technique. The subtractive clustering technique is a useful and effective approach for ANFIS modeling. The main parameter of this technique is setting an optimal clustering radius. The optimal clustering radius was searched manually by gradually increasing the clustering radius from 0 to 1. For each selected radius, the RMSE and R^2 were computed and the ANFIS model with the highest overall accuracy was selected as the optimal model. The number of epochs and error tolerance were set to 100 and 0, respectively. In our case, the cluster radius of 0.2 and fuzzy system with three rules, with RMSE of 0.373 and 0.854 considered as the favorite system. The generated rules are:
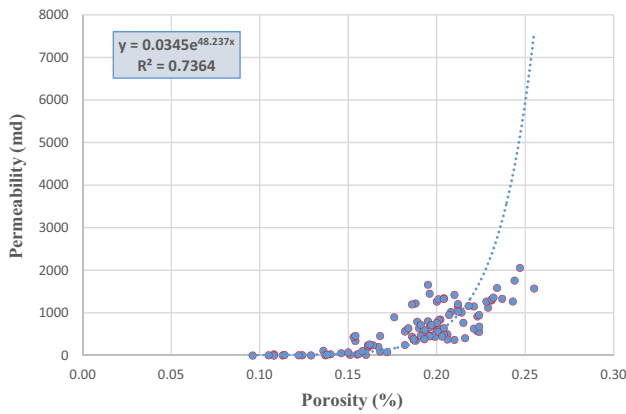
Fig. 4 Linear regression model for porosity–permeability relationship in the study area

1. **If** porosity is moderate **then** permeability is moderate
2. **If** porosity is low **then** permeability is low
3. **If** porosity is high **then** permeability is high.

The fuzzy inference for this model with three Gaussian membership functions is shown in (Fig. 5). The testing data was then passed through ANFIS inference system, (Fig. 6), to predict permeability values. The RMSE and R^2 for testing phase were 0.380 and 0.844, respectively.

The M5 decision tree model was build using Weka 3.6 software. Weka is an open source machine learning software written in Java (Witten and Frank 2005). It is freely distributed across (http://www.cs.waikato.ac.nz/ml/weka/) web page. The default parameters of M5 technique were set to their default values; pruning factor 4.0 and smoothing option. The RMSE and R^2 of applying M5 model for the training phase were 0.378 and 0.848. For the M5 model, the following rule was extracted from M5 algorithm.

M5 pruned model tree: (using smoothed linear models)

Porosity ≤ 17 : LM1 (29/58.279%)

Porosity > 17 : LM2 (75/26.416%).

LM1 = Linear model #1

Log Permeability = $0.3234 \times$ Porosity $- 3.0979$.

LM2 = Linear model #2

Log Permeability = $0.0993 \times$ Porosity $+ 0.929$.

After creating the M5 model for the training data, the testing data was passed through the model to create predict permeability and compare it with measured testing data. The RMSE and R^2 for this experiment were 0.366 and 0.832, respectively.

The comparison between the three methods used in this study, Table 2, revealed that the best technique for computing reservoir permeability was ANFIS. Results also revealed that the M5 model performed better than the linear regression model for both low and high values predictions. The performance of M5 model was very close to
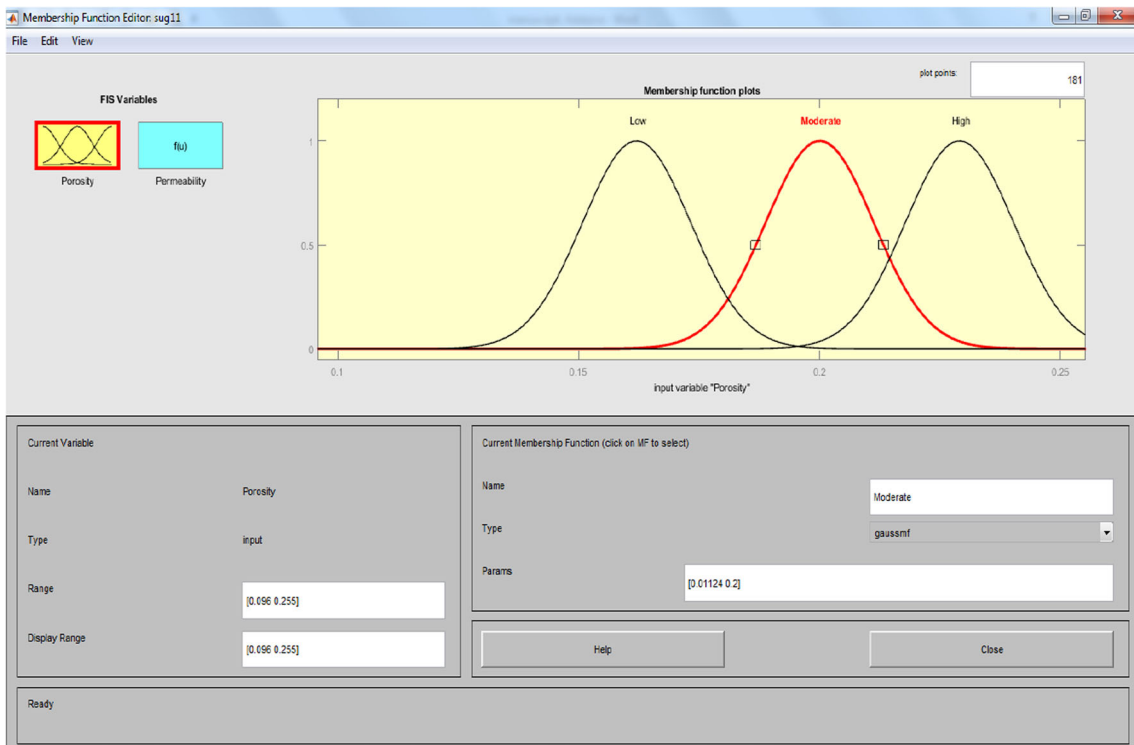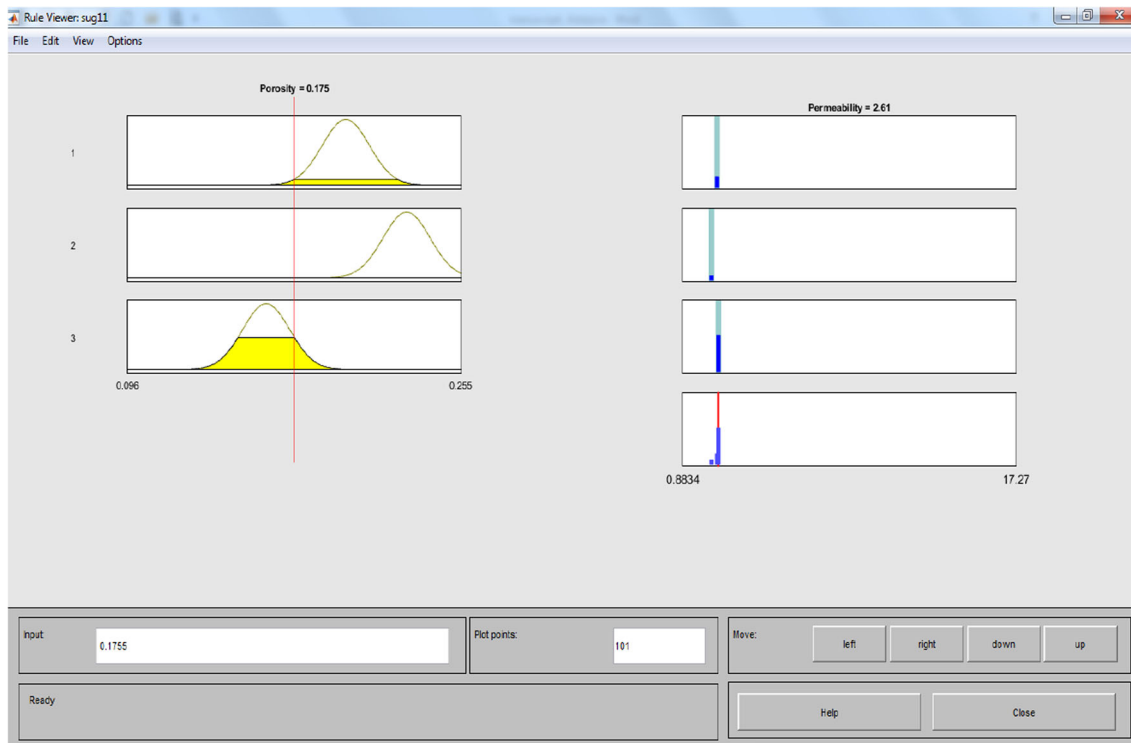

Fig. 5 Membership functions

Fig. 6 Developed fuzzy inference system

Table 2 A comparison among the performance of used approaches

Used methods	Training phase		Testing phase	
	RMSE	R^2	RMSE	R^2
Linear regression	0.401	0.736	0.432	0.480
ANFIS	0.373	0.854	0.380	0.844
M5	0.378	0.848	0.366	0.832

that of ANFIS in both training and testing phases. The simple structure and very fact training M5 algorithm make this technique the best choice for build predictive model. The other benefit of M5 is the generated tree is very easy to understand even from those people who unfamiliar with petrophysics terminology. The final results indicated that both ANFIS and M5 technique were very good models for predicting permeability from porosity for Zubair heterogeneous reservoir in south of Iraq.

Conclusions

In this study, the abilities of three different techniques, namely, linear regression, ANFIS, and M5 decision tree techniques were used for predicting reservoir permeability from core porosity measurements for upper part of Zubair Formation in Rumila oil field in southern Iraq. The study revealed that simulation of this relationship is possible for this heterogeneous reservoir. The ANFIS model perform better than M5 and linear regression models for both low and high values. The M5 model results were very similar to ANFIS for both training and testing data. The complex structure of ANFIS if it compared with a very simple structure and very easy to interpret M5 model make this technique is a best choice for predicting reservoir permeability form core porosity in the study area.

References

Ali AJ, Nasser ME (1989) Facies analysis of the lower Cretaceous oil-bearing Zubair Formation in southern Iraq. Mod Geol 13:225–242

Aqrawi AAM, Goff JC, Horbury AD, Sadooni FN (2010) The Petroleum Geology of Iraq: Scientific Press, Iraq, p 424

Bacanli UG, Firat M, Dikbas F (2009) Adaptive neuro-fuzzy inference system for drought forecasting. Stoch Environ Res Risk Assess 23(8):1143–1154. doi:10.1007/s00477-008-0288-5

Chehrazi A, Rezaee R (2012) A systematic method for permeability prediction, a Petro-Facies approach. J Petrol Sci Eng 82–83:1–16. doi:10.1016/j.petrol.2011.12.004

Coates GR, Menger S, Miller D (1997) Applying total and effective NMR porosity to formation evaluation": SPE Paper 38736, SPE Annual Technical Conference and Exhibition

Cobaner M (2013) Reference evapotranspiration based on Class A pan evaporation via wavelet regression technique. Irrig Sci 31(2):119–134

Dunnington HV, Wetzel R, Morton DM (1959) Mesozoic and Palaeozoic. In: van Bellen RC, Dunnington HV, Wetzel R, Morton DM (eds) Lexique stratigraphique international III Asie, international geological congress commission on stratigraphy (Sous la direction de Louis Dubertret, Paris) fascicule 10a, Iraq. Centre National de le Recherche Scientifique, Paris, p 333

Jamil AK (1978) Hydrochemical and hydrodynamic zones and the probable direction of water flow within the Zubair reservoir of Zubair and Rumaila oil fields (southern Iraq), vol 25. Geologisches Jahrbuch, Reihe D: Mineralogie, Petrographie, Geochemie, Lagerstättenkunde, pp 199–211

Jang S (1993) Adaptive network-based fuzzy inference system". IEEE J 23(3):665–685

Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing. Prentice-Hall, Englewood Cliffs

Jassim SZ, Goff JC (2006) Geology of Iraq. Dolin, Prague and Moravian Museum, Brno **431p**

Nayak PC, Sudheer KP, Rangan DM, Ramasastri KS (2004) A neuro-fuzzy computing technique for modeling hydrological time series". J Hydrol 291(1–2):52–66

Owen RMS, Nasr SN (1958) Stratigraphy of the Kuwait Basrah area. In: Weeks LG (ed) Habitat of oil. Tulsa, APPG, pp 1252–1278

Pramanik N, Panda RK (2009) Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction. Hydrol Sci J 54(2):247–260. doi:10.1623/hysj.54.2.247

Quinlan JR (1992) Learning with continuous classes". In: Proceedings of A192, 5th Australian join conference on artificial intelligence, Singapore. 343–348

Witten IH, Frank E (2005) Data mining. Morgan Kaufmann, USA

Zoveidavianpoor M, Ariffin S, Seyed RS (2013) Prediction of compressional wave velocity by an artificial neural network using some conventional well logs in a carbonate reservoir. J Geophys Eng 10(4):96–107. doi:10.1088/1742-2132/10/4/045014