A Committee Machine with Intelligent Systems for Estimation of Total Organic Carbon Content from Petrophysical Data: an Example from Kangan and Dalan Reservoirs in South Pars Gas Field, Iran

Ali Kadkhodaie-Ilkhchi[1*]; Hossain Rahimpour-Bonab[1]; M.Reza Rezaee[2]

1. Department of Geology, College of Science, University of Tehran, Tehran, Iran

2. Department of Petroleum Engineering, Curtin University of Technology, Perth, Australia

Abstract

Total Organic Carbon (TOC) content present in reservoir rocks is one of the important parameters which could be used for evaluation of residual production potential and geochemical characterization of hydrocarbon bearing units. In general, organic rich rocks are characterized by higher porosity, higher sonic transit time, lower density, higher gamma-ray, and higher resistivity than other rocks. Current study suggests an improved and optimal model for TOC estimation by integration of intelligent systems and the concept of committee machine with an example from Kangan and Dalan Formations, in South Pars Gas Field, Iran. This committee machine with intelligent systems (CMIS) combines the results of TOC predicted from intelligent systems including fuzzy logic (FL), neuro-fuzzy (NF), and neural network (NN), each of them has a weight factor showing its contribution in overall prediction. The optimal combination of weights is derived by a genetic algorithm (GA). This method is illustrated using a case study. One hundred twenty-four data points including petrophysical data and measured TOC from three wells of South Pars Gas Filed were divided into eighty-seven training sets to build the CMIS model and thirty-seven testing sets to evaluate the reliability of the developed

[1*] Corresponding author: Tel. /fax: +98 21 664 91 623.
E-mail addresses: akadkhoda@khayam.ut.ac.ir
akadkhoda@khayam.ut.ac.ir (A. Kadkhodaie-Ilkhchi), rahimpor@khayam.ut.ac.ir (H. Rahimpour-Bonab),
mrezaee@exchange.curtin.edu.au (M.R. Rezaee)
[1] Address: Department of Geology, College of Science, University of Tehran, Iran
Tel. /Fax: +98 21 66491623.
[2] Department of Petroleum Engineering, Curtin University of Technology, ARRC Building, 26 Dick Perry Avenue, Kensington, WA 6151, Australia.
Tel: +618 9266 7980; Fax: +618 9266 7063

model. The results show that the CMIS performs better than any one of the individual intelligent systems acting alone for predicting TOC.

1. Introduction

Hitherto, several researches have tried to make a quantitative and qualitative correlation between well log responses and organic carbon richness of different rocks. Among them, Beers (1945), Swanson (1960), Fertle (1988), Schmoker (1981) and Hertzog et al. (1989) used gamma-ray spectral log to identify organic rich rocks. Schmoker and Hester (1983) proposed the use of density log for estimating organic matter content. A method involving combination of resistivity, density and sonic logs has been introduced by Meyer and Nederlof (1984). This method discriminates between source rocks and non-source rocks without any effort to quantify organic richness from combination of various logs. Passey et al. (1990) invented a technique called ΔlogR. This technique employs the overlaying of porosity logs (Sonic, Density and Neutron) and resistivity log for identifying and calculating total organic carbon. Huang and Williamson (1996) applied neural network modeling for source rock characterization. Lately, Kamali and Mirshady (2004) used ΔlogR and neuro-fuzzy techniques for determining TOC from well log data.

As it is evident from the previous studies, they have mainly focused on one or more techniques, independently. A committee machine has a parallel structure that produces a final output by combining the results of individual experts using an optimization technique (Haykin, 1991; Sharkey, 1996). The experts may be neural networks, empirical formulas, or other algorithms (Chen and Lin, 2006). So, the model accuracy could be increased in comparison with the best individual expert.

The Iranian part of South Pars Gas Field, the world's largest non-associated gas accumulation, is located in the Persian Gulf, between Qatar and Iran at about 100 km from Iranian shoreline. The Upper Permian to Lower Triassic Dalan and Kangan Formations (equivalent of Khuff formation) are two main condensate and gas bearing reservoirs units in this field (Aali et al., 2006, Rahimpour-Bonab, 2007). This study

integrates intelligent systems and the concept of committee to develop an improved and more accurate model for TOC prediction in reservoir intervals with an example from Kangan and Dalan Formations in South Pars Gas Field.

2. Methodology: Committee machine with intelligent systems (CMIS)

The proposed methodology, CMIS, consists of two major steps: at the first stage TOC parameter will be predicted from petrophysical data using intelligent systems including fuzzy logic, neuro-fuzzy and neural networks. At the next step, a CMIS will be constructed using genetic algorithms for TOC prediction. The inputs of CMIS are the outputs of the mentioned intelligent systems. The methodology described in this study provides an improved and novel model for predicting TOC parameter using petrophysical data from two points of view; they are, in use of committee machine concept for predicting TOC parameter and thus reaping the benefit of all works, and genetic algorithms for determining the contributions (weights) of individual algorithms in constructing CMIS. It is clear that many components of the method described in this study are based on other researcher's works which are not novel in their own right. For example, neural networks and fuzzy logic are well known techniques. Overall, the integrated technique described in this study could be considered as an efficient and instrumental way for predicting TOC parameter.

2.1. Intelligent systems

There are several intelligent systems which depending on the problem to be solved, could be used for modeling and prediction in different disciplines of science. In this study, genetic algorithms, fuzzy logic, neural network and neuro-fuzzy methods are employed to construct a committee machine for modeling and predicting TOC from petrophysical data. Following section, presents a brief description of the intelligent systems used in this study.

Fuzzy logic (FL): The basic theory of fuzzy sets was first introduced by Zadeh (1965). In recent years, it has been shown that uncertainty may be due to fuzziness (possibility) rather than probability. Fuzzy logic is considered to be appropriate to deal with the nature of uncertainty in system and human errors, which were not considered in existing reliability theories (Nikravesh and Aminzadeh, 2003). Generally, geological data are not

clear-cut and habitually are associated with uncertainties. For example, prediction of core parameters from well log responses is difficult and is usually associated with error (Kadkhodaie et al., 2006). Fuzzy logic derives useful information from this error and applies it as a powerful parameter for increasing the accuracy of the predictions.

Fuzzy inference system (FIS) is a method to formulate inputs to an output using fuzzy logic (Matlab user's guide, 2004). Takagi-Sugeno method (1985) is a FIS in which output membership functions are constant or linear and are extracted by a clustering process. Each of these clusters refers to a membership function. Each membership function generates a set of fuzzy if-then rules for formulating inputs to outputs. A schematic diagram of FIS is shown in figure 1.

Back propagation neural networks (BP-NN): Neural network is an intelligent tool for solving complex problems. A BP-NN is a supervised training technique that sends the input values forward through the network then computes the difference between calculated output and corresponding desired output from the training dataset. The error is then propagated backward through the net and the weights are adjusted during a number of iterations, named epochs. The training ceases when the calculated output values best approximate the desired values (Bhatt and Helle, 2002). A flowchart of training procedure in a supervised neural network is shown in figure 2.

Neuro-fuzzy (NF): Hybrid neuro-fuzzy systems combine the advantages of fuzzy systems (which deal with explicit knowledge) with those of neural networks (which deal with implicit knowledge). On the other hand, fuzzy logic enhances generalization capability of a neural network system by providing more reliable output when extrapolation is needed beyond the limits of the training data. A schematic diagram of information flows in a NF system is shown in figure 3.

Genetic algorithms (GA): GAs are effective search methods based on the principles of natural selection and genetics. They were developed by John Holland (1975) to simulate some of the processes observed in natural evolution. Generally, a GA works on a set of potential solutions for a specific problem encoded into chromosome-like data structures. Some of these solutions, chosen on the basis of their performance in solving the problem, are used to create a new set of potential solutions through the use of operators. A GA uses this process repeatedly until a particular criterion is met. GAs are often described in

biological terms. Potential solutions are called chromosomes and represented by binary strings or floating point numbers. A set of chromosomes is called a population and a problem to be solved is represented by a fitness function. Choosing the individuals to be reproduced is performed in a process called selection which is based on the fitness values assigned to chromosomes. Genetic operators such as crossover and mutation are operators used to create a new population. Crossover permits the exchange of information among individuals in the population and provides the innovative capability of a GA. Mutation ensures desirable diversity (Reformat, 1997). A general flowchart of a GA is shown in figure 4.

2.2. Committee machine with intelligent systems (CMIS)

A CMIS consists of a group of intelligent systems which combines the outputs of each system and thus reaps the benefits of all works, with little additional computation. Thus, performance of the model could be better than the best single network. A schematic diagram of CMIS is shown in figure 5. There are different ways of combining the intelligent systems outputs in the combiner. The simple ensemble averaging method is the most popular one (Naftaly et al., 1997, Chen and Lin, 2006). Proper combination of the intelligent systems contribution (weight) in a CMIS could be obtained by a GA.

The section below describes the fundamentals of our CMIS with regard to the works of Bates and Granger (1969), Haykin (1991), Geman et al. (1992), Naftaly et al. (1997), Huang et al. (2001), Ligtenberg and Wansink (2001), Bhatt and Helle (2002), Lim (2005), and Chen and Lin (2006):

Assumption is that there are N trained intelligent systems with output vector o_i which are used to predict target vector T. The prediction error could be written as

$$e_i = o_i - T ,$$ (1)

$i = 1, ..., N$

The expectation of the squared error for the ith intelligent system o_i is

$$E_i = \xi[(o_i - T)^2] = \xi[e_i^2],$$ (2)

in which $\xi[.]$ is the expectation. The average error for each of the intelligent systems acting alone is

4

$$E_{avg} = \frac{1}{N} \sum_{i=1}^{N} E_i = \frac{1}{N} \sum_{i=1}^{N} \xi[e_i^2], \tag{3}$$

Applying the averaging method, output vector o_i of the CMIS is

$$O_{CMIS} = \frac{1}{N} \sum_{i=1}^{N} o_i . \tag{4}$$

Therefore, the CMIS has the prediction squared error:

$$E_{CMIS} = \xi[(O_{CMIS} - T)^2] = \xi[(\frac{1}{N} \sum_{i=1}^{N} o_i - T)^2] = \xi[(\frac{1}{N} \sum_{i=1}^{N} e_i)^2]. \tag{5}$$

Considering Cauchy's inequality:

$$(a_1 b_1 + a_2 b_2 + ... + a_n b_n)^2 \leq (a_1^2 + a_2^2 + ... + a_n^2).(b_1^2 + b_2^2 + ... + b_n^2) \tag{6}$$

and applying it to the E_{CMIS}

$$E_{CMIS} = \xi[(\frac{1}{N} \sum_{i=1}^{N} e_i)^2] \leq \frac{1}{N} \sum_{i=1}^{N} \xi[e_i^2] = E_{avg}. \tag{7}$$

which indicates that the CMIS gives smaller errors than the average of all the intelligent systems.

3. Physical relationships between TOC and input petrophysical data

There is a logical relationship between petrophysical data used and total organic carbon content present in reservoir rocks (Fig. 6a-e). According to Fig. 6a-c, petrophysical data including GR, NPHI and DT show a direct relationship with TOC. Generally, Rlld increases as TOC decreases. This relationship is reverse for RHOB data. But according to Fig.6d-e, these relationships may not be seen obviously for RHOB and Rlld data. These could occur due to reasons such as rock heterogeneities, mineralogy changes, variations in the fluid content and saturation. Section below describes physical fundaments of these relationships, briefly.

Gamma ray tool measures the radioactivity of various formations. Generally, organic rich rocks have high concentrations of radioactive elements including Potassium, Thorium, and Uranium and increase the reading of gamma ray log.

Neutron log reading is a response of hydrogen atoms concentration in rocks. The volume of organic matter in the formation has a direct relationship with hydrogen atoms content and porosity of the rock. Thus, neutron porosity increases in the organic rich intervals.

The sonic transit time (DT) is the reciprocal of the velocity of the compressional wave and is a function of formation lithology, porosity, type and distribution models of fluids (water, gas, oil, kerogen, etc.). With apparent DT value increase TOC content tends to elevate (Kamali and Mirshady, 2004).

Density log measures the bulk density of the formation, a response of fluids and matrix constituent minerals density. Organic matters have a low density (about 1 gr/cm^3) and their concentration tends to reduce the bulk density of the rock.

The resistivity log indirectly measures rock resistivity through variations in fluid saturation. Because, fluid content is a major control on the rock resistivity. Generally, organic matter bearing layers have higher resistivity than the other rocks. Specially, it is true when kerogen becomes mature and generates hydrocarbon filling pore spaces.

4. Case study

In this study intelligent systems including NN, FL, NF and a CMIS were used to predict TOC from petrophysical data including GR, NPHI, DT, RHOB, and Rlld. For this purpose, one hundred twenty-four samples from the logged intervals of Kangan and Dalan reservoirs of South Pars Gas Field were collected for Rock-Eval pyrolysis and measuring TOC content. In the Rock-Eval pyrolysis method during a programmed temperature heating (in a pyrolysis oven) in an inert atmosphere (helium) a small sample (~100 mg) is heated. In this experiment geochemical parameters of the rock, from which TOC is extracted, is determined, quantitatively.

The dataset were divided into eighty-seven training sets to build the intelligent models, and thirty-seven testing sets to evaluate the reliability of the models.

4.1. Predicting TOC by intelligent systems

Takagi-Sugeno FIS (TS-FIS) model: For construction of a TS-FIS model, at the first stage it is necessary to classify input and output dataset into groups, named cluster. In this study, a subtractive clustering method which is a useful and effective way to fuzzy logic modeling (Chiu, 1994; Jarrah and Halawani, 2001), is used for extraction of clusters and fuzzy if-then rules. The details of subtractive clustering could be found in Chiu (1994), Chen and Wang (1999), Jarrah and Halawani (2001). The important parameter in

subtractive clustering which controls number of clusters and fuzzy if-then rules is clustering radius. This parameter could take values between the range of [0,1]. Specifying a smaller cluster (say 0.1) radius will usually yield more and smaller clusters in the data resulting in more rules. In contrast, a large cluster radius (say 0.9) yields a few large clusters in the data (Chiu, 1994) resulting in few rules.

Effectiveness of a fuzzy model is related to search for optimal clustering radius which is a controlling parameter for determing the number of fuzzy if-then rules. Few rules couldn't cover the entire domains, and more rules will complicate the system behavior and may lead to low performance of the model. Searching for optimal clustering radius is done by performing clustering process for several times on input and output data. At each time, clustering radius is chosen as a value between [0,1]. Thus, several fuzzy models with different number of if-then rules could be established. Then, the fuzzy model with the lowest mean squared error (MSE) is selected as the optimal model for output estimation problem. According to graphs of figure 7, choosing value of 0.6 for clustering radius is associated with the lowest (MSE) for the test well (e.g. 0.000469) and this generates eight fuzzy if-then rules. Thus, the TS-FIS model was established by eight membership functions (clusters) for input and output data resulting in eight rules (Fig. 8). According to figure 9a-e, which shows the generated membership functions for TOC modeling, Gaussian membership function is fitted to the extracted input clusters. Gaussian function $f(x)$ shows the normal distribution of data (x):

$$f(x) = \frac{e^{-(x-\mu)^2/\sigma^2}}{\sigma\sqrt{2\pi}}$$

(8)

where μ and σ are the parameters of normal distribution showing the mean and standard deviation of data, respectively. These Gaussian membership functions are constructed from mean and sigma values of the clusters. Mean is cluster centers and sigma is derived from sigmas = (radii * (max (data) – min (data))) / sqrt (8.0). The Gaussian membership function parameters are shown in Table 1a.

In TS-FIS, output membership functions are linear equations constructed from inputs. For example, output MF1 which is the consequent of rule no. 1 is constructed from five petrophysical inputs as below:

Output MF1= $c_{1*}GR + c_{2*}NPHI + c_{3*}DT + c_{4*}RHOB + c_{5*}Rlld + c_6$

In this equation parameters c_1, c_2, c_3, c_4 and c_5 are coefficients corresponding to GR, NPHI, DT, RHOB and Rlld inputs, respectively. Parameter c_6 is constant in each equation. These parameters are obtained by linear least squares estimation. With these explanations there will be 6 parameters for each output membership function which are shown in Table 1b.

The generated fuzzy if-then rules for formulating input petrophysical data to TOC are as below:

1. *If* (GR is mf4) *and* (NPHI is mf6) *and* (DT is mf7) *and* (RHOB is mf2) *and* (Rlld is mf5) *then* (TOC is mf1)

2. *If* (GR is mf6) *and* (NPHI is mf8) *and* (DT is mf6) *and* (RHOB is mf4) *and* (Rlld is mf7) *then* (TOC is mf2)

3. *If* (GR is mf3) *and* (NPHI is mf3) *and* (DT is mf4) *and* (RHOB is mf5) *and* (Rlld is mf2) *then* (TOC is mf3)

4. *If* (GR is mf2) *and* (NPHI is mf7) *and* (DT is mf8) *and* (RHOB is mf1) *and* (Rlld is mf3) *then* (TOC is mf4)

5. *If* (GR is mf7) *and* (NPHI is mf1) *and* (DT is mf1) *and* (RHOB is mf7) *and* (Rlld is mf8) *then* (TOC is mf5)

6. *If* (GR is mf8) *and* (NPHI is mf5) *and* (DT is mf2) *and* (RHOB is mf8) *and* (Rlld is mf1) *then* (TOC is mf6)

7. *If* (GR is mf5) *and* (NPHI is mf2) *and* (DT is mf3) *and* (RHOB is mf6) *and* (Rlld is mf6) *then* (TOC is mf7)

8. *If* (GR is mf1) *and* (NPHI is mf4) *and* (DT is mf5) *and* (RHOB is mf3) *and* (Rlld is mf4) *then* (TOC is mf8)

Figure 10 is a graphical illustration showing the processing steps to use TS-FIS for predicting TOC from input petrophysical data. The performance of the fuzzy model was measured as 6.01864E-05.

After construction of the fuzzy model, input matrix of the test data (GR, NPHI, DT, RHOB, and Rlld) was exposed to the TS-FIS model and TOC was calculated. Measured error using MSE function is 0.000469 and the R^2 between measured and FL predicted TOC is 0.768 (Fig. 11). A comparison between measured and FL predicted TOC versus depth in the test data is shown in figure 12.

NN model: A three layered NN with back propagation algorithm was used for TOC prediction. The dataset were divided into three groups including training (61 data points), validation (26 data points) and testing data (37 data points) from three wells. Similar to TS-FIS, five inputs including GR, NPHI, DT, RHOB, and Rlld logs data were used in the first layer, respectively. Number of neurons in the hidden layer was 3 and in the output layer one neuron was included for TOC data. The schematic diagram of BP-ANN designed in this study is shown in figure 13. Aforementioned three layered neural network was trained using the Levenberg-Marquardt training algorithm (TrainLM) which details of its computation process and training could be found in Boadu (1997, 1998) and

Bishop (1995). The default MSE performance function was applied to optimize weights and default bias values. Transfer function from layer one to two is TANSIG and from layer two to layer three is PURELIN. The selected epochs were 100 and the error goal was set to 0. In neural networks, training stops until a maximum number of epochs occurs or the performance goal is met. After 82 epochs of training, MSE performance function was fixed in 5.39785E-05 (Fig. 14). Thus, until 100 epochs no improvement was seen in the network performance and weights adjustment was stopped after 82 epochs. When the training and optimization of the model was finished, the input well log data of the test data were exposed to it and TOC was calculated. Measured error using MSE function is 0.000336 and the R^2 between measured and NN predicted TOC is 0.801 (Fig. 15). A comparison between measured and NN predicted TOC versus depth for the test data is shown in figure 16.

NF model: In this study, an adaptive neuro-fuzzy inference system was developed to optimize the fuzzy model. The NF model is quite similar to fuzzy model, but all input and output membership functions were derived by a grid partitioning method.

Number of input (Gaussian) and output (constant) membership functions is twenty eight. A back propagation neural network was applied to optimize and adjust the membership functions parameters. Error tolerance of the NF model was set to 0. After 4 epochs of training no change was found in the model performance and the MSE of the NF model was fixed in 5.87121E-05. A formulation of inputs to TOC using NF is shown in figure 17. Once the model was constructed, inputs matrix of the test data including GR, NPHI, DT, RHOB, and Rlld was exposed to the NF network and TOC was calculated. Measured error using MSE function is 0.000361 and the R^2 between measured and NF predicted TOC is 0.780 (Fig. 18). A comparison between measured and NF predicted TOC versus depth for the test data is shown in figure 19.

4.4. Predicting TOC by CMIS

In this part of research, a CMIS was constructed for the overall prediction of TOC by integrating the results of predicted TOC from TS-FIS, NN, and NF, each of them has a weight factor showing its contribution in overall prediction. At the first step, the CMIS was constructed by applying simple averaging method. In this approach, any one of the

intelligent systems has equal contribution in constructing CMIS, namely each of them has the weight value of 0.333. Thus, calculated TOC from CMIS was obtained by following equation:

$$TOC_{CMIS} = 0.333 * TOC_{from\ TS-FIS} + 0.333 * TOC_{from\ NN} + 0.333 * TOC_{from\ NF} \qquad (9)$$

Applying this equation has provided the R^2 value of 0.828 and MSE of 0.000271.

In the next step, a genetic algorithm was used to obtain optimal combination of the weights for constructing CMIS. The fitness function for GA was defined as below:

$$MSE_{CMIS} = \sum_{i=1}^{n} 1/n (w_1 o_{1i} + w_2 o_{2i} + w_3 o_{3i} - T_i)^2 \qquad (10)$$

This function shows the MSE of CMIS for training step predictions where w_1, w_2 and w_3 are the weight factors corresponding to TS-FIS (o_{1i}), NN (o_{2i}), and NF (O_{3i}) predictions, respectively. T_i is the target values (measured TOC) and n is the number of training data (87 samples). Parameters of applied GA are described as following:

Initial population size is 20 which specifies how many individuals are in each generations and initial range is [0,1] which specifies the range of the vectors in the initial population. The crossover function is *scattered* that creates a random binary vector and selects the genes where the vector is [1] from the first parent, and the genes where the vector is [0] from the second parent, and combines the genes to form a child. The value of crossover fraction is 0.8. This parameter specifies the fraction of the population that could be seen in the crossover children. Mutation function is *Gaussian* that adds a random number, or mutation, from a Gaussian distribution, to each entry of the parent vector. Parameters controlling the mutation are specified as the *scale value* of 1 and *shrink value* of 1. The *scale value* controls the standard deviation of the mutation at the first generation. This parameter is multiplied by the range of the initial population. *Shrink value* controls the rate at which the average amount of mutation decreases. The standard deviation decreases linearly so that its final value equals 1. Stopping generation of GA was chosen as 120. After 120 generations, change in the fitness function values over Stall generations was insignificant and the mean and best fitness values were fixed in 0.000202 and 0.000198, respectively (Fig. 20a). The best, worst and mean scores within the mentioned generations are shown in figure 20b. Finally, after 120 generations the GA derived values

for w_1, w_2 and w_3 are 0.232, 0.482 and 0.205, respectively. The diagram of CMIS designed in this study is shown in figure 21.

Then, the weights obtained from GA were applied to the test dataset predictions of intelligent systems and overall estimation of TOC by CMIS was calculated as below:

$$TOC_{CMIS} = 0.232 * TOC_{from\ TS-FIS} + 0.482 * TOC_{from\ NN} + 0.205 * TOC_{from\ NF} \qquad (11)$$

The MSE of predicted TOC from CMIS is 0.000197 that shows a significant improvement in comparison with TS-FIS, NN, NF, and R^2 between measured and CMIS predicted TOC is increased to 0.845 (Fig. 22). Also it shows some improvement in comparison with constructed CMIS by simple averaging method. A comparison between measured and CMIS predicted TOC versus depth for the test data is shown in figure 23.

5. Conclusions

Intelligent systems including TS-FIS, NN, NF and CMIS were used for the estimation of TOC from petrophysical data in South Pars Gas Field. Regarding the results of the models developed in this research, following points are concluded:

(a) Intelligent systems have been successful for making a quantitative correlation between TOC and petrophysical data. The MSE of TS-FIS, NN and NF methods for estimation of TOC in the test data are 0.000469, 0.000336 and 0.000361 which correspond to the R^2 values of 0.768, 0.801 and 0.780, respectively. Among the intelligent systems which are used, NN model has provided more accurate results than those of the others in the test data.

(b) All the three intelligent systems which were used overestimated the most extreme values of TOC. Such a problem could occur due to several reasons as follows:

- Intelligent systems used in this study try to learn the relationships between a set of input petrophysical logs and a desired output (TOC). Any unwanted error in their training data in the extreme values of TOC may have similar effects on the intelligent models. These lead to construction of the models which perform well on training data but perform poorly on test data.

- In this study, number of measured TOC data in extreme values was limited. So, there were not sufficient data in this range for training three intelligent models. This problem associated with rock heterogeneities such as changes in mineralogy, fluid content and

saturation could lead to unusual responses of the intelligent systems at extreme value (over-estimation or under-estimation).

(c) The intelligent systems which were combined in a committee machine, had a weight factor showing its contribution to constructing CMIS. Applying simple averaging method for combination of the weights has provided the R^2 value of 0.828 and MSE of 0.000271 which shows some improvement in comparison with individual intelligent systems.

(d) The optimal combination of the weights in CMIS was obtained by a GA. The GA derived weights for TS-FIS, NN, and NF experts are 0.232, 0.482 and 0.205, respectively. MSE of the CMIS for the test data is 0.000197 which corresponds to the R^2 value of 0.845. This indicates that CMIS had a significant improvement for the estimation of TOC from petrophysical data. Therefore, CMIS performs better than any one of the individual intelligent systems acting alone for TOC predicting problem. Also it has provided better results than the constructed CMIS by simple averaging method.

(e) Due to high costs of Rock-Eval pyrolysis, limited number of samples was used in this study. However, intelligent systems (especially neural network) predictions for TOC were satisfying. So, it could be concluded that when there is a logical relationship between input and output data (such as those mentioned for TOC and petrophysical data), intelligent systems could recognize the patterns even with limited data.

(f) The CMIS introduced in this study is able to estimate TOC from well log data for other wells of South Pars Gas Field which have not been cored or their TOC are not measured..

(g) CMIS has a simple and easy structure and when there are multiple ways to solve a problem, it could provide smaller errors than the average of all experts by combining the outputs of each method.

7. References

Aali, J., Rahimpour-Bonab, H., Kamali, MR, 2006. Geochemistry and origin of natural gas in world's largest non- associated gas field. Journal of Petroleum Science and Engineering 50, 163-175.

Bates, J.M., Granger, C.W.J., 1969. The combination of forecast. Operations Research Quarterly 20, 451–468.

Beers, R.F., 1945. Radioactivity and organic content of some Paleozoic shales. American Association of Petroleum Geologists Bulletin 26, 1–22.

Bhatt, A., Helle, H.B., 2002. Committee neural networks for porosity and permeability prediction from well logs. Geophysical Prospecting 50, 645–660.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford, 670 pp.

Boadu, F.K., 1997. Rock properties and seismic attenuation: neural network analysis. Pure and Applied Geophysics 149, 507–524.

Boadu, F.K., 1998. Inversion of fracture density from field seismic velocities using artificial neural networks. Geophysics 63, 534 545.

Chen, C.H., Lin, Z.S., 2006. A committee machine with empirical formulas for permeability prediction. Computers & Geosciences 32, 485-496.

Chiu, S., 1994. Fuzzy model identification based on cluster estimation. Journal of Intelligent and Fuzzy Systems 2, 267–278.

Chen, M.S., Wang, S.W., 1999. Fuzzy clustering analysis for optimizing fuzzy membership functions. Fuzzy Sets and Systems 103, 239-254.

Fertle, H., 1988. Total organic carbon content determined from well logs. Society of Petroleum Engineers Formation Evaluation 15612, 407– 419.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural Computation 4, 1–58.

Haykin, S., 1991. Neural Networks: A Comprehensive Foundation. Prentice Hall, Englewood Cliffs, New Jersey, 842 pp.

Hertzog, R., Colson, L., Seeman, B., O'Brian, M., Scott, H., Mckeon, D., Wraight, P., Grau, J., Schweitzer, J., Herron, M., 1989. Geochemical logging with

spectrometry tools. Society of Petroleum Engineers Formation Evaluation 4, 153– 162.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor, USA, 183 pp.

Huang, Z., Williamson, M.A., 1996. Artificial neural network modeling as an aid to source rock characterization. Marine and Petroleum Geology 13, 227-290.

Huang, Y., Gedeon, T.D., Wong, P.M., 2001. An integrated neural-fuzzy-genetic-algorithm using hyper-surface membership functions to predict permeability in petroleum reservoirs. Engineering Applications of Artificial Intelligence 14, 15-21.

Jarrah, O.A., Halawani, A., 2001. Recognition of gestures in Arabic sign language using neuro-fuzzy systems. Artificial Intelligence 133, 117–138.

Kadkhodaie Ilkhchi, A., Rezaee, M.R., Moallemi, S.A. 2006. A fuzzy logic approach for the estimation of permeability and rock types from conventional well log data: an example from the Kangan reservoir in Iran Offshore Gas Field, Iran. Journal of Geophysics and Engineering 3, 356-369.

Kamali, M.R., Mirshady, A.A., 2004. Total organic carbon content determined from well logs using ΔlogR and neuro fuzzy techniques. Journal of Petroleum Science and Engineering 45, 141–148.

Ligtenberg, J.H, Wansink, A.G., 2001. Neural network prediction of permeability in the EL Garia Formation, Ashtart Oilfield, Offshore Tunisia. Journal of Petroleum Geology 24, 389-404.

Lim, J-S, 2005. Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea. Journal of Petroleum Science and Engineering 49, 182-192.

Matlab user's guide, 2004. Version 4, Fuzzy Logic Toolbox, Math works, USA, 235 p.

Meyer, B.L., Nederlof, M.H., 1984. Identification of source rocks on wireline logs by density/resistivity and sonic transit time/resistivity cross plots. American Association of Petroleum Geologists Bulletin 68, 121–129.

Naftaly, U., Intrator, N., Horn, D., 1997. Optimal ensemble averaging of neural networks. Computation in Neural Systems 8, 283–296.

Nikravesh, M., Aminzadeh, F., 2003. Soft Computing and Intelligent Data Analysis in Oil Exploration. Part1: Introduction: Fundamentals of Soft Computing. Elsevier, Berkeley, USA. 744 pp.

Passey, O.R., Moretti, F.U., Stroud, J.D., 1990. A practical modal for organic richness from porosity and resistivity logs. American Association of Petroleum Geologists Bulletin 74, 1777-1794.

Rahimpour-Bonab, H., 2007. A procedure for appraisal of a hydrocarbon reservoir continuity and quantification of its heterogeneity. Journal of Petroleum Science and Engineering 58, 1-12.

Reformat, M., 1997. Application of Genetic Algorithms in Control Design for Advanced Static VAR Compensator at ac/dc Interconnection. University of Manitoba Press, Canada, 129 pp.

Schmoker, J.W., 1981. Determination of organic-matter content of Appalachian Devonian shales from gamma-ray logs. American Association of Petroleum Geologists Bulletin 65, 2165–2174.

Schmoker, J.W., Hester, T.C., 1983. Organic carbon in Bakken Formation, United States portion of Williston Basin. American Association of Petroleum Geologists Bulletin 67, 2165-2174.

Sharkey, A.J.C., 1996. On combining artificial neural nets. Connection Science 8, 299-314.

Swanson, V.E., 1960. Oil yield and uranium content of black shales. United States Geological Survey Professional Paper 356-A, 1-44.

Takagi, T., Sugeno, M., 1985. Identification of systems and its application to modeling and control. IEEE Transaction on Systems, Man and Cybernetics 15, 116–132.

Zadeh. L. A., 1965. Fuzzy Sets. Information and Control 8, 338-353.

Figure captions

Fig. 1 A schematic diagram of FIS.

Fig. 2 A flowchart of training procedure in a supervised neural network (Matlab user's guide, 2004).

Fig. 3 A schematic diagram of information flow in a NF system.

Fig. 4 A general flowchart of a genetic algorithm (from Chen and Lin, 2006).

Fig. 5 A schematic diagram of CMIS.

Fig. 6 Crossplots showing relationship between measured TOC content and GR (a), NPHI (b), DT (c), RHOB (d) and Rlld (e) in well A of South Pars Gas Field.

Fig. 7 Graphs showing clustering radius versus number of generated fuzzy if-then rules (dotted blue) and model MSE (continuous black). Choosing value of 0.6 is associated with lowest MSE (0.000469) resulting in eight fuzzy if-then rules.

Fig. 8 Diagram showing formulation of input petrophysical logs to TOC by eight rules generated by TS-FIS ("afinal: name of fuzzy model", "Sugeno: type of fuzzy model")

Fig. 9 TS-FIS generated Gaussian membership functions for GR (a), NPHI (b), DT (c), RHOB (d), and Rlld (e).

Fig. 10 A graphical illustration showing steps to formulation of petrophysical data inputs to TOC using eight fuzzy if-then rules generated by TS-FIS. *1. Fuzzify inputs:* FIS takes inputs and determines degree to which inputs belong to each membership function. *2. Truncation by fuzzy operators:* Applying fuzzy operators gives a value to antecedent of each rule, and then output membership function is truncated by this value (here "and" or "minimum" operator). *3. Aggregation*: In this step, outputs of each rule that fit into a fuzzy set are combined into a single fuzzy set. *4. Defuzzify:* TS-FIS uses a defuzzification method from output of aggregation method (in this study, a weighted average) which is a crisp numerical value.

Fig. 11 Crossplot showing correlation coefficient between measured and FL predicted TOC.

Fig. 12 A comparison between measured and FL predicted TOC versus depth in test data.

Tables

Table 1 Table showing input **(a)** and output **(b)** membership functions parameters derived by TS-FIS

Table 1a

Inputs		GR (Api)		NPHI (v/v)		DT (us/ft)		RHOB (gr/cm^3)		Rlld (Ohm-m)	
Parameters		Σ	μ	Σ	μ	σ	μ	σ	μ	σ	μ
Input MF No.	mf1	2.83	20.84	0.036	0.013	5.93	51.35	0.134	2.04	191.9	54.40
	mf2	2.83	21.42	0.036	0.019	5.93	54.41	0.134	2.17	191.9	152.4
	mf3	2.83	21.90	0.036	0.091	5.93	59.63	0.134	2.23	191.9	216.7
	mf4	2.83	22.47	0.036	0.109	5.93	66.74	0.134	2.25	191.9	260.6
	mf5	2.83	23.88	0.036	0.120	5.93	70.89	0.134	2.42	191.9	309.2
	mf6	2.83	25.31	0.036	0.134	5.93	73.16	0.134	2.52	191.9	397.9
	mf7	2.83	25.78	0.036	0.140	5.93	74.05	0.134	2.61	191.9	621.3
	mf8	2.83	30.80	0.036	0.149	5.93	76.57	0.134	2.67	191.9	916.3

Table 1b

Output		TOC (Wt%)					
		c_1	c_2	c_3	c_4	c_5	c_6
Output MF No.	mf1	-0.01	31.57	-0.046	-0.22	-0.01	0
	mf2	-0.01	-3.18	0.01	-0.27	-0.01	1.30
	mf3	-0.04	18.29	0.69	-19.42	0.01	0
	mf4	0.20	0	0.029	-2.9	-0.01	0
	mf5	0	0	0	0	0.01	0
	mf6	0	0	0.01	0	0	0
	mf7	0	0	0.11	0	-0.02	0
	mf8	-0.29	0	0.34	-14.02	0.01	11.23
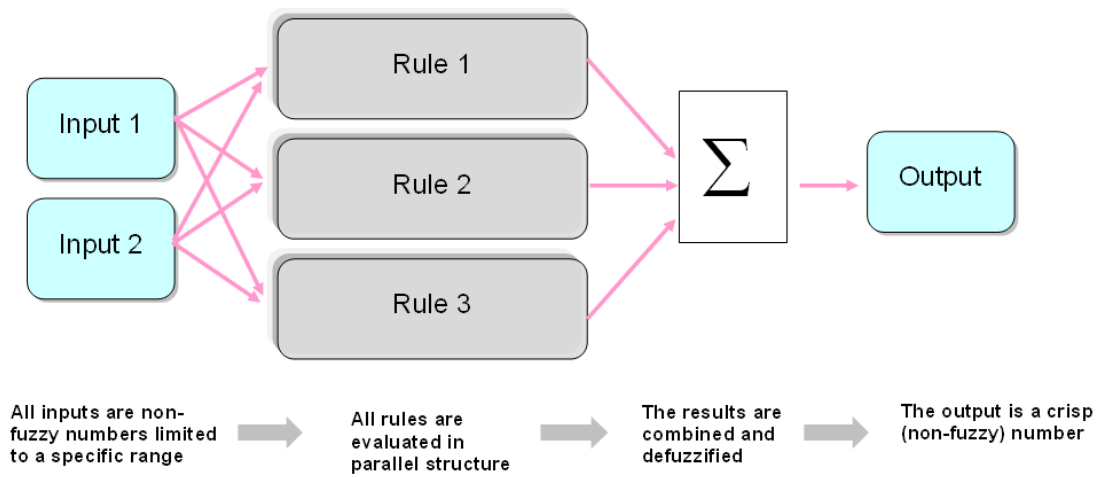
Figure 1
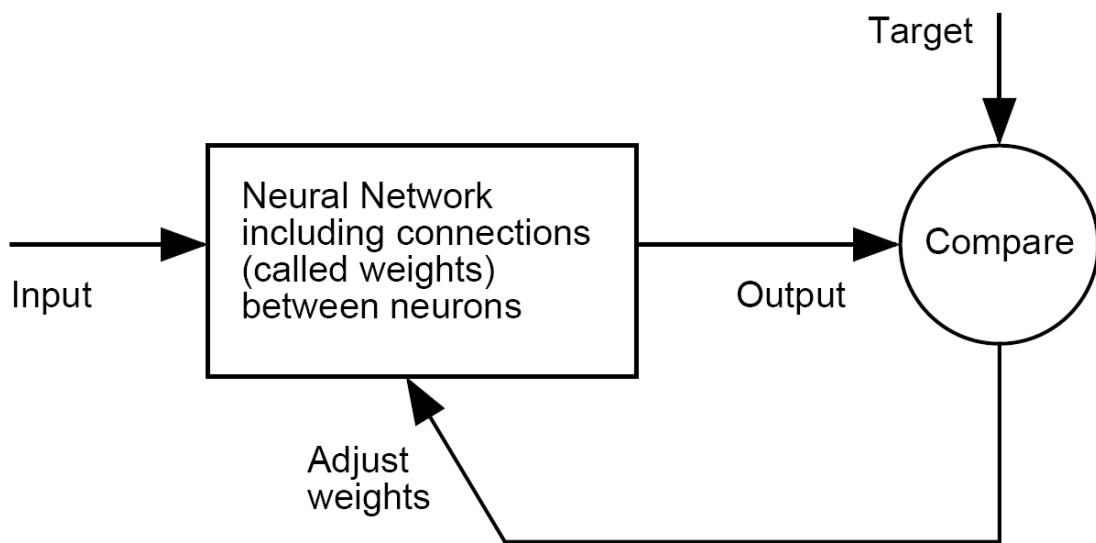
Figure 2
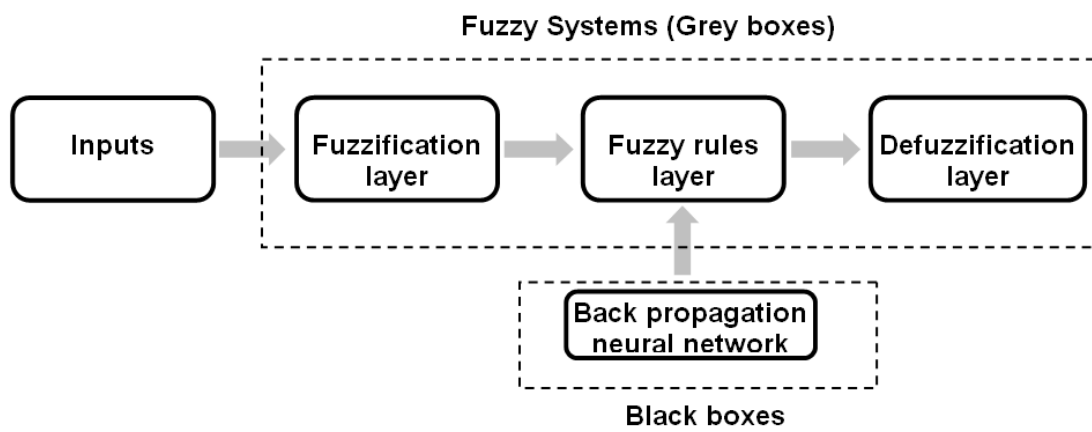
Figure 3
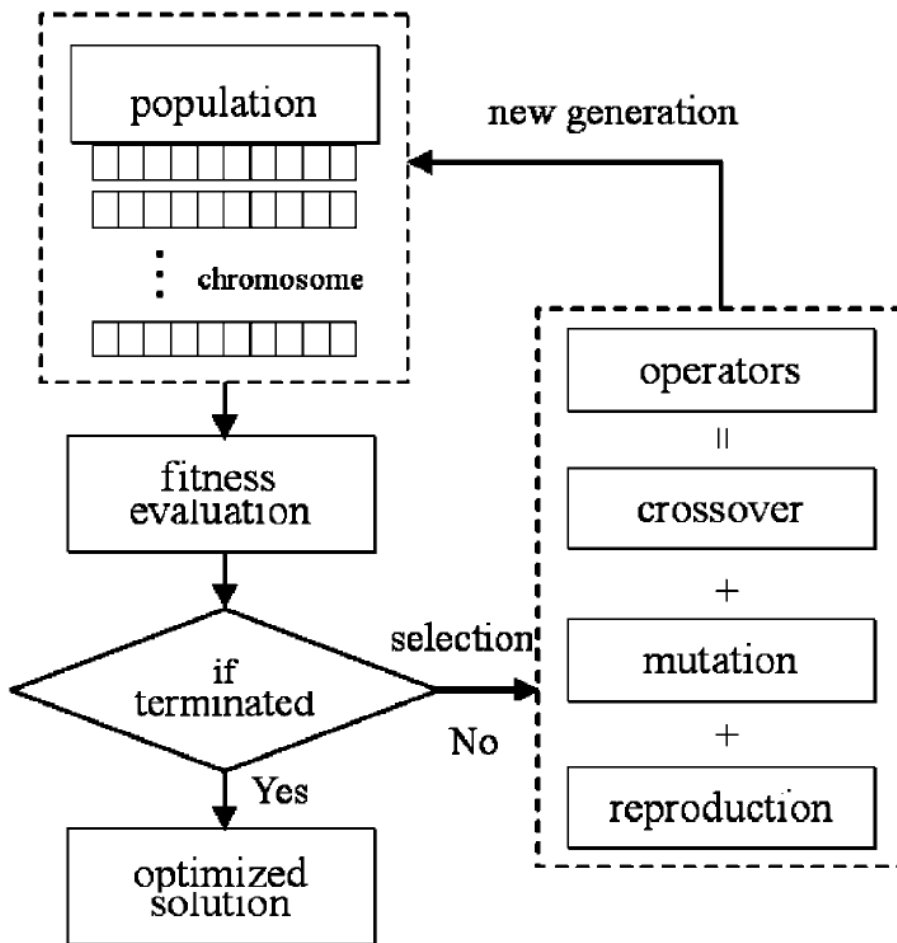
Figure 4
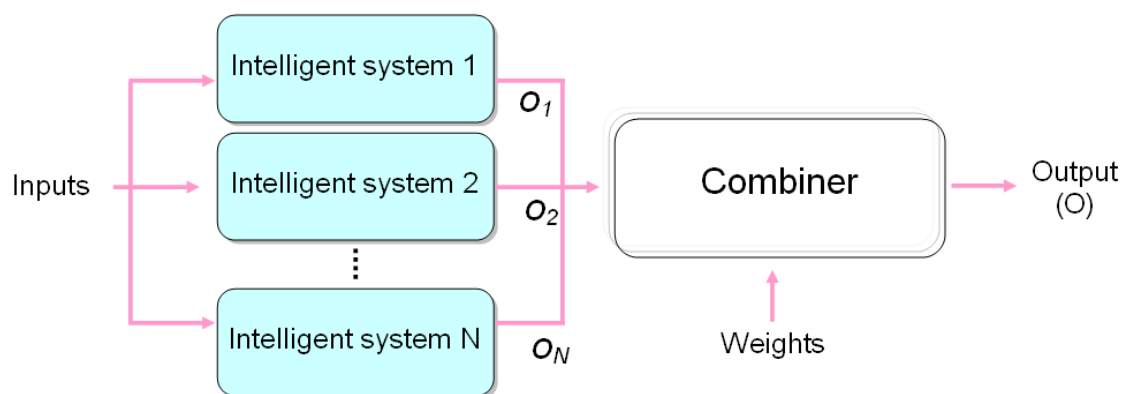
Figure 5
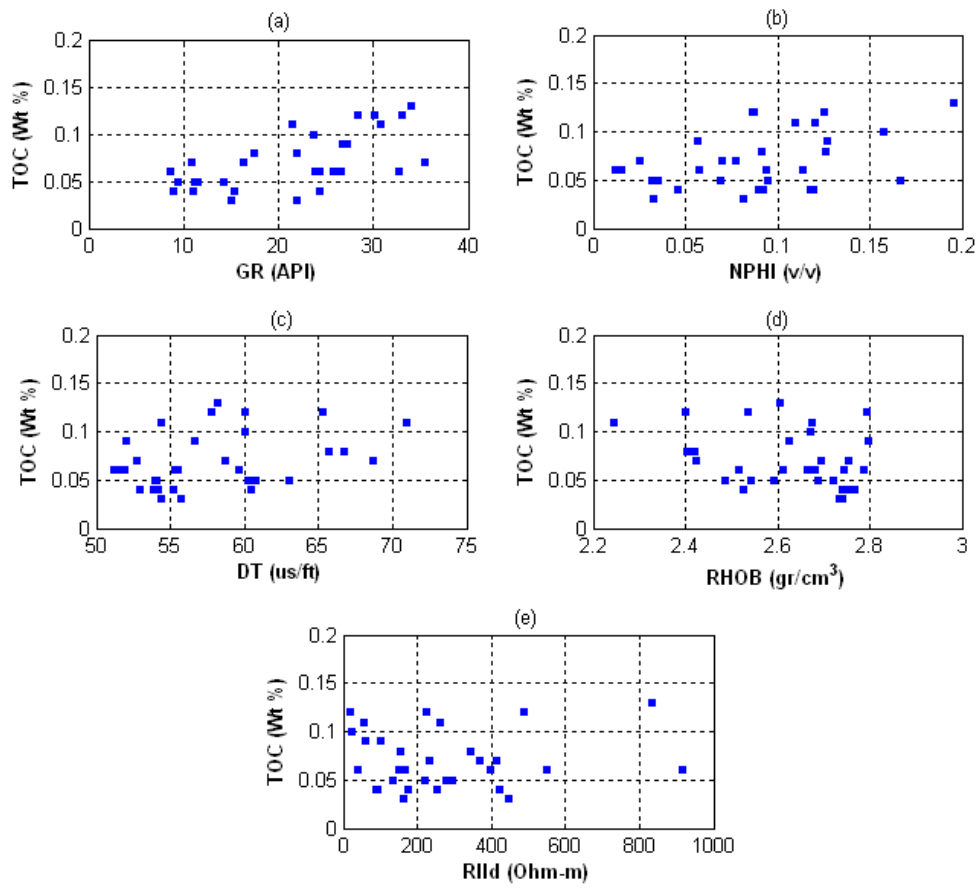
Figure 6
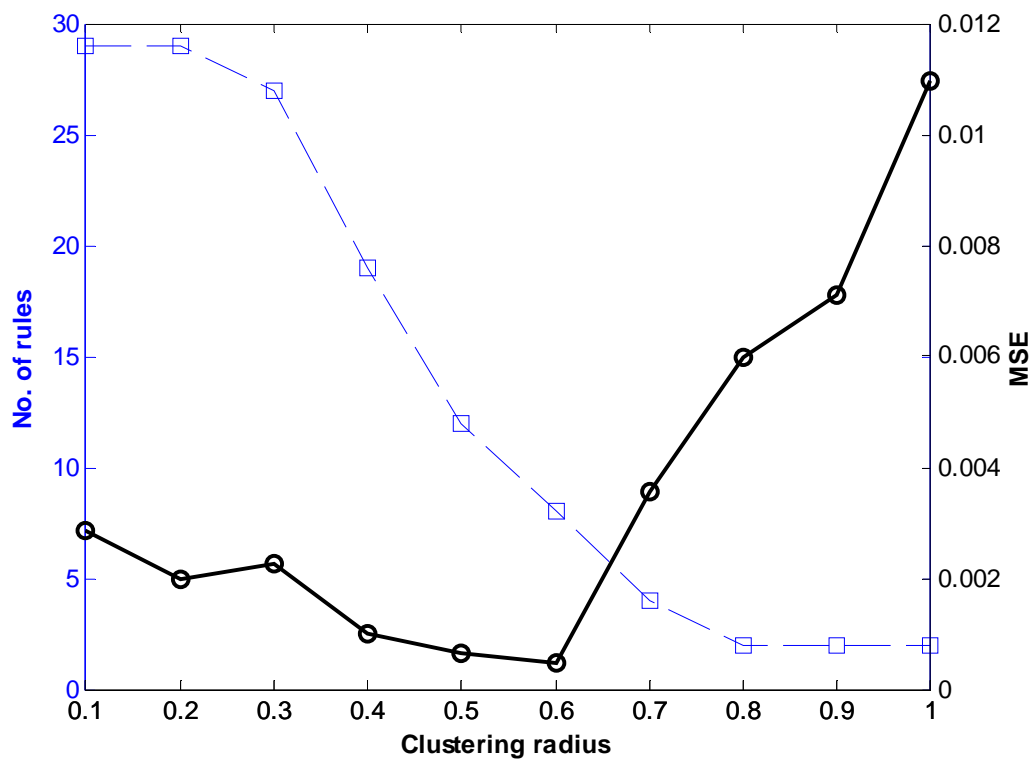
Figure 7
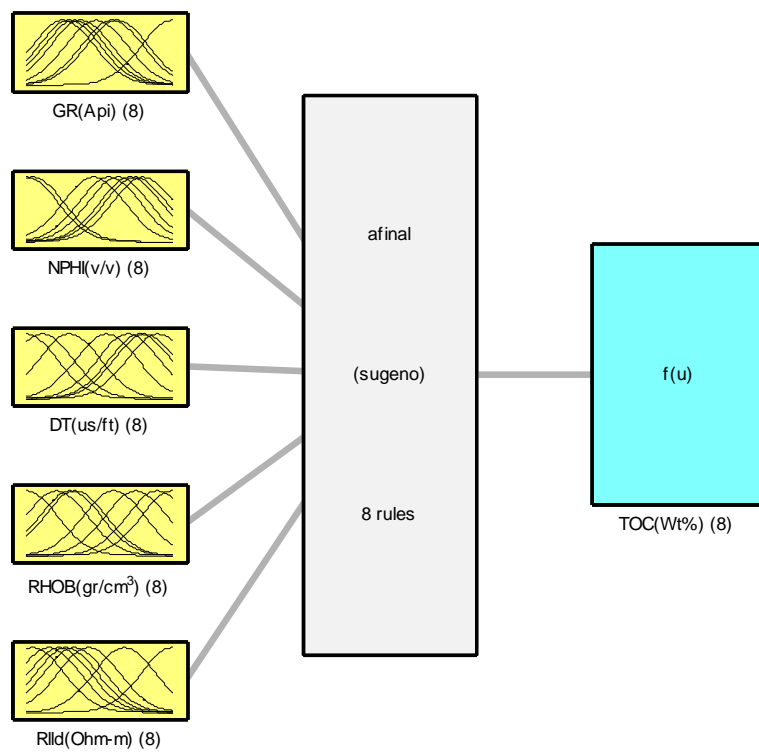
System afinal: 5 inputs, 1 outputs, 8 rules

Figure 8

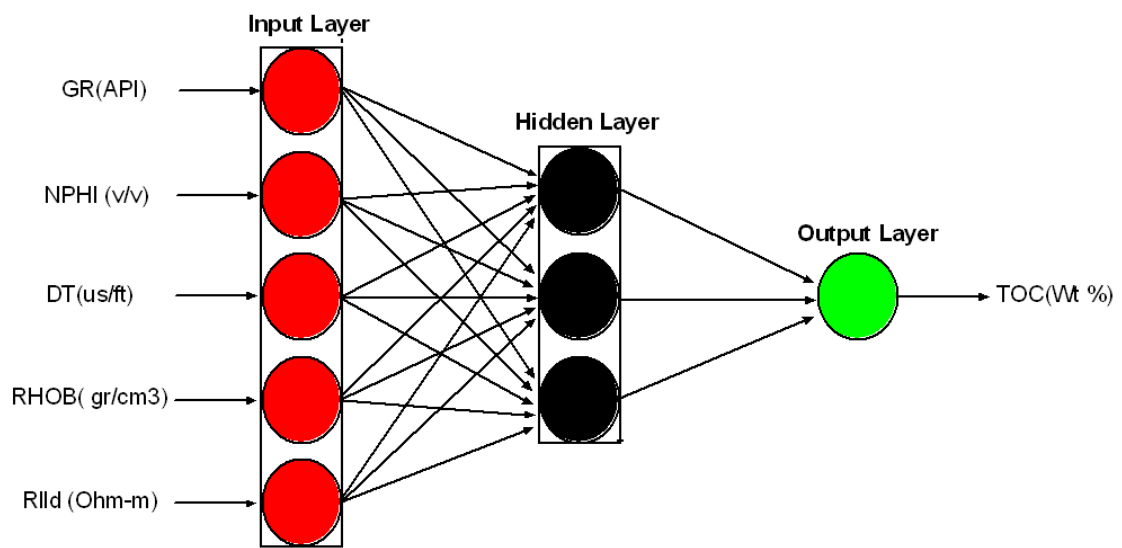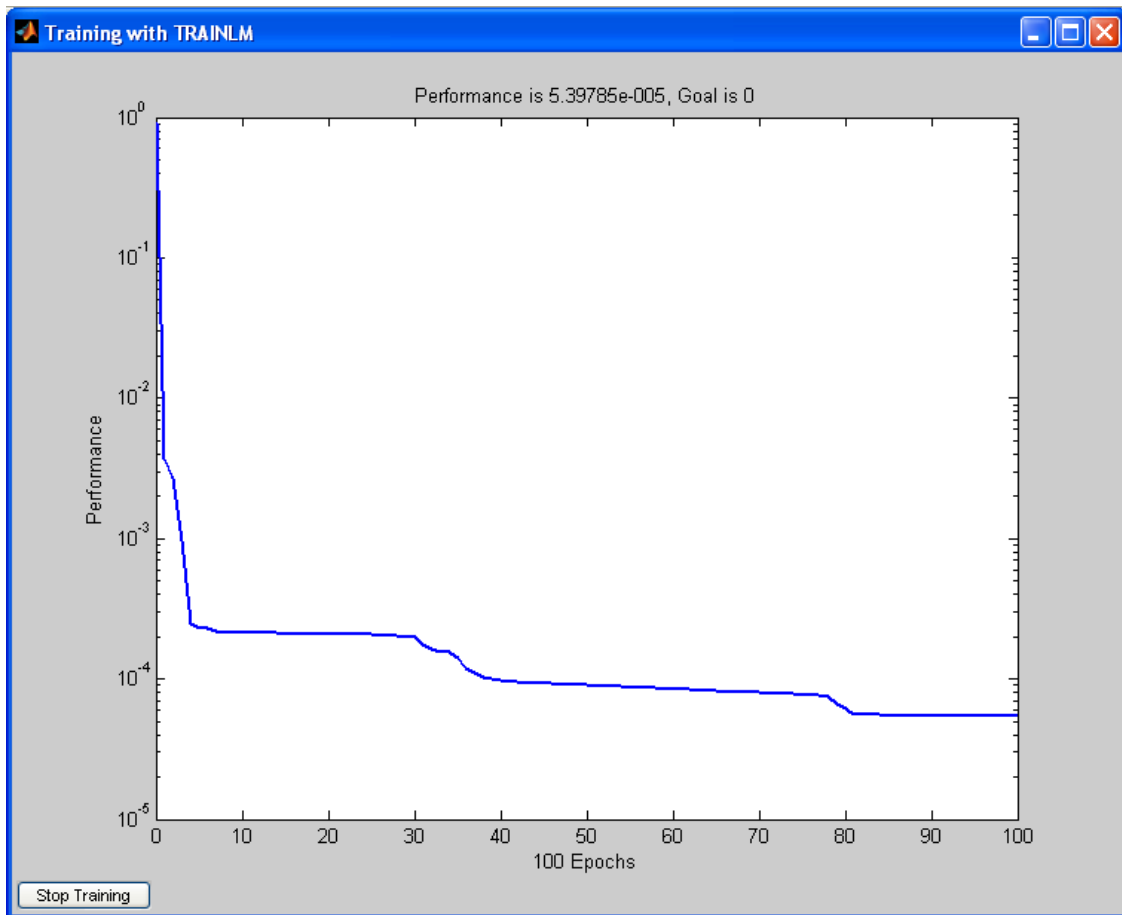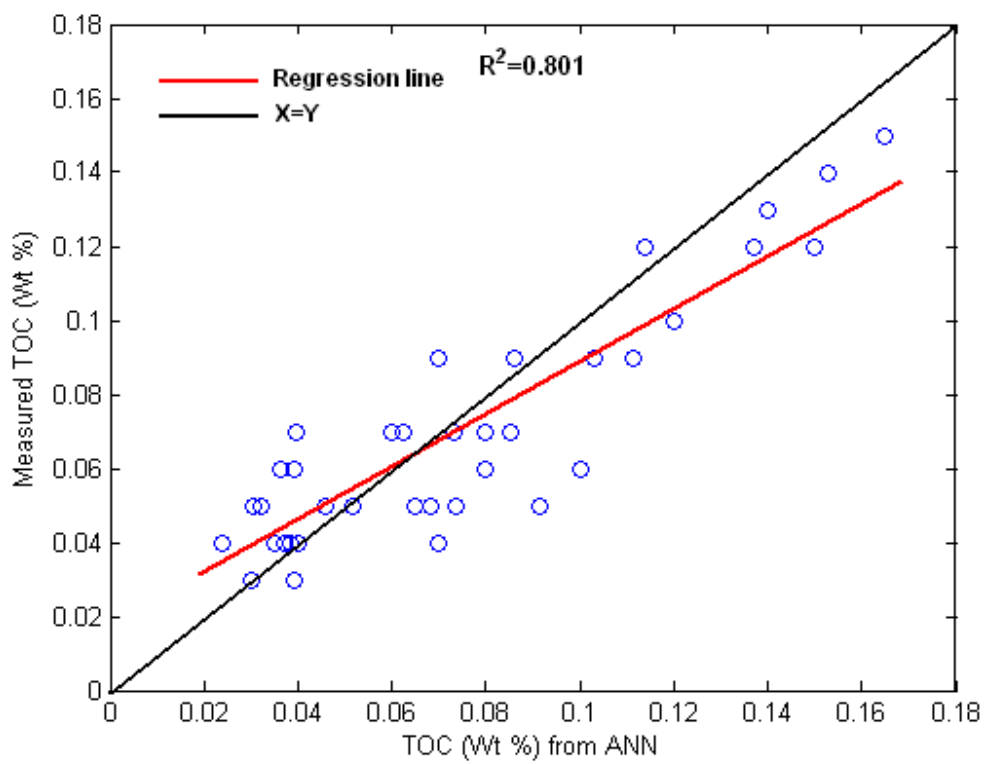Figure 9

27
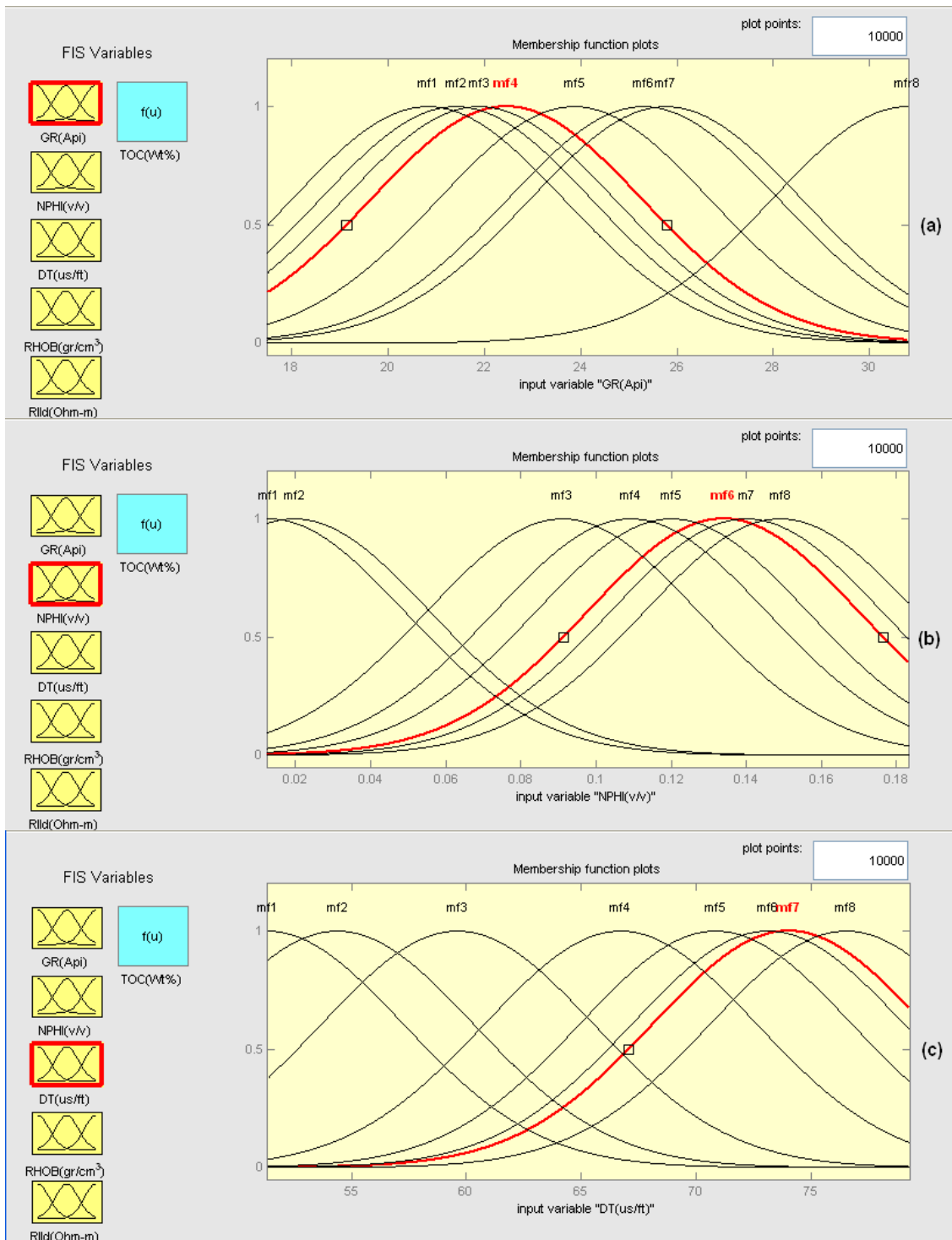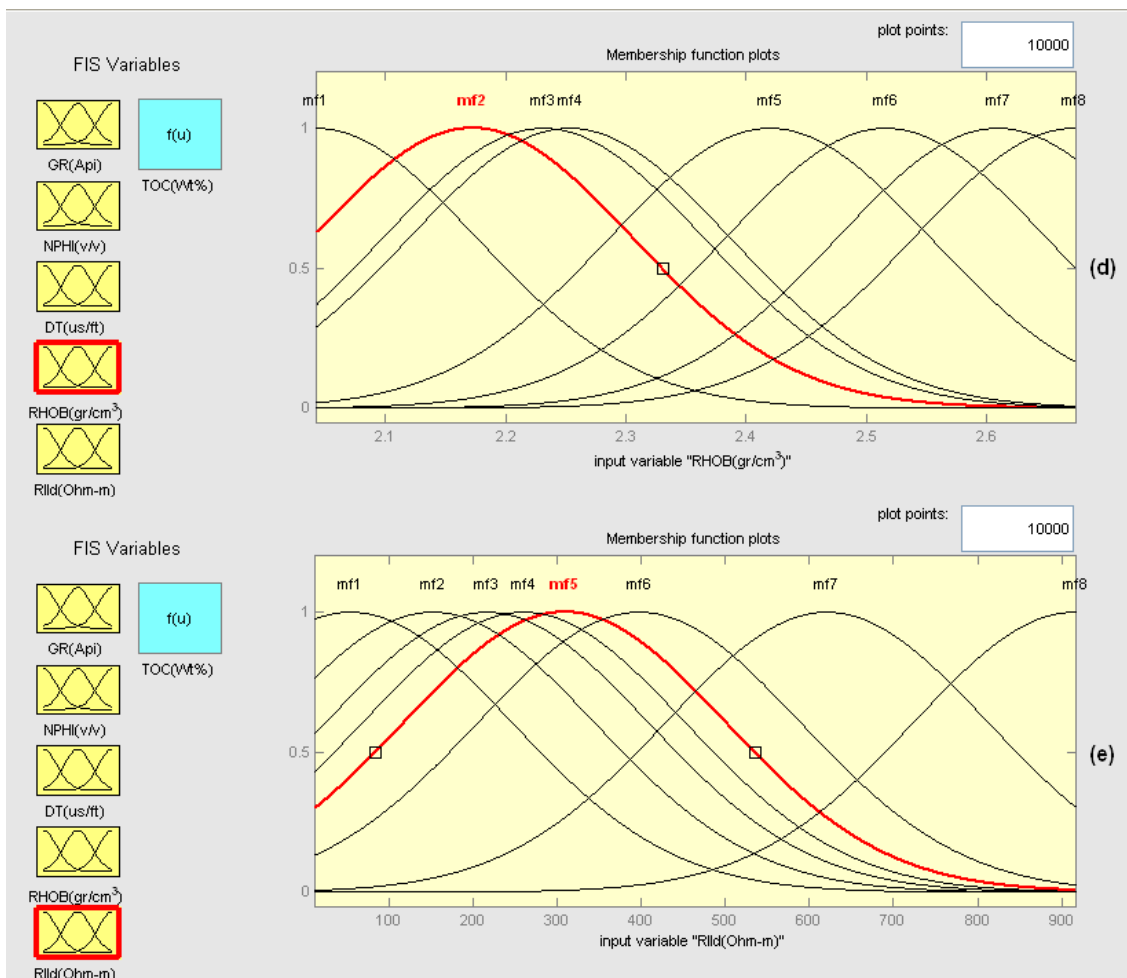
Figure 9 (continued)
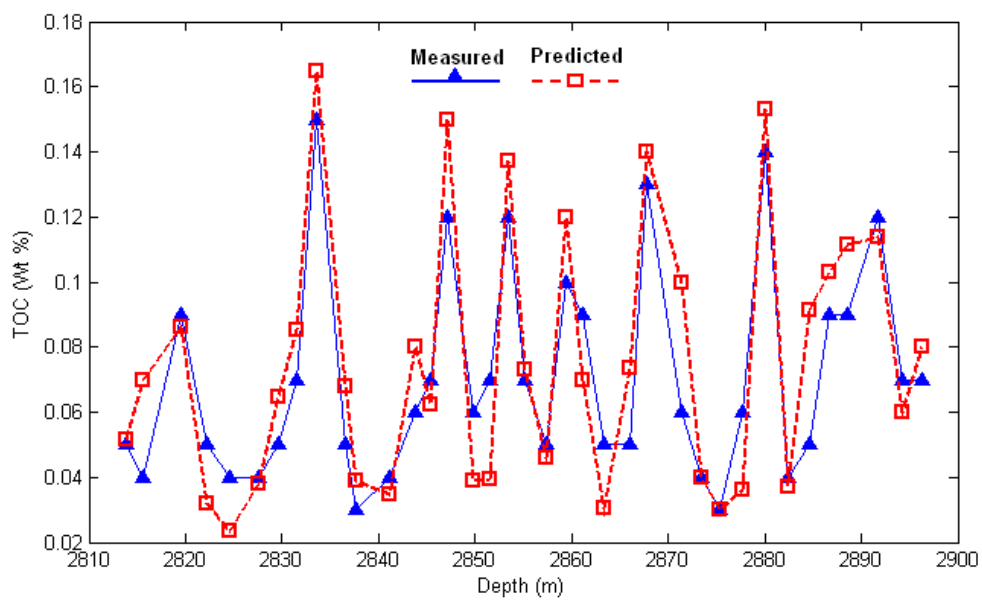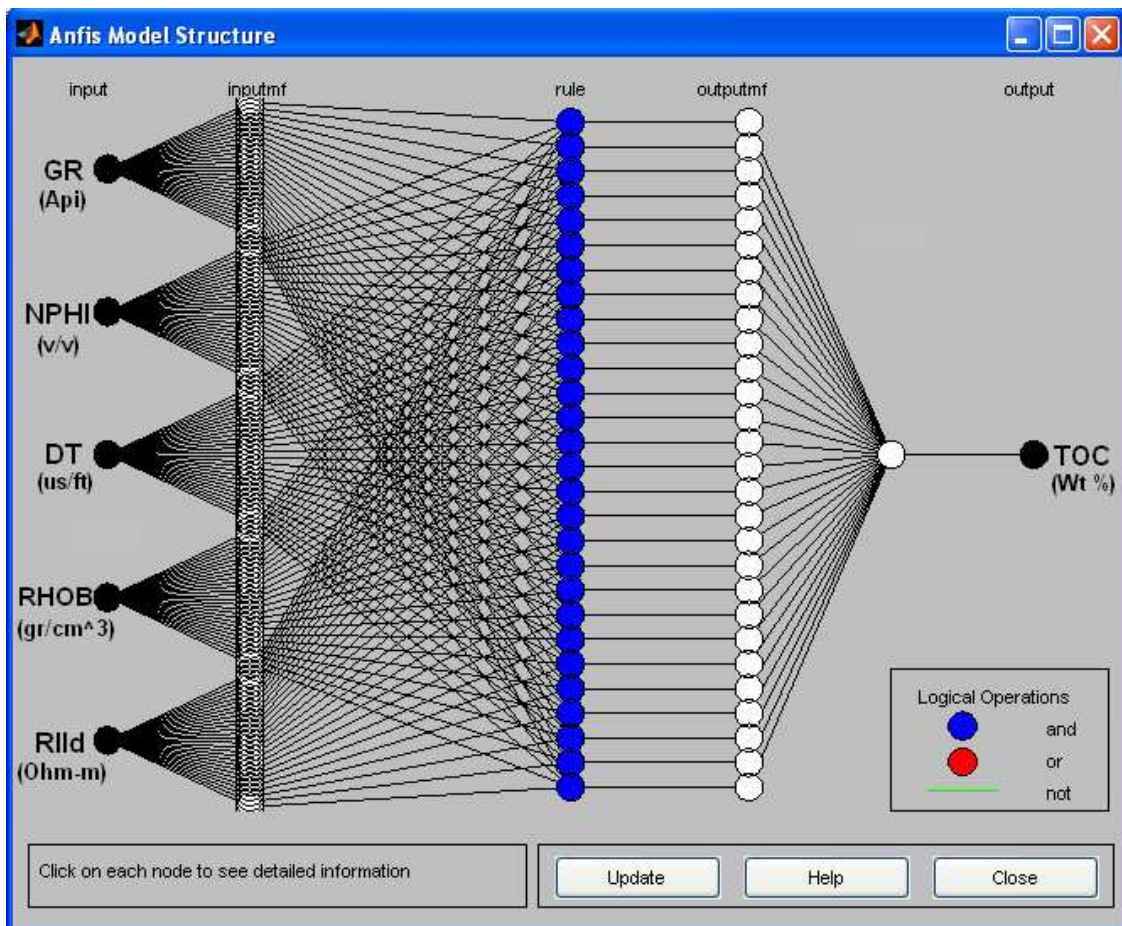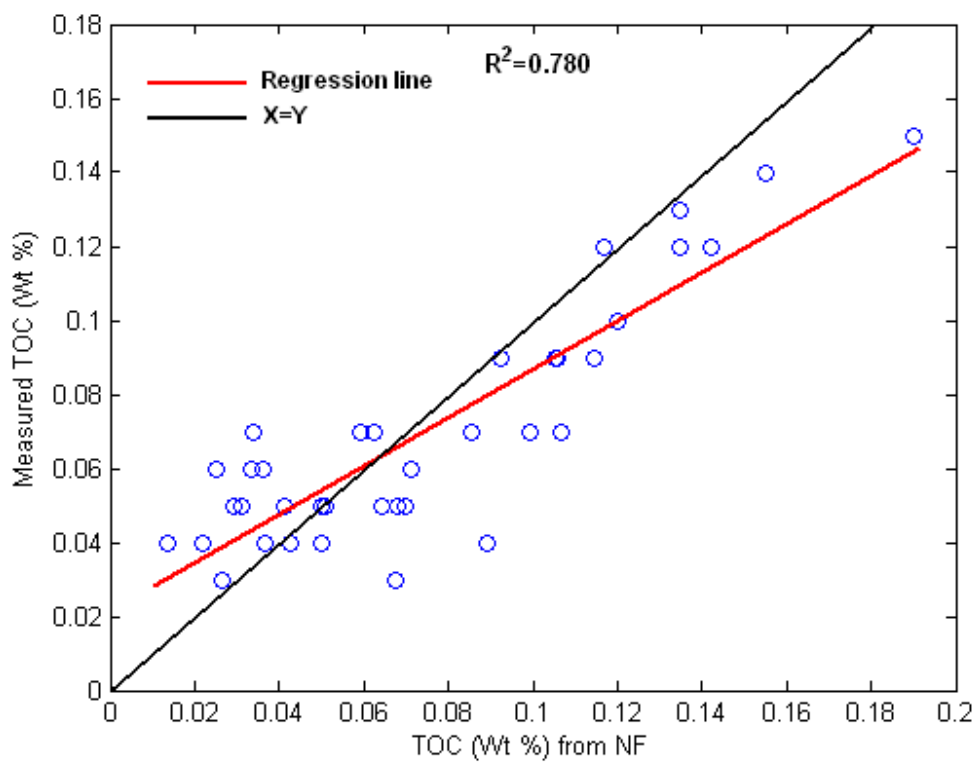
Figure 10

29
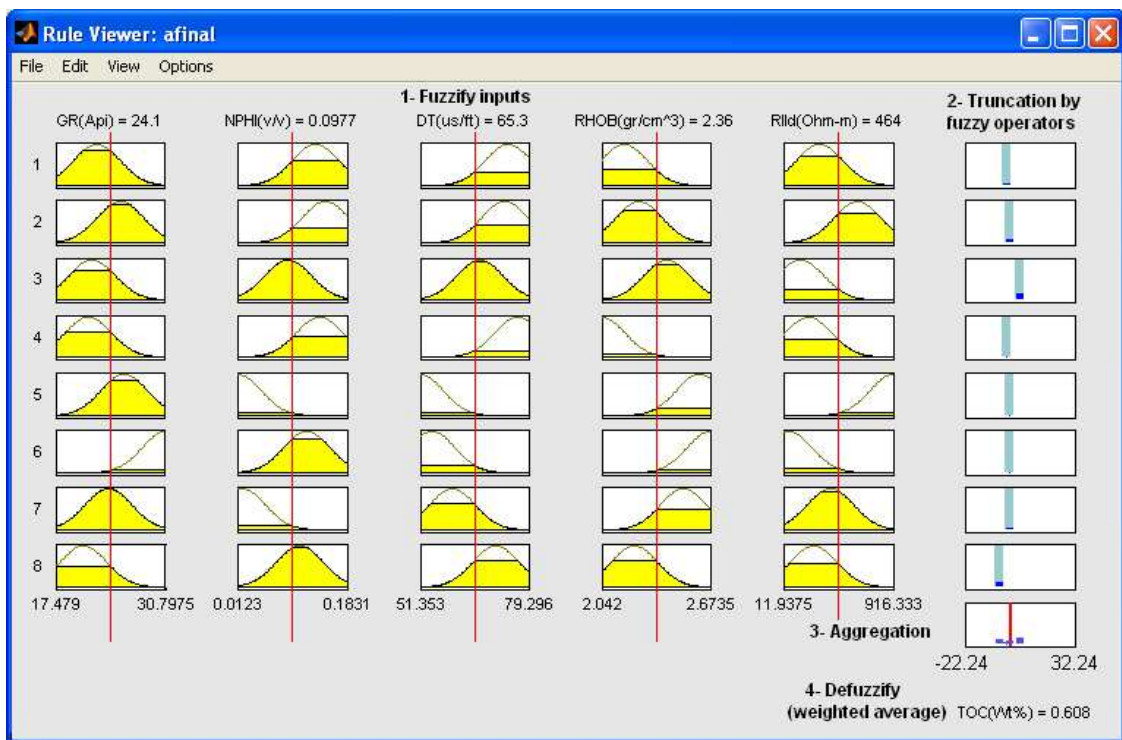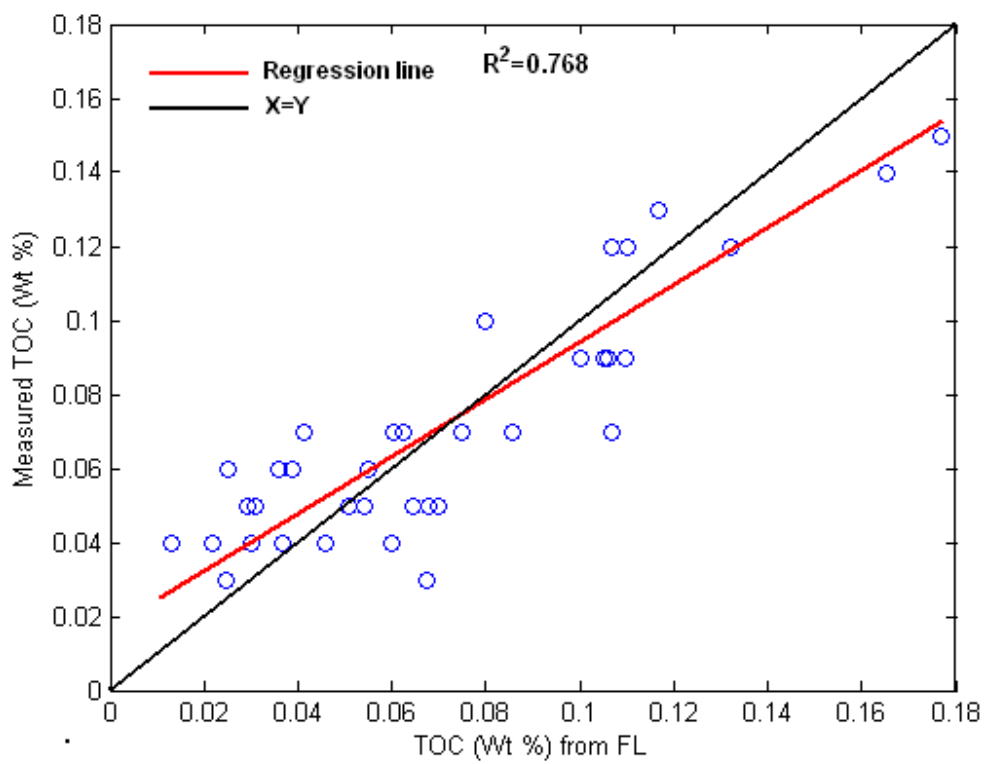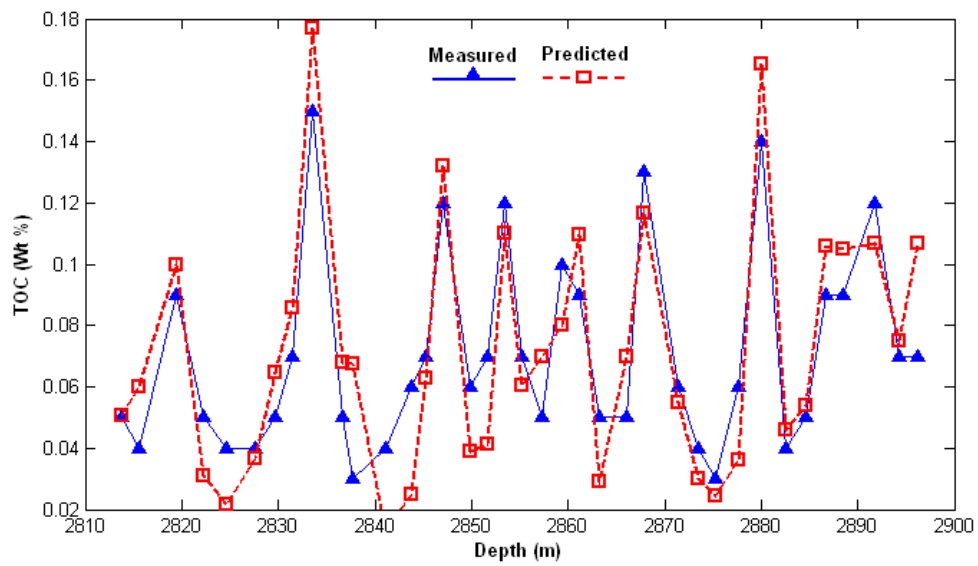
Figure 11

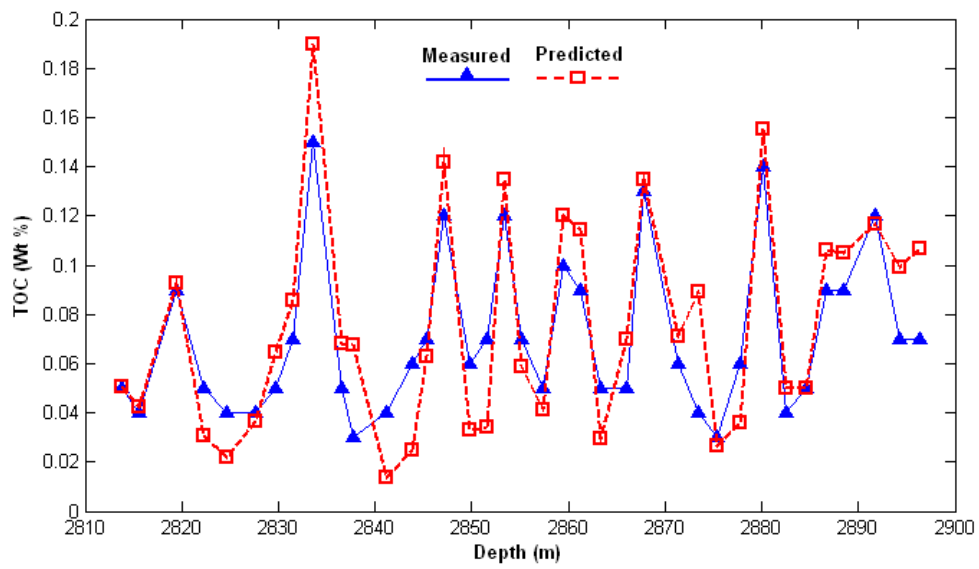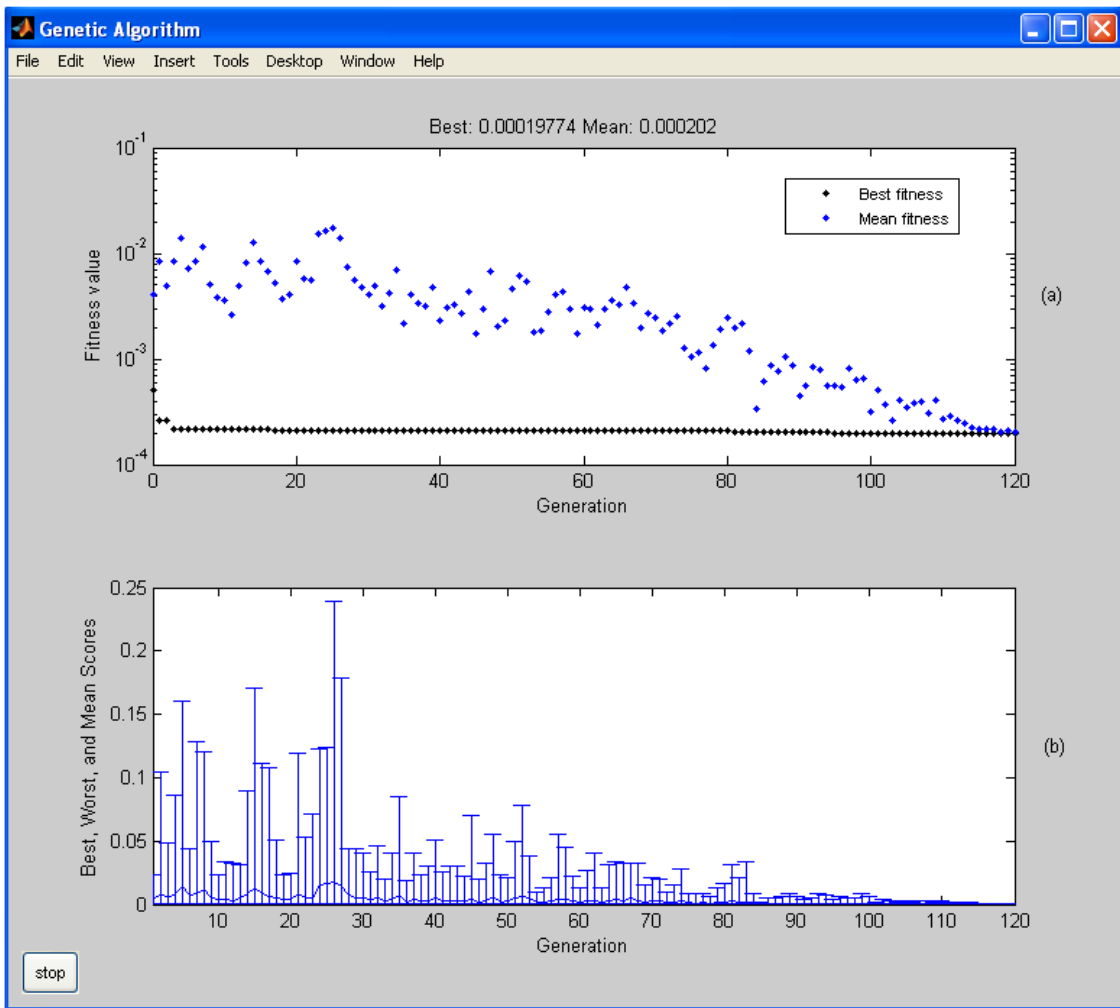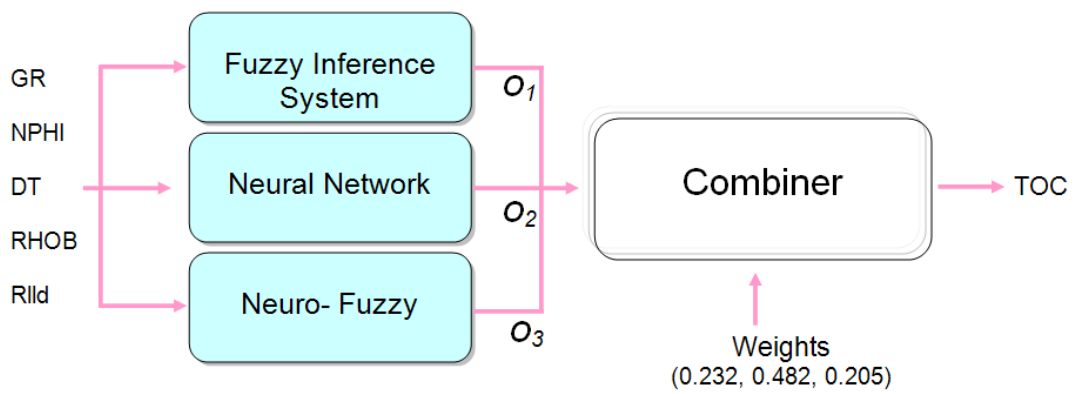Figure 12

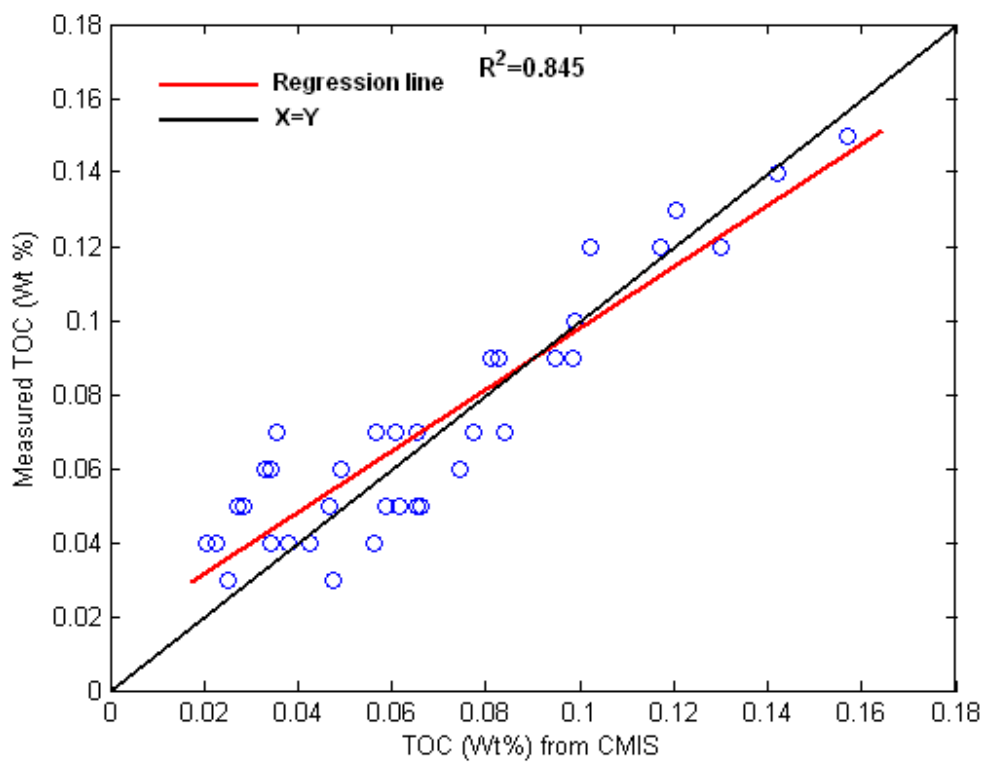Figure 13

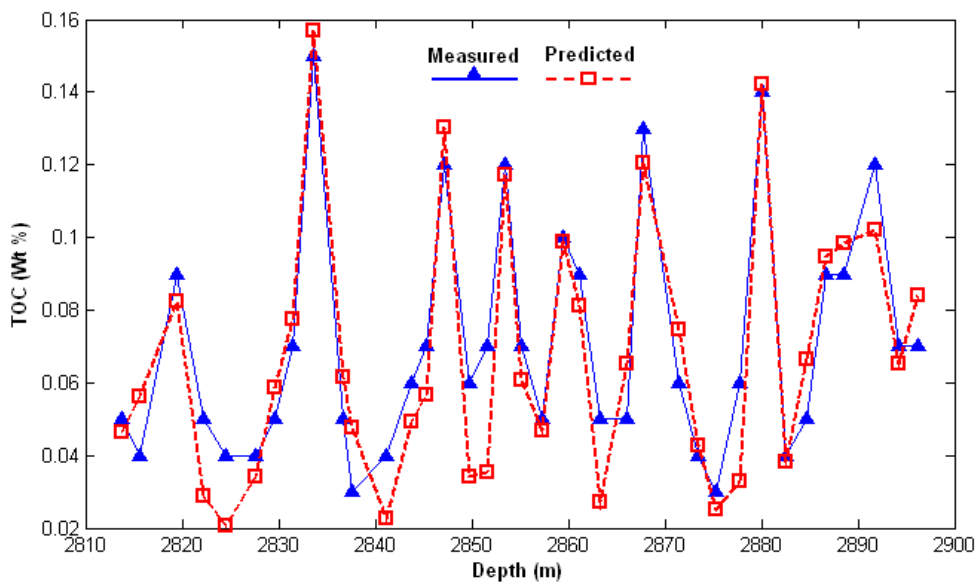Figure 14

Figure 15

Figure 16

Figure17

Figure 18

Figure 19

Figure 20

Figure 21

Figure 22

Figure 23