Electrofacies clustering and a hybrid intelligent based method for porosity and permeability prediction in the South Pars Gas Field, Persian Gulf

Ebrahim Sfidari1*, Abdolhossein Amini2, Ali Kadkhodaie2, Bahman Ahmadi3
1 Department of Geology, University College of Sciences, University of Tehran, Tehran, Iran
2 Department of Geology, Faculty of Natural Sciences, University of Tabriz, Iran
3 Department of Mechanical engineering, University of Guilan, Rasht, Iran
*Corresponding author, e-mail: ebrahimspidari@ut.ac.ir
(received: 10/04/2012 ; accepted: 28/11/2012)

Abstract
This paper proposes a two-step approach for characterizing the reservoir properties of the world’s largest non-associated gas reservoir. This approach integrates geological and petrophysical data and compares them with the field performance analysis to achieve a practical electrofacies clustering. Porosity and permeability prediction is done on the basis of linear functions, succeeding the electrofacies clustering. At the start, an unsupervised neural network was employed based on the self-organizing map (SOM) technique to identify and extract electrofacies groups. No subdivision of the data set was required for the technique on account of the natural characters of the well logs that reflect lithological character of the formations. The second step was examining a supervised neural network which is designed based on the back propagation algorithm. This technique quantitatively predicts the porosity and permeability within the determined electrofacies. The final part of the study was calibration and comparison of the electrofacies clustering results with core and petrographic data. Based on the porosity and permeability maps at different depth levels, the target reservoir is classified into six electrofacies clusters (EF1-EF6) among which the EF5 and EF4 show the best reservoir quality. The EF6 shows moderate reservoir quality, while the EF1 to EF3 show no reservoir quality. A propagation map was also prepared for each reservoir zone, regarding the conceptual depositional maps in the studied reservoir.

Keywords: Electrofacies, Self-organizing maps, Porosity/permeability prediction, South Pars Gas Field, Persian Gulf.

Introduction
Precise mapping of the reservoir properties is a key product for reservoir characterization studies. In case of availability of the conventional well logs, artificial neural networks are capable of recognition of non-linear relationships between well log data and reservoir parameters. In general, the conventional methods fail to accurately classify and estimate the petrophysical properties of the carbonate reservoirs that is mainly due to formation heterogeneity and nonlinearity of its properties. In contrast, the artificial neural networks are found as powerful tools in reservoir nonlinearity examination (Al Moqbel et al., 2011). The reservoir of the South Pars Gas Field (SPGF) is famous for having a centimeter-scaled heterogeneity (Insalaco et al., 2006; Rahimpour-Bonab et al., 2009), hence the neural network has been applied here to decrease the impact of heterogeneity on the estimation of reservoir properties such as porosity and permeability. To do this, a two-step approach has been applied in the Permo-Triassic Kangan-Dalan formations (equivalent of Khuff Formation in Saudia Arabia) in the SPGF.

At the start, the self organizing map (SOM) clustering algorithm is applied to classify the reservoir rocks. The SOM, with non-supervised learning, performs a feature projection non-linearly from a high-dimensional (input) into a low-dimensional (output) space (feature space) while consisting a 2D array of neurons in an orderly fashion (Kohonen, 1989). This method classifies and visualizes the data and proposes a useful approach for characterizing the reservoir properties. The electrofacies clusters are the ultimate output of the SOM approach.

Second part of the approach is to use a back-propagation (BP) neural networking algorithm in order to estimate the porosity and permeability within each particular electrofacies. The BP is a feed-forward supervised network that is known as one of the most popular methods for prediction of the reservoir properties (Looney, 1997). The BP architecture is a hierarchical design consisting of fully interconnected layers or rows of processing units. It is an iterative process in estimation of neural connection strengths (the weights) and
minimizing the error function (Bishop, 1995; Kosko, 1996; Haykin, 1999).
The main inputs for the neural network are well log data, among which NPHI, RHOB, DT and GR are selected for this study. To train the neural network models, seven appraisal wells were involved.

**Geological setting**
The South Pars Gas Field with its extension in Qatar (known as North Field), as world’s largest non-associated gas reservoir (Aali et al., 2006, Khoshnoodkia et al., 2011) is located in the Persian Gulf in the border of Iran and Qatar (Fig. 1). The shallow marine carbonates of Kangan and Dalan formations (Late Permian and Early Triassic in age respectively) are the main reservoir units of the SPGF. They are time equivalent of the Khuff Formation in the North Field (Alsharhan & Nairn, 1997; Kashfi, 2000; Rahimpour-bonab et al., 2010). The late Permian Dalan Formation is about 680m thick in the South Pars, where is divided into three members including Lower Dalan and Upper Dalan carbonates with Nar Anhydrite in their between (Szabo & Kheradpir, 1978). The Early Triassic Kangan Formation is about 193m thick that consists limestone, dolomite, anhydritic dolomite and thin layers of shale (Aali et al., 2006). The Early Silurian Sarchahan Formation (black shales) is considered as the source rock for most oil and gas reservoirs of the Arabian Plate including the SPGF (e.g. Mahmoud et al., 1992; Rahmani et al., 2010). General stratigraphy of the studied area is illustrated in Fig. 1.

Methods and Results
The unsupervised and supervised neural networks have been adapted for electrofacies clustering and reservoir characterization in the studied reservoir. The unsupervised method (SOM) is mainly applied for classification of electrofacies, while the supervised approach (BP) is carried out for the porosity and permeability prediction.
A suite of well logs were selected for the data analysis with regard to the quality of data and location of the appraisal wells. The input data include neutron (NPHI), density (RHOB), acoustic transmit-time (DT), gamma ray (GR) and photoelectric logs (PEF). Moreover, core data from seven appraisal wells (3400 core samples) were used to calibrate the determined permeability and porosity. Log and core data was hence correlated in the studied wells and the electrofacies clusters were

Figure 1: Location map of the South Pars Gas Field in the Persian Gulf (right) and stratigraphic position of the studied formations (left).
Electrofacies clustering and a hybrid intelligent based method for porosity and permeability

used to predict the porosity and permeability in the non-cored or blind wells.

Electrofacies Classification
As described above, a self-organizing map (SOM) neural network clustering was applied for determination and clustering of the electrofacies in the studied field. The basic concept of the SOM application is illustrated in Fig. 2.
The SOM network was designed in the Matlab software for clustering of the well log data. This method implements an ordered dimensionality-reducing map for data training, i.e. it provides projection of the multidimensional data into a 2D map and preserves the topology of this input data space (Kohonen et al., 1997).

Figure 2: A Schematic diagram representing the methodology of SOM application in the electrofacies clustering

The term “self-organizing” refers to the ability of learning and organizing data without involving the associated-dependent output values for the input pattern (Mukherjee, 1997). The SOM shares with the conventional ordination methods to develop a basic idea of displaying a high-dimensional signal manifold on a much lower dimensional network in an orderly fashion (usually a 2D space) (Astela et al., 2007). A SOM approach normally consists of neurons that are organized on a regular low-dimensional grid. The number of neurons may vary from a few dozen up to several thousands. The neurons are connected to adjacent neurons by a neighborhood relation that dictates the topology or structure of the Kohonen (2001) Map. Hence, similar objects (in our case sampling points) should be mapped close together on the grid (Astela et al., 2007). A training algorithm of SOM constructs the nodes to represent whole dataset and optimize their weights at each iteration step. In each step, one sample vector (x) from the input dataset is chosen randomly and the distance between the selected vector and all the weight vectors of the SOM are calculated. Accordingly, an optimal topology is expected.

Based on the Kohonen (2001), training algorithm of the self-organizing map network is as follows:
Let \( x_k \) (with k=1 to the number of training patterns N) be the n-dimensional training patterns. Let \( w_{ij} \) be the neuron in position (i, j). Let \( 0 \leq \alpha \leq 1 \) be the learning rate and \( h(w_{ij}, w_{mn}) \) be the neighborhood function.
This neighborhood function assumes values in \([0, 1]\), and is high for neurons that are close in output space, and small (or 0) for neurons far away. It is usual to select a function that monotonically decreasing the non-zero to a radius r (called neighborhood radius) and zero from there onwards.
Let winner be the winning neuron for a given input pattern. Then, the algorithm for training the network of each input pattern is:
1. Calculate the distance between the pattern and all neurons (\( d_{ij} = ||x_k-w_{ij}|| \))
2. Select the nearest neurons as winner \( w_{\text{winner}} \) (\( d_{ij} = \min(d_{mn}) \))
3. Update each neuron according to the rule \( w_{ij} = w_{ij} + \alpha h(w_{\text{winner}}, w_{ij}) \frac{||x_k-w_{ij}||}{d_{ij}} \)
4. Repeat the process until a certain stopping criterion is met. Usually, the stopping criterion is a fixed number of iterations. To guarantee convergence and stability of the map, the learning rate and neighborhood radius are decreased f each iteration, and then they are converged to zero (Sfidari et al., 2012).
The distance measured between pair of objects in this study is Euclidean distance. The unsupervised self-organizing map (SOM) is usually developed for the purpose of clustering dominant logfacies within the reservoir.

Classification of well log data
The SOM was trained with different numbers of map units and the optimum map size and normalization method were selected according to the minimum quantization error. The results together with the prerequisite that the number of neurons should be close to the number of the samples by equation (1), led to the selection of best unit map (Figs. 3, 4). The number of the output neurons in the SOM (the map size) is important to detect the deviation of the data. If the map size is small, it might not explain some important
differences that should be detected. Conversely, if the map size is too big, the differences are too small. The number of output neurons in the SOM can be selected using the heuristic rule suggested by Vesanto et al., (2000):

\[ m = 5\sqrt{n} \]  

Eq. (1)

where \( m \) is the map size and \( n \) is the number of the training samples. Using this formula, the map size can be efficiently determined without trial and error. Results of the SOM analysis are presented through various visualizations of the self-organized map (Fig. 3). The five component plots, one for each input variable, are presented as rectangular maps with square nodes in Fig. 3. These visualizations reveal moderate correlation between DT and NPHI, and nonlinearity of correlation between PEF with other component plans (Fig. 3).

Figure 3: Visualizations of the Component plots representing the normalized values of the five variables. Correlation between DT and NPHI can be confirmed from this SOM visualization.

The SOM segregates the log data in to six particular clusters, based on the number of samples, that establish the individual cluster and the individual reservoir type. Using K-means algorithm and the C-index (Hubert & Schultz, 1976), six clusters are also appreciated (Figs. 4, 5). The color coding red is for cluster 3, yellow for cluster 6, light purple for cluster 4, magenta for cluster 5, green for cluster 2, and blue for cluster 1. The response of the given data to the map (signal hits–BMUs number) for each cluster was summed as a cluster index value (Fig. 5).

Figure 4: Top: C-index versus cluster number. Down: k-means clusters with hit histogram numbers (data density indicated by relative size of a circle) in U-matrix (left) and U-matrix indicates the distances between neighboring neurons (right).
The results of SOM clustering on generation of the electrofacies clusters are shown in Fig. 5. Based on the SOM visualization, C-index and the understanding about reservoir non-linearity (heterogeneity), the reservoir rocks were classified into six electrofacies (EF1 to EF6) associated with particular petrophysical properties. The EF5 that shows the best reservoir quality is characterized by the low amount of density, highest DT, high amount of neutron porosity (NPHI) and lowest GR response. The EF4 with good reservoir quality is marked by medium amount of RHOB response, medium to high DT, high NPHI and low GR. The EF6 that shows moderate to good reservoir quality is marked by medium to low response on density log (RHOB), medium DT, high neutron porosity (NPHI) and moderate GR. The EF1, that is considered as a non-reservoir rock unit, is recognized by high amount of RHOB response, low DT, low NPHI and high GR. Similar to the EF1, the EF2 is marked by poor to non-reservoir quality and is characterized by high RHOB response, low DT, low NPHI, and low GR. The EF3 with low to poor reservoir quality is marked by high RHOB, medium DT, medium NPHI and low GR. The graphical pairwise characterization of well logs for each electrofacies is illustrated in Fig. 5. Vertical distribution of electrofacies extracted for one of the examined wells (well A) is shown in Fig. 6. The EF5 with the highest porosity and permeability is mainly located in the depth interval of 3150 to 3200m. The EF4 with fair to good reservoir quality is distributed in the depth intervals of 3150 to 3210, 2990 to 3010 and 2900 to 2920m (Fig. 6).

Porosity and Permeability Prediction
Various neural network types have been used for field data analysis which are discussed in the literature (e.g. Huang et al., 1996; Cuddy, 1997; Huang et al., 1999; Rezaee et al., 2006; Sfidari et al., 2012). The neural network method is widely used by computer scientists, electronic engineers, biologists and psychologists but rarely used by geologist during the last decades. However, it is becoming popular method for prediction of field
Neural networks are also known as Artificial Neural Network (ANNs), Connectionism or Connectionist Models, Multi-layer Perceptrons (MLPs) and Parallel Distributed Processing (PDP). Despite the different terms and types, there are a small group of “classic” networks which are widely...
Electrofacies clustering and a hybrid intelligent based method for porosity and permeability used, on which many others are based. The back-propagation neural networking algorithm (BP) is the most popular one that provides good results in the geological studies (Al Moqbel et al., 2011). Therefore, the BP method has been used in this investigation to predict the porosity and permeability values in the studied wells. The BP neural network is a multi-layered feed-forward neural network that comprises input, hidden and output layers (Fig. 6).

Mathematically, neuron j has two equations for summation and activation as shown below:

\[
k_{pj} = \sum_{i=1}^{k} w_{ji} x_{pi} \text{ \hspace{1cm} Eq. (2)}
\]

\[
z_{pj} = f(k_{pj} - \theta_{pj}) \text{ \hspace{1cm} Eq. (3)}
\]

Were \(x_{pi}, x_{p1}, x_{p2}, \ldots, x_{pN}\) are the input logs, \(w_{j1}, w_{j2}, \ldots, w_{jk}\) are the synaptic weights of neuron j within total of N samples that connect neuron in a layer to the neuron in the next layer, \(k_{pj}\) is the linear combiner output, \(\theta_{pj}\) is the threshold, \(f\) is the activation function (sigmoid function) for example in this paper for each node a sigmoid function is used, i.e., \(f_x = 1/(1 + e^{-x})\), and \(z_{pj}\) is the output porosity or permeability from the neuron.

In general, there are four steps in the training process:

Step 1: Preprocessing and initializing the weight values. Preprocessing data in this study achieved by the method discussed in section 3.1. Before training a network, the weights and biases must be initialized. In order to avoid the result located in flat area, random weight values are selected.

Step 2: For input sample, outputs of arbitrary node in both hidden layer and output layer (in forward propagation) are calculated.

Step 3: Errors of hidden layer and output layer (in backward propagation) are calculated.

Step 4: Weight values according to error function are adjusted until the minimal error is achieved. The errors of the nodes of the output layer are back-propagated; this process gives the error BP neural network its name (Rumelhart & McClelland, 1986).

The individual BP net is adapted here for each logfacies extracted in the clustering step. For prediction of porosity and permeability of the reservoir, the networks are trained using input/output pairs in the four wells. Three input logs (NPHI, RHOB, and DT) are feed in to the network to estimate the reservoir porosity and permeability. The correlation between cores porosity and permeability with estimated values in each logfacies is illustrated in Fig. 7 and Table 1. The overall correlation is approximately 0.908 and 0.86 for porosity and permeability, respectively.

Table 1. Correlation of predicted porosity and permeability with those available from the cores. Per.= permeability, Por.= porosity

<table>
<thead>
<tr>
<th>Facies</th>
<th>DT</th>
<th>NPHI</th>
<th>RHOB</th>
<th>MSE_Por</th>
<th>MSE_Per</th>
<th>R_Por</th>
<th>R_Per</th>
<th>Epoch</th>
<th>Distribution based on lucia classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48.1</td>
<td>.040</td>
<td>2.85</td>
<td>.111</td>
<td>.00100</td>
<td>.84</td>
<td>.912</td>
<td>12</td>
<td>Non-reservoir quality</td>
</tr>
<tr>
<td>2</td>
<td>48.3</td>
<td>.051</td>
<td>2.88</td>
<td>.019</td>
<td>.00100</td>
<td>.86</td>
<td>.910</td>
<td>6</td>
<td>Non-reservoir quality</td>
</tr>
<tr>
<td>3</td>
<td>50.1</td>
<td>.121</td>
<td>2.70</td>
<td>.008</td>
<td>.00050</td>
<td>.88</td>
<td>.930</td>
<td>8</td>
<td>Non to bad reservoir quality</td>
</tr>
<tr>
<td>4</td>
<td>57.0</td>
<td>.690</td>
<td>2.40</td>
<td>.010</td>
<td>.00073</td>
<td>.84</td>
<td>.920</td>
<td>7</td>
<td>Mainly class 3 and som class 2 with good reservoir quality</td>
</tr>
<tr>
<td>5</td>
<td>70.0</td>
<td>.112</td>
<td>1.80</td>
<td>.014</td>
<td>.00910</td>
<td>.85</td>
<td>.880</td>
<td>9</td>
<td>Class 1 and 2 with best reservoir quality</td>
</tr>
<tr>
<td>6</td>
<td>56.3</td>
<td>.153</td>
<td>2.60</td>
<td>.009</td>
<td>.00400</td>
<td>.89</td>
<td>.900</td>
<td>5</td>
<td>Wide lange distribution with good to moderate reservoir quality</td>
</tr>
<tr>
<td>Sum</td>
<td>54.9</td>
<td>.194</td>
<td>2.53</td>
<td>.010</td>
<td>.00270</td>
<td>.86</td>
<td>.908</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Validation of the results was done prior to generating the porosity and permeability values in the non-cored intervals and wells. This is to ensure that the network is generalizing not over-fitting. A good way for validation of the results is the take-one-out approach, i.e. excluding some wells in the training stage (3 wells in this study). The networks are trained by data from four wells for each electrofacies. The accuracy of the networks is evaluated by comparing the predicted results with the actual well data at the location of excluded wells (Figs. 7, 8). Once the networks are fully trained and validated, the neural network was applied to field data in order to estimate the porosity and permeability in non-cored wells or intervals.

**Discussion**

In general, the carbonate reservoirs show more scattered porosity and permeability values than their siliciclastic counterparts. Such characteristics make their lateral and vertical prediction hard from petrophysical logs. The porosity and permeability of the carbonate reservoirs are affected by depositional environment of the original facies, diagenetic overprints and their sequence stratigraphic position (Rahimpour-Bnab et al, 2008).

The porosity/permeability cross-plot for one of the studied wells (well A) is shown in Fig. 9. The original rock fabric is compared with the Lucia rock fabric classification chart (Lucia, 1999).

![Figure 7: Comparison between true measured and predicted porosity in 3 tested wells.](image-url)
Electrofacies clustering and a hybrid intelligent based method for porosity and permeability

Figure 8: Comparison between true measured and predicted permeability in 3 tested wells.

Such comparison represents a scattered plot indicating a heterogeneous formation. Fine scale fluctuations in sea level, cause of centimeter-scaled variations of depositional facies, is supposed to be the source of mentioned heterogeneity. Accordingly, the original rock fabric varies significantly in the space and results in the scattering of reservoir properties. In this study, the SOM method has been adapted to decrease the formation heterogeneity and reduce the amount of non-linearity between porosity and permeability.

Once the electrofacies determination was implemented for all studied wells, the resulted electrofacies were compared with the original rock fabric achieved from petrographical studies. In addition, each electrofacies was labeled based on

Figure 9: Porosity and permeability cross-plot for well A in the studied field. The plot shows no distinct relation between porosity and permeability, as data is scattered in all Lucia porosity-permeability classes.
its reservoir properties and was compared with the porosity and permeability obtained from the cores (Fig.10).

Fig. 10. Porosity and permeability cross-plots of the studied electrofacies. EF1 with the lowest porosity and low permeability (a), EF2 with low porosity and lowest permeability (b), EF3 with low to moderate porosity and permeability (c), EF4 with a high porosity and moderate permeability (d), EF5 with highest porosity and permeability (e) and EF6 with a wide range of porosity and permeability values (f).

Our observation showed that the EF5 is mainly characterized by grain-dominated fabrics including dolomitized and non-dolomitized grainstone, ooid grainstone and grain-dominated packstones. The last showed the best reservoir quality among the reservoir rocks. The EF4, which showed fair to good reservoir quality includes wackestones, mudstones, mud-rich packstones, and tight
Electrofacies clustering and a hybrid intelligent based method for porosity and permeability

grainstones. Dissolution is the main factor controlling the reservoir quality in the mentioned electrofacies.

The EF6, which shows a fair to moderate reservoir quality, was found in variety of forms including a tight mudstone to wackestone, anhydrite cemented grainstone, and highly cemented dolomites. The EF3 which is considered as a poor reservoir rock, consists of anhydrite and tight mudstone, highly cemented dolomites and some grainstones. The EF1 and EF2, with non reservoir quality is comprised of anhydrite, tight wackstone, mudstone and anhydrite cemented grainstones.

The studied reservoir is supposed to be deposited in a shallow water carbonate ramp (Insalaco et al., 2006). Relative sea level change is a key process which affects the porosity and permeability in the carbonates (Jacquin et al., 1991). The late transgressive (late TST) and early highstand (early HST) phase in the carbonate platform is characterized by massive barrier settings mainly composed of grain supported facies including bioclastic oolitic grainstones and packstones. These settings showed high porosity and permeability values and are found within the EF4 and EF5 rock classes. They coincide with Lucia rock fabrics classes 1 and 2 and are recognized by low RHOB, high DT and high NPHI on the electrical logs. EF1, EF2 and EF3 are supposed to be developed in the sabkha and lagoon sub-environments during the late highstand (late HST) early transgressive (early TST) phase. These are poor quality to non-reservoir rocks and are characterized by high RHOB, low NPHI, low DT and high GR on the well logs.

After application of unsupervised learning method for clustering, a supervised learning was used for porosity and permeability prediction within each electrofacies. Results from this analysis on the seven wells propagated in the framework of fence diagrams (Fig. 11).

Conclusions

A carbonate reservoir characterization based on well log data is discussed in a two-stage approach using unsupervised and supervised neural network algorithms. The unsupervised self-organizing map is developed for the aim of clustering log responses into reservoir electrofacies. This classification does not require any further subdivision of the dataset but follows the log measurements indicative of mineral and lithofacies characteristics of the rocks. On this basis the SOM clustering of the studied rocks fall in to six different classes. A good correlation between the corefacies and determined electrofacies was observed.

Use of the back-propagation neural network method for the estimation of porosity and permeability in the individual electrofacies provided satisfactory results. Validation of the
results makes certain that the application of neural network is generalizing not over-fitting. A good way for validation is found the take-one-out approach. The accuracy of the networks was evaluated by comparison of the predicted results with the actual well’s data at the excluded wells. The unsupervised and supervised learning generated more accurate results for estimation of porosity and permeability, where core data are not available.

References