Computers & Geosciences 57 (2013) 1–10

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

A hybrid framework for reservoir characterization using fuzzy ranking and an artificial neural network

Baijie Wang [a], Xin Wang [a,*], Zhangxin Chen [b]

[a] Department of Geomatics Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada T2N 1N4
[b] Department of Chemical and Petroleum Engineering, Schulich School of Engineering, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada T2N 1N4

ARTICLE INFO

Article history:
Received 22 March 2012
Received in revised form
13 March 2013
Accepted 16 March 2013
Available online 4 April 2013

Keywords:
Reservoir characterization
Fuzzy ranking
Artificial neural networks
Geographic information system

ABSTRACT

Reservoir characterization refers to the process of quantitatively assigning reservoir properties using all available field data. Artificial neural networks (ANN) have recently been introduced to solve reservoir characterization problems dealing with the complex underlying relationships inherent in well log data. Despite the utility of ANNs, the current limitation is that most existing applications simply focus on directly implementing existing ANN models instead of improving/customizing them to fit the specific reservoir characterization tasks at hand. In this paper, we propose a novel intelligent framework that integrates fuzzy ranking (FR) and multilayer perceptron (MLP) neural networks for reservoir characterization. FR can automatically identify a minimum subset of well log data as neural inputs, and the MLP is trained to learn the complex correlations from the selected well log data to a target reservoir property. FR guarantees the selection of the optimal subset of representative data from the overall well log data set for the characterization of a specific reservoir property; and, this implicitly improves the modeling and predication accuracy of the MLP. In addition, a growing number of industrial agencies are implementing geographic information systems (GIS) in field data management; and, we have designed the GFAR solution (*G*IS-based *F*R *A*NN *R*eservoir characterization solution) system, which integrates the proposed framework into a GIS system that provides an efficient characterization solution. Three separate petroleum wells from southwestern Alberta, Canada, were used in the presented case study of reservoir porosity characterization. Our experiments demonstrate that our method can generate reliable results.

1. Introduction

In the petroleum industry, reservoir properties, such as porosity, permeability and saturation, have significant impacts on reservoir simulation, enhanced oil recovery design, field operations and geological studies (Aminian and Ameri, 2005; Mohaghegh et al., 1996). For example, porosity describes the volume fraction of the pore space and is related to the hydrocarbon reserves contained in a reservoir.

Reservoir characterization is a process for quantitatively assigning reservoir properties using all the available field data. In practice, reservoir characterization is a very complex geological problem, and the ability to obtain reliable and accurate reservoir properties is crucial (Al-Bulushi et al., 2009; Lim, 2005; Mohaghegh, 2000).

Due to the importance of reservoir characterization, various methods for acquiring reliable reservoir properties have been proposed. The most direct method is core analysis, which entails laboratory analysis of rock sample cores that have been taken from a reservoir. Although core analysis results are accurate, the process is costly; in addition, this approach can only provide reservoir properties for certain discrete points (Mohaghegh et al., 1996)

Well logs record information via subterraneous sensors and provide valuable high resolution but indirect information about the mineralogy, texture, sedimentary structures and fluid content of a reservoir (Lim, 2005). For acquiring reservoir properties on a large scale, regression analysis is used to build linear or nonlinear correlations between well log data and reservoir properties. However, as reservoir heterogeneity (i.e., reservoir properties, such as porosity, permeability and saturation, change both vertically and horizontally in space) increases, these correlation analyses usually cannot provide satisfactory results (Aminian and Ameri, 2005).

Recently artificial neural networks (ANN) have shown great potential in reservoir characterization by inferring reservoir properties from well log data (Al-Bulushi et al., 2009; Aminian and Ameri, 2005; Eidsvik et al., 2004; El-Sebakhy et al., 2012; Mohaghegh et al., 1996; Mohaghegh, 2000; Olatunji et al., 2011a; Olatunji et al., 2011b; Wong et al., 1995). In this method, well log data and core analysis data at the same depth interval are paired

* Corresponding author. Tel.: +1 403 220 3355; fax: +1 403 284 1980.
E-mail address: xcwang@ucalgary.ca (X. Wang).

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2013.03.016

together to train an ANN. After proper training, the ANN can be used to predict reservoir properties based only on well log data for depth intervals where core analysis data are not available.

The essential assumption in applying ANN to reservoir characterization is that relationships between the well log variables and the reservoir properties can be correctly represented by the training data. Hence, training data selection, or representative data selection, is critical for ANN performance.

The selection of a proper set of well log variables for an ANN is not a trivial problem. A random selection or empirical selection based on limited experience for reservoir characterization may eliminate useful information, which will decrease the accuracy of the ANN. The other option is to use all available well log data. However, well log data are sometimes noisy and highly correlated, which may negatively influence the modeling performance of the ANN.

A complete well log contains approximately 20 well log variables, recording geological information from gamma ray, resistivity, spontaneous potential, sonic, and/or thermal sensors (Ellis and Singer, 2007; Brock, 1986; Wong and Gedeon, 2001). Well log variables generated from similar well logging sensors are highly correlated (Aminian and Ameri, 2005). For example, Array Induction Four Foot Resistivity A20 (AF20) and Array Induction Four Foot Resistivity A30 (AF30) are two well log variables measuring the reservoir formation resistivity. They are recorded from the same device with different measuring resolutions and thus are highly correlated with each other. Different well log variables also have various levels of relationship to a target reservoir property. For example, resistivity logs are believed to have a close relationship with saturation.

When the number of dependent or irrelevant neural inputs rises, an ANN tends to converge to local minima, decreasing the predication accuracy of ANNs in reservoir characterization (Lin et al., 1996; Lin et al., 1998). Only a few works have been conducted for selecting representative well log data for ANN-based reservoir characterization (Aminian and Ameri, 2005; Helle et al., 2001; Wong, Gedeon and Taggart, 1995). Under such circumstances, the modeling ability and predication accuracy of ANN in reservoir characterization can be jeopardized.

Fuzzy ranking (FR) is a global prioritizing technique that can automatically and quickly identify a subset of significant independent inputs for use in nonlinear system models (Lin et al., 1996). FR can identify the representative data for neural inputs mechanically without prior knowledge. In addition, this technique is tolerant of noise and has linear computation complexity, which adds robustness to the reservoir characterization system. Through the identification of a subset of representative variables from the well log, FR has the potential to improve the modeling and prediction performance of an ANN in reservoir characterization problems.

Another problem in reservoir characterization is the provision of a user-friendly system for the engineers managing the field data and performing the reservoir characterization. Digital field data for petroleum wells have accumulated for decades, most of which are in numerical form. Without proper visualization, it is difficult for engineers to work with these data. An efficient method is the visualization and access of field data on a map.

A geographic information system (GIS) is a computer system used for the capture, storage, querying, analysis and display of geospatial data (Chang, 2008). In this paper, we present the development of a GIS to help to manage oil field data. Well log and core analysis data are stored in a geodatabase and referenced to the wells locations, which can be visualized and accessed geographically via a digital map. With the proposed intelligent reservoir characterization function in GIS, the system allows efficient data loading and visualization of the reservoir results.

The main contributions of this paper are summarized in the following paragraphs.

First, we propose a novel hybrid FR-neural (fuzzy ranking—artificial neural network) framework for reservoir characterization from well logs. The proposed framework integrates a two-step FR and a multilayer perceptron (MLP) neural network. In the first FR step, a fuzzy curve ranks the well logs and selects the important well log variables related to the target reservoir property. In the second FR step, fuzzy surface further removes the highly dependent variables. The selected representative well log variables then feed an MLP with the objective to characterize a specific reservoir property. This method is applicable on a well-by-well basis. For each individual well, FR removes irrelevant and highly correlated well log data with respect to the target reservoir property, which helps to reduce the risk of local minima and over-fitting, implicitly increasing the predication accuracy of the ANN.

Second, we describe the development of an intelligent a GIS—the GIS-based FR ANN reservoir characterization solution (GFAR Solution) system. The system can manage and visualize large amounts of field data from petroleum wells on a map. It is also integrated with the proposed FR-neural framework, providing a user-friendly reservoir characterization interface.

Third, we present a set of experiments and case studies that were conducted on three separate petroleum wells located in southwestern Alberta, Canada. Well log and core analysis data extracted from these wells were used for testing the proposed FR-neural reservoir characterization framework. Comparisons with other reservoir characterization methods have been made. The results show that the FR-neural framework increased the predication accuracy and stability.

The remainder of the paper is organized as follows: some related work on reservoir characterization is presented in Section 2. The FR-neural reservoir characterization framework is described in detail in Section 3, the GFAR Solution system is presented in Section 4. The quantitative evaluation of the proposed framework is described in Section 5, and conclusion and future work are discussed in Section 6.

2. Related work

2.1. Regression analysis in reservoir characterization

Regression analysis is often used in reservoir engineering to predict a reservoir's properties from various field data. Well log data are the most widely used. Traditionally, statistical correlations, linear or nonlinear, were built between the well log data and the various reservoir properties based on assumptions of homogeneity. For example, nonlinear formulas exist that compute reservoir porosity based on resistivity logs and density logs (Ellis and Singer, 2007; Brock, 1986). However, these correlations work only for homogeneous reservoirs. As the degree of heterogeneity of the reservoir increases, these correlations lose accuracy. Predication of reservoir properties from well log data in heterogeneous reservoirs is very complex, and traditional statistical methods fail to produce accurate predications (Aminian and Ameri, 2005; Helle et al., 2001; Lim, 2005; Mohaghegh et al., 1996).

2.2. ANN in reservoir characterization

MLP neural networks with back propagation algorithms have gradually become accepted as the new intelligent reservoir characterization tools (Aminian and Ameri, 2005; Helle et al., 2001; Wong, Gedeon and Taggart, 1995). Mohaghegh et al. (1996) built a three-layer MLP neural network for porosity and permeability characterization. Three well log variables (gamma ray, bulk density and

depth induction) were used for both porosity and permeability characterization. Results showed that ANNs exceeded traditional statistical methods for accurately predicting reservoir properties for heterogeneous reservoirs. However, the selection of the three well log variables as neural inputs was not discussed.

Independently, Helle et al. (2001) concluded that using an ANN in reservoir characterization has a number of advantages over conventional methods, despite the required effort in selecting good representative training data. In their work, Helle et al. built two MLP neural networks to predict porosity and permeability. Well log variables from the sonic, density and resistivity measurements were used for neural inputs for porosity characterization; and, well log variables from the density, gamma ray, neutron and sonic measurements were used for permeability characterization. All the neural inputs were manually and selected.

The main drawback of using ANN in reservoir characterization continues to be the difficulty in selecting a representative collection of training samples.

2.3. Data selection for ANN in reservoir characterization

As a data driven approach, the modeling ability of an ANN relies heavily on the quality of its neural inputs; and, any inappropriate well log data selection decreases an ANN's performance. Several studies have shown that fuzzy analysis can help identify the optimal set of independent variables for an ANN by addressing the uncertainties of neural inputs (Mohaghegh, 2005; Mohaghegh, 2000). Lim (2005) used a fuzzy curve analysis to select neural inputs from well logs for use in ANN-based reservoir characterization. Results showed that the modeling ability with the combination of a fuzzy curve and an MLP significantly exceeded multiple regression analysis. However, the data correlation problem among well logs was not discussed. Additionally, Lim (2005) did not give a comprehensive evaluation of the proposed method, because the proposed ANN model was trained and tested using the same data set.

3. A hybrid framework for reservoir characterization

In this section, we propose the novel FR-neural framework for reservoir characterization. This framework is based on FR and an ANN. As shown in Fig. 1(a), the new hybrid framework includes two stages: FR and pattern recognition.

The FR stage selects the important well log variables related to the target reservoir property as the neural inputs for ANN and helps in improving the pattern recognition performance of the ANN. In the pattern recognition stage, an MLP neural network is trained to learn the desired complex relations between the selected well log variables and a reservoir property. After verification, it further predicts the reservoir property values for intervals where core analysis data are nonexistent.

The FR and pattern recognition steps are discussed further in Sections 3.1 and 3.2, respectively.

3.1. Fuzzy ranking stage

FR stage includes a fuzzy curve (FC) step and a fuzzy surface (FS) step, as shown in Fig. 1(b) and (c), respectively, both of which work to select the best well log variables for the ANN. The FC ranks all potential well log variables according to their relevance to the target reservoir properties and eliminates irrelevant variables. The FS identifies the dependency among well log variables selected by the FC and eliminates the redundant, highly correlated variables.

3.1.1. Fuzzy curve

An FC is based on the assumption that the most important input variable plays the most important role in approximating the output (Lin et al., 1998). It simulates the relationship between each potential input and the output by building an FC function, and meaningful inputs are selected based on the closeness between the simulated relationship and the real relationship.

In our case, all well log variables serve as candidate inputs, and the target reservoir property serves the output. The well log variables are denoted by $X=\{x_i \mid i=1,2,\ldots,n\}$ and the target reservoir property by y. Each well log variable has m data points $(x_{ik},\ k=1,2,\ldots,m)$ and the corresponding reservoir property values $(y_k,\ k=1,2,\ldots,m)$. With no prior knowledge about the relationship between x_i and y, the objective of the FC is the selection of a subset of variables from X, $SX=\{x_s \mid s=1,2,\ldots l,$ and $l\le n\}$, so that a nonlinear relation exists between SX and y

$$y\approx f(SX),\quad SX=\{x_s|s=1,2,\ldots,l \text{ and } l\le n\} \tag{1}$$

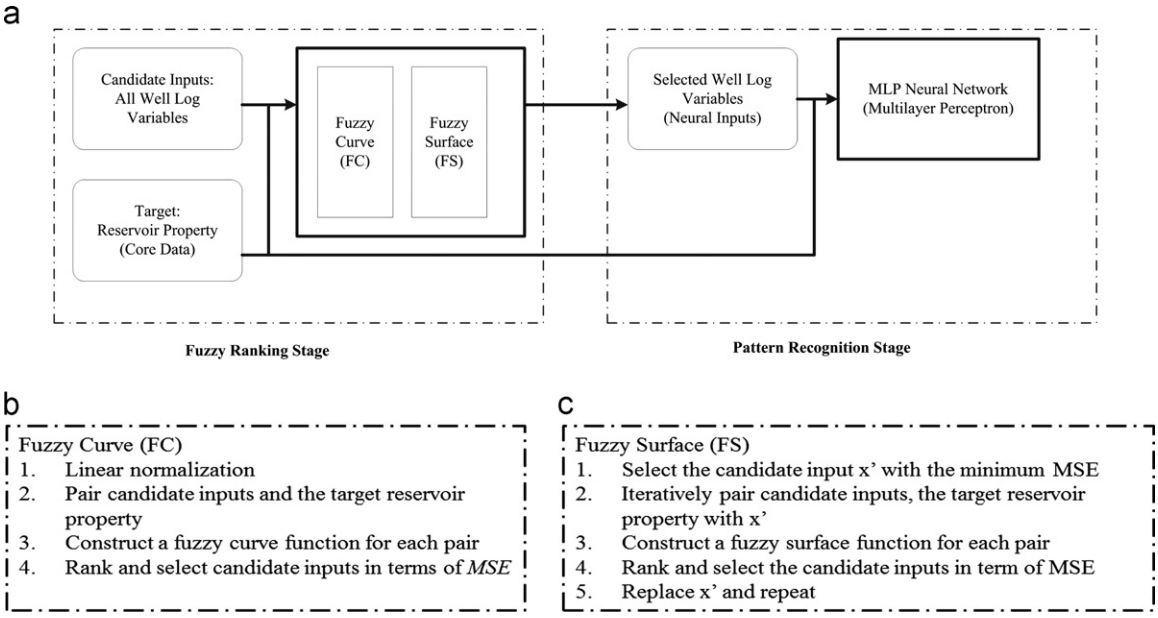

Fig. 1. The proposed FR-neural reservoir characterization framework.

The unselected variables tend to have a random correlation with y.

Since different well log variables have different value ranges, the first step of the FC is normalization of all the candidate well log variables, as shown in the following equation:

$$x_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{2}$$

The FC then builds fuzzy membership functions for every data point in the x_i–y space. In this study, the fuzzy membership functions are in a Gaussian form, as shown in Eq. (3), where b is a constant that controls the span of the Gaussian function

$$\mu_{ik}(x_i) = \exp\left(-\left(\frac{x_{ik} - x_i}{b}\right)^2\right) \tag{3}$$

The Gaussian fuzzy membership function gives a prediction of the target reservoir property, y, when the well log variable, x_i, changes slightly in a neighborhood close to x_{ik}. For example, the density porosity sandstone scale (DPSS) is a well log that measures the density porosity. Fig. 2(a) shows the scatter plot of data points between the normalized DPSS using Eq. (2) and reservoir porosity, denoted by '+'. The Gaussian fuzzy membership functions, $\mu_{ik}(x_i)$, are built for each data sample, x_{ik}, in the normalized DPSS. Fig. 2 (a) shows the fuzzy membership function curves, $y_k\mu_{ik}(x_i)$, for three points, where y_k is the reservoir porosity value at the same depth interval as x_{ik}.

Next, the FC integrates all the fuzzy membership functions forming a fuzzy curve, $c_i(x_i)$. Specifically, $c_i(x_i)$ defuzzifies all the Gaussian fuzzy membership functions, $\mu_{ik}(x_i)$, by normalized summation, as in Eq. (4), which is an approximation of the relationship between x_i and y

$$c_i(x_i) = \frac{\sum_{k=1}^{m} y_k \times \mu_{ik}(x_i)}{\sum_{k=1}^{m} \mu_{ik}(x_i)} \tag{4}$$

With the example in Fig. 2(a), all Gaussian functions are weight averaged using Eq. (4), where the weight is the target, y_k, corresponding to x_{ik}. The solid line in Fig. 2(b) shows the FC of the DPSS to the reservoir porosity.

The FC is a weighted local average of y_k along each x_i axis, where the size of a local neighborhood is controlled by b in Eq. (2). When b is large, $\mu_{ik}(x_i)$ is equal to approximately 1 for all x_{ik}, so that $c_i(x_i) \approx \sum_{k=1}^{m} y_k/m$ equals the average of y_k at every x_{ik}. When b is small, $\mu_{ik}(x_i) \approx 1$ only for $x_i = x_{ik}$, while 0 elsewhere, so $c_i(x_i) \approx y_k$ only for $x_i = x_{ik}$ and $c_i(x_i) \approx 0$ elsewhere. Hence, $c_i(x_i)$ is an approximation of y based on x_i. Here, b controlling the size of the local neighborhood is critical to the approximation of y. When b is too large, $c_i(x_i)$ is not sensitive to a local change. When b is too small, $c_i(x_i)$ loses the average information. As suggested by Lin et al. (1998), b was empirically chosen to be 0.08 in this study.

The more information x_i contains, the closer the approximation c_i is to output y. $MSEc_i$ is the normalized mean square error to measure the distance from c_i to y, as shown in Eq. (5), where $var(y)$ is the variance of y used to scale the mean square error

$$MSEc_i = \frac{\sum_{k=1}^{m} (c_i(x_{i,k}) - y_k)^2}{m \times \mathrm{var}(y)} \tag{5}$$

Thus, the last step of FC is the sorting of the well log variables in ascending order in terms of $MSEc_i$. The most important well log variable is the one with the smallest $MSEc_i$ value. Specifically, if x_i is a random noise and has no relation with the output, then the c_i to x_i tends to be flat, which also results in a high mean square error.

After FC, irrelevant well log variables to the output are eliminated and the remaining ones serve as the candidate inputs for the FS step.

3.1.2. Fuzzy surface

An FS identifies and eliminates the highly correlated variables from the results selected from the FC step. The FS is based on the assumption that two independent variables do a better job of approximating the output than two correlated variables (Lin et al., 1998).

For two selected variables from the FC, x_i and x_j, the FS is defined in Eq. (6), where $\mu_i$and$\mu_j$are Gaussian fuzzy membership functions of x_i and x_j, as defined by Eq. (3). s_{ij} computes the weighted average of output y_k based on the information from the combination of x_i and x_j

$$s_{ij}(x_i, x_j) = \frac{\sum_{k=1}^{m} y_k \times \mu_{ik}(x_i) \times \mu_{jk}(x_j)}{\sum_{k=1}^{m} \mu_{ik}(x_i) \times \mu_{jk}(x_j)} \tag{6}$$

The more information contained between x_i and x_j, the better the approximation of $s_{ij}(x_i,x_j)$ to output y. The normalized mean square error function in Eq. (7) is used to compute the distance from $s_{ij}(x_i,x_j)$ to y, where $var(y)$ is the variance of y. The $MSEs_{ij}$ from two independent variables (x_i and x_k) will be smaller than from two correlated variables (x_i and x_j)

$$MSEs_{ij} = \frac{\sum_{k=1}^{m} (s_{ij}(x_{ik}, x_{jk}) - y_k)^2}{m \times \mathrm{var}(y)} \tag{7}$$

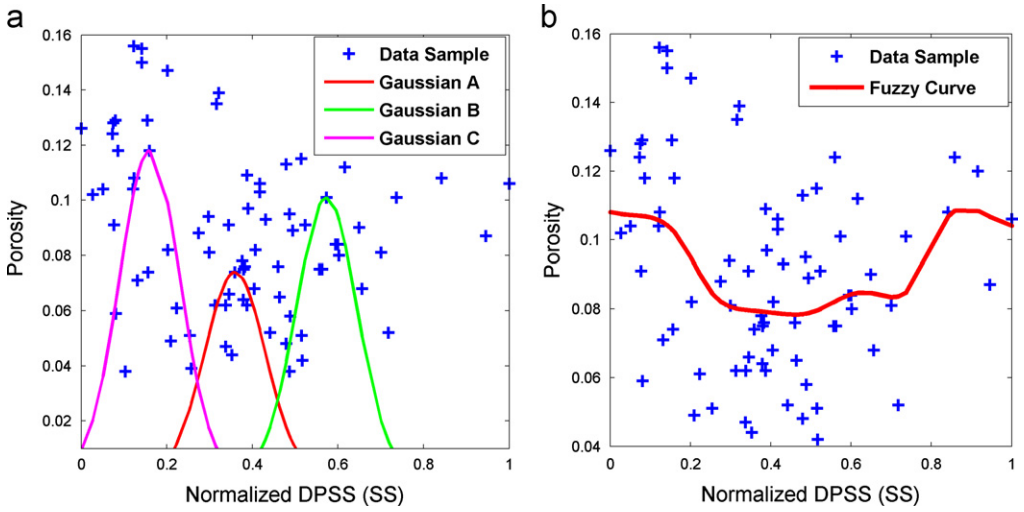

Fig. 2. (a) Gaussian fuzzy membership functions in the DPSS-porosity space and (b) fuzzy curve of DPSS to porosity.

In the following paragraphs and Fig. 3, we summarize the application of FR in the selection of neural inputs for reservoir characterization.

Assume that the original well log contains n candidate variables, $X=\{x_i \mid i=1,2,\ldots n\}$, and the target reservoir property is y. FR works as illustrated in Fig. 3.

Fig. 3 shows the pseudo code of FR in the FR-neural reservoir characterization framework. It uses three inputs: $X=\{x_i \mid i=1,2,\ldots n\}$ containing n well log variables, y as the target reservoir property, and $\alpha\%$ as the elimination threshold in the FC process. FR starts with the FC ranking n candidate well log variables in the candidate list (CL). In line 5, the variable with the smallest $MSEc_i$ is selected as the most important variable, denoted by $\bar{x}$, and is added to the selected list (SX). Meanwhile, the $\alpha\%$ variables in CL with the highest $MSEc_i$ are eliminated.

After the FC step, $\bar{x}$, is paired with each of the remaining $(1-\alpha\%)$ $n-1$ candidate variables in the CL, and FS is applied to rank them. In line 11, the variable with the highest $MSEs_{ij}$ is eliminated from the CL as a dependent variable; and, the variable with the smallest $MSEs_{ij}$ is selected, added to SX, and assigned to $\bar{x}$. Using $\bar{x}$, the FS process is repeated until the number of variables in the CL is smaller than 2. In the end, the variables in the SX form the final selected results and are returned and used as neural inputs in the subsequent pattern recognition step.

3.2. Pattern recognition stage

The pattern recognition step implements the MLP neural network to simulate the complex relationship between the selected well log variables and the target reservoir property. The assumption for reservoir characterization using an MLP is that the relationship between the well log variables and a reservoir property is correctly represented by the data. After the FR step, representative training data for reservoir characterization are selected as neural inputs, helping to improve the pattern recognition performance of the MLP.

The MLP has one input layer, one hidden layer and one output layer (as shown in Fig. 4). Neurons in each layer are fully connected with the ones in the adjacent layers with weighted links. Each neuron applies an activation function that processes the weighted sum results and transfers to others. As a supervising learning mechanism, there is a desired target value for each pair of neural inputs. For example, the neural inputs can be the values of the selected well log variables at a certain depth interval, and the target value can be the core analysis data at the same depth interval.

As shown in Fig. 4, the difference between the network output and the desired target generates an error signal, which is transferred back to the network modifying the link weights via back propagation algorithms (Haykin, 1999). After being trained by a number of training data, the MLP cannot only generate the reservoir property very closely to the desired value for training samples, but can predict the desired target for unseen neural inputs.

Fuzzy Ranking ($X, y, \alpha\%$)

Input: (1) Candidate well log variables $X=\{x_i i=1,2,\ldots n\}$
(2) Target reservoir property y
(3) Elimination threshold $\alpha\%$ in the Fuzzy Curve step
Output: Selected well log variables $SX=\{x_s \mid s=1,2,\ldots l, \text{and} l \le n\}$

01: Initialize candidate list $CL=X$and selected list $SX=null$;
/* **Fuzzy Curve Step** */
02: **for** each variable x_i $CL \in$
03: Compute the fuzzy curve $ci(x_i)$ by Eq. (4);
04: Compute $MSEc_i$ by Eq. (5);
05: $\bar{x}$ =the variable in $CL$with the smallest $MSEc_i$, and add $x$into SX;
06: Remove $\bar{x}$and $\alpha\%$ variables with highest $MSEc_i$ from CL;
/* **Fuzzy Surface Step** */
07: **while**(SizeOf (CL) > 1) **do**
08: {**for**each variable $x_i CL \in$
09: Compute the fuzzy surface $s_{ij}(x_i,x)$ by Eq. (6);
10: Compute $MSEs_{ij}$ by Eq. (7);
11: $\bar{x}$ =the variable in $CL$with the smallest $MSEs_{ij}$, and add x into SX;
12: Remove $\bar{x}$ and the variablewith the highest $MSEs_{ij}$from CL;}
13: **if** ($CLnull$)$\neq$
14: Add CL into SX;
15: **return** SX;

Fig. 3. Pseudo code of fuzzy ranking for well log variable selection.

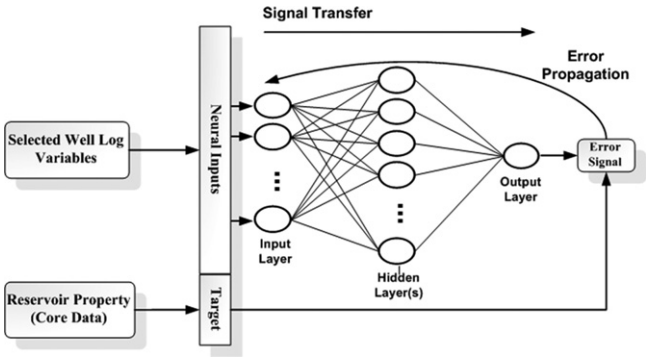

Fig. 4. The multilayer perceptron (MLP) model in the pattern recognition step for reservoir characterization.

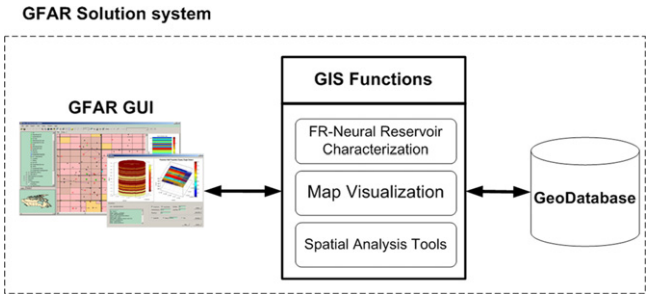

Fig. 5. GFAR solution system architecture.

4. GFAR solution

The petroleum industry is now facing the challenge of managing and processing large amounts of field data. A GIS can store these field data in a geodatabase and visualize and manage the data geographically as part of a well map. We have designed and implemented a *G*IS-based *F*R *A*NN *r*eservoir characterization solution, called the GFAR solution.

Fig. 5 shows the architecture of the GFAR solution, which consists of three parts: a geodatabase, GIS functions and the GFAR graphical user interface (GUI). The main GIS was created using C# and ESRI ArcObjects (ESRI ArcObjects Webpage, 2012), which have the ability to retrieve data from diverse data formats (*.mxd map file, *.lyr layer file, *.shp shape file) and visualize them on a map. We describe each component in detail in the following subsections.

4.1. Geodatabase

The geodatabase contains spatial data and non-spatial data. The spatial data includes the spatial information of petroleum wells, such as the geographical coordinates of wellheads. Non-spatial data include the unique well identification (UWI), well name, well type, recovery date, and owners. The digital well log and core analysis data are also part of the non-spatial data. The non-spatial data is related to the spatial data with the primary key of UWI.

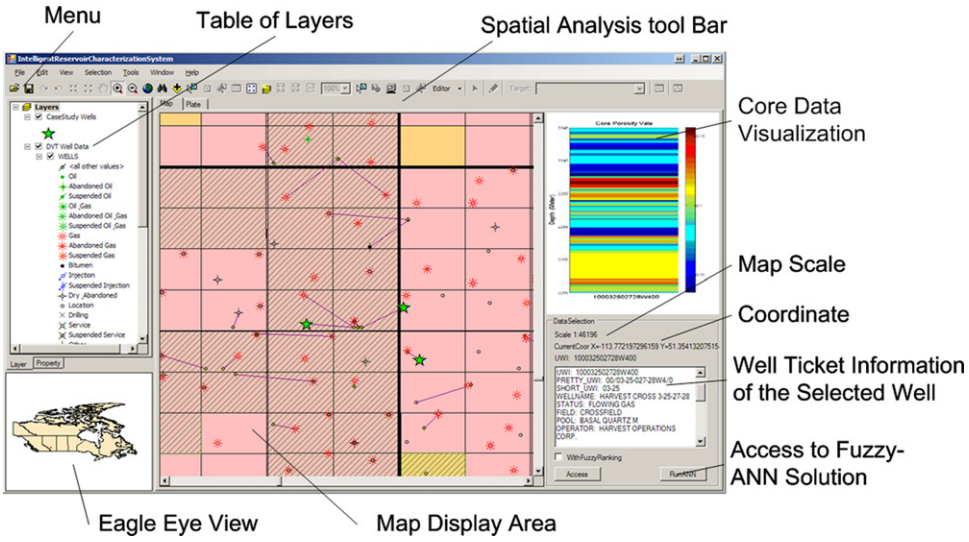

Fig. 6. Graphical user interface of the GFAR solution system.

4.2. GFAR interfaces and functions

As shown in Fig. 6, the interfaces of the GFAR solution include a map area, bird's eye view, layer table, menu, toolbar and well information viewer. These interfaces provide users with access not only to geospatial operational tools but also to reservoir characterization functions. The geospatial operation functions include map visualization, map navigation, feature selection, and distance measurement. For example, as shown in Fig. 6, the map visualization component visualizes petroleum wells and other geographical information, such as provincial boundaries and main cities.

The system also provides a reservoir characterization function, where the system can query the corresponding digital well log and core analysis data from the geological database and then perform reservoir characterization.

We conducted a detailed case study using the FR-neural framework in the GFAR solution system, as discussed in Section 5.

5. Case study

5.1. Data description

The proposed FR-neural framework for reservoir characterization is demonstrated using three petroleum wells in southwestern Alberta, Canada. As shown in Fig. 6, the three case-study wells, which are represented by pentagrams, are visualized on the well map area in the GFAR Solution system. The three wells are denoted as W-1, W-2 and W-3.

The corresponding digital well log data and core analysis data were extracted from the geodatabase. Table 1 shows the overall data description for the three wells. Each well has 20–23 well log variables serving as candidate neural inputs to the MLP, with the porosity values from the core analysis data serving as the target data. For instance, W-1 had 21 well log variables, each of which ranged from 278.4 m to 2329.8 m below the surface. W-1 also had 79 data samples of porosity data, ranging from 2240.0 to 2265.4 m below the surface.

W-1 is used as an example in this paper to illustrate the use of the proposed FR-neural framework. Table 2 shows that W-1 had 21 well log variables organized into seven categories based on the well logging sensors used. Note that not all candidate well log variables had the same degree of relationship to the target reservoir porosity; therefore, a subset of well log variables correlated to the porosity needed to be selected.

Table 1
Overall data description.

Well ID	Candidate inputs: all well log variables		Target data: porosity (core data)	
	No. of variables	Depth (unit: meter)	No. of data samples	Depth (unit: meter)
W-1	21	278.4–2329.8	79	2240.0–2265.4
W-2	23	291.8–2199.2	76	2109.0–2135.6
W-3	20	2126.0–2333.2	76	2232.0–2254.3

Table 2
Well log variables list for W-1.

Category	Well log variables
Resistivity logs	AF20, AF30, AF90, AT90, AT9C
Caliper logs	CAL2, HCAL
Sonic logs	DPDL, DPLS, DPSS
Thermal logs	CFTC, CNTC
Neutron logs	NPDL, NPLS, NPSS, GR, GDEV
Spontaneous potential	SP
Density logs	HDRA, PEFZ, RHOZ

5.2. Well Log Variables selection

To identify the representative data from W-1's well log, 21 well log variables and core porosity records, with the same depth intervals, were paired together. This formed a data space (x_i, y), $i=1,2,\ldots,21$, where x_i denotes for the 21 well log variables and y denotes for the target reservoir porosity.

The first step was the identification of a subset of well log variables containing the most useful information for characterizing porosity. The FC ranked all 21 candidate inputs along with ascending $MSEc_i$. Parameter b in Eq. (4) was empirically chosen as 0.08. As listed in Table 3, the FC ranked results show that CAL2 had the smallest $MSEc_i$ (0.78725) and, thus, was ranked as the most important well log variable for the target reservoir porosity. In contrast, HCAL had the largest $MSEc_i$, thereby ranking as the least important well log variable.

Visual observations of fuzzy curves can also be indicative of the different levels of connection of the candidate inputs to the target reservoir porosity. As previously mentioned, a flatter curve means that the corresponding variable contains less or more random

Table 3
Fuzzy curve (FC) ranking result.

Candidate well log	$MSEc_i$	Rank	Candidate well log	$MSEc_i$	Rank	Candidate well log	$MSEc_i$	Rank
CAL2	0.787250	1	GR	0.842015	8	GDEV	0.875072	15
DPSS	0.796621	2	CNTC	0.842649	9	PEFZ	0.882191	16
DPDL	0.796866	3	SP	0.848234	10	AT9C	0.886251	17
RHOZ	0.796887	4	CFTC	0.858982	11	HDRA	0.904673	18
DPLS	0.796922	5	NPSS	0.870224	12	AF30	0.908325	19
AF90	0.804379	6	NPDL	0.871355	13	AF20	0.941702	20
AT90	0.841790	7	NPLS	0.872583	14	HCAL	0.951503	21

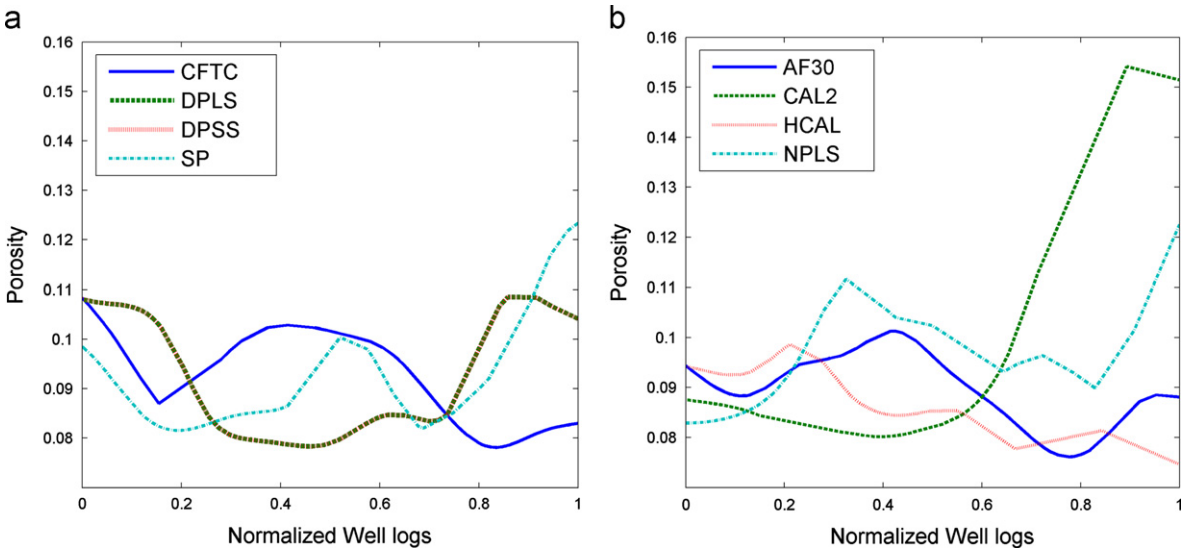

Fig. 7. Fuzzy curves for partial well log variables: (a) AF30, CAL2, HCAL, NPLS; (b) CFTC, DPLS, DPSS, SP.

information for the target output. Fig. 7 shows fuzzy curves for partial well log variables. HCAL, as shown in Fig. 7(b), had the flattest fuzzy curve, again demonstrating that it contained the least information of all variables.

If the elimination threshold for the FC was set at 30%, the last seven variables – HCAL, AF20, AF30, HDRA, AT9C, PEFZ and GDEV – should be removed from candidate inputs.

In the next step, the FS was implemented to remove the highly dependent variables from the candidate well log variables.

In the FC step, CAL2 was selected as the most important well log variable for the porosity characterization. Therefore, CAL2 was used as the reference variable in the FS analysis, where the $MSEs_{ij}$ was calculated between CAL2 and each of the remaining 13 candidate inputs. Table 4 shows the first iteration result with ascending $MSEs_{ij}$ from the FS.

From Table 4, the CAL2/GR pair had the minimum $MSEs_{ij}$; thus, GR was identified as the second important variable. GR was the recorded signal via the gamma ray well logging tool and had minimum dependence with CAL2 from the caliper log; therefore, the paired well log variables of CAL2 and GR were selected. On the other hand, AF90 was discarded, because the CAL2/AF90 pair showed the highest $MSEs_{ij}$ value. After the first iteration, 11 candidate inputs remained. In the second iteration, GR was used as the reference variable to evaluate the remaining 11 inputs.

Table 5 shows the results after the FS step, which stopped after seven iterations when only one variable was left in the candidate list. Ultimately, eight variables were selected, 35% of the 21 original well log variables. These formed the final selection result.

Fig. 8 shows partial well log variables versus depth (from 1400 to 1600 m below the surface) for W-1, and the well log variables

Table 4
The first iteration FS ranking result.

Candidate well log variables	$MSEs_{ij}$	Rank	Candidate well log variables	$MSEs_{ij}$	Rank
(CAL2, GR)	0.455055	1	(CAL2, CNTC)	0.497867	8
(CAL2, CFTC)	0.466716	2	(CAL2, NPSS)	0.521086	9
(CAL2, SP)	0.467368	3	(CAL2, NPDL)	0.523705	10
(CAL2, DPSS)	0.481365	4	(CAL2, NPLS)	0.525688	11
(CAL2, DPDL)	0.481389	5	(CAL2, AT90)	0.533946	12
(CAL2, RHOZ)	0.481453	6	(CAL2, AF90)	0.560741	13
(CAL2, DPLS)	0.481508	7			

from the same category were paired together. Visual observation shows that the well log variables from the same category were highly correlated. For example, high correlation could be found for the pairs of DPLS/DPSS and CFTC/CNTC. After the FS, DPLS had been removed from the DPLS/DPSS pair, and CNTC has been removed from the CFTC/CNTC pair.

The eight selected well log variables covered all seven categories shown in Table 2, which indicated that all seven categories played roles in characterizing reservoir porosity. Caliper log, gamma ray and sonic log variables are always empirically selected when using statistical correlations for porosity characterization. These variables were also selected in the FR step, showing the success of FR in feature well log variable selection.

5.3. Reservoir porosity characterization

Eight well log variables from the previous FR step served as neural inputs to the MLP in the pattern recognition step, and the

Table 5
Fuzzy surface (FS) ranking results and final selection result.

Iteration	Reference variable	Ranked sequence by ascending $MSEs_{ij}$	Selected variables	Eliminated variables
2	GR	AT90, DPSS, DPDL, DPLS, RHOZ, CNTC, SP, CFTC, NPSS, NPDL, NPLS	CAL2, GR, AT90	AF90, NPLS
3	AT90	CFTC, CNTC, RHOZ, DPSS, DPDL, DPLS, NPSS, SP, NPDL	CAL2, GR, AT90, CFTC	AF90, NPLS, NPDL
4	CFTC	SP, DPDL, DPSS, DPLS, RHOZ, NPSS, CNTC	CAL2, GR, AT90, CFTC, SP	AF90, NPLS, NPDL, CNTC
5	SP	RHOZ, DPSS (SS), DPLS, DPDL, NPSS	CAL2, GR, AT90, CFTC, SP, RHOZ	AF90, NPLS, NPDL, CNTC, NPSS
6	RHOZ	DPDL, DPSS, DPLS	CAL2, GR, AT90, CFTC, SP, RHOZ, DPDL	AF90, NPLS, NPDL, CNTC, NPSS, DPLS
7	DPDL	DPSS	CAL2, GR, AT90, CFTC, SP, RHOZ, DPDL, DPSS	AF90, NPLS, NPDL, CNTC, NPSS, DPLS

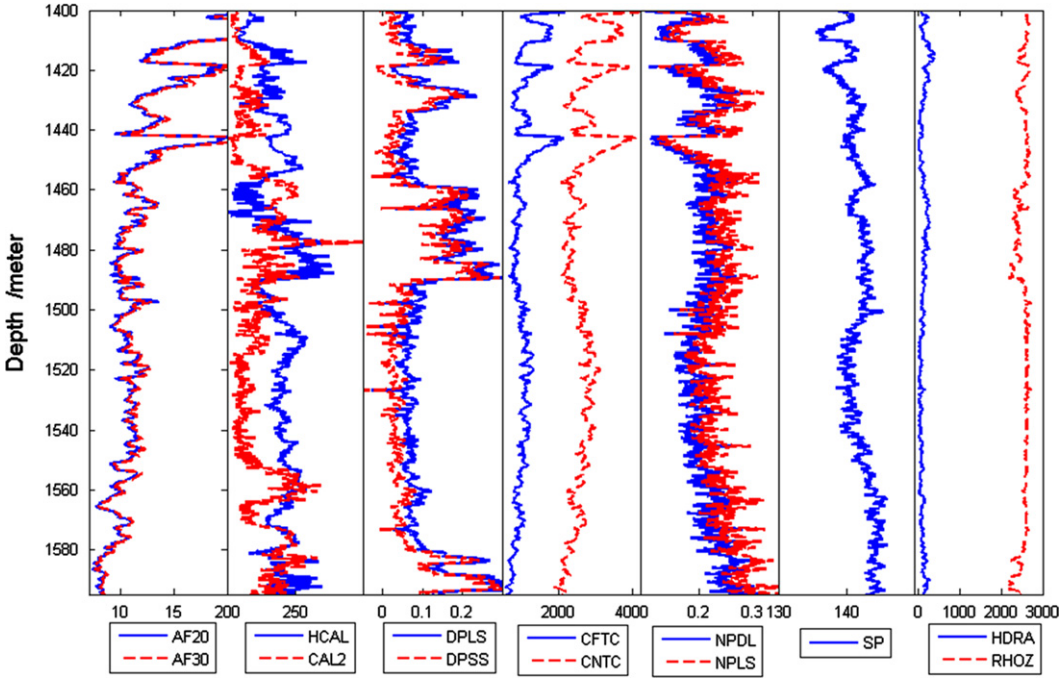

Fig. 8. Partial well log variables versus depth. Each graph shows the comparison between two variables from the same category shown in Table 2.

porosity values in the core analysis data served as the target data for the MLP's supervised learning process. Data were randomly separated into training data (60%), validation data (20%) and test data (20%).

The input layer was automatically initialized with the eight feature well log variables. Ten hidden neurons constructed the hidden layer of the MLP, and the output layer included only one neuron, in terms of the target reservoir porosity. Transfer functions for the input, hidden and output layers were tan-sigmoid, tan-sigmoid and log-sigmoid, respectively. The designed MLP was trained repeatedly (10 times) using the Levenberg–Marquart algorithm (Marquardt, 1964; Chen et al., 2003), and the final neural network was the one with the highest value correlation coefficient, R^2, of the test data.

Fig. 9 shows the final results for the porosity characterization from the MLP using the well log variables selected by FR, where the round points denote the training and validation samples and the square points denote the test samples. In Fig. 9(a), the curve is the estimated reservoir porosity from the MLP. Note that, all the training, validation and test data match well with the estimated porosity curve. Fig. 9(b) evaluates the results with the linear regression analysis, which shows the R^2 for the test data was 0.9504. Fig. 9 demonstrates that the proposed method is capable of predicting accurate porosity values given unseen well log data.

5.4. Performance comparison

To evaluate the efficiency of representative well log data selection using FR, the reservoir characterization results from the MLP using FR results were compared with results from four other control neural inputs, namely random selection, empirical selection, no selection, and principle component analysis (PCA). The first control group of neural inputs contained eight randomly selected well log variables, and the MLP network structure was 8–10-1 (i.e., eight variables (neurons) in the input layer, 10 neurons in the hidden layer and one neuron in the output layer), the same as used in the FR case.

The second group used the empirical selection result from Helle et al. (2001), where three well log variables selected from the sonic, density and resistivity categories served as neural inputs. The MLP network structure was 3–7-1, which was consistent with the work in Helle et al. (2001). In the third group (no selection), all 21 well log variables from W-1 were imported into the MLP as neural inputs. The network structure was 21–14-1, which included four extra hidden neurons due to the augmentation of the input data.

For the fourth control group of neural inputs, PCA was applied to analyze and select well log data. All 21 well log variables were transformed to the space defined by their principle components, and eight transformed variables with a variance higher than 0.16 were

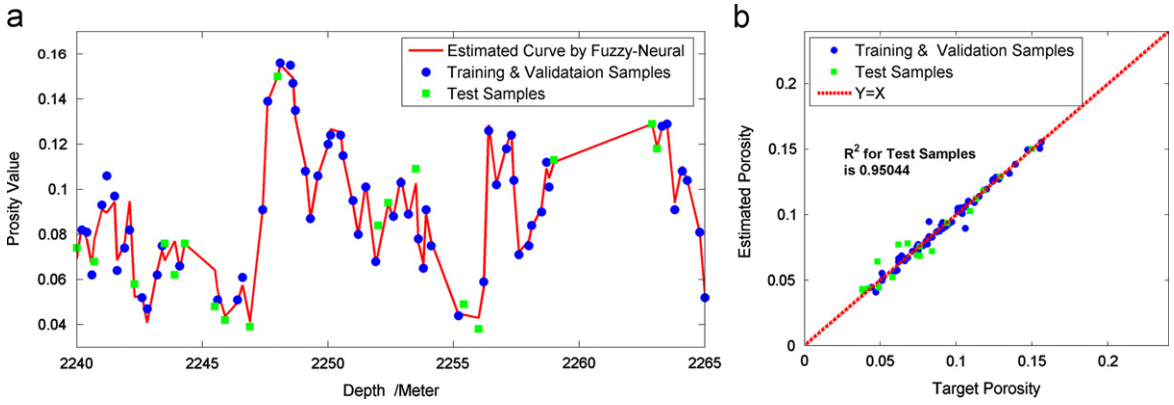

Fig. 9. (a) Estimated and core porosity values from the MLP with depth; (b) cross plot of estimated and core porosity values.

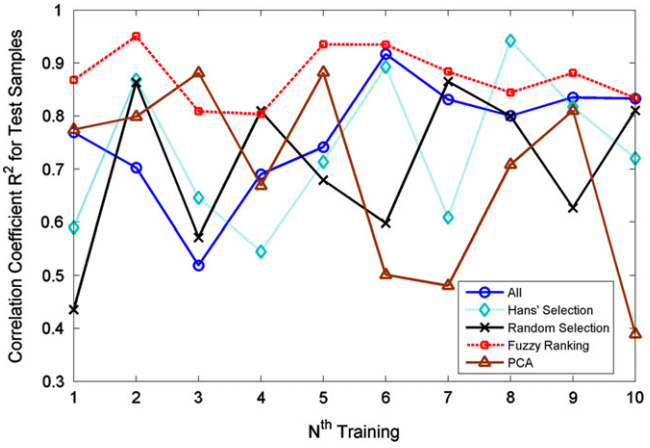

Fig. 10. Comparison of R^2 on test samples for MLPs using different neural inputs: fuzzy ranking selection, all well log variables, Helle et al.'s selection, and random selection.

Table 6
Results comparison among MLPs using neural inputs from four different methods.

Neural input selection method	Fuzzy ranking	Random selection	Helle et al.'s selection	All variables	PCA
Best R^2	**0.9504**	0.8762	0.9437	0.9108	0.9171
Avg R^2	**0.8746**	0.7058	0.7344	0.7639	0.7733
Worst R^2	**0.8037**	0.4350	0.5451	0.5188	0.5198

used as neural inputs. With the same number of neural inputs, the network structure was the same as the FR case (8–10-1).

In order to make a consistent comparison, this experiment kept the other configurations for the MLP neural networks, including the training algorithm, transfer function and training parameters, the same. The four MLP neural networks with different inputs were repeatedly trained (10 times), and the results are shown in Fig. 10.

Fig. 10 compares the predication accuracy of the five MLPs. Table 6 summarizes the comparison results. Several observations can be made. First, the overall predication accuracy of the MLP using FR results surpassed the other four methods. As shown in Table 6, the average R^2 of the test samples for the MLPs, using FR, PCA, all well log variables, Helle et al.'s selection, and random selection, were 0.8746, 0.7733, 0.7639, 0.7344 and 0.7058, respectively.

Second, the prediction results of the MLP using FR were more stable, since there was only a low chance of generating a very poor predication result, e.g., the worst R^2 for the MLP using FR was 0.8037. Therefore, the elimination of well log variables in the FR step does increase the MLP's system stability and prediction accuracy.

Table 7
Summarized results for three wells.

UWI	Selected/original well log variables	MLP structure	Avg train time (s)	Best R^2 for test samples
W-1	8/21	8–10-1	4.1462	0.9504
W-2	9/23	9–11-1	5.1892	0.9190
W-3	7/20	7–9-1	4.3908	0.9216

Following the same procedure as used on W-1 with the proposed FR-neural framework, the porosity characterization results for W-2 and W-3 were calculated, which are summarized in Table 7. Note that the number of the selected well log variables through the FR step account for approximately only 40% of the original well log variables, significantly decreasing the computational cost of the MLP. The MLP repeated the training processes 10 times for each well, and the best R^2 for the test samples was calculated for each group. The overall R^2 of these test samples was above 0.9, which shows the prediction accuracy of the proposed FR-neural reservoir characterization framework.

6. Conclusion

In this paper, a novel FR-neural framework, including FR and pattern recognition stages, is proposed for reservoir characterization. The FR step selects the representative well log variables for the target reservoir property characterization. An MLP neural network simulates the complex correlation from the well log data to the target reservoir property. After proper training, the MLP predicts the target reservoir property based on new well log data. Experiments for porosity characterization show that the proposed FR-neural framework predicted the target reservoir property more accurately and stably than previous methods.

The proposed FR-neural framework is suitable for solving most reservoir characterization problems. In the future, we will apply the proposed framework to different reservoir characterization problems. Other field data, like seismic data, will be used for the purpose of increasing the modeling capability. In addition, we will also test other ANN models, such as a radial basis function (RBF) neural network (Kanevski et al., 2009), in characterizing reservoir properties. The cross-validation method will also be applied to test the generalization ability of the trained neural network model in accurately predicting reservoir properties.

We also integrated the proposed reservoir characterization method with a GIS and designed the GFAR solution system, which

provides industrial agencies with a user-friendly reservoir characterization interface. The GFAR solution can be extendable with 3D visualization of the reservoir characterization results.

References

Al-Bulushi, N., King, P.R., Blunt, M.J., Kraaijveld, M., 2009. Development of artificial neural network models for predicting water saturation and fluid distribution. Journal of Petroleum Science and Engineering 68, 197–208.

Aminian, K., Ameri, S., 2005. Application of artificial neural networks for reservoir characterization with limited data. Journal of Petroleum Science and Engineering 49, 212–222.

Brock, J., 1986. Applied Open-Hole Log Analysis. Gulf Pub. Co. 284pp.

Chang, K.T., 2008. Introduction to Geographic Information System, fourth ed McGrawHill 450 pp.

Chen, T.C., Han, D.J., Au, F.T.K., Tham, L.G., 2003. Acceleration of Levenberg–Marquardt training of neural networks with variable decay rate. IEEE Transaction in Neural Networks 3, 1873–1878.

Eidsvik, J., Mukerji, T., Switzer, P., 2004. Estimation of geological attributes from a well log: an application of hidden Markov chains. Mathematical Geology 36, 379–397.

Ellis, D.V., Singer, J.M., 2007. Well Logging For Earth Scientists, second ed. Springer 692pp.

El-Sebakhy, E., Abdulraheem, A., Ahmed, M., Al-Majed, A., Awal, M., Azzedin, F., Raharja, P., 2012. Functional networks as a novel approach for prediction of permeability in a carbonate reservoir. Expert Systems with Applications 39 (12), 10359–10375.

ESRI ArcObjects Webpage, URL: ⟨http://edndoc.esri.com/arcobjects/9.2/welcome.htm⟩ (accessed 12 Febraury, 2012).

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, second ed. Prentice-Hall 842pp.

Helle, H.B., Bhatt, A., Ursin, B., 2001. Porosity and permeability predication from wireline logs using artificial neural networks: a north sea case study. Geophysical Prospecting 49, 431–444.

Kanevski, M., Pozdnoukhov, A., Timonin, V., 2009. Machine Learning for Spatial Environmental Data, Theory, Applications and Software. EPFL Press 377 pp.

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

Lin, Y.H., Cunningham, G.A., Coggeshall, S.V., 1996. Input variable identification—fuzzy curves and fuzzy surfaces. Fuzzy Sets and Systems 82, 65–71.

Lin, Y.H., Cunningham, G.A., Coggeshall, S.V., Jones, R.D., 1998. Nonlinear system input structure identification: two stage fuzzy curves and surfaces. IEEE Transactions on Systems Man and Cybernetics. Part A—Systems and Humans 28, 678–684.

Marquardt, D.W., 1964. An algorithm for least-squares estimation of non-linear parameters. Journal of the Society for Industrial and Applied Mathematics 11, 431–441.

Mohaghegh, S., 2005. Recent developments in application of artificial intelligence in petroleum engineering. Journal of Petroleum Technology 57, 86–91.

Mohaghegh, S., 2000. Virtual-intelligence applications in petroleum engineering. Part I—artificial neural networks. Journal of Petroleum Technology 52, 64–72.

Mohaghegh, S., Arefi, R., Ameri, S., Aminiand, K., Nutter, R., 1996. Petroleum reservoir characterization with the aid of artificial neural networks. Journal of Petroleum Science and Engineering 16, 263–274.

Olatunji, S.O., Selamat, A., Abdulraheem, A., 2011a. Modeling the permeability of carbonate reservoir using type-2 fuzzy logic systems. Computers in Industry 62 (2), 147–163.

Olatunji, S.O., Selamat, A., Abdulraheem, A., 2011b. Predicting correlations properties of crude oil system using type-2 fuzzy logic systems. Expert Systems with Applications 38 (9), 10911–10922.

Wong, K.W., Gedeon, T.D., 2001. Fuzzy rule interpolation for multidimensional input space with petroleum engineering application. In: Proceedings of the Ninth IFSA World Congress, Vancouver, Canada, pp. 2470–2475.

Wong, P.M., Gedeon, T.D., Taggart, I.J., 1995. An improved technique in porosity prediction—a neural-network approach. IEEE Transactions on Geoscience and Remote Sensing 33, 971–980.