The Application of Multilayer Perceptron Neural Network in Volume of Clay Estimation: Case Study of the Shurijeh Gas Reservoir, Northeastern Iran

Golnaz Jozanikohan*[a], Gholam Hossain Norouzi[a], Fereydoun Sahabi[a], Hossein Memarian[a], Behzad Moshiri[b]

Abstract

Volume of clay is an important component in the assessment of shaly sand reservoirs, due to its significant impact on the production characteristics. The Shurijeh sandstone Formation of Lower Cretaceous age, with subordinate shales is one of the most challenging gas reservoirs to be properly characterized in the eastern Kopet-Dagh sedimentary Basin, Northeastern Iran. This paper describes the improvement achieved in estimating the volume of clay in the Shurijeh reservoir Formation, with an application to a gas producing well and another non-producing well in a joint field between Iran and Turkmenistan. A clear comparison between estimates from several conventional petrophysical methods and actual laboratory measured data of 76 core samples showed very large estimation errors; therefore an attempt has been made to improve the estimation with developing a multilayer feedforward backpropagation neural network. Six types of well logs were selected, through a sensitivity analysis, as the most relevant input data to the volume of clay (network output). Data were then standardized and randomly divided into three sets of 70% for training, 15% for validation and 15% for testing. Three different training algorithms of genetics, particle swarm optimization, and Levenberg-Marquardt were tested on a network with certain topology, and the latter was

*Corresponding author:
E-mail address: gjkohan@ut.ac.ir, (Golnaz Jozanikohan)
Telephone: (+98 21 82084291)
Fax: (+98 21 88008838)
[a]School of Mining Engineering, College of Engineering, University of Tehran, Tehran, Iran
[b]School of Electrical and Computer Engineering, Control and Intelligent Processing Center of Excellence, College of Engineering, University of Tehran, Tehran, Iran

chosen, due to the better performance. The number of hidden layer neurons, was efficiently determined 8 through a trial and error process. The developed network with the architecture of 6-8-1 was then successfully validated with 16 unseen core data. Mean squared error of $2.8069E^{-3}$ and a correlation coefficient of 0.9013, as the network criteria; clearly showed a multilayer neural network can significantly improve the estimation of volume of clay by 53% within data set from the Shurijeh Formation.

Keywords: Volume of clay, Artificial Neural Network, Core Analysis, Shaly sand reservoir, Kopet-Dagh sedimentary Basin

1. Introduction

Varying amounts of clay minerals are present in the most clastic hydrocarbon reservoirs (Fertl et al. 1982), which can greatly influence on the reservoir quality from both the petrophysical and geomechanical perspectives (SLB, 2009). Since the imprecise estimation of volume of clay (V_{cl}), can result in a big difference in the billions cubic meters of recoverable hydrocarbons, having an accurate estimate of clay mineral content is of primary importance to the reservoir evaluation studies (Asquith et al. 2004). The miscalculation risk for two important reservoir parameters, i.e., water saturation and effective porosity, drastically increases without an accurate quantitative knowledge of clay minerals present within a reservoir (Causey, 1991). Based upon the core availability, the actual value of volume of clay can be determined, using laboratory methods such as X-ray diffraction (XRD); but in the most cases, direct measurements are not always available (Adeoti et al. 2009). Volume of clay relates to almost all types of well logging data with some defined relationships (Ellis and Singer, 2008; Serra, 2008), however, no single or combination of relationships from petrophysical logs, can accurately estimate it (Causey, 1991; Dewan, 1983). The fact that few or none of conventional petrophysical methods can provide estimates as accurate as the artificial intelligence techniques is due to the complexity and the high nonlinearity in actual relationships between volume of clay and well logging data (Nikravesh and Aminzadeh, 2001). The artificial intelligence techniques, particularly neural networks, have the remarkable ability to establish a complicated mapping between non-linearly linked input and output data (Nikravesh et al. 2003). In other words, the essential requirement for using the ANN in this specific problem domain has been justified, due to the nonlinear nature of actual relationships between volume of clay and well logging data as well as the superior ability of neural networks in solving difficult problems that do not yield to traditional algorithm approaches.

Artificial neural network (ANN), firstly introduced by McCulloch and Pitts (1943), is a powerful computational model inspired by biological nervous systems consisting of a large number of elemental units, called neurons, organized in input, hidden, and output layers (Fausett, 1993). Any neuron in the network is characterized by some features such as input weights, a threshold, and an activation function (Haykin and Network, 2004). The adjusting weights connect the neurons in different layers, so that a particular input, according to a learning algorithm, leads to a specific target output (Fausett, 1993). In the simplest type of artificial neural network, neurons are generally connected in a feedforward manner to allow data to move only in the forward direction, from input through the hidden to finally reach to the output neurons (Auer, 2008). A multilayer perceptron (MLP), firstly introduced by Rosenblatt (1958), is a feedforward artificial neural network model, consisting of multiple layers of neurons fully connected to the next neurons in each layer (Hornik et al. 1989). With adequate learning and sufficient numbers of hidden layers, the MLP networks are indeed capable of universal approximation in a precise and satisfactory sense for any type of bounded piecewise continuous functions (Funahashi, 1989; Hornik et al. 1989). The MLP network utilizes a supervised learning technique called backpropagation (BP), which is widely suggested as the most efficient procedure in the training of neural networks and used in conjunction with an optimization method such as gradient descent (Werbos, 1994). The BP method basically calculates the differences between the estimated and the measured values to define an error (Demuth and Beale, 2002). The error is then backpropagated through the network to update the weights and obtain the optimum results (Werbos, 1994). A predefined threshold for the differences controls the process, until it ends (Demuth and Beale, 2002; Yang and Rosenbaum, 2002).

It is known that with training an expert system from known fused input data, any complex type of nonlinear problems can be handled in the earth sciences (Ouenes, 2000). Recently, the neural network technique has gained the most attention as an accurate, fast method yielding results somewhat superior to the conventional methods due mainly to the independence from any prior knowledge about the nature of relationships between the input and output variables (Nikravesh and Aminzadeh, 2001). In the last two decades, neural networks have been widely used in solving many different problems of upstream oil industry such as estimation of fundamental reservoir properties (Aminzadeh et al. 1999; Lim, 2005; Liu and Liu, 1998; Ouenes, 2000; Verma et al. 2012; Walls et al. 2000; Wiener et al. 1991), prediction of petrophysical properties from well log data (Fung et al. 1997; Quirein et al. 2000), prediction of complex lithologies (Benaouda et al. 1999; Bueno et al. 2006; Rogers et al. 1992; Wang and Zhang, 2008), generating synthetic well logs (Rolon et al. 2009), oil field production forecast (Chen and Lang, 2003; Liu et al. 2008), predicting temperature profiles in producing oil wells (Farshad et al. 2000) as well as ranking reservoirs in order of exploitation priority (Li et al. 2008).

The present study aims to improve the volume of clay estimation in the Shurijeh gas reservoir Formation from a suite of well logging data (input data) and core measured volume of clay (desired target) using a multilayer perceptron artificial neural network (MLP) approach. The originality of this paper lies in the 1) comparison of conventional petrophyscial methods for estimation the volume of clay with modern techniques based on the artificial intelligence in a given formation which is rare in the literature; 2) based on the extended literature review, this is the first research in which artificial neural network is used for estimating the most critical parameter in the reservoir evaluation studies, volume of clay, and benefits from large quantitative mineralogical core data sets (76 samples) that can be used for volume of clay estimations in other wells located in the area studied.

2. Geological setting

The study area in this research is located in the Kopet-Dagh tectono-sedimentary unit, Northeastern Iran (Fig. 1). The Kopet-Dagh sedimentary Basin stretches for 700 km along the Southeastern corner of the Caspian Sea in the west to the Iran-Turkmenistan border in the east (Stoklin, 1968). In this area, a thick Mesozoic-Tertiary sedimentary sequence, deposited in a deep, narrow sedimentary trough and was folded in the young alpine Neogene-Quaterner phases to form the northern end of the Alpine-Himalayan mountain belt in Iran (Eftekharnejad et al. 1991). Table 1 shows the Cretaceous lithostratigraphic sequence in the area under study.

The Shurijeh Formation, hosting main sweet gas reservoirs in the eastern Kopet-Dagh Basin, primarily consists of mixed red bed sediments, including sandstone, siltstone, and claystone which have been deposited in a variety of continental, coastal, and marine environments through Upper Jurassic to Lower Cretaceous times (Moussavi-Harami and Brenner, 1992). Based on its varying lithology, the Shurijeh is divided into three parts of upper, middle and lower (NIOC, 1986). The upper part is mainly composed of coarse to medium grained sandstone and siltstone alternating with thin beds of partly gypsiferous, silty clay to claystone (NIOC, 1986). The middle part is composed of hard quartzitic sandstone interbedded with thin layers of siltstone and silty claystone (NIOC, 1986). The lower part consists of gypsiferous claystone alternating with thin beds of very fine grained partly glauconitic sandstone and hard anhydrite (NIOC, 1986). The Shurijeh Formation is barren of indicator fossils, but it is inferred by some researcher to be of Neocomian age (Afsharharb, 1979; Kalantari, 1969).

3. Data set

3.1. Core data

The conventional and artificial intelligence approaches, presented in this paper are based on data collected from two wells, approximately 8.5 kms distance from each other in the eastern end of the Gonbadli structure, Iranian part of Kopet-Dagh Basin (Fig. 1). Both wells were drilled in the Shurijeh Formation, one of which produces sweet gas and the other is non-producing. The core samples were selected from almost every half meter of the Shurijeh Formation. 20 samples were taken from the interval of 3202.8-3210 m in the gas producing well, and 56 samples were taken between depths of 3180 m and 3207.55 m in the non-producing well. The inner parts of each core sample were used for laboratory analysis to protect the results from possible contaminations effects which normally found on the outer parts of samples. The cores were then crushed (<mm) and mixed to obtain a homogeneous mixture. To increase the number of data points for developing the neural network in next steps, 1024 data points were added to the data set from estimating the volume of clay, using a geostatistical approach introduced by Jozanikohan et al. (2014a), exclusively for the Shurijeh Formation. Several model-checking algorithms were successfully validated this specific geostatistical model. The mentioned relationship was further verified based on the data obtained from all of the instrumental methods (e.g. XRD, XRF, FTIR, and thermal analysis) in the Shurijeh cored intervals. The relationship has no bias (Σ estimated volume of clay –

mean of laboratory measured volume of clay=0.0) and the minimum possible estimation error.

3.2. Wireline logging data

A complete set of wireline logging data, including, natural gamma ray log and its spectral component data (potassium, thorium, and uranium), density, sonic, neutron, resistivity (laterolog deep resistivity, laterolog shallow resistivity, and micro spherically focused resistivity), and photoelectric was available from the Shurijeh Formation in the gas producing well and other non-producing well in one of eastern Kopet-Dagh fields. The logging vertical sampling interval was 15 cm to acquire 712 data samples from depths of 3182.9 m down to 3254 m in the gas producing well and 458 data from depths of 3174 m down to 3219.7 m in the non-producing one. Schlumberger environmental correction chart (2009), GR-1, was used to correct the effect of borehole size and mud weight from natural gamma ray and its spectral component data. Since the depths of core samples did not precisely match with the depths of well log data, the average of the two nearest neighbor logging data was calculated for a given core sample to have core data at equal intervals corresponding to the depth in the logging data.

3.3. Input data selection, using correlation criteria

The input data for the MLP network should be accurately selected with a great care in order to only assign the key variables to the network and reduce the model uncertainty (Simpson, 1990). To attain the most affective factors on the volume of clay, firstly a sensitivity analysis was carried out on the complete set of logging data, using cross correlations. The Pearson correlation coefficient has been used extensively in selecting proper inputs for the ANN (Wilby et al. 2003; Guyon and Elisseeff, 2003). In the Pearson correlation matrix, linear correlation between two variables is expressed by a value, (r), between +1 and −1 (Kenney and Keeping, 1954). The upper and lower limits are assigned to total positive and negative correlations, respectively. The linear relationship between X_i and Y_i is defined in terms of Pearson coefficient as (Kenney and Keeping, 1954):

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{1}$$

where, n is the number of data points, X_i and Y_i are the sample paired data, and $\bar{X}$ and $\bar{Y}$ designate the mean of X_i and Y_i, respectively (Kenney and Keeping, 1954).

Based on data of the Pearson correlation coefficient matrix for the Shurijeh well logging data given in Table 2, the relative importance/effectiveness of input well logging parameters in relation to the volume of clay is shown in Fig. 2. It is clear from Figure 2 that the thorium, natural gamma ray, and the sonic logs are the most influential factors on the volume of clay, unlike the photoelectric, uranium, potassium, laterolog deep and shallow resistivity. The last five mentioned logs are the least sensitive parameters affecting the volume of clay and could be excluded from the input variables to build a more compact network. Six types of well logging data, comprising, thorium (THOR), natural gamma ray (GR), sonic (DT), density (RHOB), micro spherically focused resistivity (MSFL), and neutron (NPHI), were chosen to be used in this study.

4. Methods

4.1. Core analyses

The Shurijeh clay mineral content in a selection of 76 core samples was measured by quantitative X-ray diffraction (XRD) technique. The non-clay minerals were separated according to the procedures described by Moore and Reynolds (1989). The oriented slides were then prepared from the less than four micron size fraction and exposed to Ni-filtered Cu

K_α radiation generated by a Bruker AXS, D8 Advance X-ray diffractometer and scanned from 4 to 40 degrees 2θ with a speed of 1.2 degrees per minute at 40 kV and 30 mA. Based on the patterns of air-dried, glycolated, heated and treated with the hydrochloric acid, clay minerals were consisted mainly of illite, magnesium rich chlorite, kaolinite with smaller amounts of glauconite, montmorillonite and mixed layer clays such as illite/montmorillonite and montmorillonite/chlorite. The percentages of individual clay minerals were determined, using a system of simultaneous linear equations from elemental X-ray fluorescence analysis of the clay mineral fraction of core samples. The accuracy of quantification was further checked based on data obtained from some other instrumental techniques such as FTIR and thermal analysis. The weight percent clay of each sample was then defined as the sum of all different types of clay minerals respective mean concentrations. Detailed quantitative XRD analysis of all core samples yielded clay mineral contents varied from 5 to 32.5 wt% with an average value of 9.2 wt% in the gas producing well to 13.3 wt% in the non-producing well. Since all well logs essentially respond to volume percent (SLB, 2009), weight percent clay from XRD analysis should be at first converted into volume percent clay (V_{clay}), using the following formula (Ellis and Singer, 2008; Hawkins and Engineer, 1995):

$$V_{clay} = \frac{WT_{clay} \times \rho_{sample} \times (1 - \emptyset_t)}{\rho_{clay}} \tag{2}$$

where, V_{clay} is volume percent clay, WT_{clay} represents the clay weight fraction, ρ_{sample} stands for the sample density, ϕ_t denotes the total porosity, and ρ_{clay} is referred to the clay density (Ellis and Singer, 2008; Hawkins and Engineer, 1995).

The grain densities of all core samples were measured by helium pycnometery test with some variability due to the mineralogical complexity of the Shurijeh Formation. The average density value of the gas producing and the non-producing well were 2.71 g/cm^3 and 2.73

g/cm^3 respectively, both with a standard deviation of only 0.04 g/cm^3, less than 1% of the average. The major clay mineral in the Shurijeh Formation is illite and ρ_{clay} was considered to be 2.71 g/cm^3. The core total porosity data were available from previous reports (NIOC, 1986). Having known all the needed parameters, volume percentage of clay minerals for each core sample was calculated using the Eq. (2). Resulting volume percent clay varied from 4.1% to 32.2% with an average value of 7.3% in the gas producing well to 12.6% in the non-producing well.

4.2. Conventional petrophysical methods for estimating the volume of clay

The clay presence affects the behavior of almost all well logs, hence, volume of clay can be obtained from the response of each log with certain advantages and inherent limitations (Asquith et al. 2004; Dewan, 1983; Ellis and Singer, 2008; Serra, 2008). Eqs. 3 to 7 show the conventional petrophysical relationships for estimating the volume of clay from the natural gamma ray log and its spectral component (potassium, thorium, and uranium), density, sonic, neutron, and resistivity well logging data, respectively (Dresser Atlas, 1982; Fertl and Chilingarian, 1990; Quirein et al. 1982; Rukhovest and Fertl, 1981). The definition of symbols and constant values used in the conventional relationships is given in Table 3.

$$I_A = \frac{A_{log} - A_{min}}{A_{max} - A_{min}} \tag{3}$$

$$V_{\mathrm{cl}} = I_A \times \left(\frac{\rho_b}{\rho_{b_{cl}}}\right)^3 \tag{4}$$

$$V_{\mathrm{cl}} = \frac{\varphi_{\mathrm{DT}}}{\varphi_{\mathrm{DT}_{\mathrm{cl}}}} \tag{5}$$

$$V_{\mathrm{cl}} = \frac{\varphi_{\mathrm{N}}}{\varphi_{\mathrm{N}_{\mathrm{cl}}}} \tag{6}$$

$$V_{cl} = \left(\frac{R_{cl}}{R_t} \times \frac{\mathrm{R_{lim}} - \mathrm{R_t}}{\mathrm{R_{lim}} - \mathrm{R_{cl}}}\right)^{\frac{1}{1.5}} \tag{7}$$

Among the conventional methods, modified natural gamma ray log proved to be the most accurate clay indicator in the both cased and open boreholes (Fertl et al. 1982; Hilchie, 1982; Serra, 2008). The other types of well logging data may give either over/under estimations of volume of clay (Causey, 1991; Poupon and Gaymard, 1970). In the simplest form, the response of natural gamma ray log and its spectral components can be expressed as the linear function expressed in the Eq. (3) (Poupon and Gaymard, 1970), however, some variations in bore-hole condition can perturb this linear equation. Therefore, it should be modified, using one of the empirically derived calibration equations. Based on the extensive research of Jozanikohan et al. (2014b), the equation introduced by Steiber (1973), Eq. (8), provides the best modification for the linear ray index (I_A) in the Shurijeh Formation.

$$Stieber\ modification:\quad V_{cl} = \frac{I_A}{3-2I_A} \tag{8}$$

The Shurijeh volume of clay was computed from each of the conventional Eqs. 3 to 7, using up to six types of well logging data and the obtained results from Eq. (3) were modified by Eq. (8). The resulting estimates were then compared with the actual volume of clay measured in the laboratory to calculate the mean squared error (MSE) (Eq. 9) and the correlation coefficient, R (Eq. 1). Table 4 sorts several conventional estimation methods of volume of clay in the Shurijeh Formation in order of efficiency.

$$MSE = \frac{\sum (x_{i_{meas}} - x_{i_{est}})^2}{N} \tag{9}$$

where, $x_{i_{meas}}$ and $x_{i_{est}}$ represent the measured and estimated outputs, and N is the number of data pairs (Demuth and Beale, 2002).

4.2.1. Estimation of volume of clay, using multiple linear regression

The regression of a scalar dependent variable, Y, on one or more independent variables denoted X, is the computation of the most probable value of Y for each value of X, based on a finite number of possibly noisy measurements of X and the associated values of Y (Kenney and Keeping, 1954). To demonstrate the difference between neural network and other conventional approaches such as multiple linear regression in estimating the volume of clay, an exercise was performed on the Shurijeh data. The most important logs, affecting on the volume of clay (i.e. THOR, GR, DT, RHOB, MSFL, and NPHI) were used for conducting a multiple linear regression. To check the linearity of relationships between volume of clay and mentioned well logging data, several cross plots of volume of clay versus each well logging data were first generated (Fig. 3). The relationships between volume of clay (dependent variable) and the six types of well logging data (independent variables) were precisely analyzed, using standard multiple linear regression. Values of standard multiple regression parameters are expressed in Eq. (10) .To assess how well the regression model represents the data set, the MSE (Eq. 9) and the correlation coefficient, R (Eq. 1), were used and the results were added to Table 4.

$$V_{cl} = -0.29\,THOR + 0.17GR - 0.05DT - 7.20RHOB - 0.01MSFL - 51.55NPHI + 25.27 \qquad (10)$$

4.3. Estimation of volume of clay, using a multilayer perceptron neural network

Six wireline well logging data, including thorium, natural gamma ray, sonic, density, micro spherically focused resistivity, and neutron data, chosen from the results of sensitivity analysis, were used as network input. The data of core measured volume of clay as well as estimated values from the exclusive geostatistical based relationship for calculating the volume of clay in the Shurijeh Formation, introduced by Joanikohan et al. (2014a), were set as desired targets. The range of variation in the input and output data is shown in Table 5. In order to prevent an estimator to dominate the whole model, input/output data were first normalized, using Eq. (11) to fall in the range of zero and unity.

$$Normalized\ value = \frac{Original\ value - Minimum\ value}{Maximum\ value - Minimum\ value} \quad (11)$$

Out of 76 core samples, 16 samples were kept out to validate the final network. The remaining 60 core data and 1024 estimated data points were then divided randomly, but statistically consisted into three subsets; 70% (770 samples) of data for training, 15% (165 samples) for validation and the rest 15% (165 samples) for testing. Data from both wells were uniformly distributed among three sub data sets. Each of training, validation and testing data sets were employed to tune the synaptic weights of the network, avoiding the overfitting of the model, and the evaluation of the network capability for accurate estimation, respectively.

4.3.1 Selection of the optimal network architecture

To select the optimal transfer function for the network, several types of output transfer functions, including log-sigmoid (logsig), tan-sigmoid (tansig), triangular basis (tribas), saturating linear (satlin), linear (purelin), symmetric hard-limit (hardlims) and hard-limit (hardlim) were tested on a certain network with fixed topology (e.g. 6-5-1). As it can be seen from the data in Table 6, the results of log-sigmoid transfer function, showed better performance compared to the rest of the functions. Considering the input data were in the range of [0 1], another reasonable reason for choosing the log-sigmoid was to have a transfer function which provides outputs in the range of zero and unity. Hence, log-sigmoid one of the most commonly used activation function in use with MLP networks (Demuth and Beale, 2002), was used as activation function and the linear (purelin) was chosen for output layer transfer function in this study. The sigmoid functions are easy to calculate, which is helpful for calculating the weight updates in certain training algorithms (Demuth and Beale, 2002). The log-sigmoid function was defined in Eq. (12).

$$f(x) = \frac{1}{1 + e^{-\beta}} \quad (12)$$

where, β is the slope parameter (Demuth and Beale, 2002).

It is very important to choose the proper algorithm for training a neural network, so the suitable training algorithm of this study was selected through the comparison of results of three different training algorithms of genetic (GA), particle swarm optimization (PSO), and Levenberg-Marquardt (LMA) for a certain network with fixed topology (e.g. 6-6-1). The GA and PSOA are stochastically global optimization methods in which the PSOA lies somewhere between genetic algorithms and evolutionary programming (Fletcher, 2013). Both algorithms of genetic (GA) and particle swarm optimization (PSOA), developed by imitating the process of biological evolution and have been used to train the neural networks in cases that trapping in the local minimums make it impossible to solve the problem by conventional optimization methods (Van Rooij et al., 1996 and Vonk et al., 1997). However, it is reported that the GA and PSOA may not always produce optimal or near optimal solutions (Fletcher, 2013). The genetic algorithm was first introduced in oil and gas industry by Bush and Carter (1996) and Velez-Langs (2005) for estimation of reservoir characteristics; but they made no comparison between their results and other methods. To evaluate the performance of the GA, PSOA, and LMA algorithms, the MATLAB R2014a software was used for implementation each of the mentioned algorithms in this study. Figure 4 shows the differences in network performance depending on the application of three different algorithms on a certain network with fixed topology. As it can be seen from Figure 4, applying the LMA algorithm improved the estimation accuracy, in terms of a significant decrease in MSE error. Hence, the backpropagation learning algorithm, the most widely used learning algorithms for the training of the neural networks (Haykin and Network, 2004; Lim, 2003), was used with the Levenberg-Marquardt training algorithm (LMA) for its ability to train a multilayered neural network so that it can learn any arbitrary mapping of input to output with rendering the best performance over other BP algorithms (Press et al. 1992). Moreover, the particular application of LMA has been proved in generating computational models of oil and gas reservoirs given the observed data (Gharib Shirangi, 2014).

There is no generally accepted standard for selection of number of hidden layers or neurons for MLP networks. In theory, it is proved that one single hidden layer network with enough neurons can provide the precision needed for estimation as a universal approximator (Bose and Liang, 1996), however, to improve the network estimation capability to unseen data, it is necessary to use as few hidden nodes as possible (Bose and Liang, 1996). Therefore, multiple networks with one and two hidden layer(s) and varying numbers of neurons in hidden layer(s) were iteratively tested to achieve the optimum performance. The number of neurons varied between 1 and 10 in networks with one hidden layer, while for networks with two hidden layers, neuron numbers were changed from 3 to 10 and 1 to 8 in the first and second hidden layers, respectively. For the performance assessment of tested MLP networks, two statistical parameters, mean squared error (MSE) (Eq. 9) and the correlation coefficient, R (Eq. 1), were employed. The results of different network architectures, having either one or two hidden layer(s) were compared in Table 7. Having compared the estimation results of volume of clay, using ANN networks (Table 7) with the conventional methods (Table 4), the superiority of the neural network approach was demonstrated. The ANN method produced more accurate results with high R-value and low MSE (Table 7) than the conventional estimation methods which are currently being used by petrophysicists for estimation of volume of clay (Table 4). Having achieved the least MSE and the highest R-value, topology of 6-8-1 had the best performance among all other tested networks (Table 7). Therefore, the final structure of the multilayer feedforward backpropagation network used in this study consisted of an input layer with six neurons, one hidden layer with 8 neurons, and the output layer with one neuron (Fig. 5). Six types of well logging data, shown in Fig. 5, were arrayed one succeeding the other to form the input layer. The hidden and output layer neurons were each connected to all of the units in the preceding layer as shown in Fig. 5. The calculation in the designed network started at the input layer, moves forward to the next layer to determine the outputs of each neuron. Many sets of input/output pairs with small random assigned values were used for initial weights. The sum of weighted inputs and a bias passed from the previous layer and then using a nonlinear activation function, log-sigmoid, an output was generated. The calculated network responses were compared with the actual laboratory measured core values and an average of all the mean squared errors were computed. In order to minimize the errors in the next stages, the weights and bias values of the network were updated at each stage. The procedure was repeated until a minimum overall error was obtained. The final obtained weights of the optimum ANN model are given in Table 8.

For the optimum MLP model gained for estimating the volume of clay in the Shurijeh Formation, the best validation performance occurred at epoch 62 with MSE and R values of $2.8069E^{-3}$ (Fig. 6), and 0.9013 (Fig. 7), respectively. Based on regression analysis (Fig. 7), the chosen architecture (6-8-1) is well capable to estimate volume of clay in the Shurijeh Formation. The progress of other training variables, such as the gradient magnitude, the number of validation checks, etc. is shown in Fig. 8. Distribution of network errors between the laboratory data and estimated values from the optimum MLP network is demonstrated in Fig. 9. The excellent performance of the MLP model was easily evaluated by the high value of correlation coefficient (R=0.9013) and the great match of estimated and measured values for volume of clay (Fig. 10). The optimum network was successfully validated by presenting 16 unseen core data from each well to it and achieving MSE of $4.5793E^{-3}$ and R-value of 0.8999. Finally, the optimum MLP network was employed to estimate the volume of clay in the whole interval of the Shurijeh Formation in both gas producing and non-producing wells (Fig. 11).

5. Results

There is no surprise that the volume of clay in the Shurijeh Formation is highly influenced by the natural gamma ray, and thorium logs, mostly due to the fact that clay minerals are naturally radioactive or have radioactive ions associated with them. Compared to the natural gamma ray and thorium logs, the potassium log did not show the same high influence on volume of clay due to the proved existence of K-deficient clays such as kaolinite and chlorite in the Shurijeh Formation by XRD analysis. Since the uranium has little association with clay minerals in general, and it is more indicative of natural fractures or organic matter in the formation, it did not have a high impact on the volume of clay. Among the resistivity logs, micro spherically focused resistivity (MSFL) was related more to the volume of clay, while both the laterolog deep (LLD) and shallow resistivity (LLS) logs did not show a good correlation with volume of clay. The reason for this inapplicability might lie in the vertical resolution of each resistivity log. The MSFL log has a very high vertical resolution, which makes it to be an excellent candidate for delineating thin shaly beds of the Shurijeh Formation. According to the laboratory results, the average amount of volume of clay in the non-producing well was 1.73 orders of magnitude larger than that was in the gas producing well.

The conventional petrophysical methods did not release good estimations for volume of clay in the Shurijeh Formation (Table 4). Having considered the results of Table 4 were derived from the linear relationships between volume of clay and well logging data (Eqs. 3 to 7), while the actual relationship between volume of clay and well logging data were not linear (Fig. 3), the low R-value in Table 4 could be justified. It seems that getting the accurate results from the conventional methods, due to the complexity and the high nonlinearity in the actual relationships between the volume of clay and well logging data (Fig. 3), indeed, is somewhat impossible. However, among all the conventional petrophysical methods, modified thorium log produced reasonable estimates which were not still satisfactory.

The multiple linear regression analysis showed all the independent variables (six types of well logging data), with low or high significance level, had an impact on the final effect. Although the performance of multiple linear regression was better than the conventional methods used in petrophysics, but the lower correlation coefficients (R = 0.59) could still justified the need for further improvements by other methods based on artificial intelligence techniques, e.g. ANN.

The normalization of input log responses and output core measured and estimated volume of clay is an important issue which should be considered in the experimental design, to avoid one estimator dominating the model. It is also necessary to set apart some core data points to be able to validate the applicability of the final neural network model.

In comparison to the other training algorithms (i.e. genetic and particle swarm optimization algorithms), LMA algorithms produced better results with less MSE. The reason might lies in the fact that repeated application of Levenberg–Marquardt algorithm (LMA), from random starting positions makes it possible to find the global minimums as well as local ones (Fletcher, 2013). As it can be seen from Figure 4, GA could not accurately estimate the volume of clay, probably because it is not as precise as the LMA which is able to find a solution even if it starts very far off the final minimum. The MLP network was suitably trained with LM algorithm, using 60 data sets and 1024 estimated data points, validated, and tested with the remaining 16 data sets.

It is clear from Table 7 that the networks with one hidden layer had better performance than the two hidden layers networks with the same number of neurons in the first hidden layer. In order to determine the structure of optimum MLP network, the number of hidden neurons varied from 1 to 10 and the both MSE and R-value were measured for each network. A 6-8-1 topology yielded the most satisfactory performance with the least possible error (MSE = $2.8069E^{-3}$ and R = 0.9013), compared to the rest of developed networks listed in Table 7. The good results of training and the exceptional ability of used training algorithm, i.e., LMA, in properly modeling the estimation were demonstrated in Figure 7. The validation and test curves were almost similar, indicating there was not an overfitting problem. The output for training, testing, and validation tracked the targets very well, reflecting the good fit and accuracy of the neural network approach (Fig. 7). The validation and test results also show R values that greater than 0.9, however, certain data points had poor fits. The robustness of the general methodology for volume of clay estimation in this study was further proved by presenting 16 previously unseen input data to the optimum MLP model. Achieving a low MSE ($4.5793E^{-3}$) and high R-value (0.8999) for unseen data, the proposed MLP network could accurately follow the laboratory data. The superimposed plots of estimated values from the MLP network with the laboratory measured data presented in Figs. 10 and 11, showed the ability of the network to adeptly capture the variability in the core data. As it can be seen from Figures 10 and 11, the simulated results of the MLP model were being satisfactory verified and validated by the laboratory measured core data in full range of values. It is understood from the validation results of unseen data that the MLP network with overall three layers of input/hidden/output and 8 neurons in the hidden layer, trained with LMA and BP learning algorithms, provided exceptionally consistent estimations. Hence, the proposed computational MLP network could be successfully utilized for an accurate estimation of volume of clay across the other drilled wells in the studied area.

Comparison of conventional and modern estimation techniques (Tables 4 and 7), based on artificial intelligence reflects the ability of the MLP network in providing exceptional matches with the core data. In fact, the adaptive learning of neural networks made it possible to detect the complex trends that are too difficult to be noticed either by human or other computer based estimation techniques in this study. Therefore, the most important achievement of the proposed methodology is an improvement to the estimation of volume of clay in the Shurijeh Formation. This improvement was achieved through a reasonable increase in the R-value and a significant decrease in MSE.

6. Conclusions

In order to overcome the complexity and uncertainty associated with the conventional estimation of volume of clay in the reservoirs, an artificial intelligent method based on a multilayer perceptron neural network, was proposed to estimate the volume percent of clay minerals in the Shurijeh gas reservoir Formation. Having achieved large estimation errors through conventional methods, a feedforward backpropagation MLP network was used for the estimation of volume of clay. To reduce the uncertainty of the model, a sensitivity analysis was conducted to select the most relevant well logs to the volume of clay, using cross correlations. The chosen well logging data were then used as input to the network and standardized so that different input ranges could not dominate and bias the output results. Data sets were trained with different training algorithms and LMA, the selected algorithm, was used to train data in different MLP architectures. Through a trial and error process and on the basis of the statistical parameters such as MSE and R, the MLP network with one hidden layer and the topology of 6–8–1 was chosen as the optimum network. For the mentioned MLP network, MSE and the correlation coefficient, R, were equal to $2.8069E^{-3}$ and 0.9013, respectively. The simulation results of the proposed MLP network were closely matched to the laboratory measured values and the successful results of network validation to unseen core data showed the exceptional ability of the proposed MLP network to obtain the most accurate functional complex relationships between the volume of clay and wireline well logging data in the Shurijeh Formation. It is worthy to mention that the high value of the correlation coefficient between the values of actual and MLP-estimated volume of clay in this study demonstrates the capability of the artificial intelligence methods for efficiently solving the reservoir characterization problems as a robust predictive model with a satisfactory degree of accuracy. Furthermore, it was demonstrated that using a neural network to estimate the volume of clay is far superior to using conventional methodologies such as multiple-regression or other petrophysical techniques. The accurate estimates of volume of clay derived from the proposed MLP network have had a great impact on the estimation of hydrocarbon in place and reserves of the Shurijeh in an eastern Kopet-Dagh gas field. Estimation of volume of clay, using the MLP approach developed in this study, can be useful for the petrophysicist to gain better and more accurate estimates of this critical parameter.

Altogether, the lowest achieved MSE for the proposed artificial intelligence technique showed a clear improvement in the volume of clay estimation within data sets of the Shurijeh Formation. In compared to the conventional methods, the MSE and R-value of estimation were significantly improved by 53.55% (from $5.2416E^{-3}$ in the estimation, using multiple linear regression to $2.8069E^{-3}$ with the proposed MLP network), and 52.32% (from 0.5917 in the estimation obtained from multiple linear regression to 0.9013 with the proposed ANN), respectively.

Acknowledgements

The authors are thankful to the school of Mining, college of Engineering, University of Tehran and Exploration Directorate of National Iranian Oil Company (NIOC) for their assistance in providing data and technical and financial support in the completion of this research work. The authors would also like to thank the editor and three anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper. Finally, the first author is very grateful to the Lowe family, especially Mr. Thomas Henry Lowe and Mr. David Alexander Lowe for all the help they have done with English editing the manuscript.

Nomenclature

V_{cl}- Volume of clay; %

ANN- Artificial neural network

MLP- Multilayer perceptron

BP- Backpropagation

GR- Natural gamma ray; API

POTA- Potassium log; wt%

THOR- Thorium log; ppm

URAN- Uranium; ppm

RHOB- Density; g/cm^3

DT- Sonic; US/F

NPHI- Neutron; p.u

MSFL- Micro spherically focused resistivity; Ohmm

LLD-Laterolog deep resistivity; Ohmm

LLS-Laterolog shallow resistivity; Ohmm

PEF- Photoelectric; B/N

XRD- X-ray diffraction technique

XRF- X-ray fluorescence technique

FTIR- Fourier transform infra red spectroscopy technique

WT_{clay} - Clay weight fraction; wt%

ρ_{sample} -Density of sample; g/cm^3

ϕ_t -Total porosity; V/V

ρ_{clay} - Density of clay; g/cm^3

I_A- Ray Index

A_{log}- Total ray reading in the zone of interest; API or wt% or ppm

A_{min}- Average ray response in clean (clay free) zone; API or wt% or ppm

A_{max}- Average ray response in dirty (clay rich) zone; API or wt% or ppm

ρ_b - Bulk density in the zone of interest; g/cm^3

ρ_{bcl} - Density of clay; g/cm^3

ϕ_{DT} - Sonic porosity in the zone of interest; V/V

ϕ_{DTcl} - Sonic porosity in dirty (clay rich) zone; V/V

ϕ_N - Neutron porosity in the zone of interest; V/V

ϕ_{Ncl} - Neutron porosity in dirty (clay rich) zone; V/V

R_{cl}- Resisitivity in dirty (clay rich) zone; Ohmm

R_t- Resisitivity in the zone of interest; Ohmm

R_{lim}- Resisitivity in clean (clay free) zone; Ohmm

MSE- Mean squared error

R- Correlation coefficient

$x_{i_{meas}}$ - Measured output

$x_{i_{est}}$ - Estimated output

$\bar{x}_{i_{meas}}$- Mean of measured output

$\bar{x}_{i_{est}}$- Mean of estimated output

N - Number of data pairs.

e_x - Weighted sum of the inputs for a processing unit

GA- Genetic algorithm

PSOA- Particle swarm optimization

LMA- Levenberg-Marquardt training algorithm

References

Adeoti, L., Ayolabi, E.A., James, P.L., 2009. An Integrated Approach to Volume of Shale Analysis: Niger Delta Example, Orire Field. World. App. Sci. J. 7, 448-452.

Afsharharb, A., 1979. Stratigraphy, tectonic and petroleum geology of Kopet-Dagh region, northeast Iran. PhD Dissertation, College of Science and Technology, University of London, England.

Aminzadeh, F., Barhen, J., Glover, C.W., Toomarian, N.B., 1999. Estimation of reservoir parameter using a hybrid neural network. J. Petrol. Sci. Eng. 24, 49-56.

Asquith, G.B., Krygowski, D., Gibson, C.R., 2004. Basic well log analysis. 2nd edition, American Association of Petroleum Geologists Publication, Tulsa, Oklahoma, USA.

Auer, P., Burgsteiner, H., Maass, W., 2008. A learning rule for very simple universal approximators consisting of a single layer of perceptrons. Neural. Networks. 21, 786–795.

Benaouda, D., Wadge, G., Whitmarsh, R.B., Rothwell, R.G., MacLeod, C., 1999. Inferring the lithology of borehole rocks by applying neural network classifiers to downhole logs: an example from the Ocean Drilling Program. Geophys. J. Int. 136, 477-491.

Bose, N.K., Liang, P., 1996. Neural Network Fundamentals with Graphs, Algorithms, and Applications. McGRaw-Hill Series in Electrical and Computer Engineering, McGraw-Hill, USA.

Bueno, E.O.J., Perez, I.C., Escamilla, G., 2006. Applications of artificial neural networks and dipole sonic anisotropy in low-porosity, naturally fractured, complex lithology formations in the Southern Land Region of Mexico. In First International Oil Conference and Exhibition in Mexico held in Cancun, Mexico.

Bush, M. D., Carter, J. N., 1996. Application of a modified genetic algorithm to parameter estimation in the petroleum industry. Intelligent Engineering Systems through Artificial Neural Networks conference, 6, 397-402.

Causey, G.L., 1991. Computer determination and comparison of volume of clay derived from petrophysical and laboratory analysis. Master Dissertation, Texas Tech University, Lubbock, Texas, USA.

Chen, M.F., Lang, Z.X., 2003. Application of a modified gray model in oilfield production forecast. Xinjiang. Petrol. Geol. 24, 246-248.

Demuth, H., Beale, M., 2002. Neural network toolbox for use with MATLAB, User's guide, Version 4. The Math Work, Inc, Natick.

Dewan, J.T., 1983. Essentials of modern open hole log interpretation. PennWell Books Publication, Tulsa, Oklahoma, USA.

Dresser Atlas, 1982. Well logging and interpretation techniques. Dresser Atlas Industries, Houston, TX., USA.

Eftekharnejad, J., Behroozi, A., 1991. Geodynamic significance of recent discoveries of ophiolites and late Paleozoic rocks in NE Iran (including Kope-dagh). Abhandlungene der geologischen Bundesanstalt. 38, 89–100.

Ellis, D.V., Singer, J.M., 2008. Well logging for earth scientists. 2nd edition. Springer-Verlag, Heidelberg, Germany.

Farshad, F.F., Garber, J.D., Lorde, J.N., 2000. Predicting temperature profiles in producing oil wells using artificial neural networks. Eng. Computation. 17, 735-754.

Fausett, L.V., 1993. Fundamentals of Neural Networks: Architectures, Algorithms and Applications. 1st edition, Pearson publication, India.

Fletcher, R. 2013. Practical methods of optimization. John Wiley & Sons, New York, USA.

Fertl, W.H., Chilingarian, G.V., Yen, T.F., 1982. Use of natural gamma ray spectral logging in evaluation of clay minerals. Energ. Source. Part A. 6, 335-360.

Fertl, W.H., Chilingarian, G.V., 1990. Type and distribution modes of clay minerals from well logging data. J. Petrol. Sci. Eng. 3, 321-332.

Funahashi, K., 1989, On the approximate realization of continuous mappings by neural networks. J. Neural. Networks. 2, 183–192.

Fung, C.C., Wong, K.W., Eren, H., 1997. Modular artificial neural network for prediction of petrophysical properties from well log data. IEEE. T. Instrum. Meas. 46, 1295-1299.

Gharib Shirangi, M. 2014. History matching production data and uncertainty assessment with an efficient TSVD parameterization algorithm. J Petrol Sci Eng, 113, 54-71.

Guyon, I., and Elisseeff, A. 2003. An Introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157-1182.

Hawkins, J., Engineer, S.R., 1995. Integration of Core Data for Calibrated Shaly-Sand Log Analysis. SCA Conference, Paper No. 9201, Bilar Bohol (Diocese of Tagbilaran), Philippines.

Haykin, S., Network, N., 2004. Neural Networks: A comprehensive foundation. Prentice Hall International, USA.

Hilchie, D.W., 1982. Advanced electric log analysis. Douglas W. Hilchie Inc., Golden Colorado, USA.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural. Networks. 2, 359–366.

Jozanikohan, G., Norouzi, Gh.H., Sahabi, F., Memarian, H., 2014a. A review of petrophysical relationships for estimation the volume of clay, Case study: Shurijeh reservoir Formation in an eastern Kopet-Dagh basin field. Accepted for publication in Iranian Journal of Mining Engineering.

Jozanikohan, G., Norouzi, Gh.H., Ojani, H., Sahabi, F., 2014b. Petrophysical estimation of volume of clay in D_1 and C_2 subdivisions of Shurijeh Formation, a case study from one of eastern Kopet-Dagh basin fields. The 32nd National and the 1st International Geosciences Congress: Petroleum Geology, Ahwaz, Iran.

Kalantari, A., 1969. Foraminifera from the middle Jurassic-Cretaceous successions of Koppet-Dagh region (NE Iran). National Iranian Oil Company, Geological Laboratories, Iran.

Kenney, J.F. and Keeping, E.S., 1951. Mathematics of Statistics. Pt. 2, 2nd ed. Princeton, NJ: Van Nostrand, USA.

Li, W., Wu, X., Shi, J., 2008. Application of a neural network system combined with genetic algorithm to rank coalbed methane reservoirs in the order of exploitation priority. Petrol. Sci. 5, 334-339.

Lim, J.S., 2003. Reservoir permeability determination using artificial neural network. J. Korean. Soc. Geosystem. Eng. 40, 232– 238.

Lim, J.S., 2005. Reservoir properties determination using fuzzy logic and neural networks from well data in offshore Korea. J Petrol. Sci. Eng. 49, 182-192.

Liu, Z., Liu, J., 1998. Seismic controlled nonlinear extrapolation of well parameters using neural networks. Geophysics. 63, 2035-2041.

Liu, Z.B., Ren, B.S., Zhao, M.A., 2008. Functional simulation for oilfield output forecast based on time-varying system. J. Southwest. Petrol. U. (Sci. Tech. E.). 30, 181-184.

McCulloch, W., Pitts, W., 1943. A Logical Calculus of Ideas Immanent in Nervous Activity. B. Math. Biophys. 5, 115–133.

Moore, D.M., Reynolds, Jr. R.C., 1989. X-ray diffraction and the identification and analysis of clay minerals, Oxford University Press (OUP), UK.

Moussavi-Harami, R., Brenner, R.L., 1993. Diagenesis of non-marine petroleum reservoirs: the Neocomian (lower Cretaceous) Shurijeh Formation, Kopet-Dagh Basin, NE Iran. J. Petrol. Geol. 16, 55-72.

Nikravesh, M., Aminzadeh, F., 2001. Past, present and future intelligent reservoir characterization trends (editors' view points). J. Petrol. Sci. Eng. 31, 67–79.

Nikravesh, M., Aminzadeh, F., Zadeh, L.A., 2003. Soft computing and intelligent data analysis in oil exploration. Developments in petroleum sciences 51, Elsevier.

NIOC (Exploration Directorate), 1986. Gonbadli geological well completion report. National Iranian Oil Company records, Tehran, Iran.

Ouenes, A., 2000. Practical application of fuzzy logic and neural networks to fractured reservoir characterization. Comput. Geosci. 26, 953-962.

Poupon, A., Gaymard, R., 1970. The evaluation of Clay Content from Logs. SPWLA 11th annual logging symposium, Society of Petrophysicists and Well-Log Analysts, Paper G.

Press, W.H., Teukolsky, S.A., Vetterling, A.W.T., Flannery, B.P., 1992. Numerical Recipes in C: the art of Scientific Computing. Cambridge University Press, USA.

Quirein, J.A., Gardner, J.S., Watson, J.T., 1982. Combined natural gamma-ray spectral/litho-density measurements applied to complex lithology. 57th Annual Fall Technical Conference and Exhibit., SPE of AIME, New Orleans, USA, Paper SPE 11143.

Quirein, J.A., Hampson, D., Schuelke, J.S., 2000. Use of Multi-Attribute Transforms to Predict Log Properties from Seismic Data. In: Petrophysics meetings Geophysics-Exploring the Synergies between Surface and Borehole Geoscience.

Rogers, S.J., Fang, J.H., Karr, C.L., Stanley, D.A., 1992. Determination of lithology from well logs using a neural network. AAPG. Bull. 76, 731-739.

Rolon, L., Mohaghegh, S.D., Ameri, S., Gaskari, R., McDaniel, B., 2009. Using artificial neural networks to generate synthetic well logs. J. Nat. Gas. Sci. Eng. 1, 118-133.

Rosenblatt, F., 1958. The Perceptron: A Probalistic Model For Information Storage And Organization In The Brain. Psychol. Rev. 65, 386–408.

Rukhovest, N., Fertl, W.H., 1981. Digital shaly sand analysis based onWaxman-Smith model and log-derived clay typing. In: Trans. SAID/SPWLA 7th European Annual Logging Symposium: 21-37.

Serra, O., 2008. The well logging handbook. Editions Technip Publication, Paris, France.

Simpson, P.K., 1990. Artificial neural system: foundation, paradigms, applications and implementations. Pergamon Publication, NewYork, USA.

SLB (R&D), 2009. Schlumberger Chart Book. Schlumberger Educational Services (Research and Development Section), France.

Steiber, R.G., 1973. Optimization of shale volumes in open hole logs. J. Petrol. Tech. 31, 147– 162.

Stoklin, J., 1968. Structural history and tectonic of Iran. Am. Assoc. Petrol. Geol. Bull. 52, 1229–1258.

Van Rooij, A.J., Johnson, R.P., Jain, L.C., 1996. Neural network training using genetic algorithms. World Scientific Publishing Co., Inc., Singapore.

Velez-Langs, O., 2005. Genetic algorithms in oil industry: An overview. J Petrol Sci Eng. 47, 15-22.

Verma, A.K., Cheadle, B.A., Routray, A., Mohanty, W.K., and Mansinha, L., 2012. Porosity and permeability estimation using neural network approach from well log data. GeoConvention vision conference, Canada.

Vonk, E., Jain, L.C., Johnson, R.P., 1997. Automatic Generation of Neural Network Architecture Using Evolutionary Computation.World Scientific Publishing Co. Pvt. Ltd.,Singapore.

Walls, J., Taner, T., Taylor, G., Smith, M., Derzhi, N., Carr, M., Drummonds, J., McGuire, D., Morris, S., Bregar, J., 2000. Seismic reservoir characterization of a midcontinent fluvial system using rock physics, post stack seismic attributes and neural networks: A case history. 69th Annual International Meeting, Society of Exploration Geophysics, 1437-1439.

Wang, K., Zhang, L., 2008. Predicting formation lithology from log data by using a neural network. Petrol. Sci. 5, 242-246.

Werbos, P.J., 1994. The Roots of Backpropagation. From Ordered Derivatives to Neural Networks and Political Forecasting. John Wiley & Sons, New York, USA.

Wiener, J.M., Rogers, J.R., Moll, R.F., 1991. Predicting carbonate permeabilities from wireline logs using a back-propagation neural network. 61th Annual International Meeting, Society of Exploration Geophysics.

Wilby, R.L., Abrahart, R.J., Dawson, C.W., 2003. Detection of conceptual model rainfall-rainoff process inside an artificial neural network. Hydrolog. Sci. 48, 163-181.

Yang, Y., Rosenbaum, M.S., 2002. The artificial neural network as a tool for assessing geotechnical properties. Geotech. Geolo. Eng. 20, 149–168.

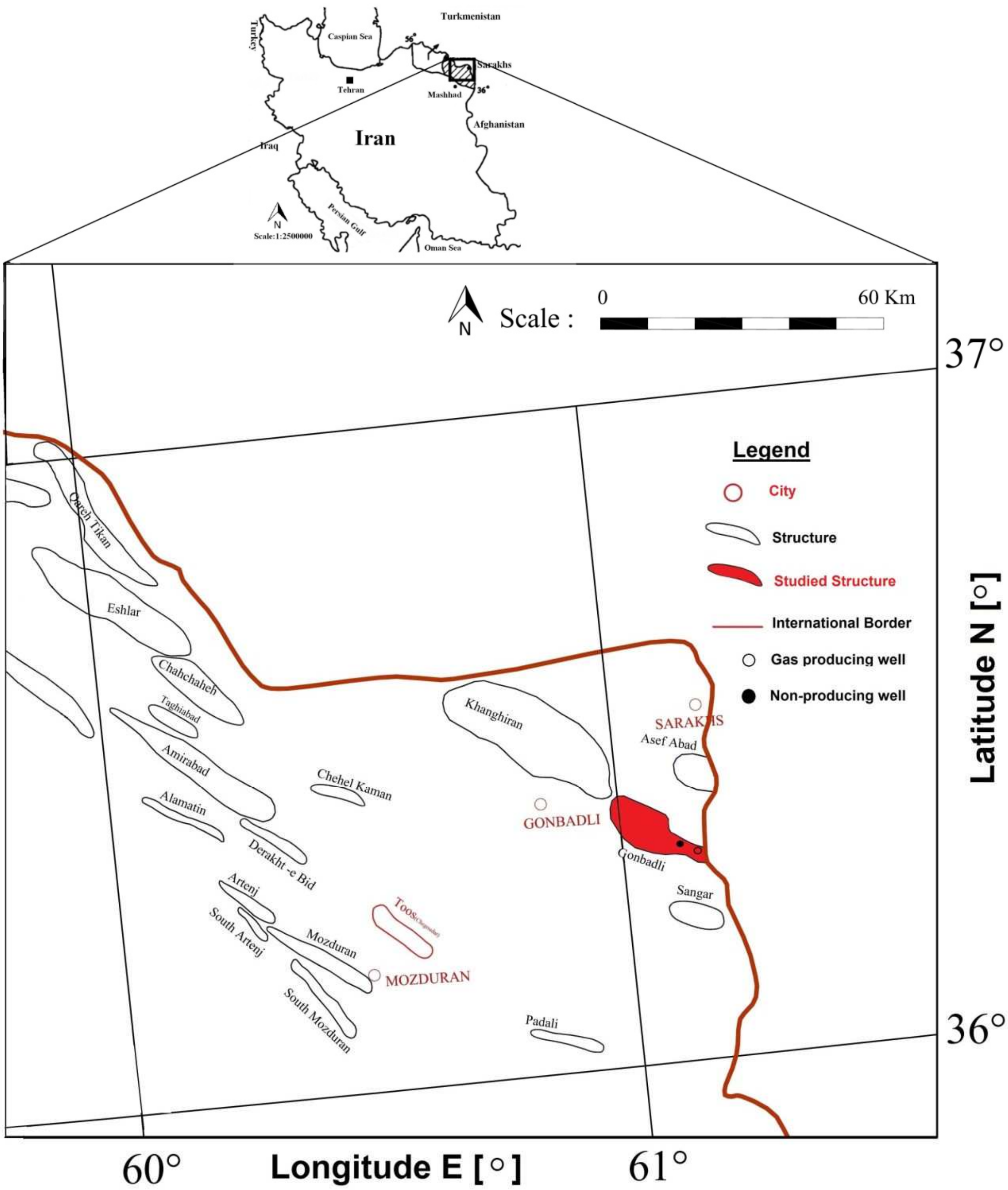

Fig. 1 Geographic location map of studied area.

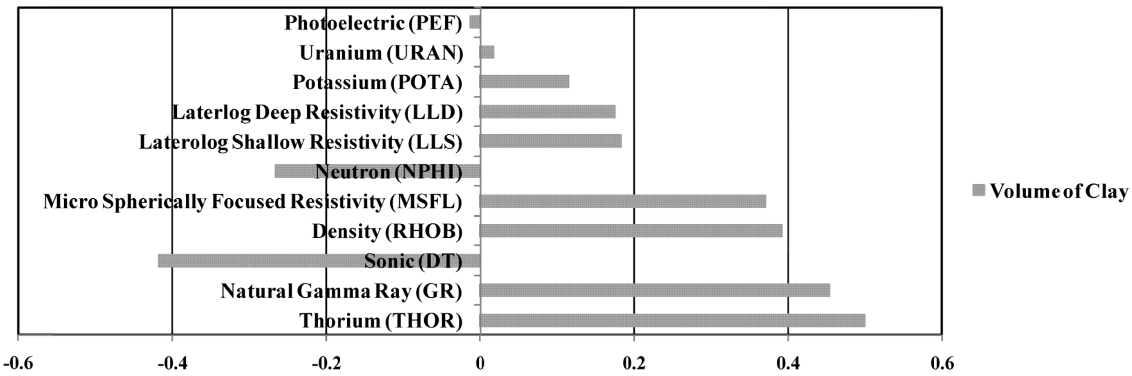

Fig. 2 Results of sensitivity analysis, using cross correlation.

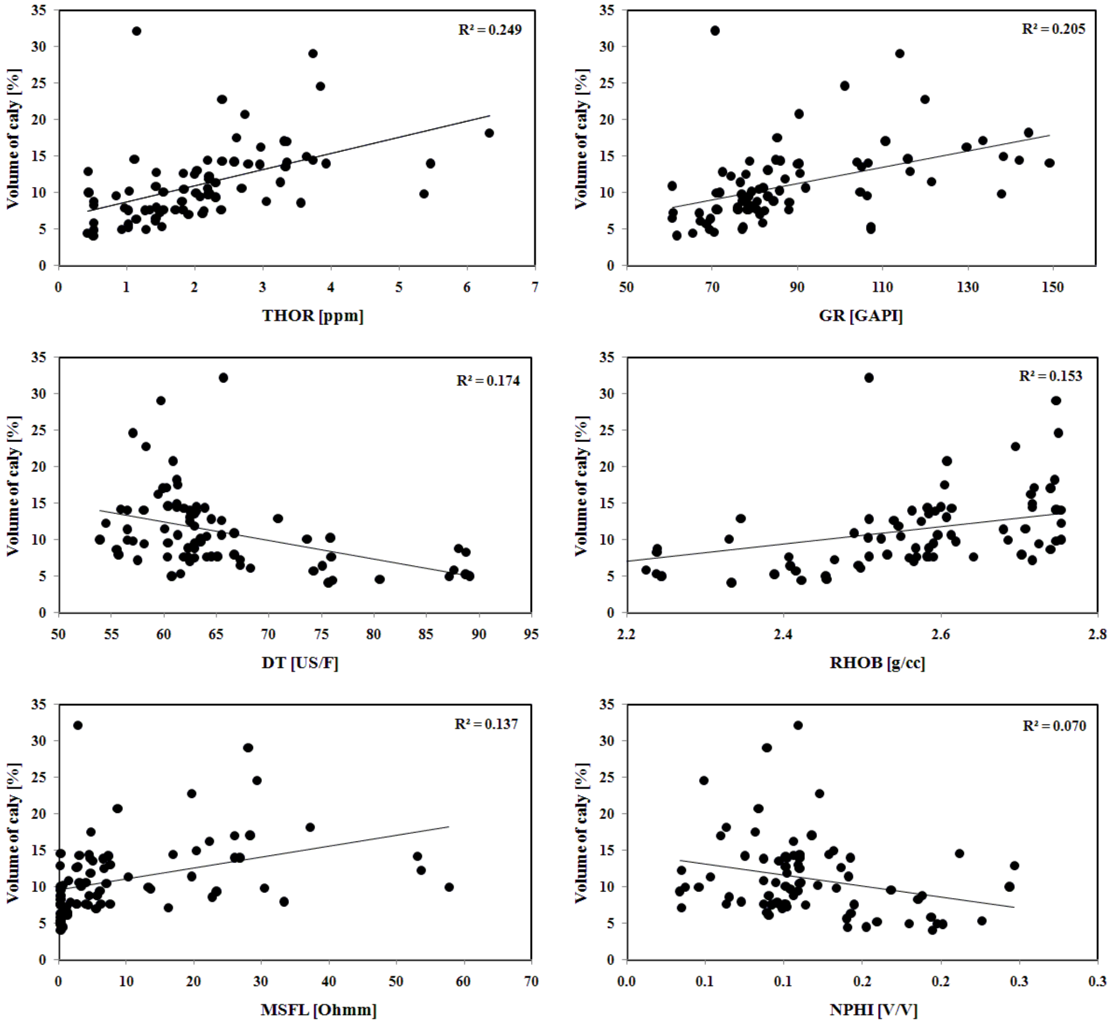

Fig. 3 The scatter plots between volume of clay and petrophysical logs in the Shurijeh Formation.

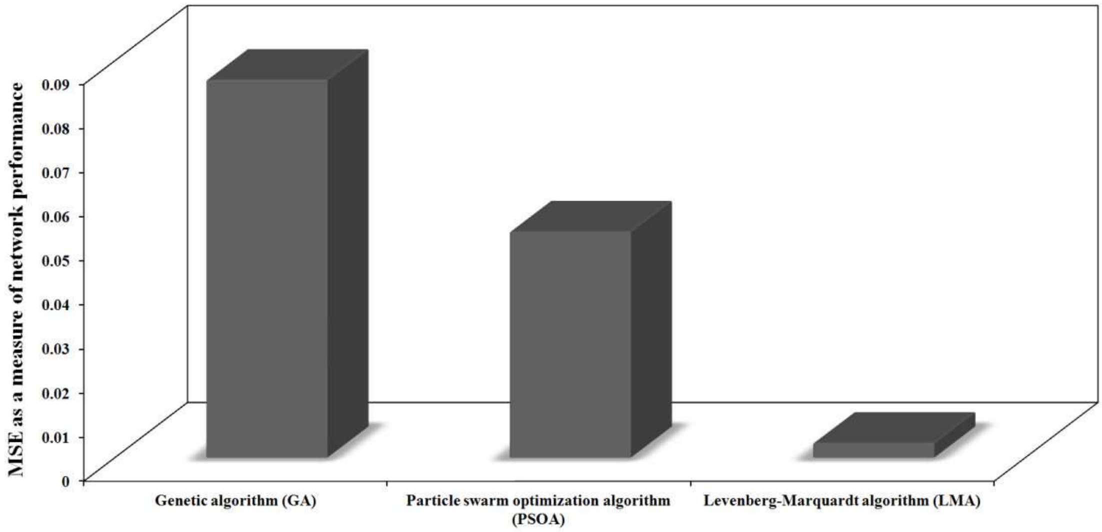

Fig. 4 Schematically comparing the performances of neural network with certain topology (e.g. 6-6-1) with different learning algorithms.

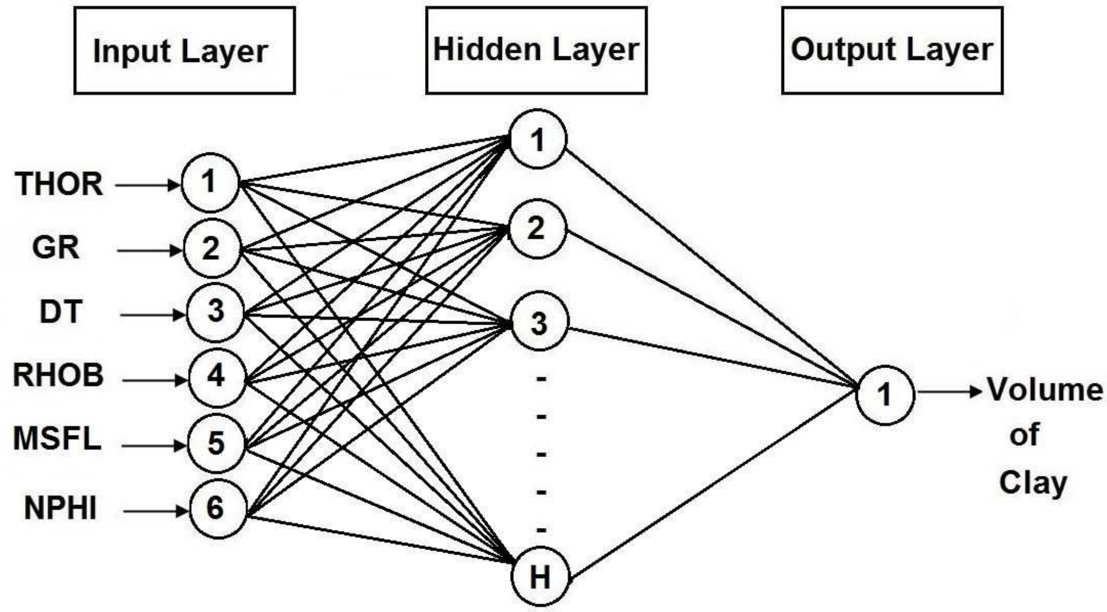

Fig. 5 The schematic architecture of the optimum MLP model for estimating the volume of clay (V_{cl}) in the Shurijeh Formation.

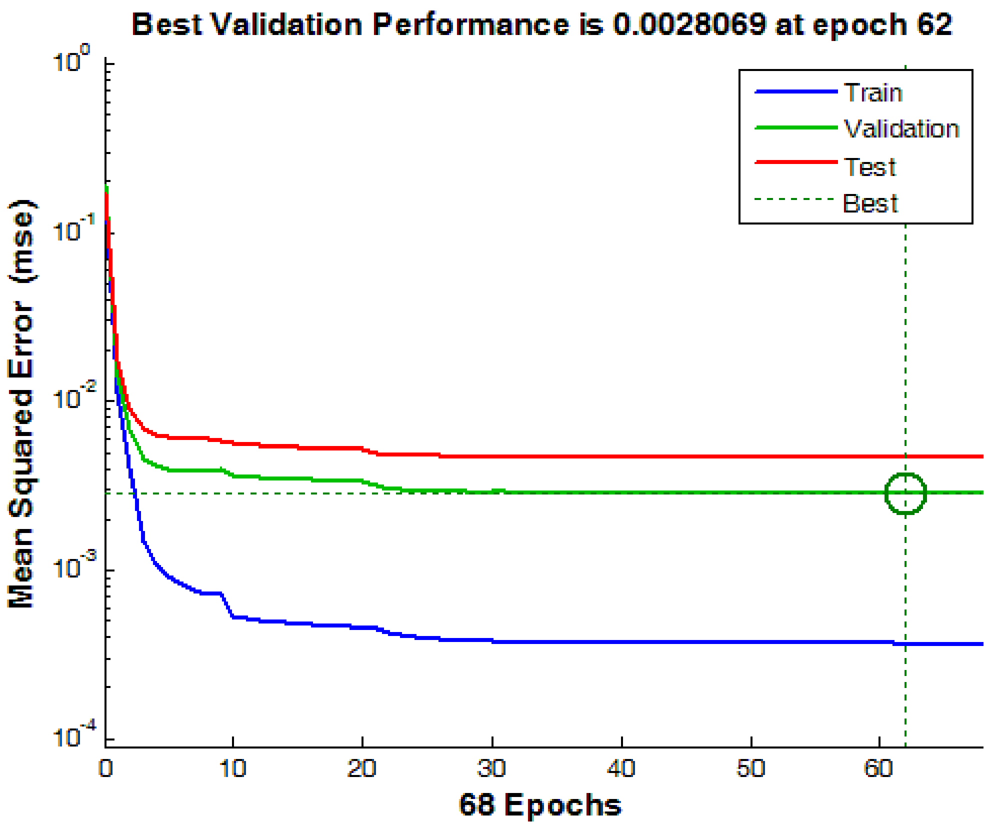

Fig. 6 The Performance of optimum MLP with 6-8-1 architecture.

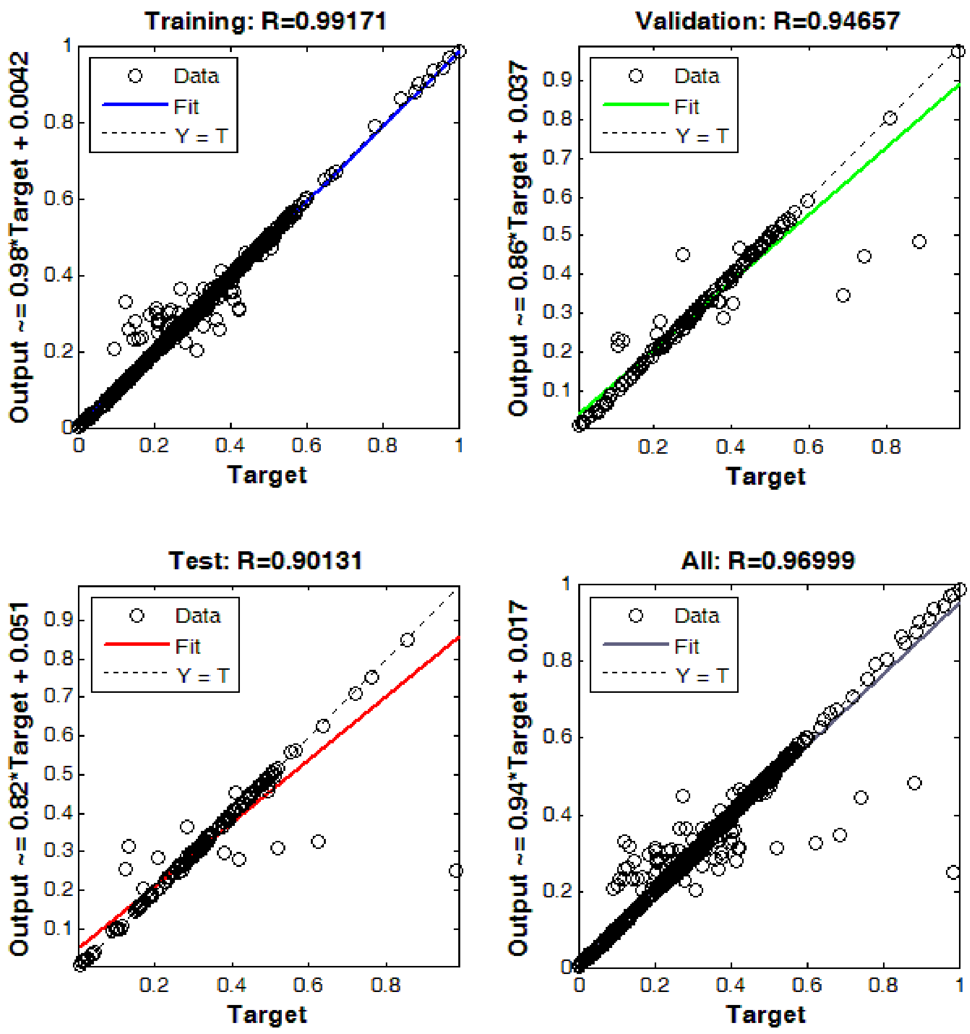

Fig. 7 R-values of training, validation and test data sets for optimum MLP network. The dashed line in each plot represents the perfect result – outputs = targets. The solid line represents the best fit linear regression line between outputs and targets.

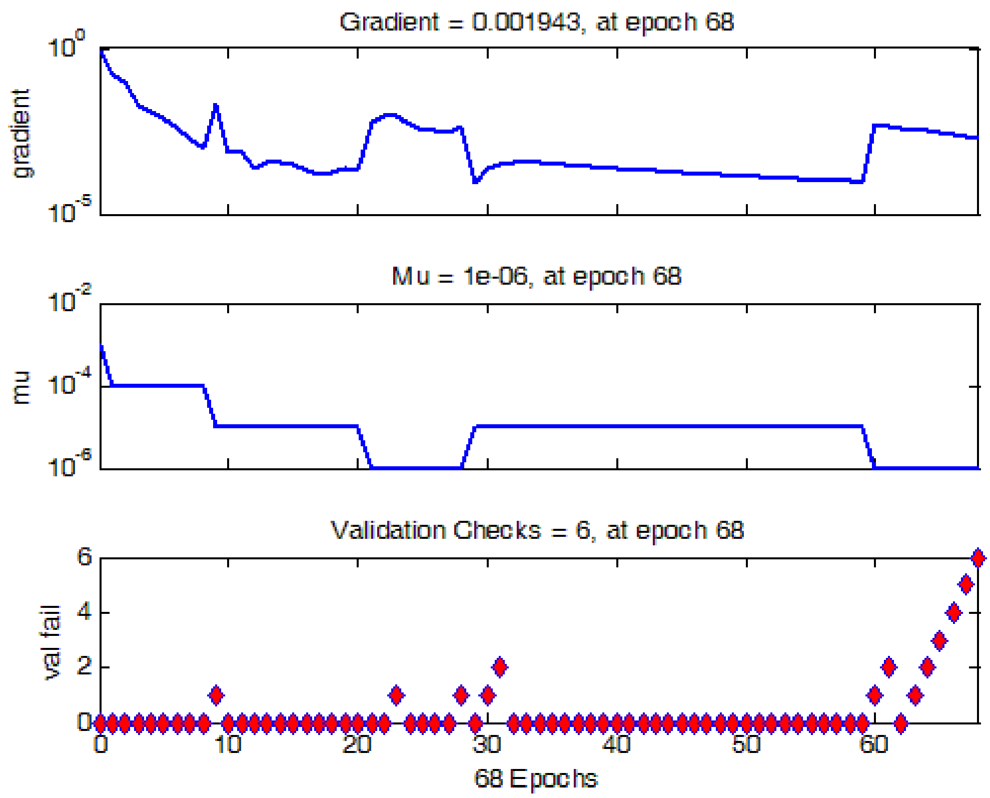

Fig. 8 The training state for the optimum MLP network, 6-8-1 architecture.

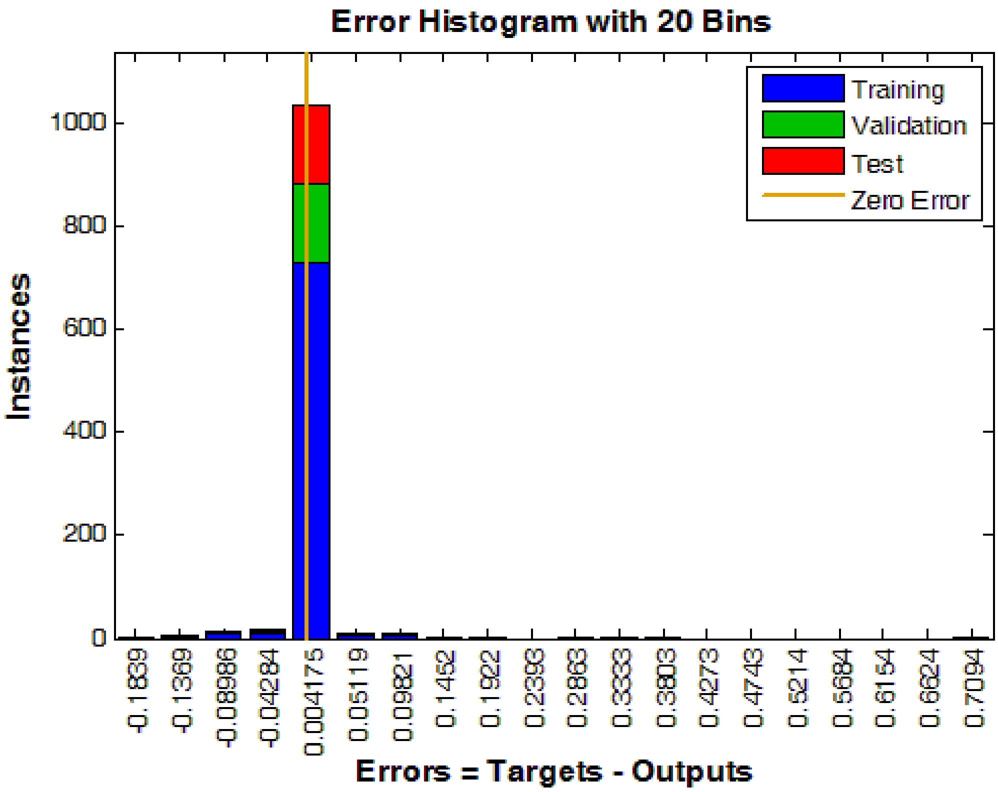

Fig. 9 The error histogram for the optimum MLP network.

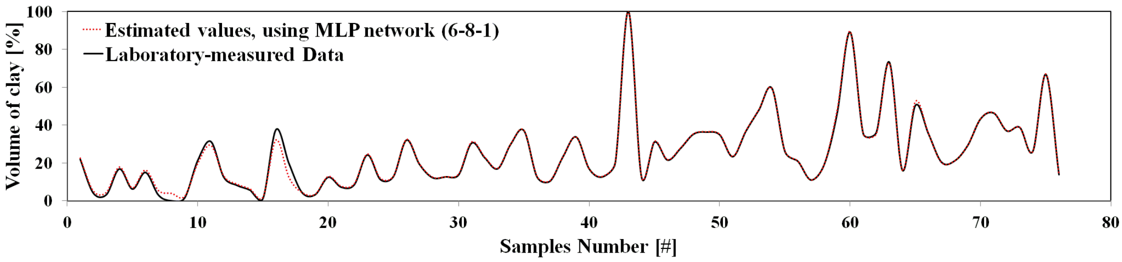

Fig. 10 The excellent ability of optimum network to reproduce the laboratory results.

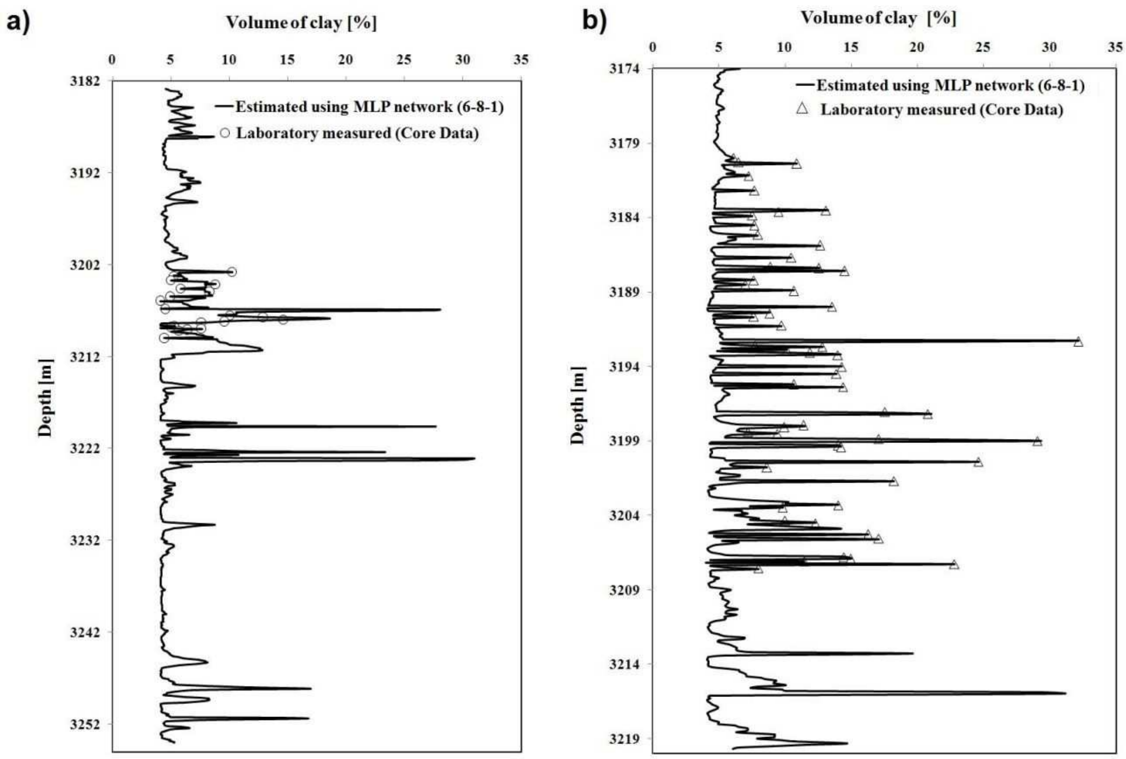

Fig. 11 Estimated the Shurijeh volume of clay by the optimum network for a) gas producing well b) non-producing well.

Table 1

Cretaceous lithostratigraphic sequence in the Kopet-Dagh sedimentary Basin (NIOC, 1986)

Era	Period		Formation	Lithology
Mesozoic	Cretaceous	Upper	Kalat	Limestone
			Neyzar	Sandstone
			Abtalkh	Silty and sandy marl
			Abderaz	Shale, Silty and chalky marl
			Aitamir	Silty shale and glauconitic sandstone
		Lower	Sanganeh	Silty calcareous shale and sandstone
			Sarcheshmeh	Silty shale to marl
			Tirgan	Limestone
			Shurijeh	Shaly sandstone

Table 2

The Pearson correlation matrix for well logging data from the Shurijeh reservoir Formation

	GR	POTA	THOR	URAN	DT	NPHI	RHOB	LLD	LLS	MSFL	PEF	V_{cl}
GR	1.00											
POTA	0.35	1.00										
THOR	0.57	0.67	1.00									
URAN	0.67	-0.12	0.12	1.00								
DT	-0.35	0.36	-0.62	-0.11	1.00							
NPHI	0.19	0.59	-0.45	0.37	0.61	1.00						
RHOB	0.26	-0.48	0.72	-0.31	-0.66	-0.75	1.00					
LLD	0.18	-0.15	0.37	0.07	-0.52	-0.42	0.47	1.00				
LLS	0.20	-0.14	0.38	0.07	-0.52	-0.43	0.49	0.99	1.00			
MSFL	0.40	-0.13	0.64	0.05	-0.57	-0.48	0.67	0.89	0.90	1.00		
PEF	0.22	0.27	-0.28	0.48	0.17	0.54	-0.46	0.01	0.01	-0.12	1.00	
V_{cl}	0.45	0.11	0.50	0.02	-0.42	-0.27	0.39	0.17	0.18	0.37	-0.01	1.00

Table 3

The definition of symbols used in conventional relations of volume of clay estimations (Dresser Atlas 1982)

Symbol	Defenition	
V_{cl}	Volume of clay	
I_A	Ray index	
A_{log}	Total ray reading in the zone of interest	
A_{min}	Average ray response in clean (clay free) zone	GR_{min}=22.242 API; $THOR_{min}$=1.144 ppm
A_{max}	Average ray response in dirty (clay rich) zone	GR_{max}=247.728 API; $THOR_{max}$=28.346 ppm
ρ_b	Bulk density in the zone of interest	
ρ_{bcl}	Bulk density in dirty (clay rich) zone	2.75 g/cm^3
φ_{DT}	Sonic porosity in the zone of interest	
φ_{DTCl}	Sonic porosity in dirty (clay rich) zone	0.334 V/V
φ_N	Neutron porosity in the zone of interest	
φ_{Ncl}	Neutron porosity in dirty (clay rich) zone	0.479 V/V
R_{cl}	Resistivity in dirty (clay rich) zone	0.193 Ohmm
R_t	Resistivity in the zone of interest	
R_{lim}	Resistivity in clean (clay free) zone	694.717 Ohmm

Table 4

The MSE and R-value results of conventional methods

Estimation Relationship	MSE	R-value
Multiple linear regression	$5.2416E^{-3}$	0.5917
Thorium log, using Eqs. (3) and (8)	$5.9619E^{-3}$	0.4796
Density log, using Eq. 4	$6.6029E^{-3}$	0.4691
Natural Gamma Ray log, using Eqs. (3) and (8)	$6.6972E^{-3}$	0.4310
Sonic log, using Eq. 5	$9.4489E^{-3}$	-0.4200
Neutron log, using Eq. 6	$3.2077E^{-2}$	-0.2660
Micro spherically focused resistivity log, using Eq. 7	$8.2077E^{-2}$	-0.1881

Table 5

Input and output data used in MLP network

Data type	Parameter	Symbol	Unit	Range	Mean	Standard Deviation
Input	Thorium	THOR	ppm	3.24-23.96	6.53	3.43
	Sonic	DT	US/F	53.92-89.15	65.19	8.73
	Natural Gamma Ray	GR	API	60.63-149.15	89.27	21.73
	Density	RHOB	g/cm^3	1.95-2.75	2.55	0.18
	Micro Spherically Focused Resistivity	MSFL	Ohmm	0.26-57.72	10.28	13.42
	Neutron	NPHI	p.u	3.35-24.66	11.51	4.80
Output	Volume of Clay	V_{cl}	%	4.10-32.20	11.18	5.36

Table 6

Results of applying different transfer functions on a certain network with fixed topology (e.g. 6-5-1)

Transfer functions	Train data set		Test data set		Validation data set	
	MSE	R-Value	MSE	R-Value	MSE	R-Value
Log-sigmoid (logsig)	$4.43110E^{-4}$	0.9838	$3.83544E^{-3}$	0.8863	$5.14784E^{-3}$	0.9382
Tan-sigmoid (tansig)	$4.68175E^{-4}$	0.9809	$3.99807E^{-3}$	0.8850	$5.30102E^{-3}$	0.9341
Triangular basis (tribas)	$5.38419E^{-4}$	0.9773	$5.72114E^{-3}$	0.8722	$5.98179E^{-3}$	0.9228
Saturating linear (satlin)	$5.80120E^{-4}$	0.9710	$5.98128E^{-3}$	0.8705	$6.01012E^{-3}$	0.9211
Linear (purelin)	$4.92560E^{-4}$	0.9781	$4.06781E^{-3}$	0.8839	$5.48139E^{-3}$	0.9308
Symmetric hard-limit (hardlims)	$5.79364E^{-4}$	0.9721	$5.90217E^{-3}$	0.8712	$6.00974E^{-3}$	0.9234
Hard-limit (hardlim)	$5.0008E^{-4}$	0.9774	$4.17238E^{-3}$	0.8869	$5.69867E^{-3}$	0.9297

Table 7

The MSE and R-value results for different MLP models

Topology	Train data set		Test data set		Validation data set	
	MSE	R-Value	MSE	R-Value	MSE	R-Value
6-3-1-1	5.56122E^{-3}	0.9763	5.98225E^{-3}	0.8693	6.00268E^{-4}	0.9301
6-1-1	5.12760E^{-3}	0.9779	5.76667E^{-3}	0.8704	5.82218E^{-4}	0.9318
6-2-1	4.92721E^{-4}	0.9787	5.00010E^{-3}	0.8791	5.76606E^{-4}	0.9322
6-4-2-1	4.86206E^{-4}	0.9791	4.88922E^{-3}	0.8805	5.50019E^{-3}	0.9341
6-3-1	4.60256E^{-4}	0.9808	4.25131E^{-3}	0.8818	5.37289E^{-3}	0.9353
6-4-1	4.53603E^{-4}	0.9821	4.19103E^{-3}	0.8823	5.25172E^{-3}	0.9378
6-5-3-1	4.48175E^{-4}	0.9830	4.00121E^{-3}	0.8859	5.22114E^{-3}	0.9380
6-5-1	4.43110E^{-4}	0.9838	3.83544E^{-3}	0.8863	5.14784E^{-3}	0.9382
6-6-4-1	4.37284E^{-4}	0.9840	3.50397E^{-3}	0.8895	5.10116E^{-3}	0.9395
6-6-1	4.32472E^{-4}	0.9844	3.32784E^{-3}	0.8921	5.09120E^{-3}	0.9401
6-7-5-1	4.10602E^{-4}	0.9853	3.25419E^{-3}	0.8973	5.00017E^{-3}	0.9428
6-7-1	3.90392E^{-4}	0.9872	3.12721E^{-3}	0.8987	4.91616E^{-3}	0.9439
6-8-6-1	3.82971E^{-4}	0.9881	2.95048E^{-3}	0.8990	4.79415E^{-3}	0.9448
6-8-1	3.65274E^{-4}	0.9917	2.80685E^{-3}	0.9013	4.63702E^{-3}	0.9466
6-9-1	3.84273E^{-4}	0.9878	3.56048E^{-3}	0.8953	4.73828E^{-3}	0.9457
6-9-7-1	3.91105E^{-4}	0.9875	3.63032E^{-3}	0.8947	4.78451E^{-3}	0.9450
6-10-8-1	3.98147E^{-4}	0.9869	3.69020E^{-3}	0.8941	4.80205E^{-3}	0.9448
6-10-1	4.34038E^{-4}	0.9846	3.78991E^{-3}	0.8934	4.86348E^{-3}	0.9446

Table 8

The final weights of optimum ANN model

+1.1132	+1.2217	+0.8756	+1.1452	+0.7503	-0.7669
-0.1213	-0.2580	-1.7918	+0.1748	+0.8368	+0.6445
+0.0514	+0.3986	+0.3342	+0.1808	+0.0043	-0.3321
-0.2267	-0.3228	+1.1709	-0.1710	-0.8916	+0.3813
+0.3625	+0.2780	-2.0177	+1.3722	-0.9783	-1.2381
-0.9894	+1.1609	+0.6247	-0.2532	+0.6989	-0.9422
+0.0625	-0.4393	-0.3623	-0.1375	-1.6289	+0.1992
+0.2419	+0.6920	-0.4751	+0.3525	-0.8441	-0.7508