SPE-180439-MS

Pseudo Density Log Generation Using Artificial Neural Network

Wennan Long, University of Southern California; Di Chai, University of Kansas; Fred Aminzadeh, University of Southern California

This paper was prepared for presentation at the SPE Western Regional Meeting held in Anchorage, Alaska, USA, 23–26 May 2016.

Abstract

Reservoir characterization is often a demanding and complicated task due to the nonlinear and heterogeneous physical properties of the subsurface environment. Those issues can be overcome accurately and efficiently by the use of computer-based intelligence methods such as Neural Network, Fuzzy Logic and Genetic Algorithm. This paper will describe how one integrates a comprehensive methodology of data mining techniques and artificial neural network (ANN) in reservoir petrophysics properties prediction and regeneration. Density log, which acts as a powerful tool in petrophysical properties indication, is often run over just a small portion of the well due to economic considerations, the borehole environment or operation difficulties. Furthermore, missing log data is common for old wells, and wells drilled by other companies. Working towards a resolution to these challenges, we will demonstrate successfully constructed automatic system which includes well logging data preprocessing, data mining technologies and ANN prediction. Based on one field case study, this methodology was proficient and stable in pseudo-density log generation.

Introduction

In reservoir engineering, well log involves three significant aspects — reservoir identification, oil-in-place (OIP), and recoverable oil in place. Each is a key element of the reservoir description; efficiency is lost when the well log information is missing or incomplete. The classic method of handling this problem is to re-log the well using various logging tools. However, most openhole log measurements are very difficult to obtain once a well has been cased, which leads to unnecessary economic issues, especially when shutting down the well is inevitable. Also, any well log data mismatch, (the discrepancy between the original and re-logged data), due to petrophysical properties changes, can cause serious problems when estimating reservoir characteristics, such as porosity and permeability from those well logs. Consequently, several alternative technologies have been applied to generate pseudo well logs since 1980.

ANN, a machine-learning model that provides the potential to establish a multidimensional, nonlinear and complex model, can be a powerful tool with which to analyze experimental, industrial and field data (Aminzadeh, 1996). The task of predicting petrophysical properties such as porosity and permeability dramatically affect the results of reservoir characterization. Using multivariate statistical analysis and nonparametric regression, electrofacies were characterized and permeability was predicted in carbonate

reservoirs (Lee et al., 1999). Associated with fuzzy logic, intelligent software was developed (Nikravesh and Aminzadeh, 2001a) that is able to identify the nonlinear relationship between seismic attributes and well logs. Distributed in a spatially nonuniform and nonlinear geologic manner, permeability and porosity are estimated from well log data using ANN, and validated using core data (Verma et al., 2012). In addition, the probabilistic-neural-network (PNN) facies analysis was used to predict facies from well log data, as it has the potential to delineate the nonlinear relationship between lithofacies and well log data (Tang et al., 2011). However, a high degree of unknown uncertainty somehow has a great impact on the predicted result. In order to quantify uncertainty, one study introduced Adaptive Neuro Fuzzy Inference System (ANFIS) to predict porosity and sand fraction value from well logs, and solidify its advantage over ANN (Chaki et al., 2013).

Because of strong sensitivity of input data in ANN, we propose to screen input. For example, different rocks have different petrophysical responses. It is crucial to find the optimal data from one well to build the model using ANN for pseudo well log generation of a target well. Thus, the pattern recognition is inevitable. Manual stratigraphic interpretation is regarded as one approach. However, it is both labor intensive and time consuming to identify the patterns of well logs. Data mining techniques is another applicable approach to automatically process data associated with nonlinearity by using a statistical method to discover the data patterns (Nikravesh and Aminzadeh, 2001b). One application in petrophysics is facies (or electrofacies) classification, which is widely used to divide well log data in order to obtain target information, such as porosity-permeability relation (Nikravesh et al., 1999). An adjunct to artificial intelligence, clustering analysis can determine electrofacies and categorize lithological profile quite efficiently (Hassibi and Ershaghi, 1996). Nikravesh and Aminzadeh (2003) explain how the reservoir characterization is solved by mining and fusing of reservoir data in an intelligent system built by soft computing techniques. Other applications, such as permeability estimation, have been easily proceeding under cluster analysis, multivariate analysis, classification trees and feedforward neural network (Kumar et al., 2006). For additional details on the use of neural networks and other soft computing techniques in the petroleum industry see Aminzadeh and de Groot, (2006).

Porosity, one of the more important petrophysical properties, can be obtained from density log. The essence of this paper is density log generation, so in this study, a three step approach was produced. First, we applied preprocessing the log data using standardization, dimension reduction (Principal Component Analysis). Second, we proposed clustering (Model Based Clustering) to recognize the specific pattern and interpret stratigraphic information. Finally, we chose a similar pattern as input to generate target pseudo-density log using ANN. A field containing eight wells was utilized as the case study to affirm and prove the stability, efficiency and accuracy of merging data mining with ANN.

Workflow & Methodology

The three step approach will be shown in a workflow (Figure 1). It comprises three components: data preprocessing, data mining and data postprocessing. Data preprocessing contains normalization and principal component analysis (PCA). In the data mining part, clustering using Gaussian mixture model (GMM) is applied to give a 2-D qualitative indication of prediction performance across two different wells in a block. Model based clustering (MBC) is then used to fully merge the data of two good correlated wells, selected from the previous step, into separate electrofacies, which have some inner correlations with lithofacies. If a large dataset is used for clustering, which is a time consuming process, the discriminant analysis (DA) is taken into account. The primary purpose of DA is to overcome the problem of long computation time and high memory demand when MBC is applied to a large dataset. In a comparison of electrofacies graphs between two wells for validation, the best related wells were chosen if similar trends were observed. Finally, in the data postprocessing (pairwise well prediction), back-propagation neural network (BPNN) is used to train and predict actual density log in two wells.

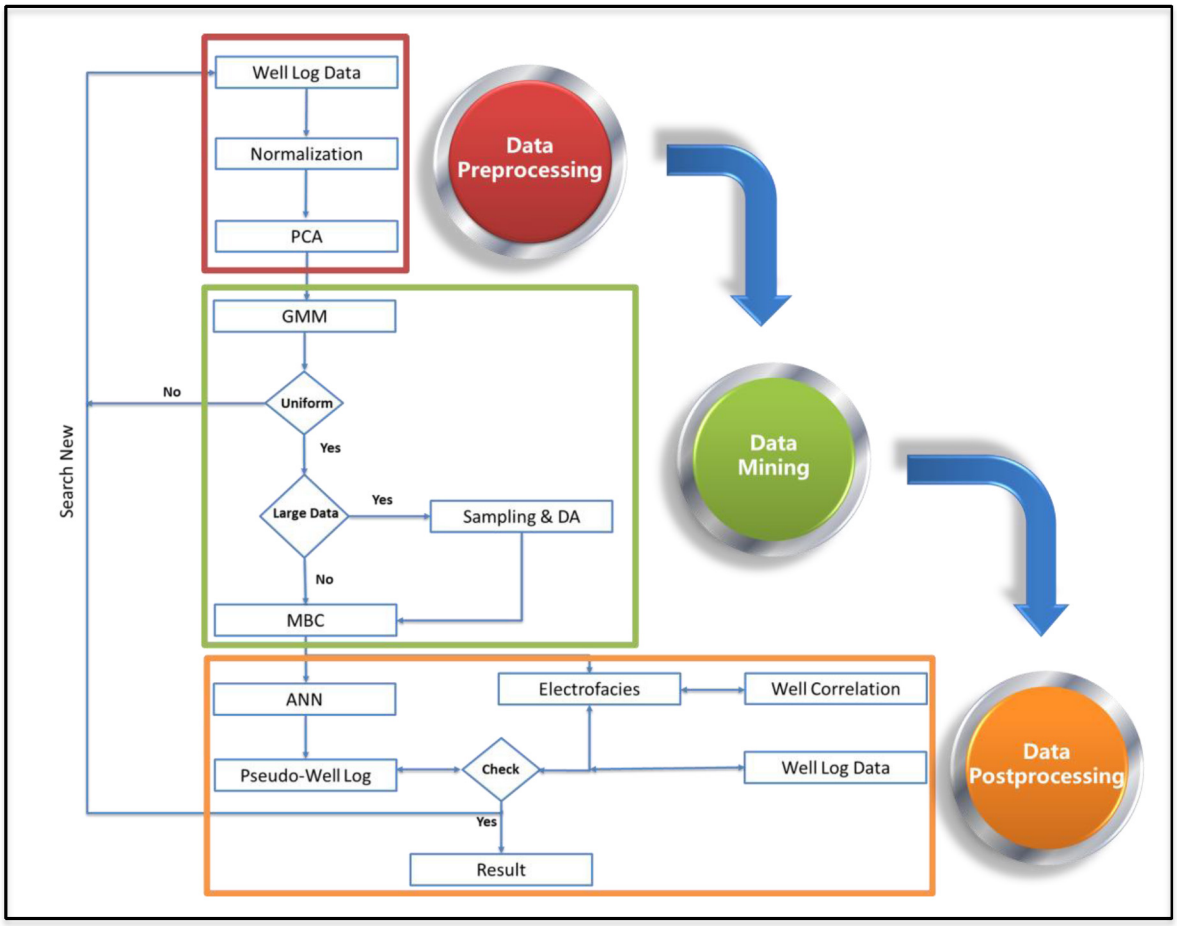

Figure 1—Systematic pseudo density log generation workflow

Well Log Data Preprocessing

Data preprocessing is the first stage of systematic pseudo-density log generation. It aims to normalize well log data and reduce the data size. It is of paramount importance based upon the facts that: (a) the quality of data has a significant impact on prediction results through ANN; and (b) the quantity of data will substantially increase the efficiency of the prediction.

Normalization

Well log data is constructed in a matrix form, whereby each row represents the depth record, and each column is the different type of well logs. Well log matrix ***(X)*** is a $m \times n$ dimension, and the number of rows m is much larger than that of the column n. Each well constructs one well log matrix and one field which has multiple wells constructs a big dataset. The initial step in the first stage is to normalize well log data. This normalization is necessary because different types of well log data have different units. For instance, the unit of Spontaneous Potential Log is millivolt, whereas the unit of Gamma Ray is API unit. Following is the normalization expression:

$$x_{i,norm} = \frac{x_i - \mu_i}{\sigma_i * \|x_i\|_\infty} \tag{1}$$

where $x_{i,norm}$ is the norm of attribute *i;* μ_i and σ_i are mean and standard deviation of attribute ***i;*** $\|x_i\|_\infty$ is L-infinity norm to scale well log data in [-1,1].

Principal Component Analysis

The principal component analysis is a statistical procedure that uses an orthogonal transformation to convert a set of well log vectors (attributes) of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PC) (Aminzadeh et al., 2000). Then, principal components constructed a new matrix which has much lower dimensions than the original. In this paper, we propose to use singular value decomposition (SVD). We consider the SVD of well log covariance matrix: $X^TX = U\Sigma V^T$, where Σ is a diagonal matrix $(r \times r)$ with the r non-null singular values ($\Sigma = diag(d_1,d_2, \ldots,d_r)$) sorted in decreasing order ($d_1 > d_2 > \ldots > d_r > 0$). The column vectors in $U_{n\times r}$ and in $V_{r\times n}$ are linearly independent of each other, which can be expressed as $U^TU = V^TV = I$. Hence, PCA allows the reduction of the dimensionality (number of the columns) of the well log data but retains most of the variability of that data. Hence, the principal components of well log matrix are generated based upon the following formula:

$$\boldsymbol{X_{PCA} = XV^T} \tag{2}$$

where X_{PCA} is $m\times r$ matrix and $r \leq n$. In the case study, we chose k number of principal components, where k is much smaller than r, and we have proved that k is an optimal number (Appendix). Figure 2 shows scatter plot of PC 1, PC 2 and PC3 from Well #1.

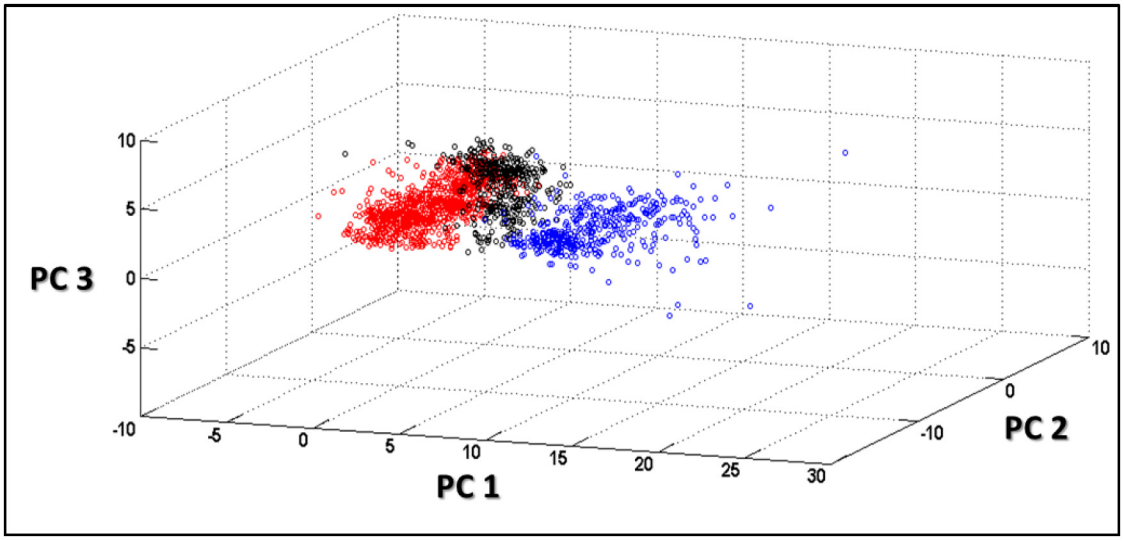

Figure 2—Scatter plot of pricinpal components of Well #1

Well Log Data Mining

Well logs can be described as a record of rock and formation properties against depth. They play an important role in understanding the petrophysical properties of the reservoir rocks. Well logs, however, are not a straightforward representation of formation, and the amount of well log data is usually extremely large. In addition, stratigraphic interpretation and classification are necessary to select appropriate wells for data postprocessing. Data mining is the computation process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems (Figure 3). It, therefore, develops into an ideal tool to take the place of conventional manual analysis, and becomes a keystone in the automatic pseudo-density log generation system.

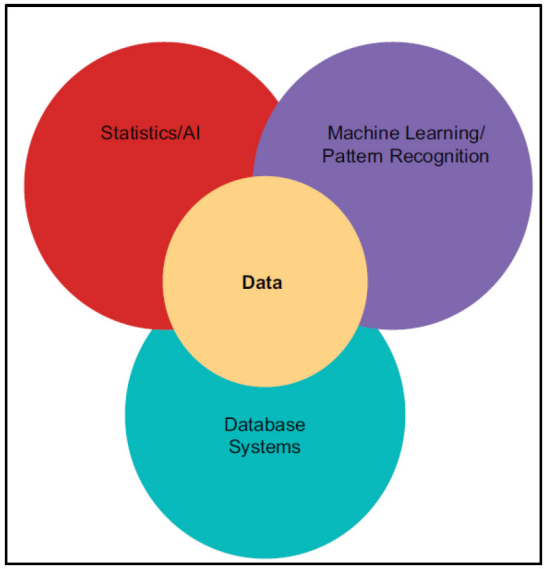

Figure 3—Data mining technologies

Lithofacies & Electrofacies

Lithofacies is a body of rock with specified characteristics. It is a distinctive rock unit that forms under certain conditions of sedimentation, reflecting a particular process or environment and petrophysical properties of that rock unit. Rock characteristics differ from one to another; therefore, different types of lithofacies have an internal different signal response on well logs. For example, gamma ray log of sandstone can reflect the clay content. If we use gamma ray log of sandstone to build a model, and subsequently use that model to predict gamma ray log of shale, severe mismatch problems arise that will ensure prediction failure. So it is of paramount importance to have correct lithofacies. Lithofacies identification is classified as three general approaches:

- Core data analysis
- Knowledge-based well log analysis by the expert system
- Electrofacies

Compared with the first and second methods, electrofacies analysis is relatively inexpensive and more efficient. The term ″electrofacies″ was first introduced by Serra and Abbott (1980). It has been defined as ″the set of log responses which characterize a bed (layer) and permits this to be distinguished from others″. Electrofacies are not equal to lithofacies, but they have a correlation with each other. Figure 4 illuminates this inner correlation. Well log data is determined by lithology to describe formation petrophysical properties. Then lithofacies are manually obtained by well log data based upon experience, while electrofacies are calculated based upon statistical analysis. Finally, electrofacies hereby have a correlation with lithofacies.

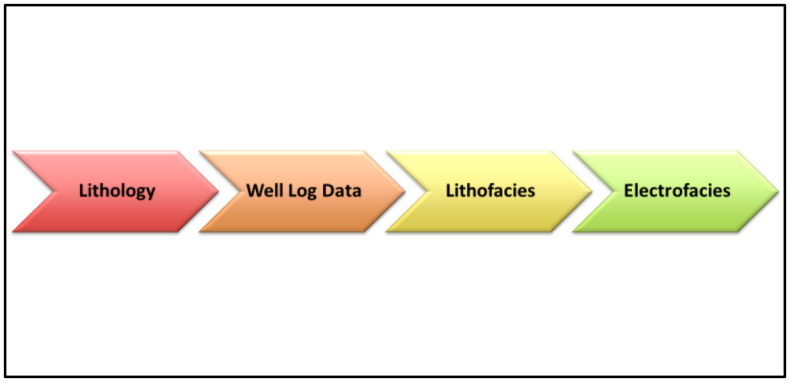

Figure 4—Correlation between lithofacies and electrofacies

Gaussian Mixture Model

We assume one well log matrix is one mixture model from the statistical point of view. A mixture model is a probabilistic model representing the presence of a subpopulation within an overall population. Here, the subpopulation is well log data from the specific type of rock. The mathematical expression of the mixture model is below:

$$\boldsymbol{P(x) = \pi_0 f_0(x) + \pi_1 f_1(x) + \cdots + \pi_j f_j(x)} \tag{3}$$

where *P(x)* is probability density function (PDF) of the overall population, *f(x)* is the PDF of subpopulation and *n* is the weight. Because different well log data from different types of rock may have different statistical properties, *f(x)* can fall into different PDF categories. However, in this paper, for simplicity, we automatically transform well log data into Gaussian distribution. When the data is converted into Gaussian distribution, the mixture model becomes Gaussian mixture model, which is a parametric probability density function represented as a weighted sum of Gaussian component densities. Thus, the *f(x)* is transformed into the following expression:

$$\boldsymbol{N\left(x|\mu_j^*, \Sigma_j^*\right) = \frac{exp\left\{-\frac{1}{2}\left(x-\mu_j^*\right)^T \Sigma_k^{*-1}\left(x-\mu_j^*\right)\right\}}{\sqrt{det\left(2\pi\Sigma_j^*\right)}}} \tag{4}$$

where μ^*_j and Σ^*_j are mean and covariance of subpopulation *j*. So the Gaussian mixture model expression is as follows:

$$\boldsymbol{P(x) = \pi_0 N(x|\mu_0^*, \Sigma_0^*) + \pi_1 N(x|\mu_1^*, \Sigma_1^*) + \cdots + \pi_j N\left(x|\mu_j^*, \Sigma_j^*\right)} \tag{5}$$

Equation (5) shows that the Gaussian mixture model has three unknowns. It is parameterized by the mean vectors $(\boldsymbol{\mu^*_j})$, covariance matrices $(\boldsymbol{\Sigma^*_j})$ and mixture weights $(\boldsymbol{\pi_j})$ from all component densities. We usually use the following symbol to collect these parameters together:

$$\boldsymbol{\theta = \{\mu_i^*, \Sigma_i^*, \pi_i\}} \qquad \boldsymbol{i = 1, 2, \ldots, j} \tag{6}$$

There are multiple ways to determine these parameters of the GMM, θ, which is the cornerstone on which to build the cluster for each different rock. In this paper, we propose to apply the expectation-maximization (EM) algorithm for estimating these parameters. The objective function of EM algorithm is the likelihood of the GMM given the training data. The well log data after PCA is both training data and observation data. For a sequence of T training vectors $X_{PCA} = \{x_1, x_2, \ldots, x_T\}$, the GMM likelihood can be written as the formula below:

$$\boldsymbol{P_\theta(x) = \sum_z P_\theta(x,z)} \tag{7}$$

where z is latent variables and P_θ is PDF of some unknown parameters θ. Therefore, the goal of likelihood maximization is shown as the following expression:

$$\boldsymbol{argmax\ P_\theta(x)} \tag{8}$$

However, equation (8) is usually difficult to maximize, but by applying numerical method, EM algorithm can give a relatively accurate approximation after several iterations. It consists of two steps:

$$\textbf{E-Step:} \qquad \boldsymbol{Q(\theta,\theta_t) = E_{\theta_t}\left(log\left(P_\theta(X,Z)\right)|X=x\right)} \tag{9}$$

$$\textbf{M-Step:} \qquad \boldsymbol{\theta_{t+1} \in argmax\, Q(\theta,\theta_t)} \tag{10}$$

where t is the time steps *(t = 0,1,2, . . ., convergence)*. Figure 5 shows the clustering results after using GMM. For visual convenience, we only chose the first and second principal components (PC1 and PC2) as inputs (observation) and overall three clusters as outputs. Thus, three different colors represent three clusters, but each color in a different figure may have a different meaning. The upper figure shows the clustering result of Well #1 and the one below shows the result of combined well log data from Well #1 and Well #2. Based on the shape of a single well and combined wells clustering results, the GMM enable us to have a qualitative indication of the similarity between the two wells. For example, Figure 6 includes all the clustering results. The first row shows single well clustering results, and the second row shows results for combined wells. In the Figure, A1 and B2 have a similar shape. It means that Well #1 has a higher probability of having a large number of the same cluster with Well #2. Then we can continue to the next stage. If the shape is obviously different, we should go back and pick another well. The main reason of the qualitative indication is that the first two principal components contain about thirty percent data information.

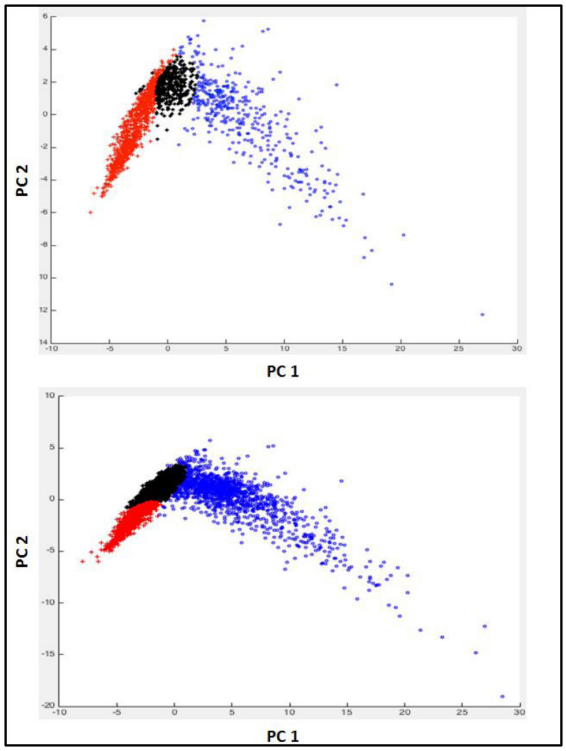

Figure 5—Well #1 GMM result and Well #1 & Well #2 GMM result

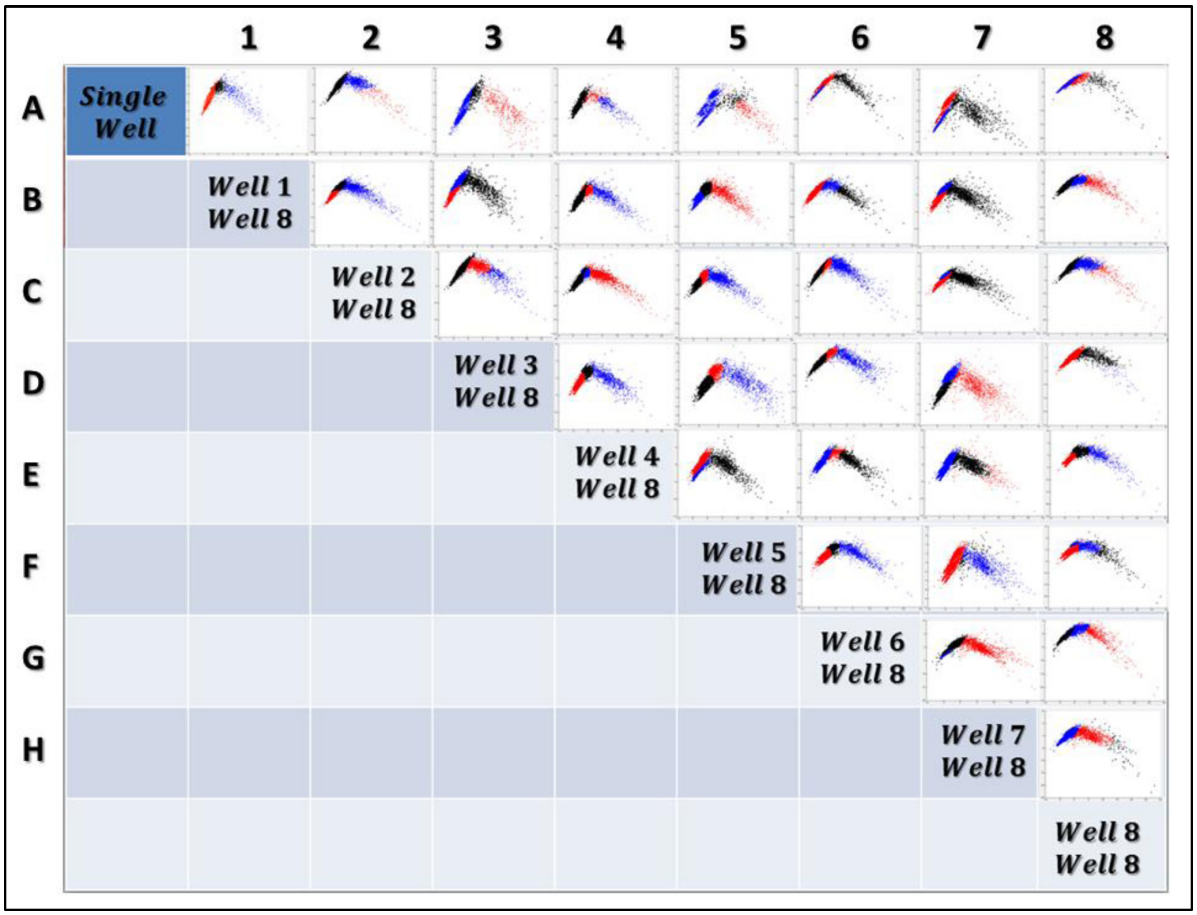

Figure 6—Eight wells gaussian mixture model clustering results

Model-Based Clustering

As with the GMM, the foundational assumption of model-based clustering is that the data is generated by a mixture of probability distributions (or mixture models) in which each part is a different cluster or model. In the GMM, the number of clusters has to be defined by the user. However, this might not be the case, because we scarcely have the luxury of accessing accurate rock properties before processing the well log data. So the easy way to set this imperative parameter in GMM is to add two components: (1) four basic models of the covariance matrix, and (2) the agglomerative algorithm. The new upgraded method is the main structure of model-based clustering.

Before applying these two components, we need to extend the likelihood function into multivariate form:

$$P_\theta(x) = \prod_{i=1}^{j} \prod_{h\in\Upsilon_i} (2\pi)^{-\frac{n}{2}} |\Sigma_i|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}(x_h-\mu_i)^T\Sigma_i^{-1}(x_h-\mu_i)\right\} \quad (11)$$

where γ_i is the set of data of observations that belongs to i^{th} group, j is the total number of clusters and n is the dimension of well log data matrix. Then we concentrate the log likelihood as:

$$log\left(P_\theta(x)\right) = -\frac{1}{2}\sum_{i=1}^{j}\left\{tr\left(W_k\Sigma_i^{-1}\right) + N_i log|\Sigma_i|\right\} - n * N * c \quad (12)$$

where $W_k = \sum_{h\in\Upsilon_i}(x_h-\mu_i)(x_h-\mu_i)^T$ is the cross-product matrix for the i^{th} group, N_i is total number of data in i^{th} group and c is constant which equal to 0.4.

The *four basic models* are scenarios which have four related criteria (Table 1). Based on the table, we have the following conclusions:

Table 1—Four basic models in MBC

Model Number	Coveriance	Model	Criterion
1	Spherical and Equal	$\boldsymbol{\Sigma_i = \sigma^2 I}$	$tr\left(\sum_{i=1}^{j} W_i\right)$
2	Spherical and Unequal	$\boldsymbol{\Sigma_i = \sigma_i^2 I}$	$\sum_{i=1}^{j} N_i W_i log\left[tr\left(\frac{W_i}{N_i}\right)\right]$
3	Ellipsoidal and Equal	$\boldsymbol{\Sigma_i = \Sigma}$	$\left\lvert\sum_{i=1}^{j} W_i\right\rvert$
4	Ellipsoidal and Unequal	$\boldsymbol{\Sigma_i = \Sigma_i}$	$\sum_{i=1}^{j} N_i W_i log\left\lvert\frac{W_i}{N_i}\right\rvert$

- The first model has diagonal and equal covariance matrices and same value in diagonal elements.
- The second model has diagonal covariance matrices but the same value in each diagonal element of the individual covariance matrix.
- The third model has equal covariance matrices which have non-zero off-diagonal elements.
- The fourth model's covariance matrices can vary among components.

The *agglomerative algorithm* is implemented by using single linkage, which is the similarity of the closest pair as the following expression:

$$\boldsymbol{Dist_{SL}}\left(\{\boldsymbol{p_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{P}}, \{\boldsymbol{q_i}\}_{\boldsymbol{j=1}}^{\boldsymbol{Q}}\right) = \boldsymbol{min}\|\boldsymbol{p_i} - \boldsymbol{q_i}\| \tag{13}$$

where $P = \sum p_i$ and $Q = \sum q_j$ are two different clusters or groups. If two clusters are close enough, the two will be merged.

Well Correlation

It enables us to do well correlation by generating electrofacies after applying model- based clustering. Before using MBC, it is necessary to change well log matrix **X** to combined well log matrix X*, which is the data from all interested wells. However, the dimension of matrix **X*** is much larger than that of **X** so it would be a serious problem for well correlation due to the low efficiency of expectation maximization algorithm. For this paper, we have proposed two steps to solve the problem: (1) sampling and (2) discriminant analysis. In the sampling part, the size of the sample $\boldsymbol{(n_{sample})}$ should be determined by following:

$$\boldsymbol{n_{sample}} = \frac{\boldsymbol{c_0 N_p}}{\boldsymbol{c_0 + (N_p - 1)}} \tag{14}$$

where N_p is the number of total population and c_0 is sampling constant which is equal to 384.16. Then, based on the stratigraphic property of the formation, we utilize systematic sampling method to select sample points. In the systematic random sampling - first, randomly choose the first item or subject from the defined population; then select proper interval ***(l)*** for sampling. The formula of the whole process is:

$$l = int(\frac{N_p}{n_{sample}}) \tag{15a}$$

$$L = \begin{cases} (1 : l), & random\ starting\ number\ i \\ i + nl, & n = 1, 2, \dots, int(\frac{N_p - i}{l}) \end{cases} \tag{15b}$$

After taking a sample from combined well log matrix **X***, we implement model-based clustering method to generate electrofacies from sampling data. For the rest of the dataset, we suggest using discriminant analysis to put data into the existing electrofacies generated from the previous step.

In terms of the discriminant analysis (DA), it is a classification method where groups or clusters from populations are known to be a priori and other new observations are classified into one of them based upon the measured characteristics. It assumes that different classes generate data based on different Gaussian distributions. For this study, we chose quadratic discriminant analysis, in which the covariance matrix can be different for each class. The classification is determined by:

$$\sum_i \neq \sum_j \tag{16a}$$

$$\delta_k(x) = arg\ \max_k[-\frac{1}{2}log|\sum_k| - \frac{1}{2}(x - \mu_k)^T\sum_k^{-1}(x - \mu_k) + log(\pi_k)] \tag{16b}$$

We will classify the sample into the groups that have the largest quadratic score function. Figure 7 shows the result of well correlation of eight wells. From the Figure, the total number of electrofacies is three. So we use three different colors to represent different electrofacies. With computer assistance, the efficiency of well correlation has substantially increased. To facilitate future improvement, we can use core analysis to calibrate well correlation results.

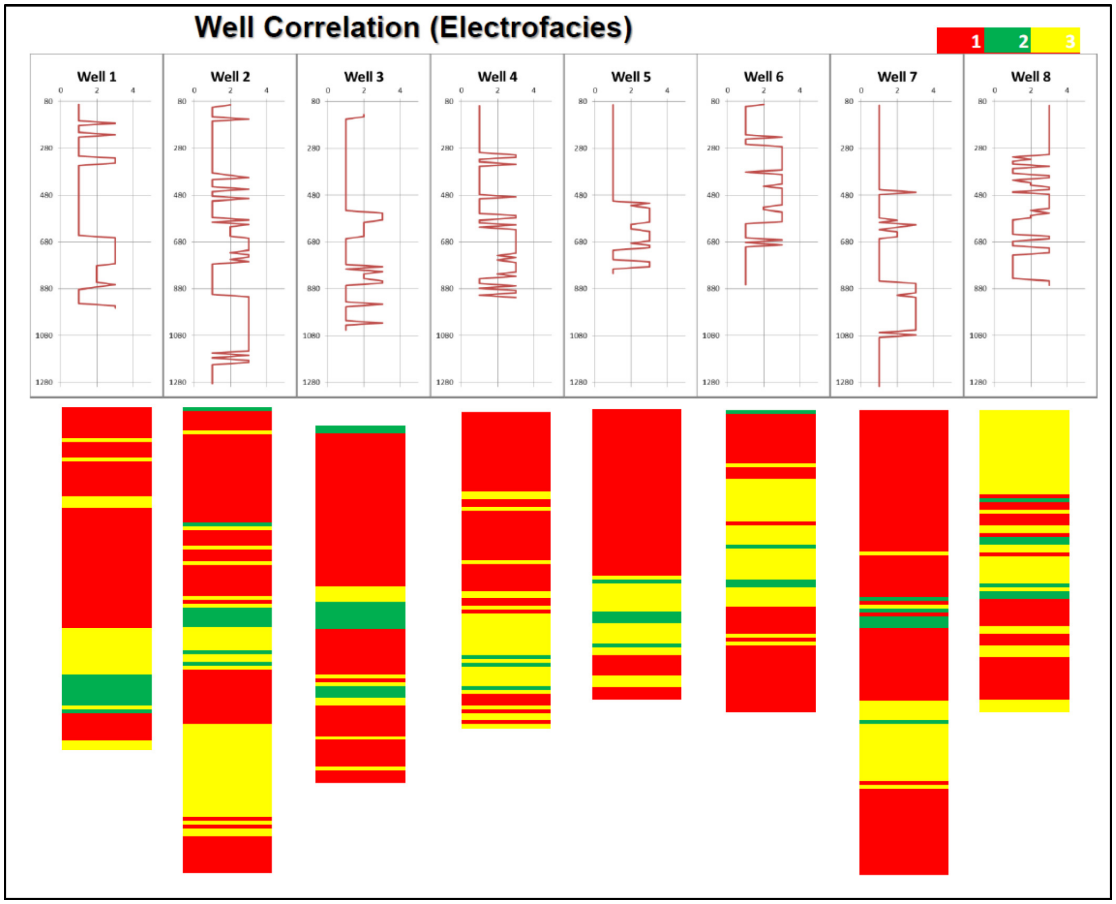

Figure 7—Well correlation result

Well Log Data PostprocessingIn the data postprocessing part, we have already selected the optimal well to build a model using the artificial neural network. ANN is a biologically inspired dynamic computation system that processes data, and learns in an inherently parallel and distributed fashion. By discerning the nature of the dependency between input and output variables, it is capable of extracting and recognizing the underlying dominant patterns and structural relationships among data. Once properly trained, the network can implicitly classify new patterns and generalize an output based upon the learned patterns.

Typically, ANN is arranged into three types of layers, namely: input, hidden and output (Figure 8). Each layer is comprised of a different number of processing elements (PE) which are massively interconnected. There are a number of designed factors which must be considered when constructing a neural network model; they include the selection of neural network architecture, the learning rule, the number of processing elements in each layer, and the type of transfer function. In terms of the architecture of ANN, feedforward back-propagation is a common scheme for training the network. It attempts to repeatedly update the matrix of weights of the inputs until the global minimum of the output error is finally achieved. The output error is described by mean squared error function (MSE) and the training algorithm is Levenberg-Marquardt, which is the fastest back-propagation algorithm.

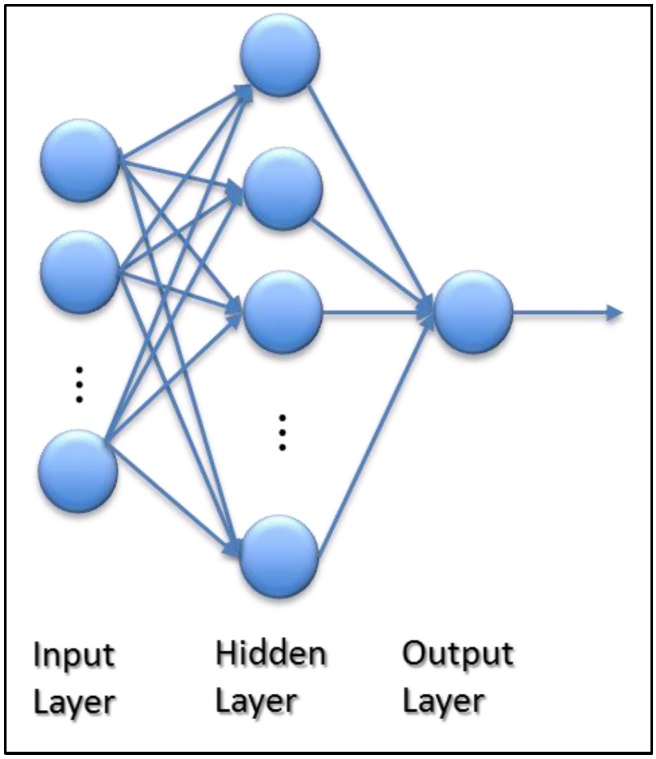

Figure 8—Structure of artificial neural network

Field Application & Results

In our case study, we have a total of eight wells in one field, each of which has 38 different types of well logs. Table 2 shows the list of those logs. After the data mining process, we have an ideal well to use as a training tool for the artificial neural model; we subsequently employ the trained neural network model to generate density log from the target well. This process is called pairwise well prediction. The learning rule is the gradient descent with momentum weight and bias learning function. The momentum (α) is an added parameter to the generalized delta rule to prevent the learning process from converging to a local

minimum. It normally varies within 0 to 1. The learning speed of a neural network is decided by the parameter of learning rate (η). Generally, it is case specific, and smaller than 1. In the most cases, the number of PEs in the input and output layers are given by the number of each dimension. But the number of hidden layers and number of PEs in each one of them is somewhat arbitrary. One rule of thumb states that the number of hidden layer neurons should be about 75% of the input variables (Al- Fattah et al., 2001). Incorporating that tenet with trial and error, we decided to use one-hidden-layer architecture (Figure 9) and specified 40 neurons for that single hidden layer.

Table 2—List of types of well logs

Abbreviation	Description
SP	Spontaneous Potential
GR	Calibrated Gamma-Ray
CALI	Caliper
A010	Resistivity, One Foot
A020	Resistivity, One Foot
A090	Resistivity, One Foot
AT10	Resistivity, Two Foot
SRES	Resistivity, Two Foot
AT20	Resistivity, Two Foot
AT30	Resistivity, Two Foot
MRES	Resistivity, Two Foot
AT60	Resistivity, Two Foot
AT90	Resistivity, Two Foot
DRES	Resistivity, Two Foot
COND	Resistivity, Two Foot
AF10	Resistivity, Four Foot
AF20	Resistivity, Four Foot
AF30	Resistivity, Four Foot
AF60	Resistivity, Four Foot
AF90	Resistivity, Four Foot
TNPH	Thermal Neutron Porosity
RHOB	Standard Resolution Formation Density
PEF	Standard Resolution Formation Photoelectric Factor
RMIN	Micro Inverse Resistivity
RMNO	Micro Normal Resistivity
RXO	Standard Resolution Invaded Zone Resistivity
RSO	Standard Resolution Resistivity Standoff
RMUD	Mud Fully Calibrated
TENS	Cable Tension
TOHC	Open Hole Contact Temperature
ADT SHALLOW TR.	Dielectric Scanner Transverse Shallow Water Filled Porosity
ADT NEAR MED TR.	Dielectric Scanner Transverse Near-Medium Water Filled Porosity
ADT FAR MED TR.	Dielectric Scanner Transverse Far-Medium Water Filled Porosity
ADT DEEP TR.	Dielectric Scanner Transverse Deep Water Filled Porosity
ADT SHALLOW LG.	Dielectric Scanner Transverse Shallow Water Filled Porosity
ADT NEAR MED LG.	Dielectric Scanner Transverse Near-Medium Water Filled Porosity
ADT FAR MED LG.	Dielectric Scanner Transverse Far-Medium Water Filled Porosity
ADT DEEP LG.	Dielectric Scanner Transverse Deep Water Filled Porosity

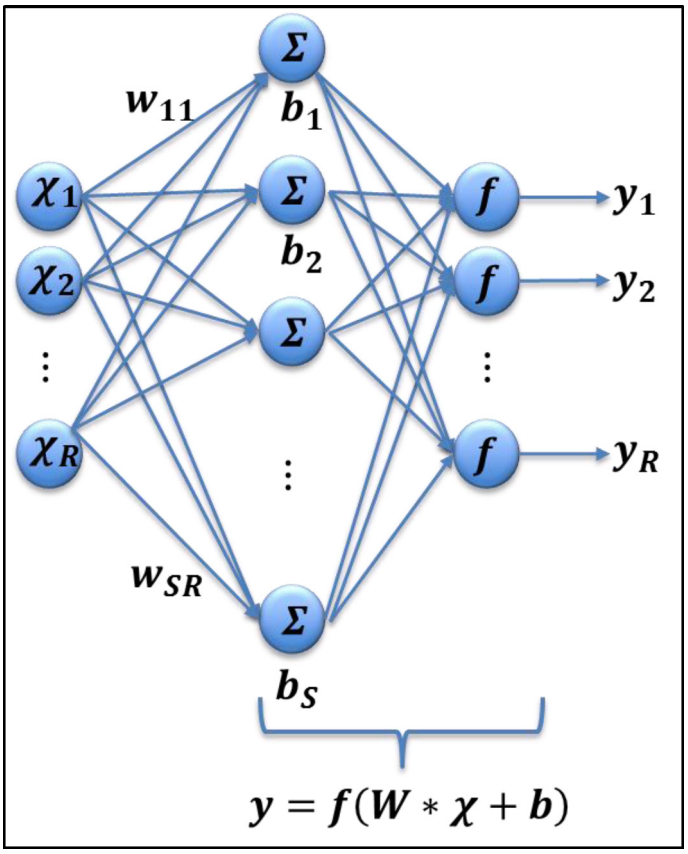

Figure 9—Architecture of one layer neural network

After data preprocessing, the input and output numerical values are normalized within −1 to 1. Initially, weights of inputs are assigned randomly but are updated after each iteration. A weighted sum of input variables at each PE is then modified by hyperbolic tangent sigmoid transfer function, given by:

$$f(S_i) = \frac{1 - e^{-2S_i}}{1 + e^{-2S_i}} \tag{16}$$

where χ is the input data, $S_i = \Sigma_j w_{ji}\chi_j$ and w is the weights. The dataset introduced into the neural network is further divided into training subset, validation subset and test subset in order to avoid overfitting during the training process. The ratio for each is 0.7, 0.15 and 0.15, respectively. After building up the model, the next critical part is prediction or pseudo well log generation using an existing model. The formula to generate well log is as follows:

$$y = f(W * \chi + b) \tag{17}$$

where y is the output, b is the bias and W is weight matrix. We have implemented a comprehensive prediction based on the obtained well log data. Figure 10 shows the correlation matrix of pairwise prediction results. The well number in the blue box is derived from the training well, while those in the red box come from target wells. For example, the number in the green circle means that using Well #2 as a training model to predict Well #3, the correlation is 0.85. The diagonal number indicates whether the training model has been built successfully. The average correlation is above 0.7 which is suitable for pseudo well log analysis. The diagonal number is the validation result. It is strictly one means that the model has been built successfully. Additionally, the correlation matrix has symmetric properties. For example, the correlation result of using Well #2 to predict Well #3 is similar to the result of using Well #3 to predict Well #2. This property can be further used to validate the prediction result. Figure 11 is the

pseudo-density log of Well #3 using both the model of Wells #1 and #2 from left and right. The orange line is the prediction and the blue line is the reference. From Fig. 11, we can realize small variances between different models (Well #1 and Well #2).

Training Well

Predicted Well	#1	#2	#3	#4	#5	#6	#7	#8
#1	1	0.85	0.78	0.64	0.75	0.66	0.77	0.65
#2	0.86	1	0.74	0.8	0.8	0.7	0.8	0.65
#3	0.85	0.77	1	0.67	0.75	0.7	0.67	0.6
#4	0.67	0.83	0.67	1	0.8	0.81	0.8	0.68
#5	0.78	0.84	0.80	0.84	1	0.8	0.71	0.63
#6	0.67	0.74	0.73	0.83	0.82	1	0.74	0.65
#7	0.79	0.83	0.68	0.79	0.73	0.84	1	0.73
#8	0.71	0.68	0.65	0.72	0.61	0.69	0.82	1

Correlation Coefficient Matrix

Figure 10—Pairwise well prediction results

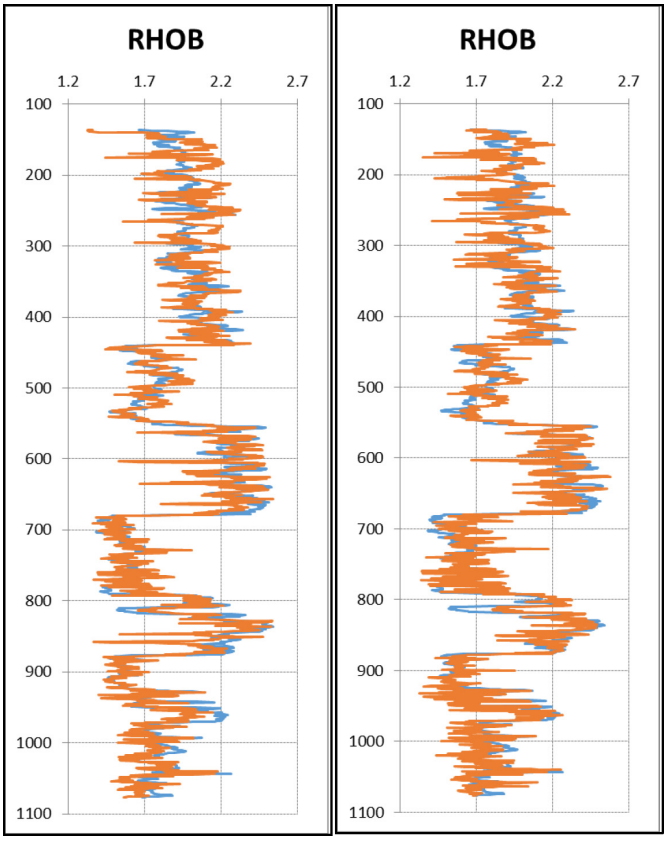

Figure 11—Pseudo density log of Well #3

Summary & Conclusions

In our paper, we created a systematic approach to generating pseudo-density log, which has been proven to work well. The system consists of three components: data preprocessing, data mining and data postprocessing. In data preprocessing, we first normalized the data, and then suggested using principal component analysis. In the next stage, we proposed the initial use of Gaussian mixture model, so as to have an intuitive scope of the property of the dataset. We subsequently introduced model-based clustering to devise a proper result without setting up the number of clusters in GMM. For well correlation, we advocated a sampling method and discriminant analysis which increase the efficiency of model-based clustering in processing a large dataset. Finally, one hidden layer with back-propagation algorithm neural network was utilized to build up the model for prediction.

The following conclusions can be drawn from this study:

1. Principal Component Analysis can be a powerful tool in dimension reduction of well log data. The approach is simple and intuitive to apply. The number of principal components has been proven in the Appendix.
2. Gaussian Mixture Model can be a natural indicator of selecting potential wells for building up a model. If clustering shape out of two wells data bears a close resemblance to the shape out of any single well data, a good relationship may exist between them and vise versa. For ease of visualization, the input should be the first two principal components after PCA.
3. Model-based clustering method can be an updated version of GMM. It incorporates the algorithm to merge the clusters if they fall into the same distribution. In MBC, instead of guessing the number of clusters, we only need to organize an upper boundary. This study proves the stability of MBC, which is of great importance in well correlation.
4. Sampling method and discriminant analysis are applied to increase the efficiency of MBC in well correlation. Theoretically, different sampling methods may result in slightly different outcomes. Herein, the sampling method we suggested is simple to implement, and works well.
5. The parameters and the structure of neural network may vary case by case. It could have some influence on the prediction result, but there would not be much disparity if the input is the same. We suggest having increased tests on the neural network before implementation.

Acknowlegdement

Authors would like to thank **L. J. (Jim) Lederhos, Lederhos Engineering Consulting, Inc.** for giving us the professional interpretation of well log dataset by applying conventional petrophysical analysis. Mr. Lederhos has also provided us the software to testify the results. Additionally, we would like to thank **Professor Donald Hill**, USC for giving us the idea of pairwise well prediction. It is of paramount importance to our paper.

Nomenclature

ANN =	Artificial Neural Network
ARD =	Auto Relevance Determination
CPV =	Cumulative Percent Variance
DA =	Discriminant Analysis
EM =	Expectation Maximization
GMM =	Gaussian Mixture Model
GR =	Gamma Ray
MBC =	Model-Based Clustering
MSE =	Mean Squared Error

PC = Pricinpal Component
PCA = Pricinpal Component Analysis
PE = Processing Element
SVD = Singular Value Decomposition

References

1. Abou-Sayed, A. 2012. Data Mining Applications in the Oil and Gas Industry. *Journal of Petroleum Technology*. 10.2118/1012-0088-JPT.
2. Al-Alwani, M., Britt, L. K. and Dunn-Norman, S. 2015. Data Mining and Statistical Analysis of Completions in an Unconventional Play: The Canadian Montney Formation. Unconventional Resources Technology Conference, San Antonio, Texas, USA, 20-22 July. 10.2118/178586-MS.
3. Al-Fattah, S.M., Startzman, R.A. 2001. Neural Network Approach Predicts, U.S. Natural Gas Production. SPE Production and Operations Symposium, Oklahoma City, Oklahoma, 24-27 March. 10.2118/67260-MS.
4. Aminzadeh, F. and de Groot, P., 2006, *Neural Networks and Soft Computing Techniques, with applications in the oil industry* EAGE Publishing.
5. Aminzadeh, F. 1996. Future Geophysical Technology Trends. *Proceedings of Stratigraphic Analysis,* Pages 1–6.
6. Aminzadeh, F., Barhen, J., Glover, C. W.N. et al. 2000. Reservoir Parameter Estimation Using a Hybrid Neural Network. *Computers & Geosciences,* Volume **26,** Issue 8, Pages 869–875. DOI: 10.1016/S0098-3004(00)00027-3.
7. Aminzadeh, F. and Chatterjee, S. 1984. Applications of Clustering in Exploration Seismology. *Geoexploration,* **23,** Pages 147–159.
8. Banchs, R. E. and Michelena, R. J. 2000. Well Log Estimates and Confidence Intervals by Using Artificial Neural Networks. SEG Annual Meeting, Calgary, Alberta, 6-11 August.
9. Banfield, J. D. and Raftery, A. E. 1993. Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics,* Volume **49,** No.3, Pages 803–821.
10. Chaki, S., Verma, A. K., Routray, A. et al. 2013. Prediction of Porosity and Sand Fraction from Well Log Data Using Neural and Neuro-fuzzy Systems: A Comparative Study. 10th Biennial International Conference and Exposition, Kochi, India, 23-25 November.
11. Chakra, N. C., Song, K., Gupta, M. M. et al. 2013. An Innovative Neural Forecast of Cumulative Oil Production from A Petroleum Reservoir Employing Higher-order Neural Networks (HONNs). *Journal of Petroleum Science and Engineering,* Volume **106,** Pages 18–33. 10.1016/j.petrol.2013.03.004.
12. De Jonge, G., Stundner, M. and Zangl, G. 2003. Automated Reservoir Surveillance through Data Mining Software. *Offshore Europe, Aberdeen, United Kingdom, 2-5 September*. 10.2118/83974-MS.
13. Fraley, C. and Raftery, A. E. 1998. How Many Clusters? Which Clustering Method? Answers via Model-Based Cluster Analysis. *The Computer Journal,* Volume **41,** No. 8, Pages 578–588.
14. Fraley, C. and Raftery, A. E. 2002. Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association,* Volume **97,** No. 458, Pages 611–631.
15. Grujic, O., Silva, C. D. and Caers, J. 2015. Functional Approach to Data Mining, Forecasting, and Uncertainty Quantification in Unconventional Reservoirs. SPE Annual Technical Conference and Exhibition, Houston, Texas, USA, 28-30 September. 10.2118/174849-MS.
16. Hassibi, M. and Ershaghi. 1999. Reservoir Heterogeneity Mapping Using an Automated Pattern Recognition Approach. SPE Annual Technical Conference and Exhibition, Houston, Texas, 3-6 October. 10.2118/56818-MS.
17. Hassibi, M., Ershaghi, and Aminzaeh, F., 1996. *High Resolution Reservoir Heterogeneity Characterization Using Recognition Technology: Development in Petroleum Science,* Vol. **51,** Pages 471–497, Elsevier Science, B.V.
18. Holdaway, K. R. 2013. Data Mining Methodologies Enhance Probabilistic Well Forecasting. SPE Middle East Intelligent Energy Conference and Exhibition, Manama, Bahrain, 28-30 October. 10.2118/167428-MS.
19. Hotelling, H. 1933. Analysis of A Complex of Statistical Variables into Principal Components. *J. Educ. Psychol.,* Volume **24,** Issue 6, Pages 417–441.
20. Jain, A. K., Murty, M. N. and Flynn, P. J. 1999. Data Clustering: A Review. *ACM Computing Surveys (CSUR),* Volume **31** Issue 3, Pages 264–323. 10.1145/331499.331504.
21. Kass, R. E. and Raftery, A. E. 1993. Bayes Factors and Model Uncertainty. *Technical Report 254, University of Washington*.
22. Kishore, K., Sharma, P. and Khanapurkar, P. 2014. State-of-the-Art Emerging Technology for Technical Data Mining and Analysis. SPE Large Scale Computing and Big Data Challenges in Reservoir Simulation Conference and Exhibition, Istanbul, Turkey, 15-17 September. 10.2118/172985-MS.

23. Kumar, B. and Kishore, M. 2006. Electrofacies Classification—A Critical Approach. Proceedings of the 6th International Conference & Exposition on Petroleum Geophysics, Kolkata, India.
24. Lee, S. H., Datta-Gupta, A. 1999. Electrofacies Characterization and Permeability Predictions in Carbonate Reservoirs: Role of Multivariate Analysis and Nonparametric Regression. SPE Annual Technical Conference and Exhibition, Houston, Texas, 3-6 October. 10.2118/56658-MS.
25. Lim, J., Kang, J. M. and Kim, J. 1997. Multivariate Statistical Analysis for Automatic Electrofacies Determination from Well Log Measurements. SPE Asia Pacific Oil and Gas Conference and Exhibition, Kuala Lumpur, Malaysia, 14-16 April. 10.2118/38028-MS.
26. Minka, T. P. 2000. Automatic Choice of Dimensionality for PCA. *Advances in Neural Information Processing Systems* **13,** Pages 598–604, MIT Press.
27. Mohaghegh, S. D. 2009. Artificial Intelligence and Data Mining: Enabling Technology for Smart Fields. *TWA SPE Journal Paper*. 10.2118/0309-014-TWA.
28. Nikravesh, M. 1998. Neural Network Knowledge-Based Modeling of Rock Properties Based on Well Log Databases. SPE Western Regional Meeting, Bakersfield, California, 10-13 May. 10.2118/46206-MS.
29. Nikravesh, M. and Aminzadeh, F. 2001a. Mining and Fusion of Petroleum Data with Fuzzy Logic and Neural Network Agents. *Journal of Petroleum Science and Engineering,* Volume **29,** Issues 3-4, May 2001, Pages 221–238. 10.1016/S0920-4105(01)00092-4.
30. Nikravesh, M. and Aminzadeh, F. 2001b. Past, Present and Future Intelligent Reservoir Characterization Trends. *Journal of Petroleum Science and Engineering,* Volume **31,** Issues 2-4, November 2001, Pages 67–79. 10.1016/S0920-4105(01)00121-8.
31. Nikravesh, M., Aminzadeh, F. and Zadeh, L. A. 2003. Intelligent Data Analysis for Oil Exploration. *Developments in Petroleum Science,* Volume **51,** ISBN: 978-0-444-50685-6.
32. Nikravesh, M., Levey, R. A., Adams, R. D. et al. 1999. Soft Computing: Tools for Intelligent Reservoir Characterization (IRESC). SEG Annual Meeting, Houston, Texas, 31 October-5 November.
33. Pearson, K. 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. *Phil. Mag., Series 6,* Volume **2,** Issue 11, Pages 559–572.
34. Serra, O. and Abbott, H. T. 1980. The Contribution of Logging Data to Sedimentology and Stratigraphic, 55th Annual Fall Technical Conference and Exhibition, Dallas, Texas.
35. Sharma, P., Mamgain, G., Bahuguna, V. K. et al. 2011. Improved Permeability Estimates in Carbonate Reservoirs Using Electrofacies Characterization: A Case Study of Mumbai High South. The 2nd South Asain Geoscience Conference and Exhibition, New Delhi, India.
36. Shokir, E.M. El-M, Alsughayer, A.A., and Al-Ateeq, A. 2005. Permeability Estimation from Well Log Responses. Canadian International Petroleum Conference, Calgary, Alberta, 7-9 June. 10.2118/2005-012.
37. Tang, H., Meddaugh, W. S. and Toomey, N. 2011. Using an Artificial-Neural-Network Method to Predict Carbonate Well Log Facies Successfully. *SPE Reservoir Evaluation & Engineering*. 10.2118/123988-PA.
38. Tipping, M. E. and Bishop, C. M. 1997. Probabilistic Principal Component Analysis. *Technical Report NCRG/97/010, Neural Computing Research Group,* Aston University.
39. Valle, S., Li, W. and Qin, S. J. 1999. Selection of the Number Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods. *Industrial & Engineering Chemi stry Research*. 10.1021/ie990110i.
40. Verma, A. K., Cheadle, B. A., Routray, A. et al. 2012. Porosity and Permeability Estimation using Neural Network Approach from Well Log Data. GeoConvention, Calgary, Alberta, Canada, 14-18 May 2012.
41. Zangl, G. and Oberwinkler, C.P. 2004. Predictive Data Mining Techniques for Production Optimization. SPE Annual Technical Conference and Exhibition, Houston, Texas, 26-29 September. 10.2118/90372-MS.

Appendix

The concept of principal component analysis (PCA) was first introduced by Karl Pearson in 1901 and independently developed by Harold Hotelling in the 1930s. Since then it has been a powerful tool to reduce dataset dimensionality by transforming an original dataset to a new one in a lower dimension subspace. In this method, choosing dimensionality in the projected space is a key issue because less information will be forthcoming if fewer principal components (PCs) than required are retained, which may lead to underfitting of the training model. In contrast, if an excess of required PCs are used, it will unavoidably increase computation expenses, and potentially generate overfitting problems in training. In the past, numerous approaches have been developed to choose the optimal number of PCs, including, but not limited to — the Kaiser rule, elbow rule on scree plot, cumulative percent variance (CPV), parallel analysis (PA), Bayesian Model Selection, RRN/RRU, and auto relevance determination (ARD). Each method has shown strengths in different cases. However, at this writing, how to determine the optimal number of PCs remains controversial. Among previous approaches, CPV, elbow test, Kaiser Rule and parallel analysis can be deemed as empirical methods. CPV determines the number of PCs based on the cumulative percent of the variance, but the benchmark of selectivity is subjective. The elbow rule is to locate the number of PCs by picking an evident turning point on the on scree plot. But sometimes a relatively smooth curve, or the presence of several elbows may make selectivity of numbers ambiguous. Kaiser Rule is a popular approach which retains PCs if corresponding eigenvalues are above 1. But generally speaking, it may overestimate the numbers of PCs. In parallel analysis, a Monte-Carlo based simulation is implemented wherein a dataset of random numbers with the same size and number of variables as the observed dataset is generated for analysis. Their eigenvalues will then be calculated and saved. This process is repeated many times prior to performing a statistical analysis to obtain the variance for each PC at a 95% confidence interval. However, this method can be quite costly in terms of computation time. Bayesian model selection, RRN/RRU and ARD have solid statistical foundations. Bayesian model selection method has been commonly applied to probabilistic PCA model since Tipping and Bishop (1997) related PCA to maximum likelihood density estimation. To solve the estimate analytically, Laplace approximation and its simplification BIC approximation (Kass and Raftery, 1993) have been adopted. That method offers a fast and statistically reliable way to pick correct dimensionality of the data (Minka, 2000) while the assumption is only Gaussian distribution. RRN/RRU method incorporates two models with Gaussian density and uniform density, respectively. But it is not prone for those two methods to get true dimensionality for a large dataset because they use a restrictive model for the subspace. ARD is also a classical method based on Bayesian inference. Nevertheless, that method seems to be computation intensive for many applications.

In our paper, we compared results of optimal PC numbers given by the above methods. In order to get an appropriate number of PCs as input to the ANN, we try to find a balance between empirical and statistical methods in terms of accuracy and computation expense. Therefore, we try to screen the candidates given those two considerations. Figure 12 and Figure 13 show simulated graphs, respectively, using our dataset with 1861 observations and 37 features. In Figure 12, x-axis value of intersection point will give us the PC number. In Figure 13, the PC number is obtained according to the evident turning point location. It is noted that the score keeps increasing with the dimensionality in (a) and (b), because corresponding eigenvalues don't level off, but continue downward. So in this case, we pick the largest possible dimensionality. All the results are listed in Table 3, from which one notices that the result from CPV may be our target. In order to validate that choice, we later carried out an experiment. Given different optimal PC numbers from previous models, we adjust our input dataset dimensions of all 8 wells to reflect that. Then we use this dataset as the input of ANN to make a pairwise prediction of bulk density. For example, provided that optimal number of PCs is 34, we pick out 34 components from the transformed dataset of all 8 wells as the input of ANN. Then we use log dataset in Well #1 as the training dataset to

predict the bulk density log value in Well #2, Well #3 and so on. Then we calculated an average correlation between predicted values and real values of the 8 wells. In Figure 14, we see the correlation between average correlation coefficients and PC numbers. We found that the PC number according to the approximate turning point matches our earlier choice. Therefore, given the computation cost and prediction accuracy, the CPV method offered the best result.

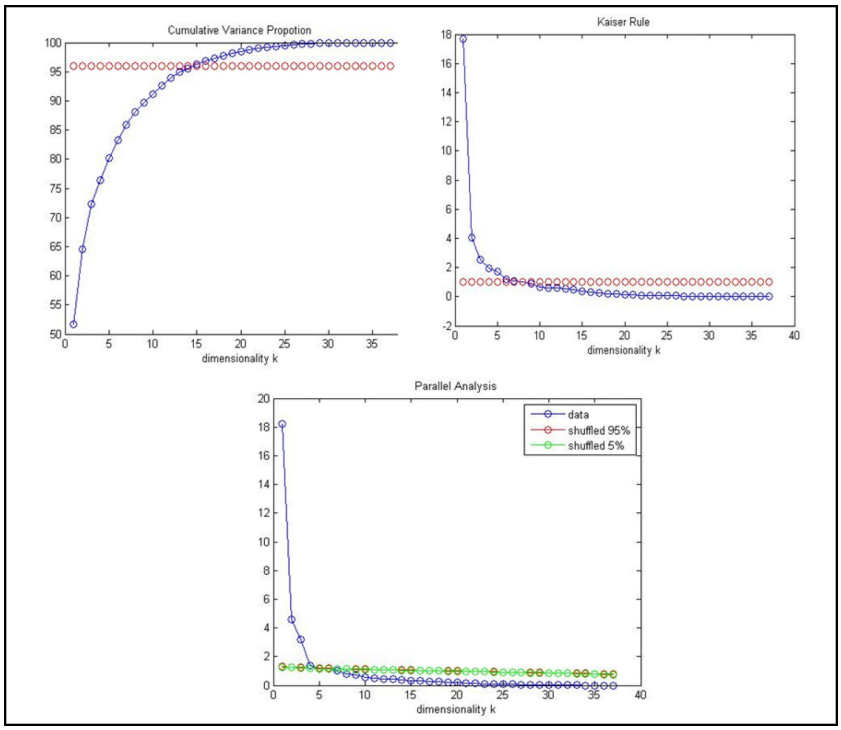

Figure 12—Principal component selection

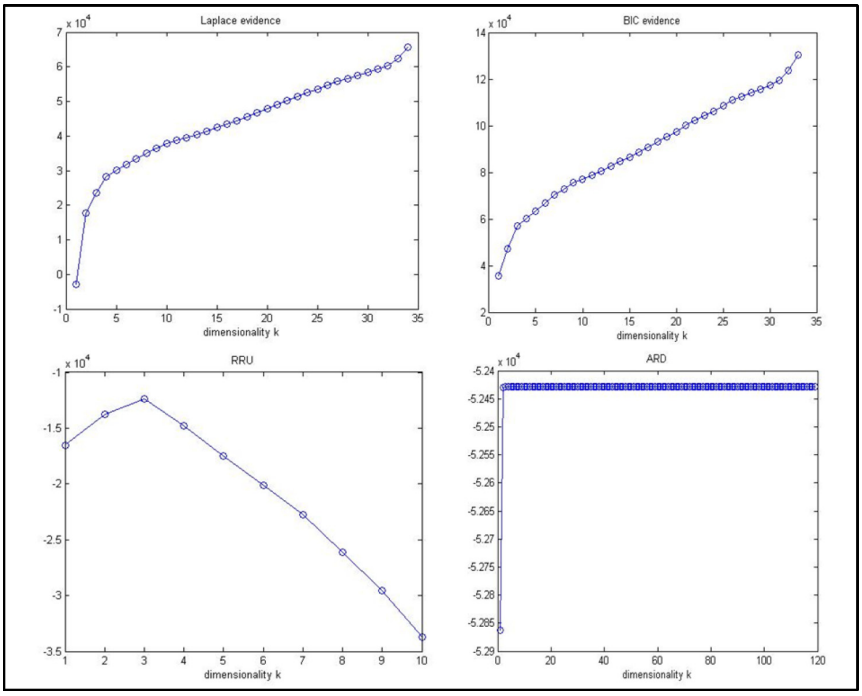

Figure 13—Principal component selection

Table 4—

Methods	Number of PCs	Methods	Number of PCs
Cumulative Variance Proportion	15	BIC Approximation	34
″Elbow″ Test	5	Laplace Approximation	33
Kaiser Rule	7	RRU	3
Parallel Analysis	6	ARD	10

Correlation K=34	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8
Well 1	1.00	0.87	0.82	0.79	0.82	0.72	0.81	0.72
Well 2	0.88	1.00	0.80	0.89	0.81	0.79	0.83	0.66
Well 3	0.84	0.83	1.00	0.79	0.82	0.79	0.72	0.65
Well 4	0.80	0.90	0.78	1.00	0.83	0.76	0.80	0.73
Well 5	0.83	0.79	0.81	0.84	1.00	0.83	0.76	0.66
Well 6	0.69	0.78	0.78	0.78	0.85	1.00	0.84	0.71
Well 7	0.79	0.85	0.73	0.79	0.77	0.84	1.00	0.75
Well 8	0.69	0.68	0.64	0.72	0.63	0.69	0.74	1.00
Average	0.81	0.84	0.79	0.82	0.82	0.80	0.81	0.74

K=15	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8
Well 1	1.00	0.85	0.78	0.64	0.75	0.66	0.77	0.65
Well 2	0.86	1.00	0.74	0.80	0.80	0.70	0.80	0.65
Well 3	0.85	0.77	1.00	0.67	0.75	0.70	0.67	0.60
Well 4	0.67	0.83	0.67	1.00	0.80	0.81	0.80	0.68
Well 5	0.78	0.84	0.80	0.84	1.00	0.80	0.71	0.63
Well 6	0.67	0.74	0.73	0.83	0.82	1.00	0.74	0.65
Well 7	0.79	0.83	0.68	0.79	0.73	0.84	1.00	0.73
Well 8	0.71	0.68	0.65	0.72	0.61	0.69	0.82	1.00
Average	0.79	0.82	0.76	0.79	0.78	0.78	0.79	0.70

K=10	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8
Well 1	1.00	0.80	0.75	0.64	0.80	0.66	0.62	0.70
Well 2	0.83	1.00	0.80	0.75	0.77	0.79	0.78	0.68
Well 3	0.77	0.83	1.00	0.50	0.51	0.48	0.64	0.57
Well 4	0.66	0.77	0.57	1.00	0.79	0.74	0.62	0.43
Well 5	0.79	0.78	0.50	0.81	1.00	0.76	0.70	0.63
Well 6	0.69	0.77	0.46	0.75	0.79	1.00	0.67	0.71
Well 7	0.58	0.81	0.66	0.61	0.70	0.69	1.00	0.74
Well 8	0.74	0.69	0.59	0.44	0.68	0.73	0.73	1.00
Average	0.76	0.81	0.67	0.69	0.76	0.73	0.72	0.68

K=7	Well 1	Well 2	Well 3	Well 4	Well 5	Well 6	Well 7	Well 8
Well 1	1.00	0.80	0.56	0.53	0.55	0.56	0.56	0.48
Well 2	0.78	1.00	0.64	0.66	0.52	0.49	0.49	0.53
Well 3	0.61	0.67	1.00	0.30	0.64	0.42	0.53	0.54
Well 4	0.55	0.73	0.33	1.00	0.67	0.68	0.66	0.52
Well 5	0.58	0.50	0.69	0.68	1.00	0.65	0.75	0.74
Well 6	0.56	0.53	0.43	0.70	0.61	1.00	0.71	0.78
Well 7	0.56	0.53	0.49	0.64	0.78	0.74	1.00	0.62
Well 8	0.50	0.56	0.51	0.49	0.71	0.80	0.64	1.00
Average	0.64	0.66	0.58	0.63	0.69	0.67	0.67	0.65

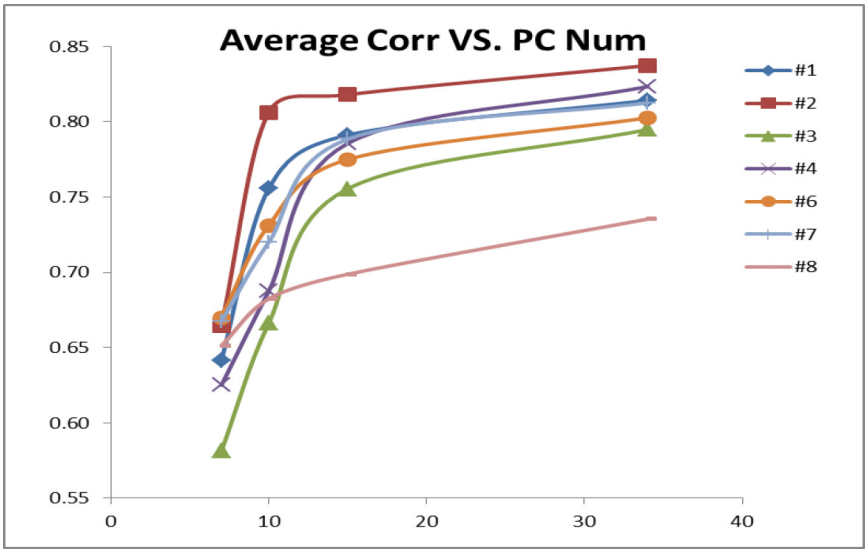

Figure 14—Average correlation between predicted values and real values of eight wells