

Petrophysical data prediction from seismic attributes using committee fuzzy inference system

Ali Kadkhodaie-Ilkhchi¹; M.Reza Rezaee^{1,2*}; Hossain Rahimpour-Bonab¹
and Ali Chehrazi^{1,3}

1. Department of Geology, College of Science, University of Tehran, Tehran, Iran
2. Department of Petroleum Engineering, Curtin University of Technology, Perth, Australia
3. Geology Division, Iranian Offshore Oilfields Company, NIOC, Tehran, Iran

Abstract

This study presents an intelligent model based on fuzzy systems for making a quantitative formulation between seismic attributes and petrophysical data. The proposed methodology comprises two major steps. Firstly, the petrophysical data, including water saturation (S_w) and porosity, are predicted from seismic attributes using various Fuzzy Inference Systems (FIS), including Sugeno (SFIS), Mamdani (MFIS) and Larsen (LFIS). Secondly, a Committee Fuzzy Inference System (CFIS) is constructed using a hybrid Genetic Algorithms-Pattern Search (GA-PS) technique. The inputs of the CFIS model are the output averages of the FIS petrophysical data. The methodology is illustrated using 3D seismic and petrophysical data of 11 wells of an Iranian offshore oil field in the Persian Gulf. The performance of the CFIS model is compared with a Probabilistic Neural Network (PNN). The results show that the CFIS method performed better than neural network, the best individual fuzzy model and a simple averaging method.

Keywords: Committee fuzzy inference system, Sugeno, Larsen, Mamdani, hybrid genetic algorithm-pattern search, probabilistic neural network, petrophysical data, seismic attributes,

* Corresponding author: Tel: +618 9266 7980; Fax: +618 9266 7063.

r.rezaee@curtin.edu.au (M.R. Rezaee) akadkhoda@khayam.ut.ac.ir (A. Kadkhodaie-Ilkhchi),
rahimpor@khayam.ut.ac.ir (H. Rahimpour-Bonab), achehrazi@iooc.co.ir (A. Chehrazi)

1 Address: Department of Geology, College of Science, University of Tehran, Iran

2 Department of Petroleum Engineering, Curtin University of Technology, ARRC Building, 26 Dick Perry Avenue, Kensington, WA 6151, Australia.

3 No. 38, Tooraj St., Vali-Asr Ave., Tehran, 19395, Iranian Offshore Oilfields Company, Iran.

Introduction

The last decade has witnessed significant advances in the study and application of expert systems in the petroleum industry. The establishment of the existence of an intelligent formulation, between two sets of data (inputs/outputs), has been the main topic of such studies. One such topic, of great interest, was to characterize how 3D seismic data can be related to lithology, rock types, fluid content, porosity, shear wave velocity and other reservoir properties. Petrophysical parameters, such as water saturation and porosity, are very important data for hydrocarbon reservoir characterization. Hitherto, several researchers have worked on predicting them from seismic data using statistical methods and intelligent systems (Nikravesh et al., 1998; Balch et al., 1999; Trappe and Hellmich, 2000; Nikravesh et al., 2001; Nikravesh and Aminzadeh, 2001; Wong and Nikravesh, 2001; Meldahl et al., 2001; Russell et al., 2002; Russell et al., 2003; Nikravesh and Hassibi, 2003; Aristimuño and Aldana, 2006; Chopra and Marfurt, 2006; Soubotcheva and Stewart, 2006). Generally, geological, petrophysical and seismic data are not clear-cut and, inherently, are associated with uncertainties. Fuzzy expert systems have become one of the modern and robust techniques for the analysis of geosciences data (Rezaee et al., 2007; Kadkhodaie et al., 2006). They are able to recognize patterns, based on the fuzzy classification of data into natural groups, which are termed as clusters. For each cluster, a membership function is fitted which is used in the process of formulating input to output data through fuzzy rules.

A Committee Fuzzy Inference system (CFIS), which is introduced in this study, has a parallel structure that produces a final result by combining the output of individual fuzzy models, using a hybrid, genetic algorithm-pattern search technique. Experts of CFIS model are the result of Sugeno, Mamdani, Larsen and simple averaging method. Each of the experts has a weight factor in the construction of the CFIS that is derived by GA-PS. The CFIS was successfully applied to the Ghar reservoir of the Iranian offshore oilfield, Persian Gulf.

1. Methodology

1.1. Fuzzy inference system

A Fuzzy Inference System (FIS) is a process of formulating, from a set of input data to a set of output data, using fuzzy sets theory. Fuzzy sets theory was first introduced

by Zadeh (1965). In fuzzy sets theory, each element may belong to a set to a degree (μ) which can take values ranging from 0 to 1 (partial membership). Each fuzzy set is represented by a membership function (MF). MFs are of several types such as Gaussian, triangular, trapezoidal, sigmoid, S-shape, Z-shape, etc. There are three main parts in a FIS. These are a) Fuzzifier, b) Inference Engine (fuzzy rule base) and c) Defuzzifier (see figure 1). In the Fuzzifier, each crisp (non-fuzzy) value passes through a membership function and takes a value between 0 and 1. The Inference Engine is the main part of the system that consists of a fuzzy rule base (RB). The fuzzy rule base may consist of multiple inputs and multiple outputs (MIMO) which, in turn, can be broken down further into a set of rule bases with multiple inputs and single output (MISO) (Lee, 2004).

$$\begin{aligned}
 \text{RB} &= \left\{ \bigcup_{i=1}^n \text{R}_{\text{MIMO}}^i \right\} = \left\{ \bigcup_{k=1}^t \bigcup_{i=1}^n [(A_i \times \Lambda \times B_i) \rightarrow C_i^k] \right\} \\
 &= \left\{ \bigcup_{i=1}^n [(A_i \times \Lambda \times B_i) \rightarrow C_1], \bigcup_{i=1}^n [(A_i \times \Lambda \times B_i) \rightarrow C_2], \dots, \bigcup_{i=1}^n [(A_i \times \Lambda \times B_i) \rightarrow C_t] \right\} \\
 &= \{ \text{RB}_{\text{MISO}}^1, \text{RB}_{\text{MISO}}^2, \dots, \text{RB}_{\text{MISO}}^t \}
 \end{aligned}
 \tag{1}$$

In cases where the *antecedents* of fuzzy rules include multiple parts, then fuzzy operators are used to connect them. The most common fuzzy operators are; *min* (\wedge); *max* (\vee); *product* (\bullet) and *not* ($-$). The *consequent* of a fuzzy rule assigns an entire fuzzy set to the output. This fuzzy set is represented by a membership function that is chosen to indicate the qualities of the consequent. If the antecedent is only partially true, then the output fuzzy set is truncated according to a process which is termed *implication*. Since decisions are based on the testing of all of the rules in an FIS, the rules must be combined in some manner in order to make a decision. *Aggregation* is the process by which the fuzzy sets, which represent the outputs of each rule, are combined into a single fuzzy set. The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule.

Defuzzification is the process of transforming the aggregation result into a *crisp* output. There are various defuzzification methods which include: centroid; bisector; large of maximum (LOM); small of maximum (SOM) and mean of maximum (MOM).

Mamdani method: The most important differences among fuzzy inference systems are the types of the output membership functions and the implication methods. In MFIS the output membership functions are fuzzy sets. After the aggregation process, there is a fuzzy set for each output variable that needs defuzzification. This method

uses the *min* operation (\wedge) as a fuzzy implication (Mamdani and Assilian, 1975; Mamdani, 1976 & 1977).

Let's suppose a rule base is given in the following form:

Ri: if x is A_i and y is B_i then z is C_i , $i = 1, 2, \dots, n$

then, $R_i = (A_i \wedge B_i) \rightarrow C_i$ is defined by $\mu_{R_i} = \mu_{(A_i \text{ and } B_i \rightarrow C_i)}(x, y, z)$

The input data $x = x_0, y = y_0$ pass through the rule, above, to produce the final output, as below (Lee, 2004):

$$\left\{ \begin{array}{l} \mu_{C_i'}(z) = [\mu_{A_i}(x_0) \wedge \mu_{B_i}(y_0)] \rightarrow \mu_{C_i}(z) \\ \mu_{C_i'}(z) = \alpha_i \wedge \mu_{C_i}(z) \\ \mu_{C'}(z) = \mu_{C_1'}(z) \vee \mu_{C_2'}(z) = [\alpha_1 \wedge \mu_{C_1}(z)] \vee [\alpha_2 \wedge \mu_{C_2}(z)] \\ \mu_{C'}(z) = \bigcup_{i=1}^n [\alpha_i \wedge \mu_{C_i}(z)] = \bigcup_{i=1}^n \mu_{C_i'}(z), C' = \bigcup_{i=1}^n C_i' \end{array} \right. \quad \text{Eq. (2)}$$

A graphical illustration of MFIS is shown in figure 2a.

Larsen method: This method uses the *product* operator for the fuzzy implication. As with the Mamdani method output, MFs are fuzzy sets (Larsen, 1980). For a rule base in following form

Ri: if x is A_i and y is B_i then z is C_i , $i = 1, 2, \dots, n$

then $R_i = (A_i \wedge B_i) \rightarrow C_i$ is defined by $\mu_{R_i} = \mu_{(A_i \wedge B_i \rightarrow C_i)}(x, y, z)$

The input data $x = x_0, y = y_0$ produce the final output, as below (Lee, 2004):

$$\left\{ \begin{array}{l} \mu_{C_i'}(z) = [\mu_{A_i}(x_0) \wedge \mu_{B_i}(y_0)] \rightarrow \mu_{C_i}(z) = [\mu_{A_i}(x_0) \wedge \mu_{B_i}(y_0)] \cdot \mu_{C_i}(z) \\ = \alpha_i \cdot \mu_{C_i}(z) \quad \text{where } \alpha_i = \mu_{A_i}(x_0) \wedge \mu_{B_i}(y_0) \\ \mu_{C'}(z) = \bigcup_{i=1}^n [\alpha_i \cdot \mu_{C_i}(z)] = \bigcup_{i=1}^n \mu_{C_i'}(z), C' = \bigcup_{i=1}^n C_i' \end{array} \right. \quad \text{Eq. (3)}$$

A graphical illustration of LFIS is shown in figure 2b.

Sugeno method: Sugeno fuzzy inference system is similar to the Mamdani method in many aspects. In the first two parts of the fuzzy inference process, fuzzifying the inputs and applying the fuzzy operator, are exactly the same. Moreover, all the lemmas expressed for Mamdani fuzzy inference system are the same for SFIS. The main difference between them is that output membership functions are either linear or constant in Sugeno method (Sugeno, 1985).

A typical rule in a Sugeno fuzzy model has the form

If Input 1 = x_0 and Input 2 = y_0 , then $z = px_0 + qy_0 + r$

For a zero-order Sugeno model, the output level is a constant ($p = q = 0$). The output level, z_i , of each rule, is weighted by the firing strength α_i of the rule. The final output of the system is the weighted average of all rule outputs, computed as follows:

$$Final\ output = \frac{\sum_{i=1}^n \alpha_i z_i}{\sum_{i=1}^n \alpha_i} \quad Eq. (4)$$

A graphical illustration of SFIS is shown in figure 2c.

1.2. Committee fuzzy inference system

The proposed methodology, CFIS, consists of two major steps. At the first stage, petrophysical data are predicted from seismic data using SFIS, MFIS and LFIS models. Then a CFIS is constructed using a GA-PS technique. A schematic diagram of the CFIS, designed in this study, is shown in figure 3. The inputs of CMIS are the outputs and average of the previously mentioned fuzzy models. In this approach, each of the individual fuzzy inference systems has a weight coefficient, in constructing CFIS, showing its contribution in the overall prediction of the output data. A GA-PS technique can extract the appropriate weights, for the mentioned models, using an integration of genetic algorithm and pattern search techniques. Genetic algorithms were first introduced by Holland (1975). Potential solutions are called *chromosomes* and are 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*. It is a method for moving from one population of chromosomes to a new population by using a kind of natural selection process together with the genetics-inspired operators of *crossover*, *mutation*, and *inversion*. The selection operator chooses those chromosomes in the population that will be allowed to reproduce, and on average the fitter chromosomes produce more offspring than the less fit ones. *Crossover* exchanges subparts of two chromosomes; *mutation* randomly changes the allele values of some locations in the chromosome; and *inversion* reverses the order of a contiguous section of the chromosome, thus rearranging the order in which genes are arrayed (Mitchell, 1999). In the pattern search technique, the algorithm searches a set of points, called a *mesh*, around the current point (the point computed at the previous step of the algorithm). The mesh is formed by adding the current point to a scalar multiple of a set of vectors called a *pattern*. If the pattern search algorithm finds a point in the mesh that improves the objective function at the

current point, the new point becomes the current point at the next step of the algorithm (Matlab user's guide, 2007).

In this study, the fitness function for minimization by GA-PS was selected as mean squared error (MSE) in predicting petrophysical data using mentioned methods. The equation for predicting final output is expressed as below:

$$Output = \beta_1 \cdot output_{SFIS} + \beta_2 \cdot output_{MFIS} + \beta_3 \cdot output_{LFIS} + \beta_4 \cdot output_{Average} \quad \text{Eq. (5)}$$

where β_1 , β_2 , β_3 and β_4 are the weight coefficients corresponding to the outputs of Sugeno, Mamdani, Larsen and simple averaging method.

The methodology, described in this study, reaps the benefit of all individual methods, and provides more accurate results. It is clear that many components of the method, described in this study, are based on the results of other researcher's works (such as Nikravesh et al., 1998; Mohaghegh et al., 1999; Mohaghegh, 2000; Hampson et al., 2001; Saggaf and Nebrija, 2003; Russell, 2003 & 2004; Kamali and Mirshady, 2004; Lim, 2005; Chen and Lin, 2006). Overall, it provides an optimal model for predicting well data from seismic attributes.

2. Application to the Iranian Offshore Oilfield

The present study focuses on application of the proposed methodology of CFIS on an Iranian offshore oilfield. For this purpose, interpreted 3D seismic data and petrophysical data from 11 wells (figure 4) of the study field were used. Seismic data were acquired in 2002 and cover a total area of approximately 242 km². Ghar Sandstone is the main reservoir unit over the study area. Seismic data quality is generally good over the entire time range with an absence of strong multiple interference. The seismic data is close to zero phase at the Ghar level. A 3D crossline showing general quality of seismic data across the study field is shown in figure 5. All of the petrophysical data were reviewed and quality controlled. Sonic and density logs were available for all wells.

3.1 Correlation of well logs to seismic data

At the first stage of study, well log data were correlated to seismic data. Synthetic seismograms were generated for each of the 11 wells: A2, A5, A9, A10, A11, A11_7H, A11_12H, A12_7H, A14_5H and A18_2H. The acoustic velocities from the sonic logs were multiplied by the bulk density values from density logs to compute

acoustic impedance logs. This impedance was converted to reflectivity, which was then converted from depth to time using a suitable time-depth relationship. Finally, the reflectivity in time was convolved with an appropriate wavelet to produce a synthetic seismogram. Depth-to-time conversion of the well logs was accomplished by applying checkshot data supplied for wells A10 and A12. It was necessary to create synthetics and extract the wavelets iteratively for the placement of the log data in time. This depth-to-time process allowed for a comparison of the well logs, and their associated tops, with the seismic data in time. Horizon interpretations and geologic well tops were used as an aid in determining a time-depth relationship for deviated wells. A sample of a well-to-seismic tie, at well A9, is shown in figure 6 where the correlation between synthetic seismogram (blue) and composite trace (red), at the well location, is 0.70.

3.2 Selection of optimal seismic attributes

Generally, the purpose of applying several statistical and intelligent models is to find linear and non-linear relationships and structures between input and output data. For this purpose, there should be a logical relationship between input and target parameter. In this section of the research investigations, physical relationships between input data (seismic attributes) and output data (petrophysical parameters) were investigated through the application of multi-regression analyses. A multi-regression analysis is a simple and practical method to find the strongest inputs for predicting a target parameter. Accordingly, multi-attributes to be used in construction of fuzzy models were chosen based on the trend obtained from regression analyses. The results of multi-regression analyses, for predicting water saturation and porosity, are shown in Table 1a, and 1b. According to Table 1, adding more attributes will improve the prediction. This does not always mean that the added attributes are predicting the true signal in the target log. The validation error can be considered as a criterion for determining when to stop adding attributes to the input set (Russell, 2004). According to Table 1a, the first four attributes of time, average frequency, filter 15/20-25/30 and dominant frequency, could be considered as the optimal inputs for predicting water saturation. The relationships between the input seismic attributes and S_w are shown in the crossplots of figure 7. Normally, in a hydrocarbon bearing interval, oil saturation decreases toward oil-water contact. Therefore, water saturation increases as time (depth) increases across the hydrocarbon bearing interval. Average frequency is a

signature of the events and effects of the abnormal attenuation due to the presence of the hydrocarbons (Taner et al., 1994). Filter 15//20-25/30 is a trapezoidal frequency filter and acts as dominant frequency. As with the average frequency, dominant frequency can indicate abnormal frequency attenuation and, thereby, indicate the presence of hydrocarbon bearing zones.

Performing a similar process, for predicting porosity, seven predictors have been proposed. These include: *inversion result*, *integrate*, *Quadrature trace*, *cosine instantaneous phase*, *integrated absolute amplitude*, *amplitude envelope* and *filter 15/20-25/30* (Table 1b). This method, proposed by Russell (2004), is efficient in determining optimal inputs for construction of fuzzy models. The relationships between the input seismic attributes and porosity are shown in the crossplots of figure 8. The physical relationships between seismic attributes and porosity are as follows:

Acoustic impedance is a product of sonic velocity and bulk density. There is an inverse relationship between velocity and bulk density. Accordingly, porosity is an inverse function of acoustic impedance. *Integrate* is the sum of the amplitudes within a window interval (Chen and Sidney, 1997). It is an indicator of an amplitude anomaly due to changes in lithology and porosity. *Quadrature trace* is calculated from a complex seismic trace analysis. It is actually a phase-delay feature and is useful in identifying vertical variation of *instantaneous phase*. Vertical variations of *instantaneous phase* relate to variations in porosity and lithology.

Cosine instantaneous phase is an attribute derived from instantaneous phase. Since its fixed bounds (-1 to +1) are easier to understand (Chen and Sidney, 1997), it can better identify variations in porosity and lithology. *Integrated absolute amplitude* is sum of all the trace amplitudes within the window interval. As with the integrate attribute, it can indicate amplitude anomalies as a result of lithology and porosity variations. The *amplitude envelope* is an indicator of the major lithology changes and of gas and liquid-accumulations (Taner et al, 1994). Accordingly, it can indicate porosity changes within a hydrocarbon bearing interval. As discussed for the water saturation, *filter 15//20-25/30* is a trapezoidal frequency filter and can indicate porosity changes due to amplitude variations.

3.3 Fuzzy clustering

Clustering of numerical data forms the basis of many fuzzy modeling and pattern classification algorithms. The purpose of clustering is to find natural groupings of

data, within a large dataset, thus revealing patterns that can provide a concise representation of the data behavior (Dubois et al., 1997). *Fuzzy c-means* and *subtractive clustering* are two powerful fuzzy clustering techniques which could be used for the construction of a fuzzy rule base.

Subtractive clustering is an effective approach to estimate the number of fuzzy clusters and cluster centers in a Sugeno fuzzy inference system (Jarrah and Halawani, 2001). In subtractive clustering, each data point is considered as a potential cluster center. Using this method, the number of effective grid points to be evaluated is simply equal to the number of data points, independent of the dimension of the problem (Chiu, 1994 & 1995). In subtractive clustering, the radius of neighborhood (cluster radius) plays an important role in construction of fuzzy inference system. It can take values between the range of [0, 1]. Specifying a smaller cluster radius will usually yield more and smaller clusters in the data (resulting in more rules). A large cluster radius yields a few large clusters in the data (Chiu, 1994).

In the present study, the dataset is a matrix of petrophysical data and corresponding seismic attributes (328 samples of Ghar reservoir from 11 wells). The dataset was divided into 252 model samples and 76 testing samples to evaluate the the reliability of this new method. The optimum number of clusters was extracted, by specifying a set of values between 0 and 1 for clustering radius, and measuring the performance of the model for test data at each stage. The result showed that, by performing subtractive clustering on the water saturation matrix (a matrix of input seismic attributes and water saturation), this process provided the lowest error ($MSE_{SFIS}=0.0136$) in the case specifying 0.15 for clustering radius, resulting in 43 clusters. For porosity, by specifying 0.4 for clustering radius, generated the lowest mean squared error ($MSE_{SFIS}=0.0035$), resulting in 50 clusters.

Fuzzy c-means (FCM) is another fuzzy clustering technique that forms the basis of Mamdani and Larsen fuzzy inference systems. This technique was originally introduced by Jim Bezdek in 1981. It provides a method of grouping data points that populate some multidimensional space into a specific number of different clusters. The *Fuzzy c-means* clustering starts with an initial guess for the cluster centers, which are intended to mark the mean location of each cluster. The initial guess for these cluster centers is most likely incorrect. Additionally, FCM assigns every data point a membership grade for each cluster. By iteratively updating the cluster centers and the membership grades for each data point, FCM iteratively moves the cluster centers to

the right location within a dataset. This iteration is based on minimizing an objective function that represents the distance from any given data point to a cluster center weighted by that data point's membership grade. The FCM output is a list of cluster centers, and several membership grades, for each data point (Matlab user's guide, 2007).

As mentioned in MFIS and LFIS, fuzzy rules are extracted through FCM. So the model matrices of water saturation and porosity were passed through FCM algorithm and cluster centers were calculated. In FCM algorithm, the number of clusters is defined by the user. However, an important question is- how many clusters are needed? To answer this question, the same method for choosing the optimum cluster radius, in subtractive clustering, is applied. That is, by specifying a number of clusters, from 1 to the number of the model data points (using a "for cycle"), and measuring the performance of the model for test data at each stage, the optimum number of clusters was calculated. Results showed that the water saturation model performs better when the number of clusters is 31 ($MSE_{MFIS}=0.0161$, $MSE_{LFIS}=0.0169$), whereas the porosity model error was the lowest when fuzzy inference system was constructed using 44 rules ($MSE_{MFIS}=0.0020$, $MSE_{LFIS}=0.0023$).

3.4 Construction of fuzzy rule base

In this section, fuzzy models' structures, for estimating water saturation and porosity based on seismic attributes, are explained. The input data for fuzzy rules generation are cluster centers extracted using subtractive clustering (for SFIS), and FCM (for MFIS and LFIS). The methodology for construction of fuzzy rule base for estimating water saturation and porosity using the cluster centers follows:

For a set of m cluster centers $\{u_1, u_2, \dots, u_m\}$ in a dimensional space M , we assume that the first N dimensions correspond to input variables and the last $M-N$ dimensions correspond to output variables. Each vector u_i could be decomposed into two component vectors v_i (inputs) and w_i (outputs). We consider each cluster center u_i as a fuzzy rule that describes the system behavior. Intuitively, each cluster center represents the rule (Chiu, 1997):

Rule i : If {input is near v_i } then output is near w_i .

Given an input vector x_0 , the degree of fulfillment of rule i is defined as Eq. (6)

$$\tau_i = e^{-\lambda \|x_0 - v\|^2}$$

Eq. (6)

Where λ is the constant. The output vector z is calculated via Eq. (7)

$$z = \left[\sum_{i=1}^m \tau_i w_i \right] \div \left[\sum_{i=1}^m \tau_i \right] \quad \text{Eq. (7)}$$

This computational model corresponds to the MFIS and LFIS employing traditional fuzzy if-then rules. Each rule has the following form:

if input₁ is A_{i1} & input₂ is A_{i2} & ... then output₁ is C_{i1} & output₂ is C_{i2} ...

where *input_j* is the *jth* input variable and *output_j* is the *jth* output variable; *A_{ij}* is an exponential membership function in the *ith* rule associated with the *jth* input and *B_{ij}* is a membership function in the *ith* rule associated with the *jth* output. For the *ith* rule, which is represented by cluster center *u_i*, *A_{ij}* is given by Eq. (8)

$$A_{ij}(Y_j) = \exp(-1/2(\text{input}_j - v_{ij})^2 / \sigma_{ij}^2) \quad \text{Eq. (8)}$$

and *C_{ij}* can be any symmetric membership function centered around *w_{ij}*, where *v_{ij}* is the *jth* element of *v_i*, *w_{ij}* is the *jth* element of *w_i*, and *σ_{ij}²* is the variance of cluster *i* in the *jth* rule.

In the this research project, in order to construction of MFIS and LFIS for estimating *S_w* and porosity, a fuzzy rule base was generated through FCM-derived input and output cluster centers. Each cluster center was used to generate a Gaussian membership function in each rule. That is, each rule is represented by a Gaussian MF which is constructed from the center and standard deviation of the corresponding cluster. So the number of membership functions and *if-then* rules for each input and output dataset is then equal to number of the clusters. As mentioned, the number of the FCM derived clusters, for water saturation, was equal to 31. Considering four inputs and one output, 31 by 5 MFs were generated participating in 31 fuzzy rules (Table 2). Accordingly, 44 by 8 MFs were constructed for the porosity model. To connect antecedents of each rule, the *min* operator was used. As mentioned, the fuzzy rule base structure for MFIS and LFIS is similar. Their main difference is in implication method. In MFIS, *min* operator was used for implication, whereas in LFIS *product* operator was used for this purpose. For the both techniques, the centroid defuzzification method was applied.

In SFIS, input MFs are of a Gaussian type. They were constructed using the cluster centers obtained from subtractive clustering (43 clusters for *S_w* and 50 clusters for porosity). However, output membership functions are linear equations constructed from inputs. For example, output MF1 of the *S_w* model, which is the consequent of rule no. 1, is constructed from four seismic attributes, as shown below:

$$\text{Output MF1} = \gamma_1 * \text{Time} + \gamma_2 * \text{Average frequency} + \gamma_3 * \text{Filter15/20-25/30} + \gamma_4 * \text{Dominant frequency} + \gamma_5 \quad \text{Eq. (9)}$$

In this equation, parameters γ_1 , γ_2 , γ_3 and γ_4 are coefficients corresponding to input seismic attributes. Parameter γ_5 is the constant of each equation. These parameters are obtained by linear least squares estimation. With these explanations, in order to estimate S_w , there will be 43 by 5 output MF parameters (Table 3). Accordingly, there will be 50 by 8 parameters showing coefficients and constants of the output MFs for the porosity estimation model. It is necessary to mention that, in this paper, only the MF parameters for water saturation model are shown. Input and output MF parameters, for the porosity models, were calculated in a similar process.

3.5 Construction of a Committee Fuzzy Interference System - CFIS

In this part of research, a CFIS was constructed for the overall prediction of petrophysical data by integrating the results of predicted data from SFIS, MFIS and LFIS, each of them having a weight factor showing its contribution in the overall prediction process. At the first step, outputs of the three fuzzy inference systems were averaged for predicting the target data, specifically; each of them has the weight value of 0.333. This output will be used as one of the experts of the CFIS.

In the second step, a genetic algorithm-pattern search tool was used to obtain an optimal combination of the weights for constructing the CFIS. The fitness function for GA-PS was defined as below:

$$MSE_{CMIS} = \sum_{i=1}^l 1/k(\beta_1 O_{1i} + \beta_2 O_{2i} + \beta_3 O_{3i} + \beta_4 O_{4i} - L_i)^2 \quad \text{Eq. (10)}$$

This function shows the MSE of the CFIS, for the training step predictions, where β_1 , β_2 , β_3 and β_4 are the weight coefficients corresponding to the outputs of Sugeno, Mamdani, Larsen and simple averaging method, respectively. O_i and L_i are output and target values, respectively, and k is the number of test data points (76 samples). The parameters of GA-PS are described as following:

The population is of a double vector type. The initial population size is 25, which specifies how many individuals are in each generation. Initial range is [0, 1]. This parameter specifies the range of the vectors in the initial population. The selection function was chosen as stochastic uniform, which chooses parents, for the next generation, based on their scaled values from the fitness scaling function. The

crossover function is scattered which 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 the crossover fraction is 0.78. This parameter specifies the fraction of the population that could be seen in the crossover children. The mutation function is Gaussian which 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. The 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. The hybrid function was chosen as pattern search. This is another minimization function that runs after the genetic algorithm terminates. Stopping the generation of GA was chosen at a value of 100. After 100 generations, change in the fitness function values over Stall generations was insignificant, and the mean fitness values for water saturation and porosity were fixed in 0.00915 and 0.00157, respectively. The results of running the genetic algorithm, with pattern search hybrid function including best and mean fitness values, average distance between individuals, fitness scaling and calculated scores for the porosity case, are shown in figure 9. Finally, the CFIS was constructed using the GA-PS derived coefficients for the results of SFIS, MFIS, LFIS and simple averaging method. That is, the final estimation of water saturation and porosity was done through Eq. (11) and Eq. (12), respectively.

$$S_{W_{CFIS}} = 0.303 * S_{W_{SFIS}} + 0.127 * S_{W_{MFIS}} + 0.098 * S_{W_{LFIS}} + 0.472 * S_{W_{Average}} \quad \text{Eq. (11)}$$

$$\varphi_{CFIS} = 0.087 * \varphi_{SFIS} + 0.249 * \varphi_{MFIS} + 0.214 * \varphi_{LFIS} + 0.450 * \varphi_{Average} \quad \text{Eq. (12)}$$

3.6. Design of a probabilistic neural network

Probabilistic neural network (PNN) is a forward feed network, built with three layers, and can be used for predicting both continuous or discrete (classification) data. It was first proposed by Specht (1990) and is based on a distance concept between objects. It is a very fast and efficient method to map a set of input data to their outputs. In PNN, for a vector of input x_i , each new output sample $O_N(x_i)$ is calculated as linear combination of the n data points in the training data using the following equation.

$$O_N(x_i) = \frac{\sum_{i=1}^n O_{Ni} \exp(-D(x, P_{x_i}))}{\sum_{i=1}^n \exp(-D(x, P_{x_i}))} \quad (\text{Eq. 13})$$

where $D(x, x_i)$ is the distance between the input point x and each of the training points, and it is calculated as follows:

$$D(x, x_i) = \sum_{j=1}^k \left(\frac{x_j - x_{ij}}{\rho_j} \right)^2 \quad (\text{Eq. 14})$$

where k is the number of input data, and ρ_j is the distance scale factor for each of the input attributes. The only parameter of the PNN that needs to be optimized is the scale factor ρ_j . In comparison with the other types of neural network, such as *multilayer perceptron* that requires many parameters to be optimized, PNN is simple, fast and efficient. The optimal value of ρ_j is obtained when the validation error is minimum. The validation result for the s^{th} target sample, which is left out of the training data, is calculated, as shown below

$$O_{Ns}(x_s) = \frac{\sum_{i=1}^n O_{Ni} \exp(-D(x_s, P_{x_i}))}{\sum_{i=1}^n \exp(-D(x_s, P_{x_i}))} \quad (\text{Eq. 15})$$

This process is repeated for all of the target samples so the mean squared error between the measured and predicted output can be calculated. More details on probabilistic neural networks can be found in Specht (1990), Masters (1995) and Hampson (2001).

For optimizing distance scale factor ρ_j range was taken between 0.10 and 3.00. The numbers of ρ_j values to try was set to 25. The optimized values of ρ_j are as below:

Water saturation cas:

Time: 0.244; Average frequency: 0.890; Filter 15/20-25/30: 2.484; Dominant frequency: 2.787; Global ρ_j : 0.342

Porosity case:

Inversion result: 0.313; integrate: 0.301; Quadrature trace: 0.804; Cosine instantaneous phase: 0.864; Integrated absolute amplitude: 0.385; Amplitude envelope: 0.121; Filter 15/20-25/30: 0.495; Global ρ_j : 0.463

4. Results and Discussion

The correlation coefficient and graphical comparison between measured and predicted water saturation for the test samples using SFIS, MFIS, LFIS and CFIS are shown in figures 10a-d and 11a-d. According to the results presented in Table 4, and figures 10 & 11, the MSE of SFIS, MFIS and LFIS models in the test data are 0.0136, 0.0161,

and 0.0169, which correspond to the correlation coefficient values of 0.868, 0.865, and 0.855, respectively. Amongst the models used, LFIS has provided more accurate results, compared to MSE of SFIS, MFIS. Applying simple averaging methods for the combination of the outputs of the three models has provided the correlation coefficient of 0.885 and the MSE of 0.0108, which shows improvement in comparison with individual models. The optimal combination of the weights in the CFIS was obtained by a GA-PS. The inputs of the CFIS are the outputs of three fuzzy models and simple averaging method. The GA-PS derived weights for SFIS, MFIS, LFIS and their average are 0.303, 0.127, 0.098 and 0.472, respectively. The MSE of the CFIS for the test data is 0.0091, which corresponds to the correlation coefficient of 0.896. That is, the CFIS shows a significant improvement for the estimation of S_w from seismic attributes. It performs better than any one of the individual intelligent systems acting alone for S_w prediction. Also it has provided a small improvement in comparison with simple averaging method. Application of the CFIS for porosity estimation from seismic attributes confirms the performance of the introduced methodology (figures 12 and 13). According to the results presented in Table 4 and figures 12 & 13, application of the MFIS results in the lowest error for porosity estimation (MSE=0.0020, CC=0.879) among fuzzy models acting alone. Application of the simple averaging method improves the results (MSE=0.0017, CC=0.891). Finally, the CFIS decreases the estimation error up to 0.0015 which corresponds to correlation coefficient of 0.899.

As it is evident from the results of this research, integrating the outputs obtained from different systems can improve the accuracy of the estimations. The simple averaging method is a kind of committee machine in which each of the inputs has equal contribution in the overall estimation. However, the CFIS, introduced in this study, is an advanced type of committee machine in which the optimal combination of weights is obtained by an optimization method, such as GA-PS. In order to evaluate performance of the designed CFIS, a neural network was used as an alternative method for estimating water saturation and porosity. Results show that a probabilistic neural network (PNN) provides more reliable results comparing to the other methods. Performance of the PNN was close to the best individual fuzzy inference system (MSE of 0.0140 for S_w and 0.0019 for porosity). That is, the CFIS method performs better than neural network and is the best individual fuzzy inference system.

Lastly, using the CFIS, constructed for this research, 3D seismic data of Ghar reservoir were converted to petrophysical data. Maps showing distribution of CFIS estimated water saturation and porosity for the Top Ghar reservoir are shown in figures 14 and 15, respectively. As the results show, the Ghar Sandstone is a high quality reservoir unit over the Iranian offshore oilfield. Porosity distribution over the reservoir is nearly uniform (mean=0.135). Towards the south east, porosity increases gradually. Water saturation in the central and north west sector of the reservoir is low (<50%), which corresponds to the hydrocarbon bearing area.

5. Conclusion

Fuzzy inference systems including Sugeno, Mamdani and Larsen were used for formulating petrophysical data to seismic attributes. Results indicate that, by the integration of different outputs into a Committee Fuzzy Inference System, using the GA-PS technique, a considerable improvement in accuracy of the target predictions can be achieved.

The CFIS performed better than any individual fuzzy models for estimating water saturation and porosity parameters. Also, it provided lower errors than the committee machine constructed through simple averaging method. The CFIS has a simple and easy structure, and when there are multiple ways to solve a problem, it can provide smaller errors when compared with the average of all experts, and with little additional computation.

The methodology introduced in this study is able to estimate petrophysical data from a large volume of 3D of seismic data. This can increase exploration success rates and reduce costs through the application of more reliable output results in hydrocarbon exploration programs.

6. Acknowledgements

The authors extend their appreciation to the National Iranian Oil Company (NIOC) for sponsorship, data preparation and their collaboration. Also, the vice-president of Research and Technology of the University of Tehran provided financial support for this research, for which we are grateful (Grant No. 6105023/1/02). We would like to thank Dr. Brian Henderson Russell for his invaluable experiences and publications in relating reservoir parameters to seismic attributes, which became a basis for

performing this research. We would like to thank Mr. Collin Williams for his constructive review of the article.

Nomenclature

A, B	Input fuzzy sets (membership functions)
α_i	Firing strength of i^{th} fuzzy rule
$\beta_1, \beta_2, \beta_3, \beta_4$	Constants in constructing output membership functions of Sugeno model
C	Output fuzzy set (membership function)
C'_i	Truncated output fuzzy set for the i^{th} rule
C'	Aggregated output fuzzy set
CC	Correlation coefficient
$D(x, x_i)$	Distance between the input point x and each of the training points
FCM	Fuzzy c-means clustering
FIS	Fuzzy inference system
GA-PS	Genetic algorithm-pattern search
k	Number of training samples in constructing CFIS and neural network
L_i	Target petrophysical data
LFIS	Larsen fuzzy inference system
LOM	Large of maximum
m	Number of clusters
MF	Membership function
MFIS	Mamdani fuzzy inference system
MIMO	Multiple inputs and multiple outputs
MISO	Multiple inputs and single output
MOM	Mean of maximum
MSE	Mean squared error
n	Number of fuzzy rules
O_i	Outputs of fuzzy models
O_N	Output of the probabilistic neural network
$O_{N\bar{E}}$	Validation result of the probabilistic neural network
PNN	Probabilistic neural networks
p, q	Coefficients of output membership functions in Sugeno fuzzy model
ρ	Distance scale factor
φ	Porosity
R^i	

r	i^{th} fuzzy rule
RB	Constant of output membership functions in Sugeno fuzzy model
γ	Fuzzy rule base
S_w	Weight coefficients of MFIS, LFIS, SFIS and average of them,
SFIS	respectively
SOM	Water saturation
t	Sugeno fuzzy inference system
τ_i	Small of maximum
σ_{ij}^2	Number of fuzzy rule base with MISO
\cup	Degree of fulfillment of rule i
u_i	Variance of cluster i in j^{th} rule
$\mu_{C_i}(z)$	Sum operator in fuzzy sets
$\mu_{C'_i}(z)$	i^{th} cluster center
$\mu_{A_i}(x_0)$	Grade of membership of element z in output fuzzy set C for i^{th} rule
$\mu_{A_i}(x_0)$	Firing strength of i^{th} fuzzy rule
μ_{R_i}	Grade of membership of element x_0 in i^{th} input fuzzy set A
v_{ij}	Grade of membership of element y_0 in i^{th} input fuzzy set B
w_{ij}	Grade of membership for i^{th} fuzzy rule
x, y	i^{th} cluster centers of j^{th} rule in input space
λ	i^{th} cluster centers of j^{th} rule in output space
z	Input data for fuzzy sets and neural network
	Constant
	Output of fuzzy set

7. References

- Aminzadeh, F., Chatterjee, S., 1985. Applications of clustering in exploration seismology. *Geoexploration* 23, 147–159.
- Aristimuño, J., Aldana, M., 2006. Fuzzy logic prediction of petrophysical parameters (porosity and velocity) at La Faja del Orinoco, Venezuela. *European Geosciences Union, Geophysical Research Abstracts*, 8, 03825.
- Balch, R.S., Stubbs, B.S., Weiss, W.W., Wo, S., 1999. Using artificial intelligence to correlate multiple seismic attributes to reservoir properties. *Society of Petroleum Engineers annual technical conference, Houston TX , ETATS-UNIS (03/10/1999)* , pp. 553-566.
- Bezdec, J.C., 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 256 pp.
- Chen, C.H., Lin, Z.S., 2006. A committee machine with empirical formulas for permeability prediction. *Computers and Geosciences* 32, 485–496.
- Chen, Q., Sidney, S., 1997. Seismic attribute technology for reservoir forecasting and monitoring. *The Leading Edge*, 16(5), 445-456.
- Chiu, S., 1994. Fuzzy model identification based on cluster estimation. *Journal of Intelligent and Fuzzy Systems* 2, 267-278.
- Chiu, S., 1995. Extracting fuzzy rules for pattern classification by cluster estimation. *Proc. 6th Int. Fuzzy Systems Association Congress (IFSA '95)*, Sao Paulo, Brazil, 2, 273-276.
- Chiu, S., 1997. *Fuzzy Information Engineering: A Guided Tour of Applications* (Chapter 9), John Wiley & Sons, 712 pp.
- Chopra, S., Marfurt, K., 2006. Seismic Attributes-a promising aid for geologic prediction: *Canadian Society of Exploration Geophysicists Recorder*, 111-121.
- Dubois, D., Prade, H., Yager, R., 1997. *Fuzzy information engineering: a guided tour of applications*. John Wiley & Sons, Inc. New York, USA, 712pp.
- Jarrah, O.A., Halawani, A., 2001. Recognition of gestures in Arabic sign language using neuro-fuzzy systems. *Artificial Intelligence* 133, 117–138.
- Hampson, D.P., Schuelke, J.S., Quirein, J.A., 2001. Use of multiattribute transforms to predict log properties from seismic data. *Geophysics* 66 (1), 220-236.

- Holland, J.H., 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, USA, 183 pp.
- 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 $\Delta \log R$ and neuro fuzzy techniques. *Journal of Petroleum Science and Engineering* 45, 141–148.
- Larsen, P.M., 1980. Industrial applications of fuzzy logic control. *International Journal of Man-Machine Studies* 12, 3-10.
- Lee, K.H., 2004. *First Course on Fuzzy Theory and Applications*. Springer, 335pp.
- 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* 49 183–92.
- Mamdani, E.H., Assilian, S., 1975. An experimental in linguistic synthesis with a fuzzy logic control. *International Journal of Man-Machine Studies* 7, 1–13.
- Mamdani, E.H., 1976. Advances in the linguistic synthesis of fuzzy controllers. *International Journal of Man-Machine Studies* 8, 669-678.
- Mamdani, E.H., 1977. Applications of fuzzy logic to approximate reasoning using linguistic synthesis. *IEEE Transactions on Computers* 26, 1182-1191.
- Masters, T., 1994. *Signal and image processing with neural networks*. John Wiley & Sons, Inc
- Matlab user's guide, 2007. *Fuzzy logic, genetic algorithm and direct search toolboxes*, Matlab CD-Room. By the math works, Inc.
- Meldahl, P., Heggland, R., Bril, B., Groot, P. de., 2001, Identifying faults and gas chimneys using multiattributes and neural networks. *The Leading Edge* 20, 474-478.
- Mitchell, M., 1999. *An Introduction to Genetic Algorithms*. Massachusetts Institute of Technology press, 158 pp.
- Mohaghegh, S., 2000. Virtual-intelligence applications in petroleum engineering: Part I. Artificial neural networks *Journal Petroleum Technology* 52, 64–73.

- Mohaghegh, S., Koperna, G., Poa, A.S., Hill, D.G., 1999. Reducing the lost of field-scale log analysis using virtual Intelligence techniques, Society of Petroleum Engineers 57454.
- Nikraves, M., Novak, B., Aminzadeh, F., 1998. Data mining and fusion with integrated neuro-fuzzy agents: Rock properties and seismic attenuation. JCIS 1998, The Fourth Joint Conference on Information Sciences, North Carolina, USA, October 23-28.
- Nikraves, M., Adams, R.D., Levey, R.A., 2001. Soft computing: tools for intelligent reservoir characterization (IRESC) and optimum well placement (OWP). Journal of Petroleum Science and Engineering 29, 239-262.
- Nikraves, M., Aminzadeh, F., 2001. Mining and fusion of petroleum data with fuzzy logic and neural network agents. Journal of Petroleum Science and Engineering 29, 221-238
- Nikraves, M., Hassibi, M., 2003. Intelligent reservoir characterization (IRESC). Industrial Informatics Proceedings. IEEE International Conference on Volume, Issue, 21-24, 369-373.
- Rezaee, M.R., Kadkhodaie-Ilkhchi, A., Barabadi, A., 2007. Prediction of shear wave velocity from petrophysical data utilizing intelligent systems: an example from a sandstone reservoir of Carnarvon Basin, Australia. Journal of Petroleum Science and Engineering 55, 201-212.
- Russell, B.H., Ross, C.P., Lines, L.R., 2002, Neural networks and AVO: The Leading Edge 21 (3), 268-277.
- Russell, B.H., Lines, L.R., and Hampson, D.P., 2003, Application of the radial basis function neural network to the prediction of log properties from seismic data: Exploration Geophysics 34, 15-23.
- Russell, B.H., 2004. The application of multivariate statistics and neural networks to the prediction of reservoir parameters using seismic attributes. Ph.D. thesis, University of Calgary, Alberta. 392pp.
- Saggaf, M.M., Nebrija, Ed. L., 2003 A fuzzy approach for the estimation of facies from wireline logs. American Association of Petroleum Geologists Bulletin, 87 1233-40.
- Soubotcheva, N., Stewart, N., 2006. Predicting lithology and porosity at the Pikes Peak heavy oilfield, Saskatchewan using 3C seismic data and well logs. Canadian Society of Exploration Geophysicists Recorder, 35-39.

- Specht, D., 1990. Probabilistic neural networks. *Neural Networks*, 3, 109-118.
- Sugeno, M., 1985. *Industrial Application of Fuzzy Control*, North-Holland, New York. 269pp.
- Taner, M.T., Schuelke, J.S., O'Doherty, R., Baysal, E., 1994. Attributes revisited. *SEG Expanded Abstracts* 13, 1104-1107
- Trappe, H., Hellmich, C., 2000. Using neural networks to predict porosity thickness from 3D seismic *First Break* 18 (9), 377-384
- Wong, P.M., Nikraves, M., 2001. Introduction: Field applications of intelligent computing techniques. *Journal of Petroleum Geology* 24 (4), 381-387.
- Zadeh, L. A., 1965. Fuzzy Sets. *Information and Control* 8, 338-353.

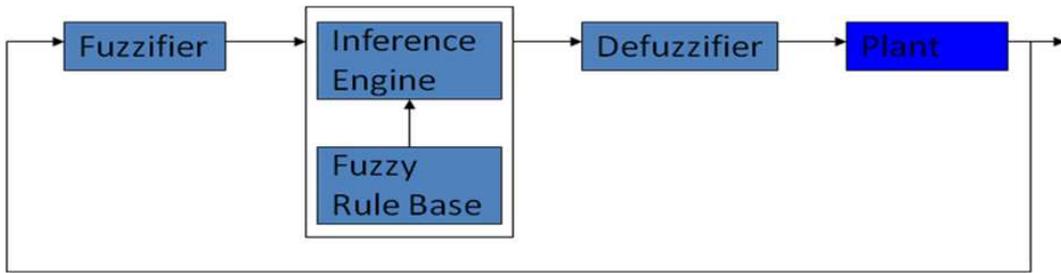


Figure 1

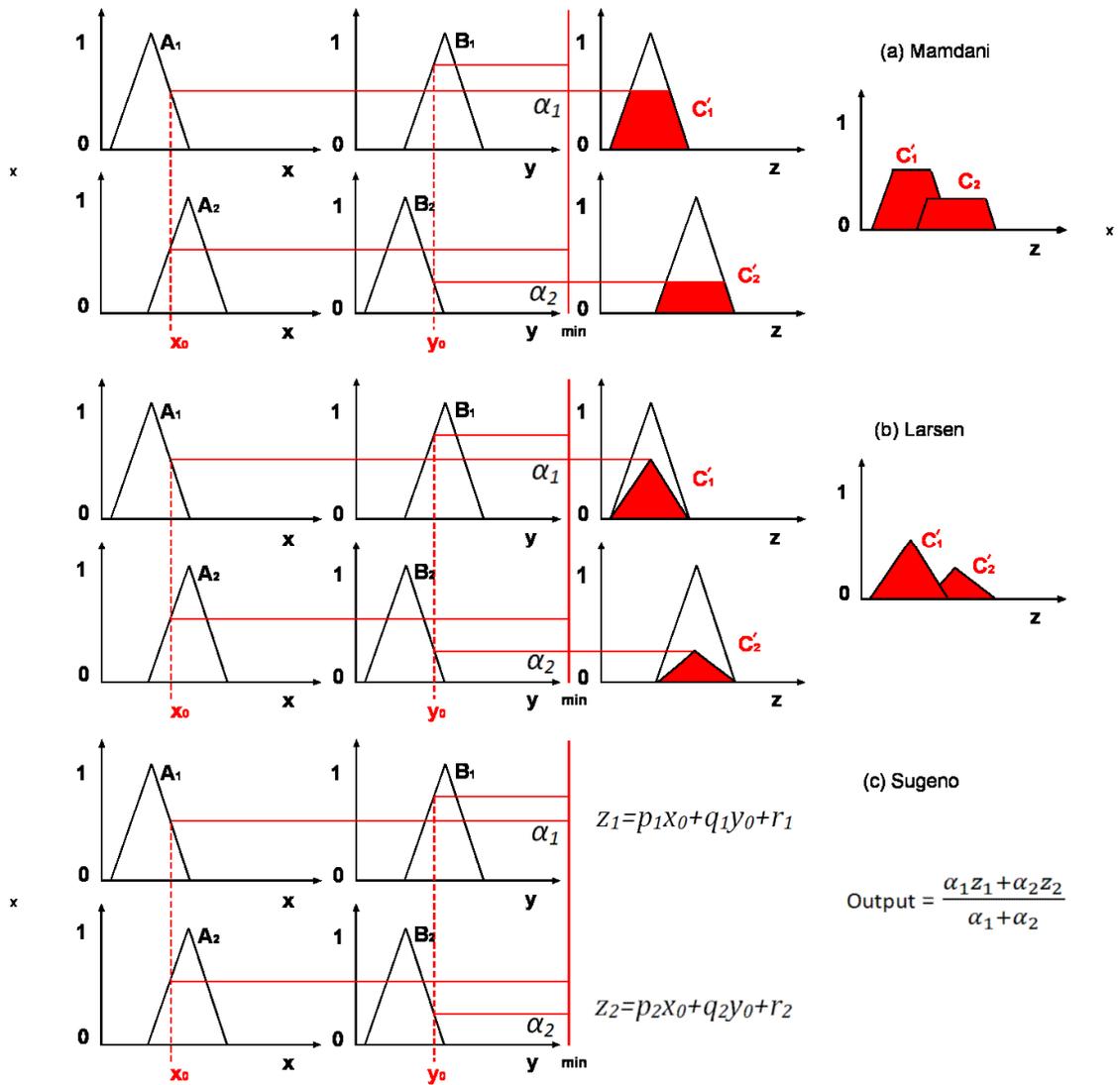


Figure 2

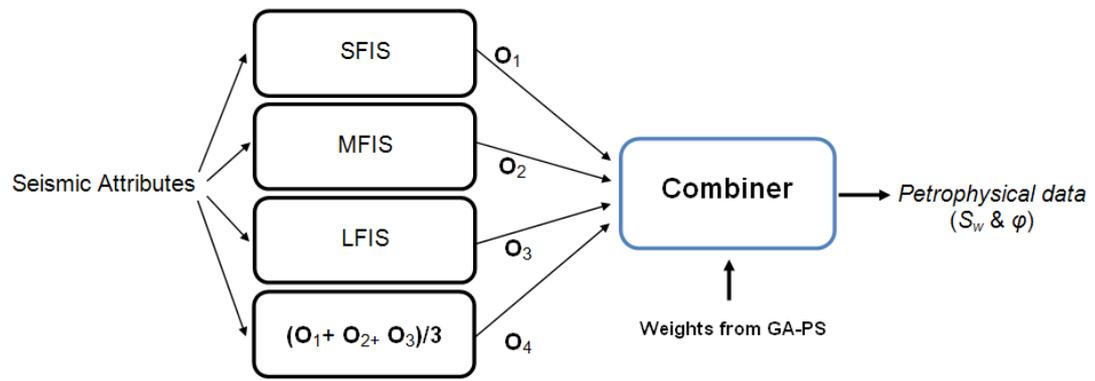


Figure 3

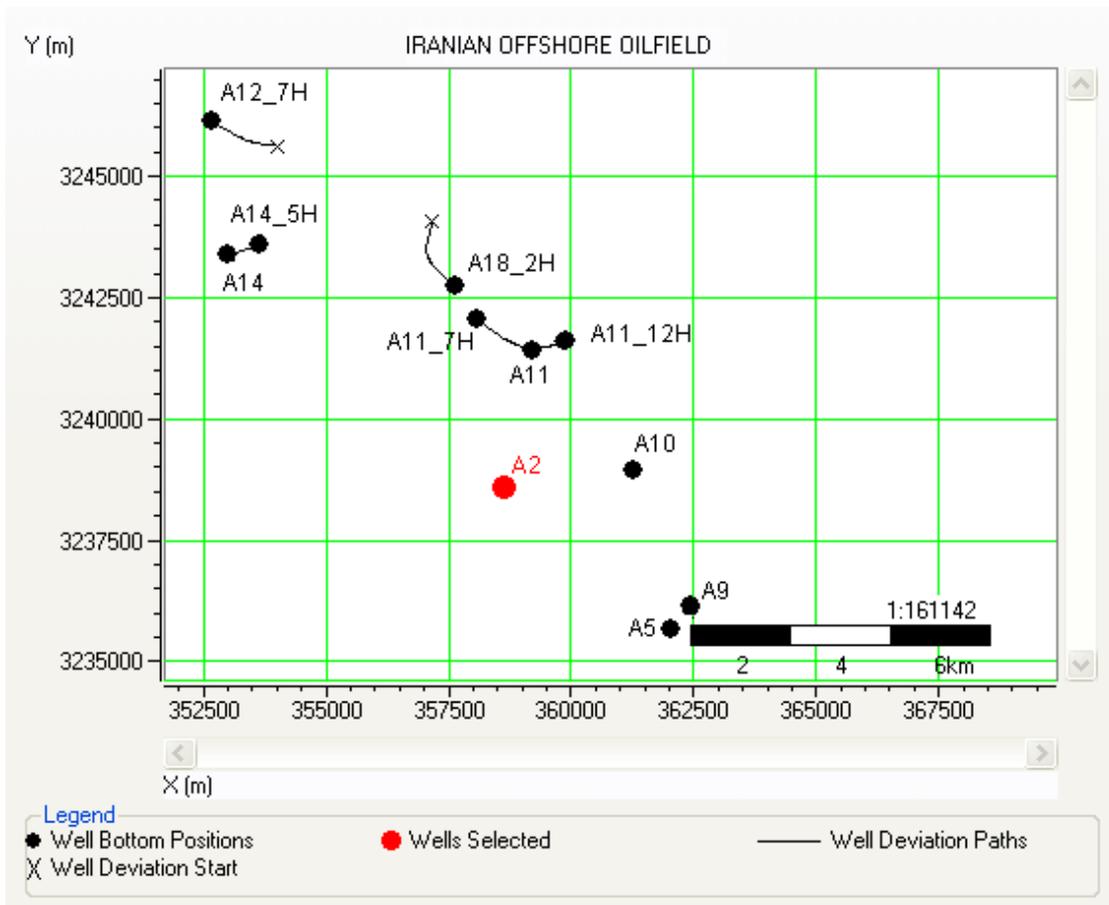


Figure 4

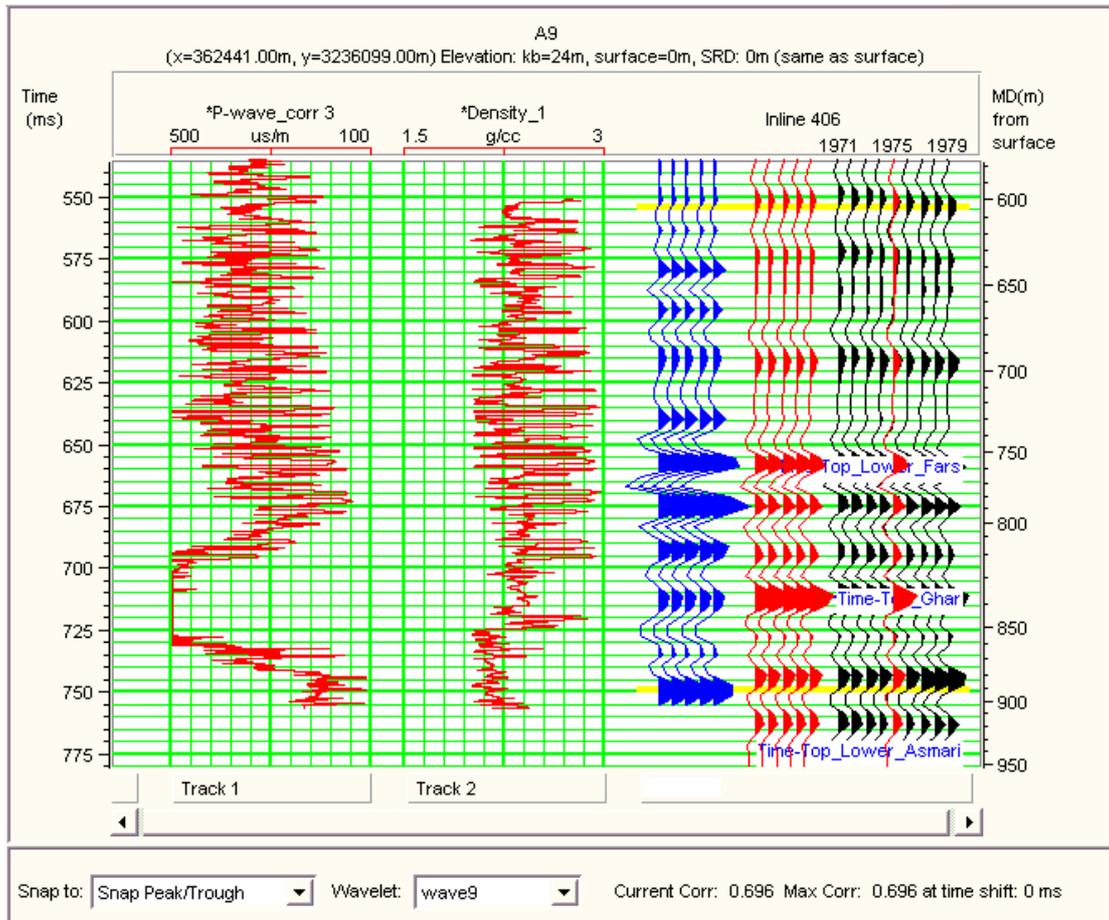


Figure 6

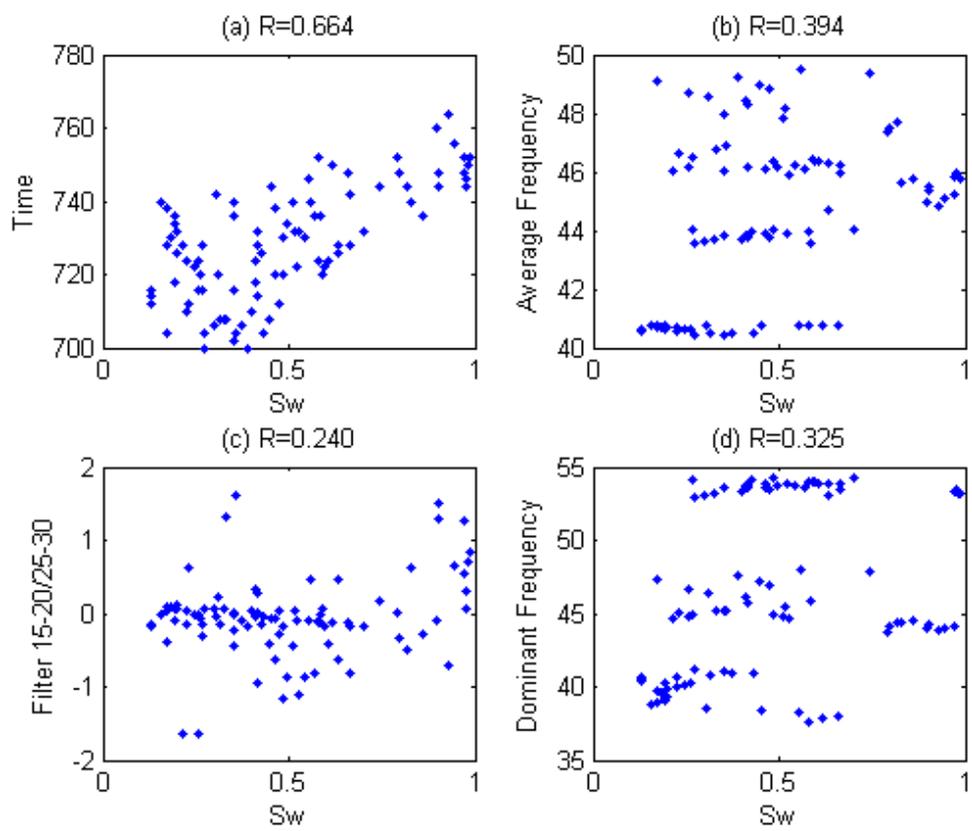


Figure 7

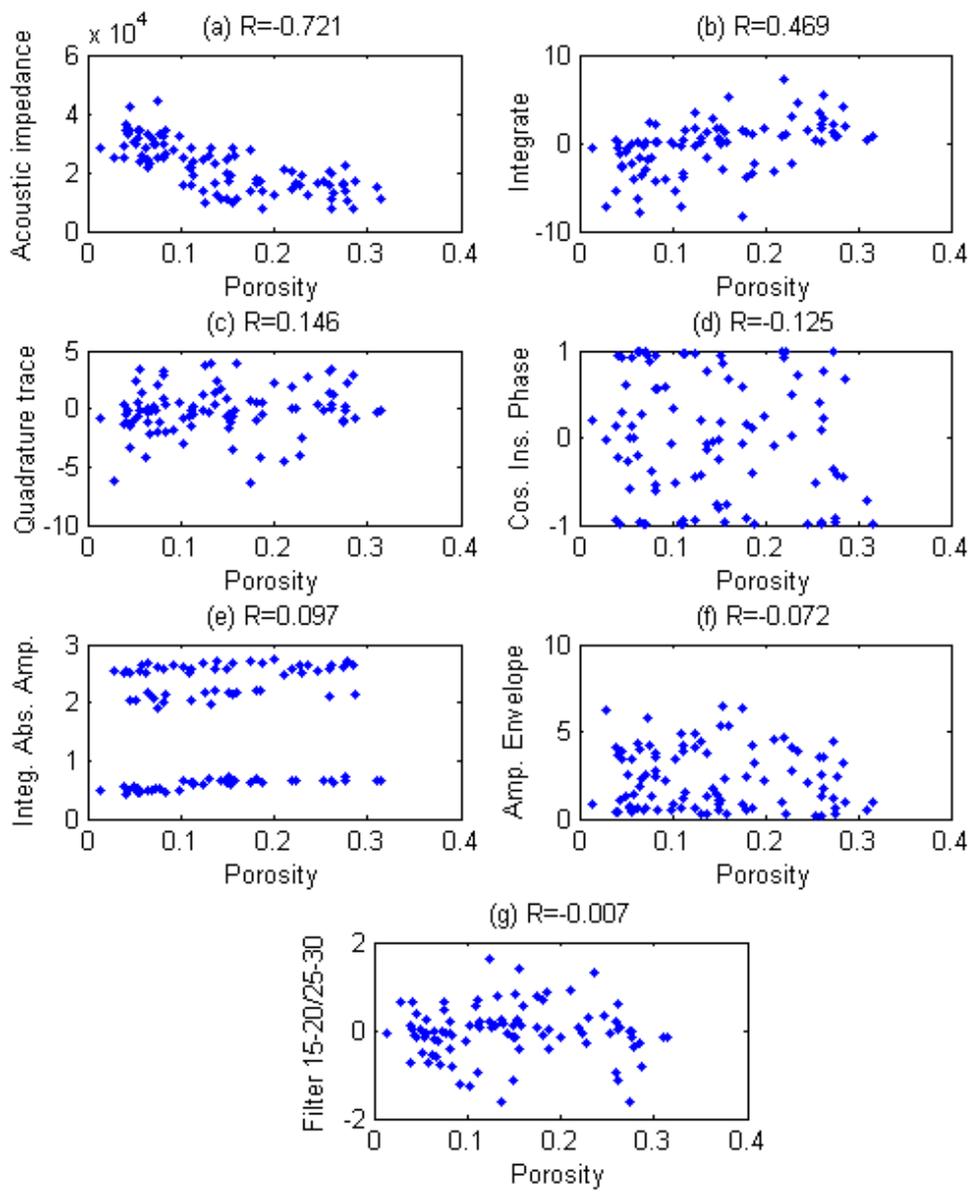


Figure 8

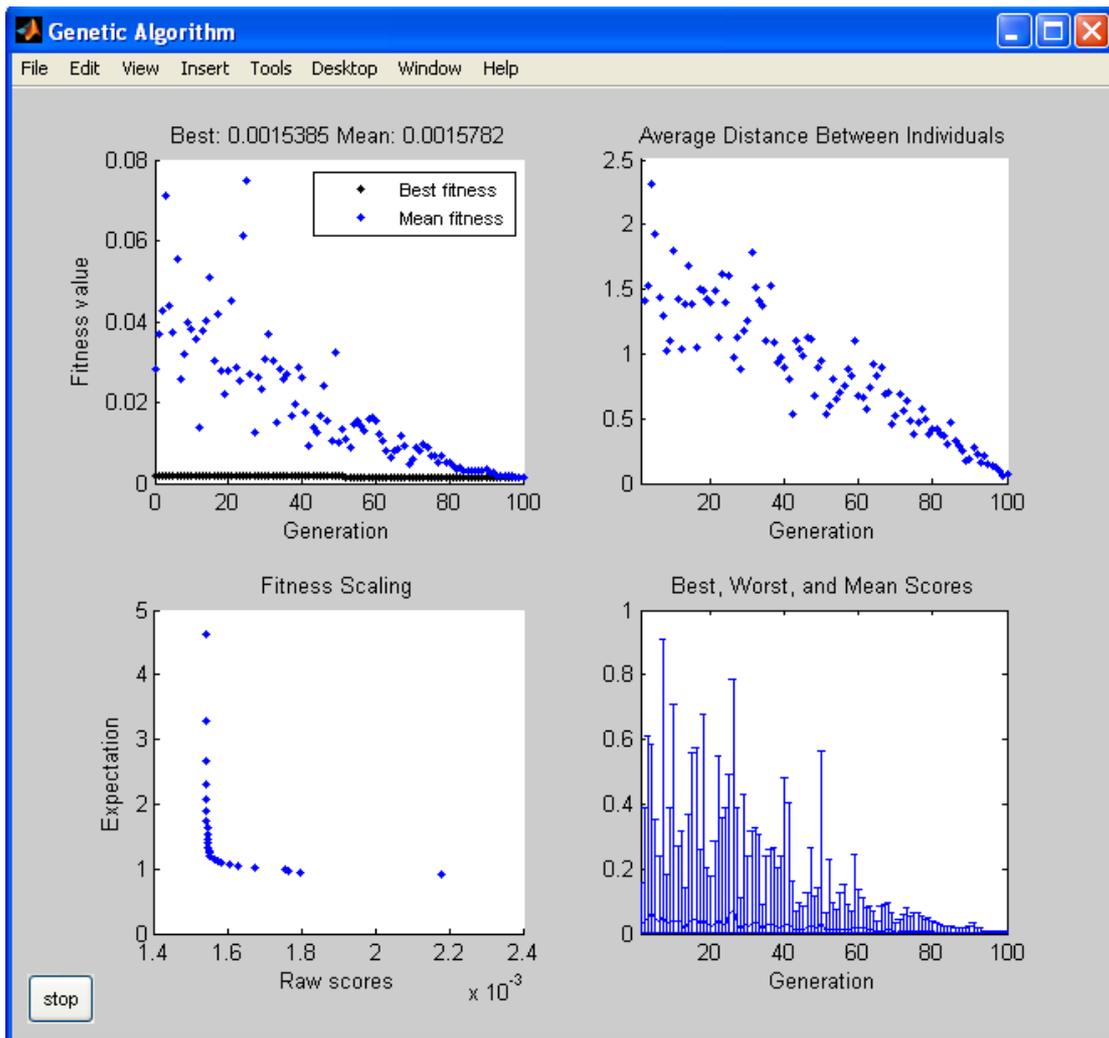


Figure 9

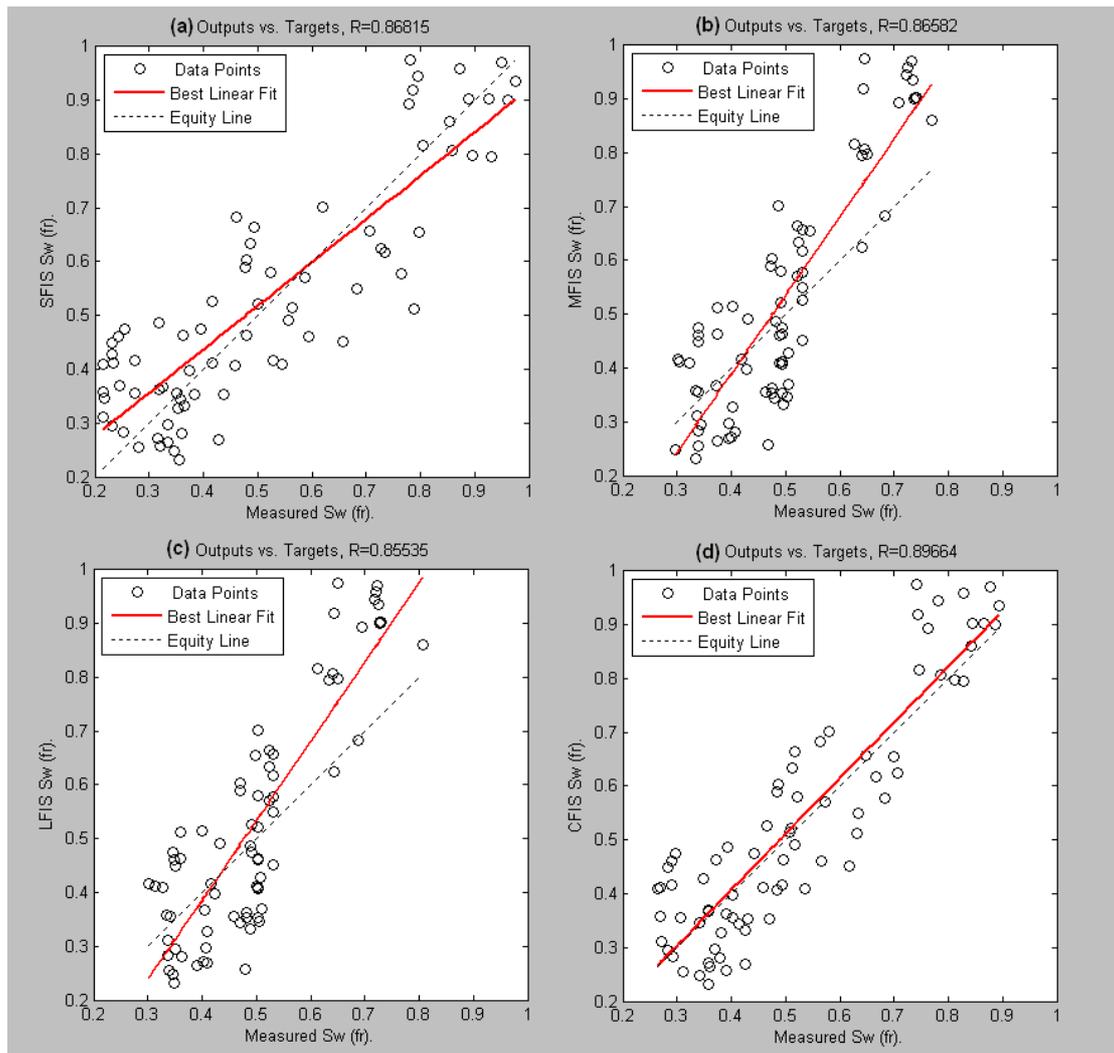


Figure 10

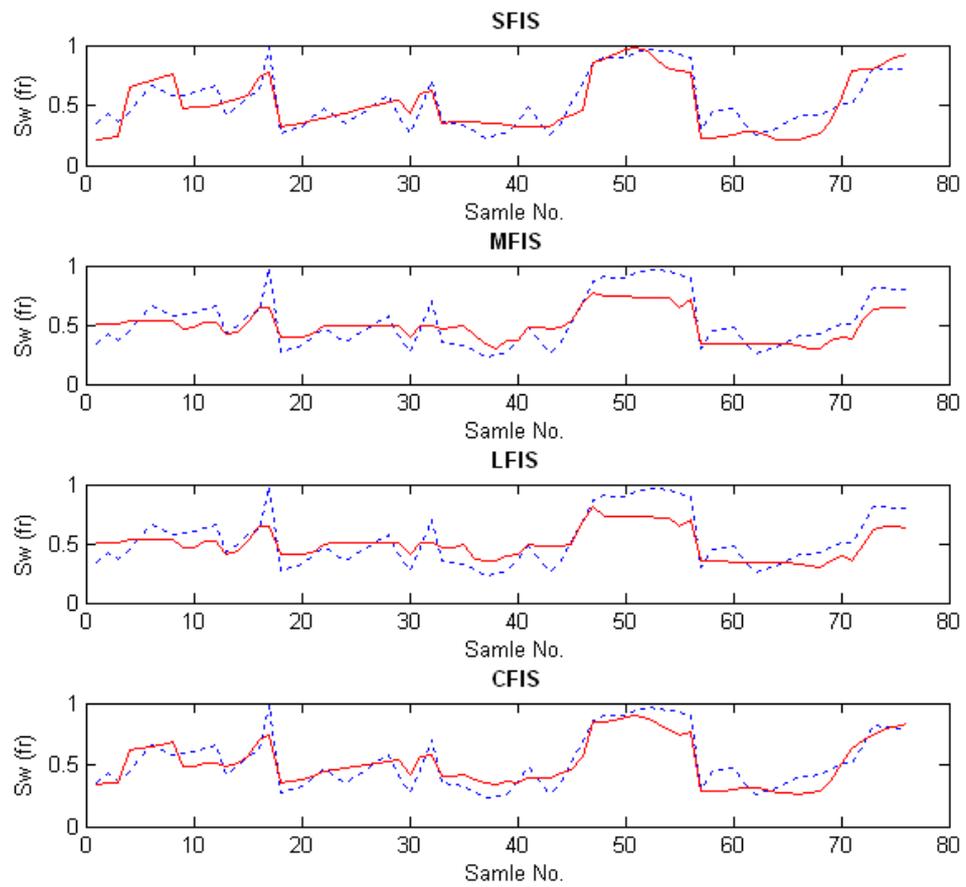


Figure 11

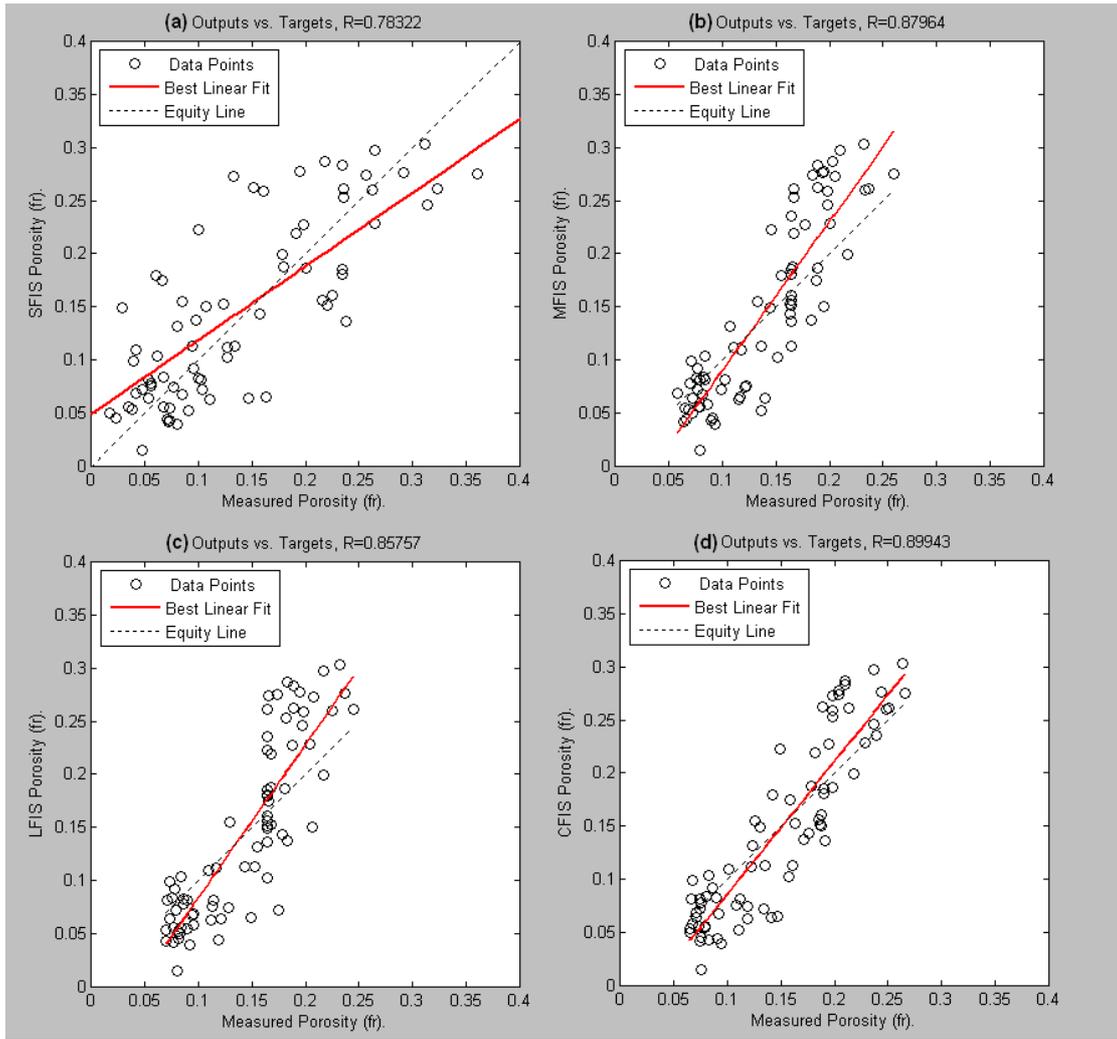


Figure 12

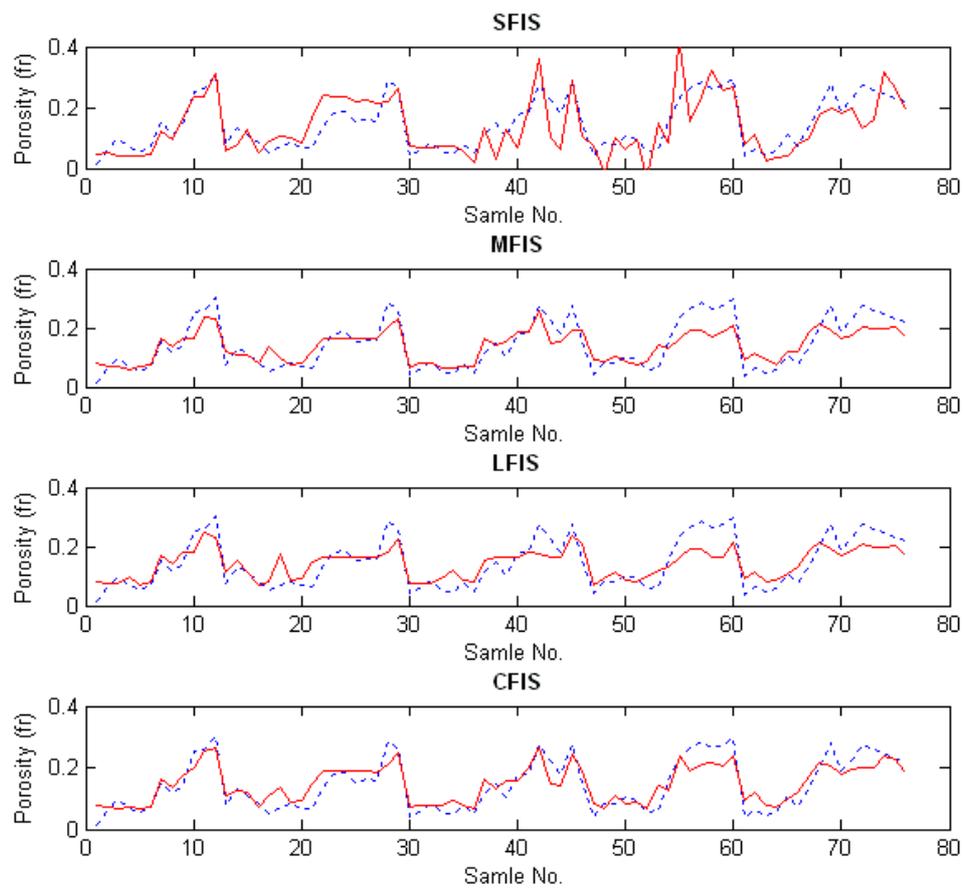


Figure 13

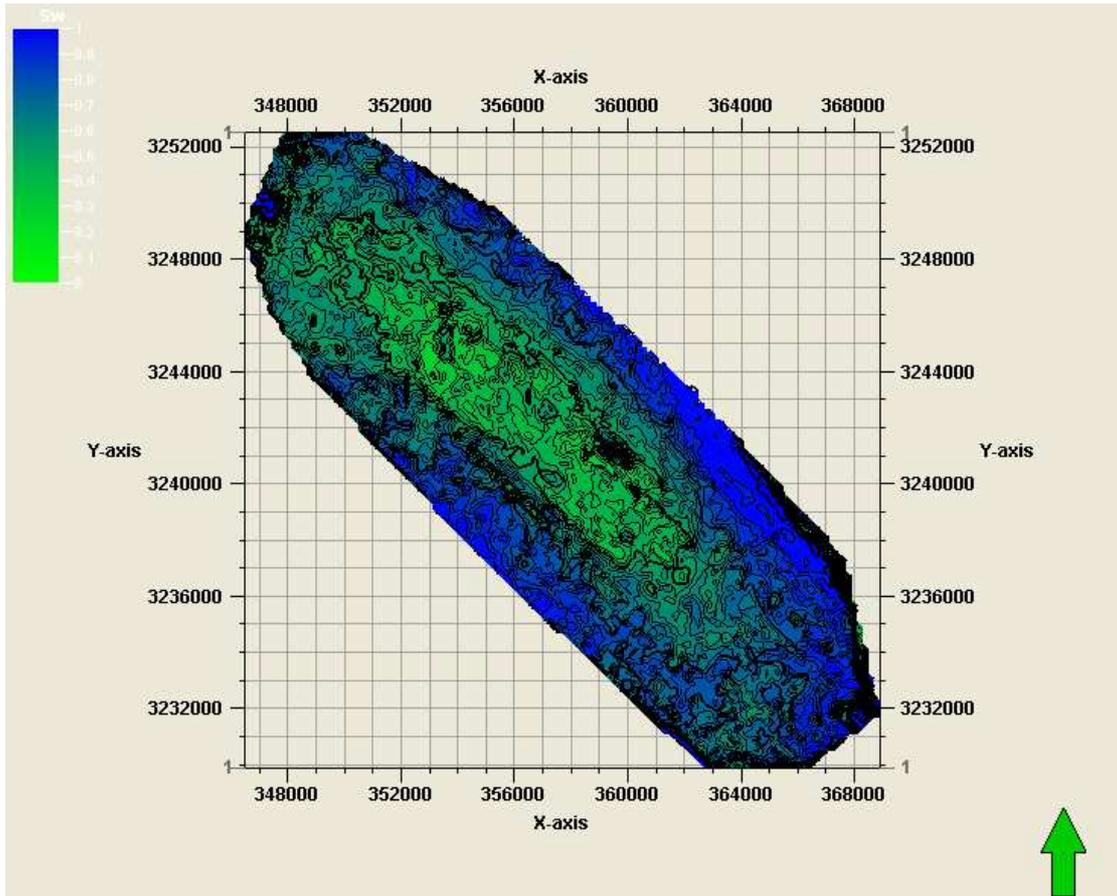


Figure 14

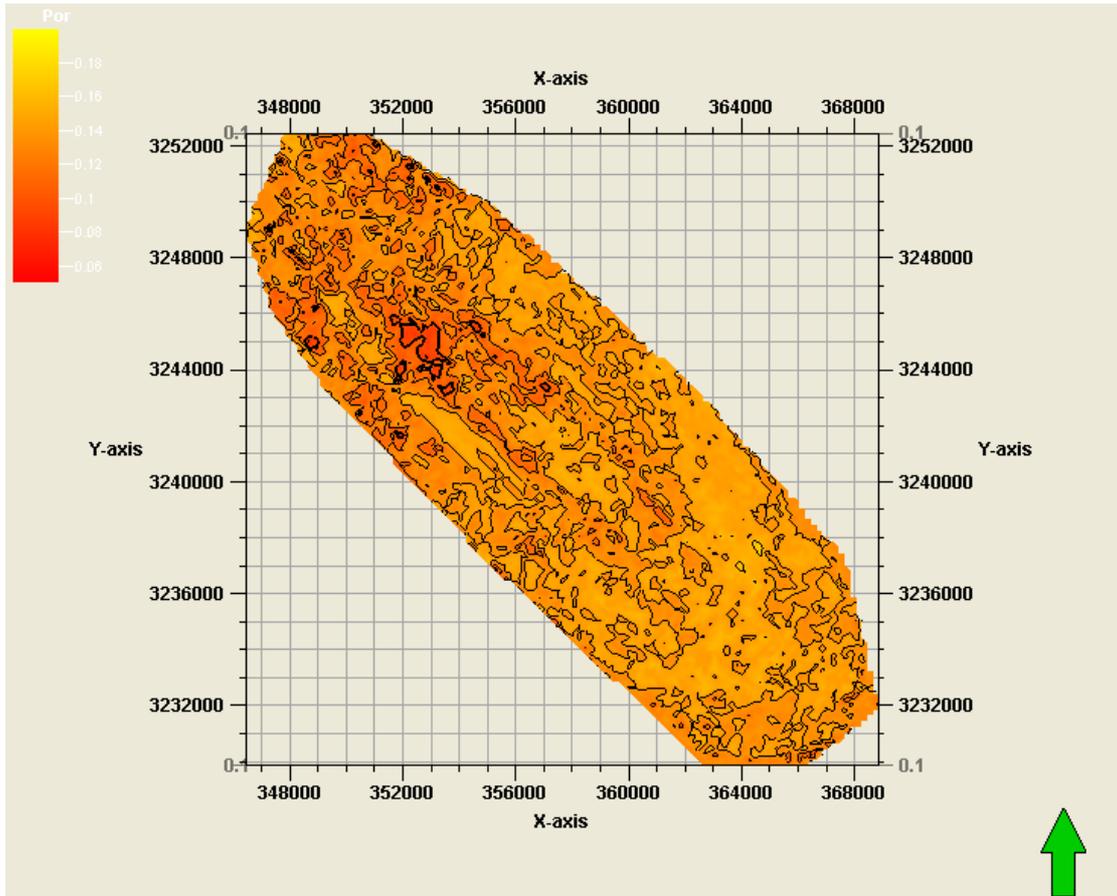


Figure 15