

Using artificial neural networks to generate synthetic well logs

Luisa Rolon^a, Shahab D. Mohaghegh^{b,c,*}, Sam Ameri^b, Razi Gaskari^c, Bret McDaniel^a

^a Chevron Corp., 1400 Smith St., Houston, TX 77002, United States

^b West Virginia University, Petroleum & Natural Gas Engineering, 345E Mineral Resources Bldg., Morgantown, WV 26506, United States

^c Intelligent Solutions, Inc. Morgantown, WV, United States

ARTICLE INFO

Article history:

Received 21 March 2009

Received in revised form

10 August 2009

Accepted 27 August 2009

Available online 28 October 2009

Keywords:

Synthetic logs

Neural networks

Reservoir characterization

Well logs

Artificial intelligence

Petrophysics

ABSTRACT

A methodology to generate synthetic wireline logs is presented. Synthetic logs can help to analyze the reservoir properties in areas where the set of logs that are necessary, are absent or incomplete. The approach presented involves the use of artificial neural networks as the main tool, in conjunction with data obtained from conventional wireline logs. Implementation of this approach aims to reduce costs to companies.

Development of the neural network model was completed using Generalized Regression Neural Network, and wireline logs from four wells that included gamma ray, density, neutron, and resistivity logs. Synthetic logs were generated through two different exercises. Exercise one involved all four wells for training, calibration and verification process. The second exercise used three wells for training and calibration and the fourth well was used for verification. In order to demonstrate the robustness of the methodology, three different combinations of inputs/outputs were chosen to train the network. In combination "A" the resistivity log was the output and density, gamma ray, and neutron logs, and the coordinates and depths (XYZ) the inputs. In combination "B" the density log was output and the resistivity, the gamma ray, and the neutron logs, and XYZ were the inputs, and in combination "C" the neutron log was the output while the resistivity, the gamma ray, and the density logs, and XYZ were the inputs. After development of the neural network model, synthetic logs with a reasonable degree of accuracy were generated. Results indicate that the best performance was obtained for combination "A" of inputs and outputs, then for combination "C", and finally for combination "B". In addition, it was determined that accuracy of synthetic logs is favored by interpolation of data. It was also demonstrated that using neural network to generate synthetic well logs is far more superior when compared to conventional approaches such as multiple-regression.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Well logging has been in use for almost one century as an essential tool for determination of potential production in hydrocarbon reservoirs. Log analysts interpret the data from the log in order to determine the petrophysical parameters of the well. However, for economical reasons, companies do not always possess all the logs that are required to determine reservoir characteristics. This paper presents an approach involving the use of an artificial neural network (ANN) as the main tool, in conjunction with data obtained from conventional wireline logs, which aim to help to solve the aforementioned problem by generating synthetic wireline logs for those locations where the set of logs that are necessary to analyze the reservoir properties, are absent or are not complete.

Artificial neural networks have been broadly used in reservoir characterization because of their ability to extract nonlinear relationships between a sparse set of data (Banchs and Michelena, 2002). Studies in this subject have used wireline measurement logs, and seismic attributes to predict reservoir properties such as effective porosity, fluid saturation and rock permeability, and to define lithofacies and predict log responses, i.e. generation of synthetic logs. In all cases it was demonstrated that ANN are powerful tool for recognition-pattern, system identification, and prediction of any variable in the future with a better correlation coefficient (R^2) over traditional statistical analysis like linear regression.

Mohaghegh et al. (1998), describe a methodology developed to generate synthetic Magnetic Resonance Imaging (MRI) logs using conventional well logs such as Spontaneous Potential, Gamma Ray, Caliper, and Resistivity for four wells located in East Texas, Gulf of Mexico, Utah, and New Mexico. The methodology incorporates an artificial neural network as its main tool. The synthetic Magnetic

* Corresponding author at: West Virginia University, Petroleum & Natural Gas Engineering, 345E Mineral Resources Bldg., Morgantown, WV 26506, United States.

E-mail address: shahab@wvu.edu (S.D. Mohaghegh).

Resonance Imaging logs were generated with a high degree of accuracy even when the model developed used data not employed during model development.

Mohaghegh et al. (1999), present an approach that involves neural-network-design software, for low cost/high effectiveness log analysis in a field scale. The cost reduction is achieved by analyzing only a group of the wells in the field. The intelligent software tool is built to learn and reproduce the analyzing capabilities of the engineer on the remaining wells. As part of this study, logs that were missed in several wells and that were necessary for analysis were generated.

Bhuiyan (2001), developed a neural network to generate synthetic Magnetic Resonance Imaging logs in order to provide information about reservoir characteristics of the Cotton Valley formation. In this work, data preparation previous to network training, involved fuzzy logic for group well logs together based on similarity criteria of the reservoir formation and to identify the most influential logs for a well. Tonn (2002), uses seismic attributes, density and sonic logs to train an ANN in order to predict the gamma ray response of the Athabasca oil sands in western Canada, in order to solve the reservoir properties and therefore chose the best location to place injection and production wells for a steam injection program.

The intention of the technique used here is not to eliminate well logging in a field but it meant to become a tool for reducing costs for companies whenever logging proves to be insufficient and/or difficult to obtain. This technique in addition, can provide a guide for quality control during the logging process, by prediction of the response of the log before the log is acquired.

The study area is located in Southern Pennsylvania (Fig. 1). The sequence of rocks recorded in the set of logs used for this work belongs to the Upper Devonian of Southern Pennsylvania (the Venango and Bradford plays), which has been producing natural gas in the area since the late 1800s. The names of the units involved from bottom to top are the following: 2nd Bradford, Speechley, Gordon, 100 Foot, and Murrysville (Figs. 2 and 3 at the end of the paper).

The 2nd Bradford and the Speechley sands (Lower zone) are quartz and feldspar cemented sandstone and- siltstone reservoirs. In general, these reservoirs have similar characteristics in all the logs used, with little discontinuity among wells. Gamma ray, density, and resistivity values are relatively constant among wells in the study area. The Speechley interval is considerably finer-grained than the 2nd Bradford (Fig. 2).

The Murrysville/100 Foot/Gordon sandstones (Upper zone) are composed of a series of conglomerates, sandstones, and shales. The

sandstones of this facies are fine- to medium-grained, well sorted, well rounded, and have good porosity and permeability. Log porosity has been measured up to 25% and permeabilities have been measured up to 250 md. The conglomerate and coarse-grained sandstone of the upper shoreface and the foreshore tend to have more cement and therefore a lower porosity and permeability (McBride, 2004). Gamma ray, density, and resistivity values are more or less constant, although changes in thickness occur from location to location (Fig. 3).

2. Some fundamentals on well logs

2.1. Gamma-ray log

This log (Fig. 4) measures the natural gamma-ray emission of the various layers penetrated in the well, a property related to their content of radiogenic isotopes of potassium, uranium and thorium. The tool may detect gamma ray energies of less than 0.5 to more than 2.5 mV. Thorium, uranium, and potassium (particularly potassium) are common elements in clay minerals and some evaporites. In terrigenous clastic successions the log reflects the "cleanliness" (lack of clays) or "shaliness" (high radioactivities) (see Figs. 4 and 5) on the API scale of the rock, averaged over an interval of depth. Because of this property, gamma-ray log patterns mimic vertical sand-content or carbonate-content trends.

It must be emphasized that the gamma-ray reading is not a function of grain size or carbonate content, only the proportion of radioactive elements, which may be related to the proportion of shale content. The Gamma Ray (GR) tool emits gamma rays into sedimentary formations to an average penetration of approximately one foot. When gamma rays pass through a formation, they experience successive Compton-scattering collisions with formation atoms (Cant, 1992). These collisions cause gamma rays to lose energy (Bassiouni, 1994). Subsequently, formation atoms absorb this energy through the photoelectric effect.

The amount of absorption is a function of the formation density. Therefore, the radioactivity level shown on the GR log will be different for two formations that have the same amount of radioactive material per unit volume but with different densities. The lower density formation will appear to be more radioactive.

Cased hole use is an important application of the GR log because the tool can be run in completion and work-over operations. GR logs are also used for interpretation of depositional environments.

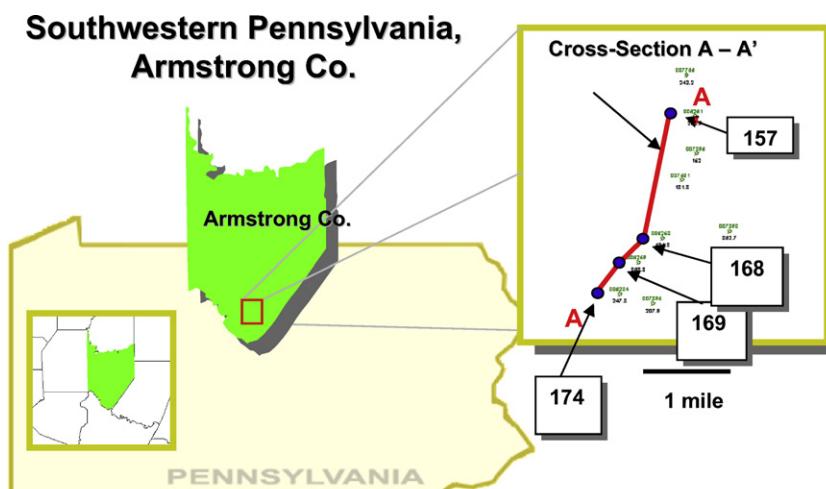


Fig. 1. Location of the study area and wells analyzed (157, 168, 169, 174). A-A' represents the line of the cross-section.

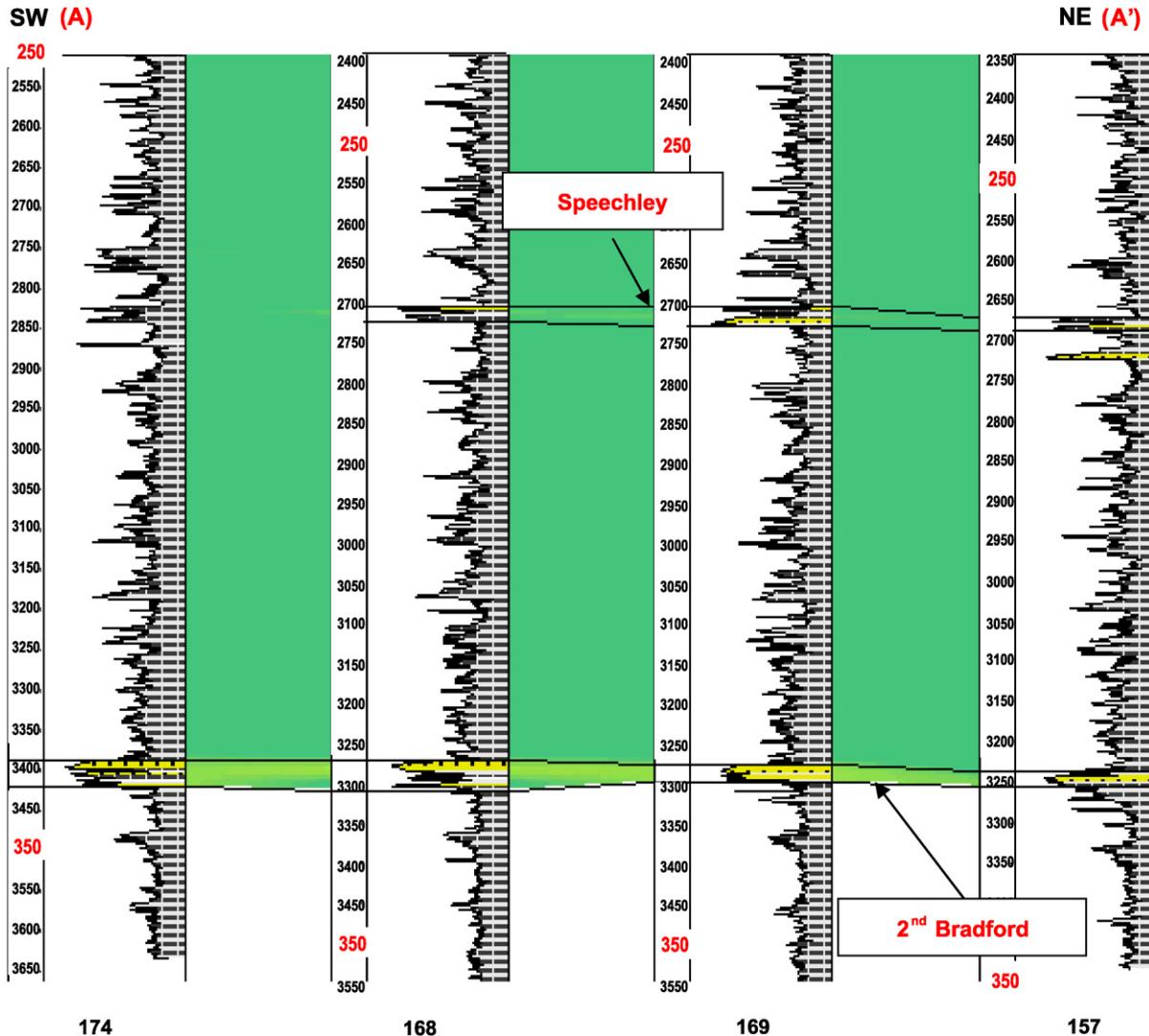


Fig 2. Gamma Ray correlation of Speechley and 2nd Bradford (Lower zone).

2.2. Porosity logs

There are three types of logging tools that are used to estimate the amount of pore space in a rock: the density, neutron, and acoustic velocity (or sonic) tool. For the scope of this paper, only density and neutron tools will be explained.

2.2.1. Density log

The formation density log is a porosity log that measures electron density of a formation. It can assist the geologist to: (1) identify evaporite minerals, (2) detect gas-bearing zones, (3) determine hydrocarbon density, and (4) evaluate shaly-sand reservoirs and complex lithologies.

The density logging device, the gradiomanometer or a nuclear fluid densimeter, is a contact tool which consists of a medium-energy gamma ray source that emits gamma rays into a formation, which is scattered back to the detector in amounts proportional to the electron density of the rock. The electron density, in most cases, is related to the density of the solid material, and the amount of density of pore fluids by determining the percentage, or holdup (record of the fractions of different fluids present at different depths in the borehole) of the different fluids.

A density derived porosity curve is recorded in tracks 2 and 3 along with the bulk density (ρ_b) and correction ($\Delta\rho$) curves. Track #1 contains a gamma ray log and caliper. Formation bulk density (ρ_b) is a function of matrix density, porosity, and density of the fluid in the pores (salt, mud, fresh mud, or hydrocarbons). To determine porosity, the density of the matrix (ρ_{ma}) and/or density of the fluid in the borehole must be known. Table 1 shows some common values of ρ_{ma} .

The formula for calculating density porosity is:

$$\phi_{den} = \frac{(\rho_{ma} - \rho_b)}{(\rho_{ma} - \rho_f)}$$

where ϕ_{den} is density derived porosity, ρ_{ma} is the matrix density, ρ_b is the bulk density, and ρ_f is the average density of the fluids in pore spaces.

Density of formation water ranges from 0.95 g/cc to 1.10 g/cc approximately depending on temperature, pressure and salinity. Average density of oil is slightly lower than these values and varies over an equally wide range. According to Bassiouni (1994), the investigation ratio of the tool is shallow, therefore it investigates the invaded zone and ρ_f can be expressed by:

$$\rho_f = S_{xo}\rho_{mf} + (1 - S_{xo})\rho_h,$$

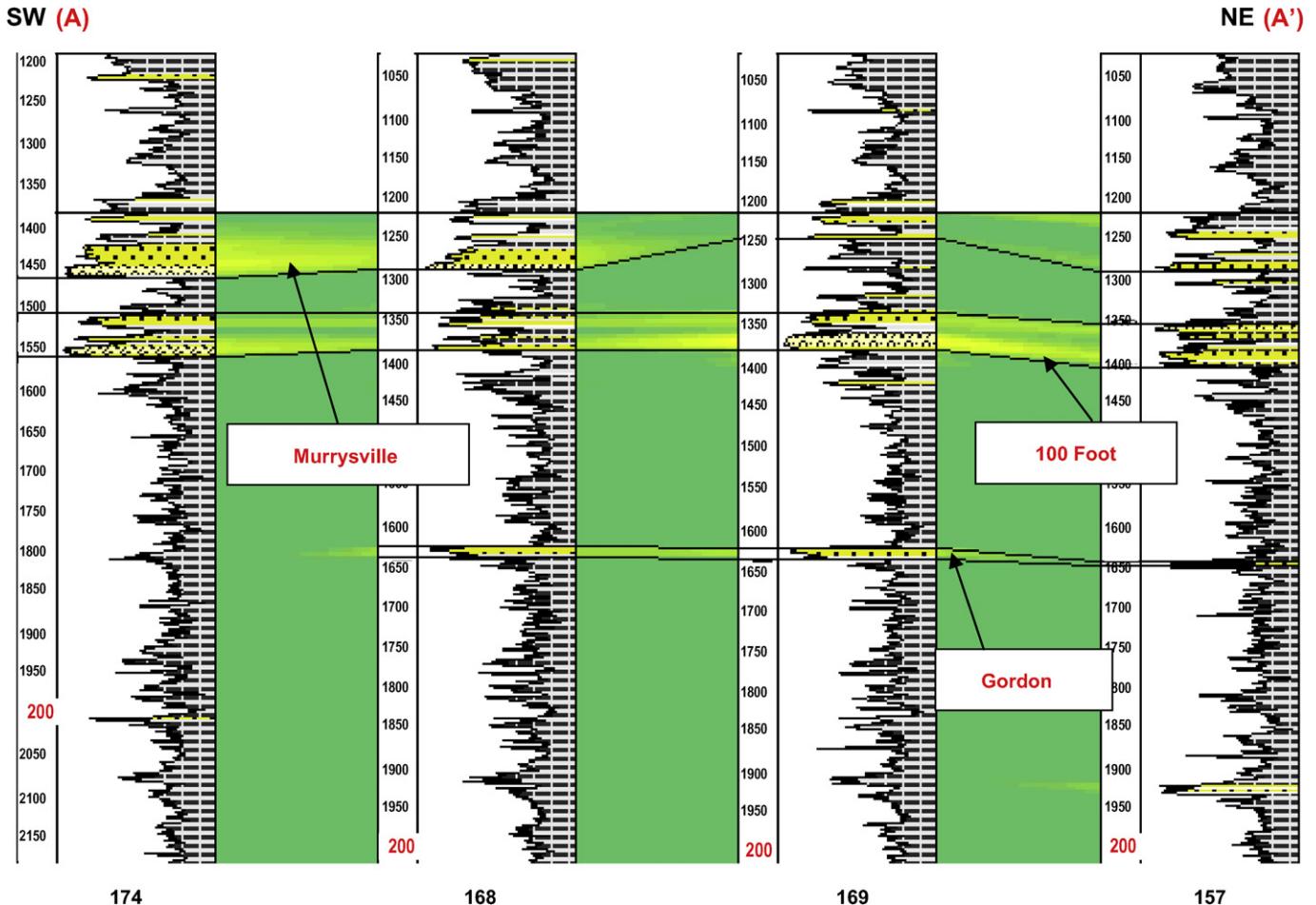


Fig. 3. Gamma Ray correlation of Murrysville/100 Foot/Gordon Sands (Upper zone).

where ρ_{mf} is the mud-filtrate density, S_{xo} is the mud-filtrate saturation in the invaded zone, and ρ_h is the invaded zone hydrocarbon density. In water-bearing zones where S_{xo} is equal to 1, ρ_h can be assumed to be equal to ρ_{mf} . For gas-bearing formations, it is reasonable to assume that ρ_{mf} is equal to 1.00. Therefore for practical reasons, 1.00 can be used as a general value for the term ρ_f (Bassiouni, 1994).

2.2.2. Neutron log

The neutron, a particle constituent of the atom, exhibits a high penetrating potential because of its lack of electric charge. Because of its penetration power, the neutron plays an important role in well logging applications.

Because hydrogen is responsible for most of the slowing-down effect, measuring the concentration of epithermal neutrons indicates the hydrogen concentration in the material. In shale-free, water-bearing formations, the hydrogen concentration reflects the porosity and lithology.

The neutron log measures the hydrogen concentration or hydrogen index in the rock. The tool emits neutrons of a known energy level, and measures the energy of neutrons reflected from the rock. Because energy is lost most easily to particles of similar mass, the hydrogen concentration can be delineated.

2.2.3. Resistivity log

This log records the resistance of interstitial fluids to the flow of an electric current, either transmitted directly to the rock through

an electrode, or magnetically induced deeper into the formation from the hole, as it is the case with induction logs. The term "deep" refers to horizontal distance from the well bore. Varying the length of the tool and focusing the induced current measure resistivities at different depths into the rock. Several resistivity and induction curves are commonly shown on the same track.

Resistivity logs are used for evaluation of fluids within formations. They can also be used for identification of coals (high resistance), thin limestones in shale (high resistance) and bentonites (low resistance). In turn, for wells where few types of logs have been run, the resistivity log may be useful for picking tops and bottoms of formations, and for correlating between wells. Freshwater-saturated porous rocks have high resistivities, so the log can be used in these cases to separate shales from porous media (sandstones and carbonates).

3. Some fundamentals on artificial neural networks

An artificial neural network is a computing parallel scheme based on the biological neural system configuration (Mohaghegh et al., 2002; Poulton, 2002; Faucett, 1994; White et al., 1995). A neuron in the brain (Fig. 6) is a unique piece of equipment that consists of three types of components called dendrites, cell body or soma, and axon. Dendrites are the sensitive part of neuron that receive signal from other neuron. Soma calculates and sums the signals and transmitted

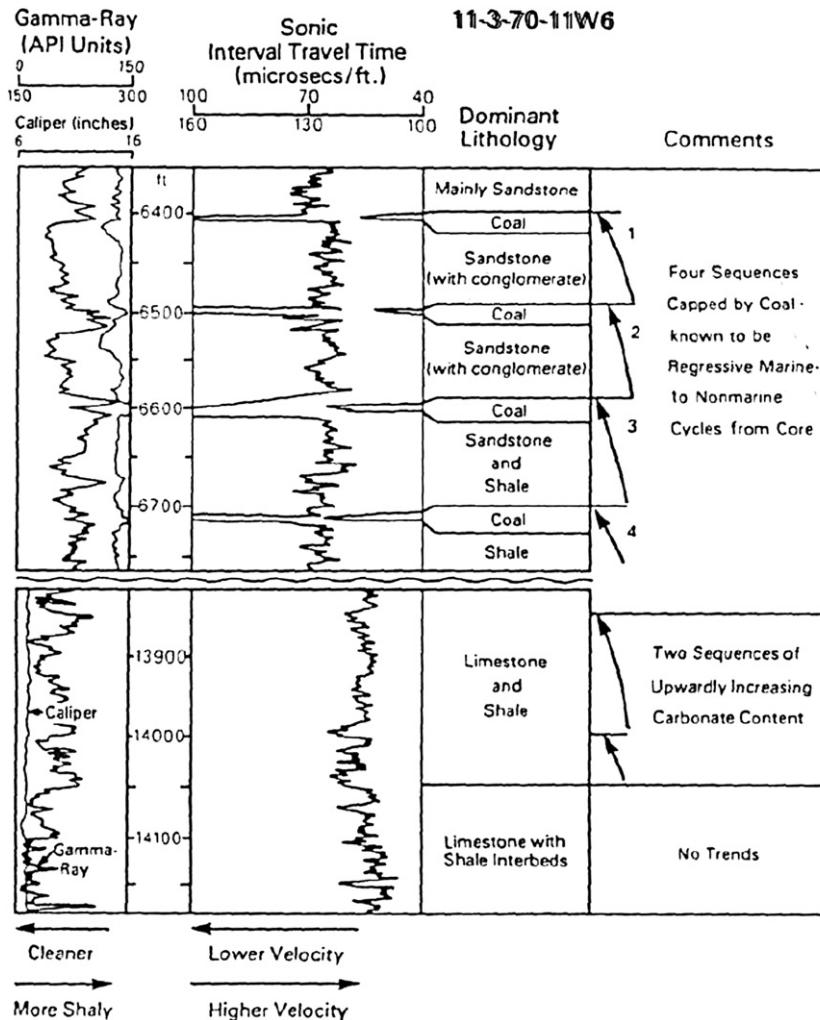


Fig. 4. Gamma ray and sonic logs from the Alberta basin, and their response to different lithologies (adopted from Cant, 1992).

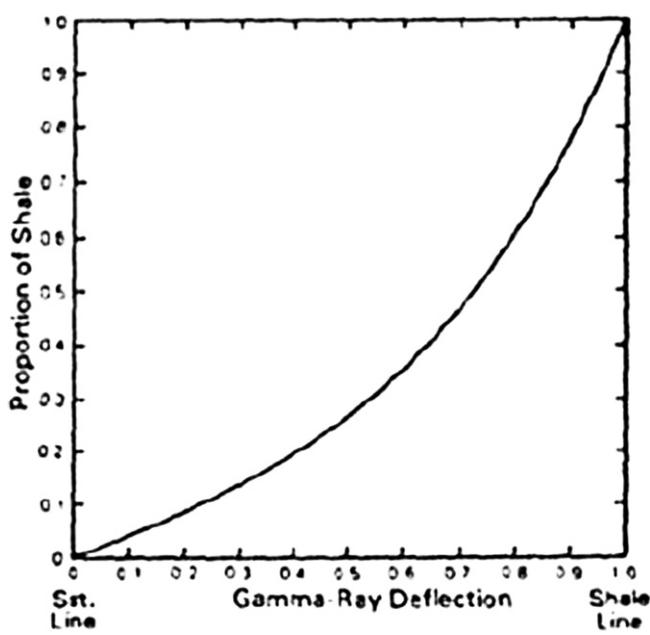


Fig. 5. Relationship between gamma-ray deflection and proportion of shale (adopted from Cant, 1992).

to other cells through axon. The biological neuron carries information and transfers to other neuron in a chain of networks.

The process of growing and learning in the human brain is one of creating neural connections through an associative process based on patterns received from the senses. The artificial neuron imitates the functions of the three components of the biological neuron and their unique process of learning.

The process of learning in an artificial neural network is similar to the process that is performed in the human brain. Artificial neural networks are systems mathematically designed to receive, process, and transmit information. Information processing occurs at the neurons. The simple neuron (Fig. 7) consists of an input layer, activation function, and output layer. The input layer receives signals from the external environment (or other neuron). The activation function is the internal neuron that calculates and sums the input signals.

Table 1
Common values of matrix density for different type of rocks.

Rock type	Matrix density (g/cm^3)
Sand or sandstone	2.65
Limestone	2.71
Dolomite	2.87
Anhydrite	2.98

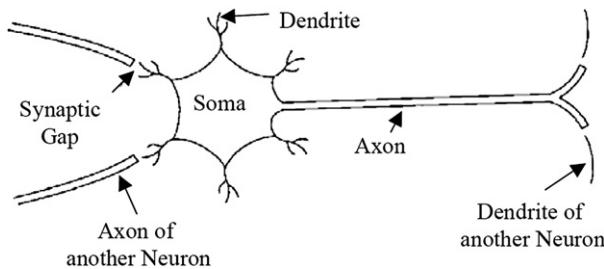


Fig. 6. Sketch of a biological neuron (adopted and modified from [Faucett, 1994](#)).

These signals are then transmitted to an output layer and retransmitted. The input layer, activation function, and output layer in artificial neuron are similar to the function of dendrites, soma, and axon in the biological neuron. Still, as seen in Fig. 7, an artificial network has many inputs but only one output.

3.1. Mechanics of neural networks

Artificial neural networks are generalizations of mathematical models of human cognition ([Faucett, 1994](#)). They function based on the following assumptions:

- Information processing occurs at many simple elements (neurons).
- Signals are passed between neurons over connection links.
- Each connection link has an associated weight, which, in a typical neural net, multiplies the transmitted signal.
- Each neuron applies an activation function to its net input to determine its output signal.

Assume we have n input units, X_1, \dots, X_n with input signals x_1, \dots, x_n . When the network receive the signals (x_i) from input units (X_i), the net input to output (y_{inj}) is calculated by summing the weighted input signals as follows:

$$\sum_{i=1}^n x_i w_{ij}$$

The matrix multiplication method for calculating the net input is shown in the equation below:

$$y_{inj} = \sum_{i=1}^n x_i w_{ij}$$

where w_{ij} is the connection weights of input unit x_i and output unit y_j .

Table 2
Common activation functions.

Function	Definition
Identity	x
Logistic	$\frac{1}{1+e^{-x}}$
Hyperbolic	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$
Exponential	e^{-x}
Softmax	$\frac{e^{x_i}}{\sum_i e^{x_i}}$
Unit sum	$\frac{x}{\sum_i x_i}$
Square root	\sqrt{x}
Sine	$\sin(x)$

The network output (y_j) is calculated using the activation function $f(x)$. In which $y_j = f(x)$, where x is y_{inj} . The computed weight from the training is stored and will become the information for the future application. [Table 2](#) shows the most common types of activation functions.

Neural networks can be divided into three architectures, namely single layer, multilayer network ([Fig. 8](#)) and competitive layer. The number of layers in a net is defined based on the number of interconnected weights in the neuron. A single layer network consists only of one layer of connected weights, whereas, multilayer networks consist of more than one layer of connection weights. The network also consists of an additional layer called a hidden layer.

Multilayer networks can solve more complicated problems than those solved by a single layer network. Both networks are also called “feed forward” networks where the signal flows from the input units to the output units in a forward direction. For example, a recurrent network is a feedback network with a closed-loop signal from the unit back to itself.

3.2. Learning mechanisms for artificial neural networks

Human beings are capable of learning with the aid of a teacher or on their own. The first case is considered supervised learning, the latter is unsupervised learning. In supervised learning, a teacher supplies both the material to be learned and corrects the student when the response to the material is incorrect. In unsupervised learning, the student receives the material to be learned and has to draw his/her own conclusion as to what the material means. The most basic division of artificial neural networks is whether they learn in a supervised or unsupervised mode. In supervised mode, common in most applications, the user supplies a set of patterns that the neural network should learn. These patterns are then associated with responses dealing with a classification or an

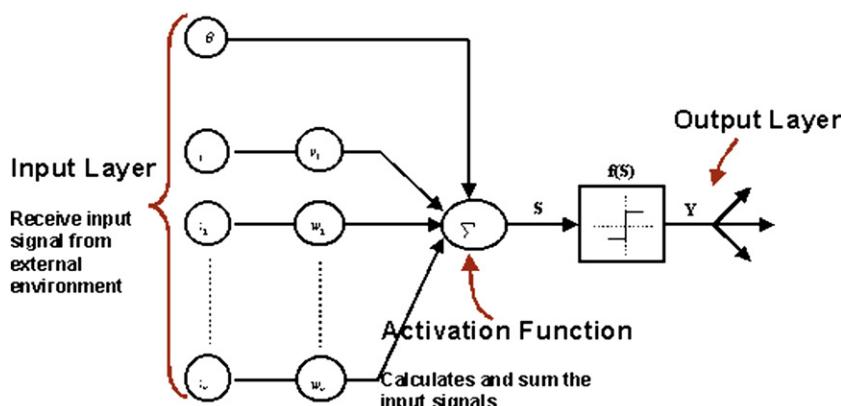


Fig. 7. Structure of a simple artificial neuron (adopted and modified from [Faucett, 1994](#)).

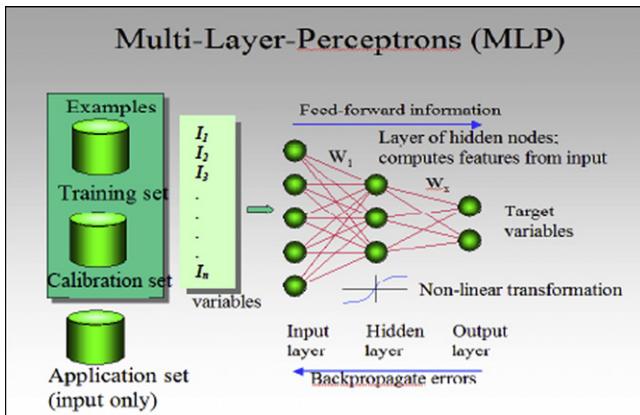


Fig. 8. Architecture of a multilayer perceptron (adopted and modified from Amizadeh, 2001).

estimation of a parameter value. In the unsupervised mode, the network is supplied with the set of patterns to learn, but it is not supplied with a prior classification or parameter value (Poulton, 2002).

While the architecture is important in defining how the network will function, the specific learning algorithm used to train the network defines how it will learn. The learning algorithm involves the input signals, the weights, the activation function (table 2), and the outputs. The learning algorithm specifies what mathematical operations will be performed by each neuron, how errors are calculated, and how connection weights adapt.

Learning starts with the assignment of random weights. The output is then calculated and the error is estimated. This error is used to update the weights until the stopping criterion is reached (stopping criteria is usually the average error). Neural networks learning algorithms can be divided into two main groups that are supervised and unsupervised.

3.2.1. Supervised learning

Supervised learning (Fig. 9) learns based on the target value or the desired outputs. During training the network tries to match the outputs with the desired target values. This method has two sub varieties called auto-associative and hetero-associative. In auto-associative learning, the target values are the same as the inputs, whereas in hetero-associative learning, the targets are generally different from the inputs. One of the supervised neural network models that are most commonly used, is the Backpropagation Network, which uses the Backpropagation-learning algorithm. Backpropagation algorithm is one of the well-known algorithms in

neural networks. Backpropagation algorithm has been popularized in 1980s as a euphemism for generalized delta rule. Backpropagation of errors or generalized delta rule is a decent method to minimize the total squared error of the output computed by the net (Faucett, 1994).

3.2.2. Unsupervised learning

Unsupervised learning (Fig. 9) method is not given any target value. A desired output of the network is unknown. During training the network performs some kind of data compression such as dimensionality reduction or clustering. The network learns the distribution of patterns and makes a classification of that pattern where, similar patterns are assigned to the same output cluster. Kohonen network is the best example of unsupervised learning network. According to Sarle (1997), Kohonen network refers to three types of networks that are Vector Quantization, Self-Organizing Map and Learning Vector Quantization.

3.2.3. Training the network

Training the network is time consuming. It usually learns after several epochs, depending on how large the network is. Thus, a large network requires more training time than a small one. Basically, the network is trained for several epochs and stopped after reaching the maximum epoch. For the same reason minimum error tolerance is used provided that the differences between network output and known outcome are less than the specified value (see for example Pofahl et al., 1998). We could also stop the training after the network meets certain stopping criteria. During training the network might learn too much. This problem is referred to as overfitting. Overfitting is a critical problem in most all-standard neural networks architectures. Furthermore, neural networks and other artificial intelligence machine learning models are prone to overfitting (Lawrence et al., 1997). One of the solutions is early stopping (Sarle, 1997), but this approach needs more critical intention, as this problem is harder than expected (Lawrence et al., 1997). The stopping criterion is also another issue to consider in preventing overfitting (Prechelt, 1998). To crack this problem during training, a validation set is used instead of a training data set. After a few epochs, the network is tested with the verification data. The training is stopped as soon as the error on verification set increases rapidly higher than the last time it was checked (Prechelt, 1998).

4. Methodology

The methodology applied in this work builds upon the approaches presented by Bhuiyan (2001), and Mohaghegh et al. (2002) with three major modifications. Bhuiyan (2001), developed a neural network to generate synthetic Magnetic Resonance Imaging (MRI) logs. Mohaghegh et al. (2002), used an intelligent software to build, learn, and reproduce the analyzing capabilities of the engineer on the remaining wells and also to generate those logs that were missed and that were necessary for analysis. Following three items significantly distinguishes this work with those in the literature:

1. The robustness of the general methodology of generating synthetic well logs using neural networks is further proved by applying the technology to a new field in the north-east of the United States where new depositional environment and geology prevails when compared to the previous studies that were applied in Texas.
2. The results of generating synthetic well logs using neural networks (Figs. 15–20) are compared with more conventional technologies such as multiple-regression (Fig. 13) and the superiority of neural network technology is demonstrated.

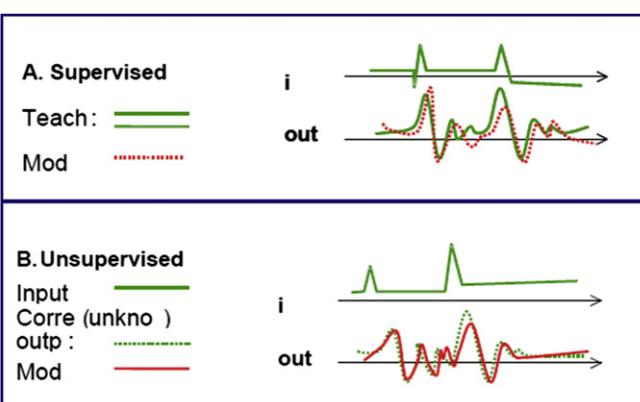


Fig. 9. Supervised and unsupervised scheme (adopted and modified from Jaeger, 2002).

ID	DEPTH	LAT	LONG	RILD	DEN	NPRL	GRGC	DNND
157	2000	40.5859	79.4719	32.87	2.70	8.79	144.66	15775.31
157	2000	40.5859	79.4719	31.73	2.71	9.08	145.10	15718.19
157	1999	40.5859	79.4719	30.91	2.71	9.38	142.85	15628.33
157	1999	40.5859	79.4719	30.82	2.71	9.58	141.16	15647.24
157	1998	40.5859	79.4719	31.57	2.71	9.61	142.10	15765.67
157	1998	40.5859	79.4719	32.53	2.69	9.43	142.63	15928.85

Fig. 10. Segment of the matrix prepared for well 157.

3. General Regression Neural Network is introduced as the most appropriate neural network paradigm for generation of synthetic well logs as compared with previous paradigms used in the studies cited in the literature (mostly Backpropagation networks).

Different neural network architectures were attempted in order to find, build, apply, and examine the best model for the data set. Outputs generated by the model were evaluated in terms of R-squared (R^2). R^2 , the relative predictive power of a model, is a descriptive measure between 0 and 1, which is defined as:

$$R^2 = 1 - \frac{SSE}{SS_{YY}},$$

where

$$SSE = \sum (y - \hat{y})^2$$

$$SS_{YY} = \sum (y - \bar{y})^2$$

y = actual value

\hat{y} = the predicted value of y ,

and

\bar{y} = the mean of the y values.

R^2 values can be interpreted, as indicators of how good are the results produced by the network. The closer R^2 is to one the better the model is. However, it is the user who ultimately decides if the network is working properly or not.

5. Data preparation

The first step of the data preparation was to identify the depth of the producing units in the area: the Murrysville, the 100 Foot, the Gordon, the Speechley, and the 2nd Bradford formations. Taken in account the sedimentary characteristics of these formations, the studied interval can be divided into two segments with different lithologic and petrophysic characteristics. In addition, this division allowed better visualization of the actual logs and the logs generated by the network. The chosen intervals were from 1000 to 2000 feet and from 2500 to 3500 feet. Data from logs were input in a spreadsheet, in order to prepare a matrix for use during the development process. The matrix for each well contained the well name, the depths, the longitude, the latitude, and the values of the resistivity (RILD), density (DEN), gamma ray (GRGC), and neutron (DNND) logs. Fig. 10 is an example of the arrangement used to prepare the matrices.

6. Neural network model development

Development of the neural network model was completed using four wells that included gamma ray, density, neutron, and resistivity logs. Different training algorithms were attempted until the best results in terms of R^2 and matching of the synthetic logs generated by the network versus the actual logs were achieved. Although in previous studies similar to this, others have used the Backpropagation algorithm^{2,3}, it was found that for this particular field using a Generalized Regression Neural Network (GRNN) would provide the best results.

General Regression Neural Network (GRNN) is Donald Specht's term for a neural network invented by him in 1990 (Specht, 1991). The General Regression Neural Network (GRNN) is a memory-based network that provides estimates of continuous variables and converges to the underlying regression surface. GRNNs are based on the estimation of probability density functions, feature fast training times and can model nonlinear functions.

The GRNN is a one-pass learning algorithm with a highly parallel structure. It is that, even with sparse data in a multidimensional measurement space, the algorithm provides smooth transitions from one observed value to another. The algorithmic form can be used for any regression problem in which an assumption of linearity is not justified. GRNN can be thought as a normalized RBF (Radial Basis Functions) network in which there is a hidden unit centered at every training case. These RBF units are usually probability density functions such as the Gaussian. The only weights that need to be learned are the widths of the RBF units. These widths are called "smoothing parameters". The main drawback of GRNN is that it suffers badly from the curse of dimensionality.

GRNN cannot ignore irrelevant inputs without major modifications to the basic algorithm. So GRNN is not likely to be the top choice if there are more than 5 or 6 no redundant inputs.

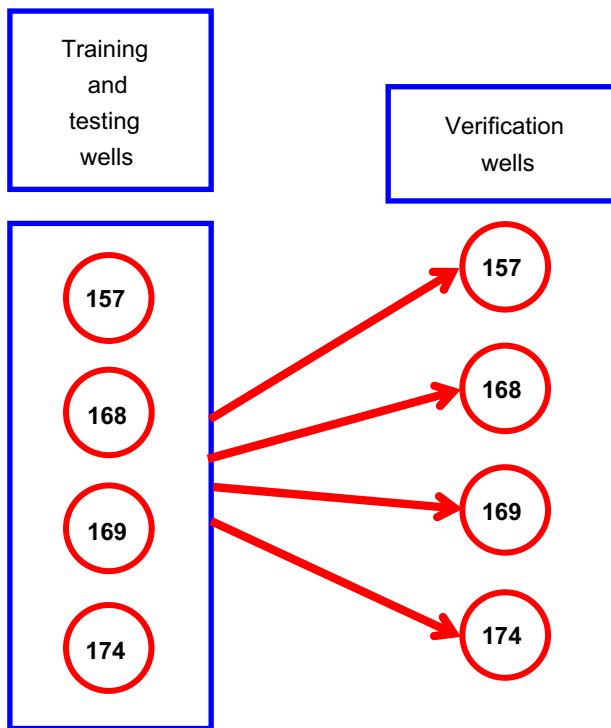


Fig. 11. Schematic diagram showing distribution of wells used for training, calibration and verification data sets through exercise 1.

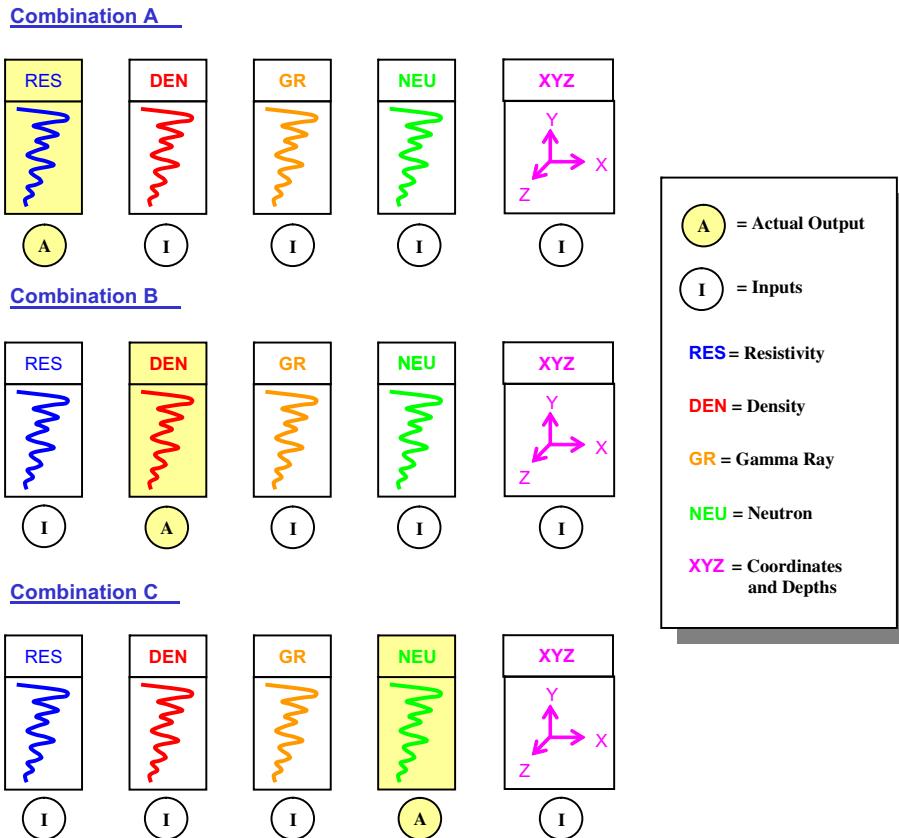


Fig. 12. Different combinations of inputs/outputs used for development of neural network model.

The regression of a dependent variable, Y , on an independent variable, X , is the computation of the most probable value of Y for each value of X based on a finite number of possibly noisy measurements of X and the associated values of Y . The variables X and Y are usually vectors.

In order to implement system identification, it is usually necessary to assume some functional form. In the case of linear regression, for example, the output Y is assumed to be a linear function of the input, and the unknown parameters, a_i , are linear coefficients. The procedure presented in Donald F. Specht's article (Specht, 1991) does not need to assume a specific functional form.

A Euclidean distance is estimated between an input vector and the weights, which are then rescaled by the spreading factor. The radial basis output is then the exponential of the negatively weighted distance.

The GRNN equation is:

$$Y(X) = \frac{\sum_{i=1}^n Y_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}$$

The estimate $Y(X)$ can be visualized as a weighted average of all of the observed values, Y_i , where each observed value is weighted exponentially according to its Euclidian distance from X . $Y(X)$ is simply the sum of Gaussian distributions centered at each training sample. However the sum is not limited to being Gaussian.

In this theory, σ is the smoothing factor and can be seen as the spread of the Gaussian bell. The network here used consisted of three layers: an input layer made up of 7 neurons, a hidden layer made up of 7000 neurons, and finally an output layer consisting of only one neuron. The smoothing factor applied was kept at 0.122.

Training, calibration and verification were carried out through two different exercises that are described as follows:

6.1. Exercise 1: four wells combined

In this exercise the entire set of data, consisting of 4 wells, was used during training, calibration and verification of the network and then each one of these wells were used to verify the trained network as shown in Fig. 11.

The data brought into the network as inputs/outputs pairs were the locations of the wells (in terms of latitude and longitude), Depths, Deep Induction (RILD) log values, Density (DEN) log values, Gamma Ray (GRGC) log values, and Neutron (DNND) log values.

Combinations of different inputs/outputs pairs were chosen to train the network (Fig. 12); at each combination one of the logs aforementioned was predicted from the other information. In combination "A" the resistivity log was used as an actual output while the density, the gamma ray, the neutron, and the coordinates and depths (XYZ) were used as inputs, in combination "B" the density log was used as an actual output while the resistivity, the gamma ray, the neutron, and XYZ were used as inputs, and in combination "C" the neutron log was used as an actual output while the resistivity, the gamma ray, the density, and XYZ were used as inputs. The percentages used for training, calibration and verification were 80%, 15%, and 5% respectively. Total of three combinations were used for exercise 1.

6.2. Exercise 2: three wells combined, one well out

Differing from exercise 1, this exercise used only three wells for training and development of the network while the fourth well,

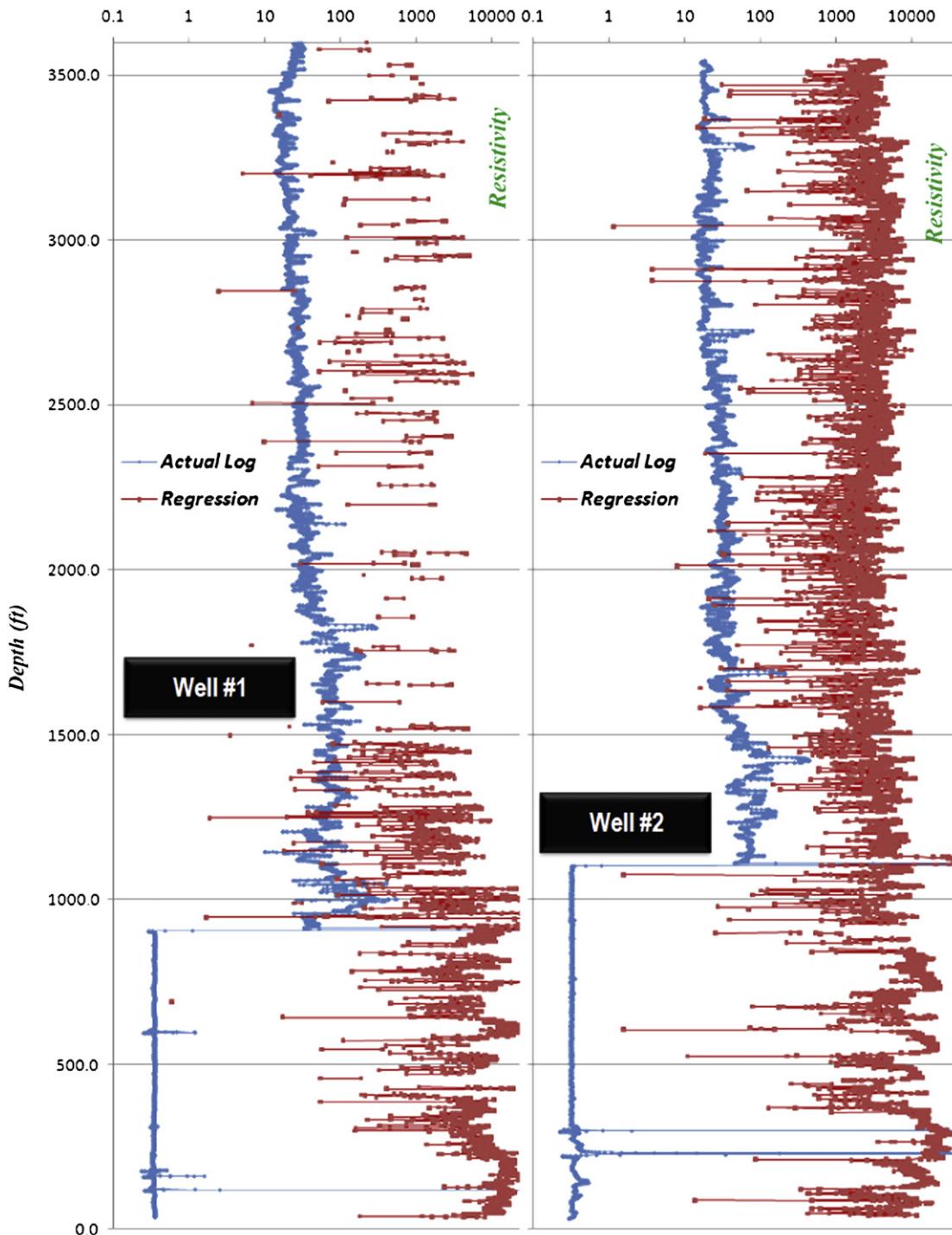


Fig. 13. Results of generating synthetic resistivity logs using multiple linear regression.

never used during training and calibration (blind well), was selected to generate synthetic logs out of the other three wells (verification). Therefore, wells 157, 168, and 169 were combined to generate logs in well 174; wells 157, 168, and 174 were combined to generate logs in well 169; wells 157, 174, and 169 were combined to generate logs in well 168; and wells 174, 169, and 168 were combined to generate logs in well 157. The percentages used were distributed 85% for training and 15% for calibration. Fig. 14 represents the combinations of wells used through this exercise (there were in total four possible combinations). Combinations of inputs/output used in exercise 1 were repeated for this exercise.

7. Results

The values of R^2 obtained for the training, calibration and verification data set, reflect the performance of the network at each of these stages. During training, the network uses a data set consisting of inputs and outputs, during calibration, the data set consists of a similar number of inputs and outputs, but in this case they are used to validate the network by verifying how well the network is performing on data that were never seen before. At this phase, the partially constructed network is checked at certain intervals of training by applying the calibration data set. Finally the verification



Fig. 14. Combinations of wells for training, testing and verification used in exercise 2.

set is used to prove the ability of the network to provide accurate results on the unseen data.

Although the most important criteria is to determine that the network is capable of generating logs with a certain degree of accuracy is indeed the degree of matching between the plots of the actual logs with the plots of logs generated by the network. It was observed that the best matching between actual and synthetic generated logs was obtained for high values of R^2 (higher than 0.7). Oppositely, matching was poor when low values of R^2 were obtained.

8. Generating synthetic logs using multiple-regression

To demonstrate the difference between using neural networks and more conventional methodologies such as multiple-regression to generate synthetic logs, an exercise was performed using the methodology described in “Exercise 1” (Fig. 11) where data from all four wells were combined during the analysis. The results for two of the wells are shown in Fig. 13 (combination A from Fig. 12). Similar results for other combination and for all wells were achieved for the other two wells. These results should be compared with results from neural networks that are presented in the following sections (Figs. 15–24).

9. Southern Pennsylvania logs

9.1. Exercise 1

R^2 values and correlation coefficients were obtained through exercise 1 for the data set of Southern Pennsylvania, at the upper and lower zones ranged from 0.79 to 0.96. These high values are also reflected in the correlation between the actual logs and the

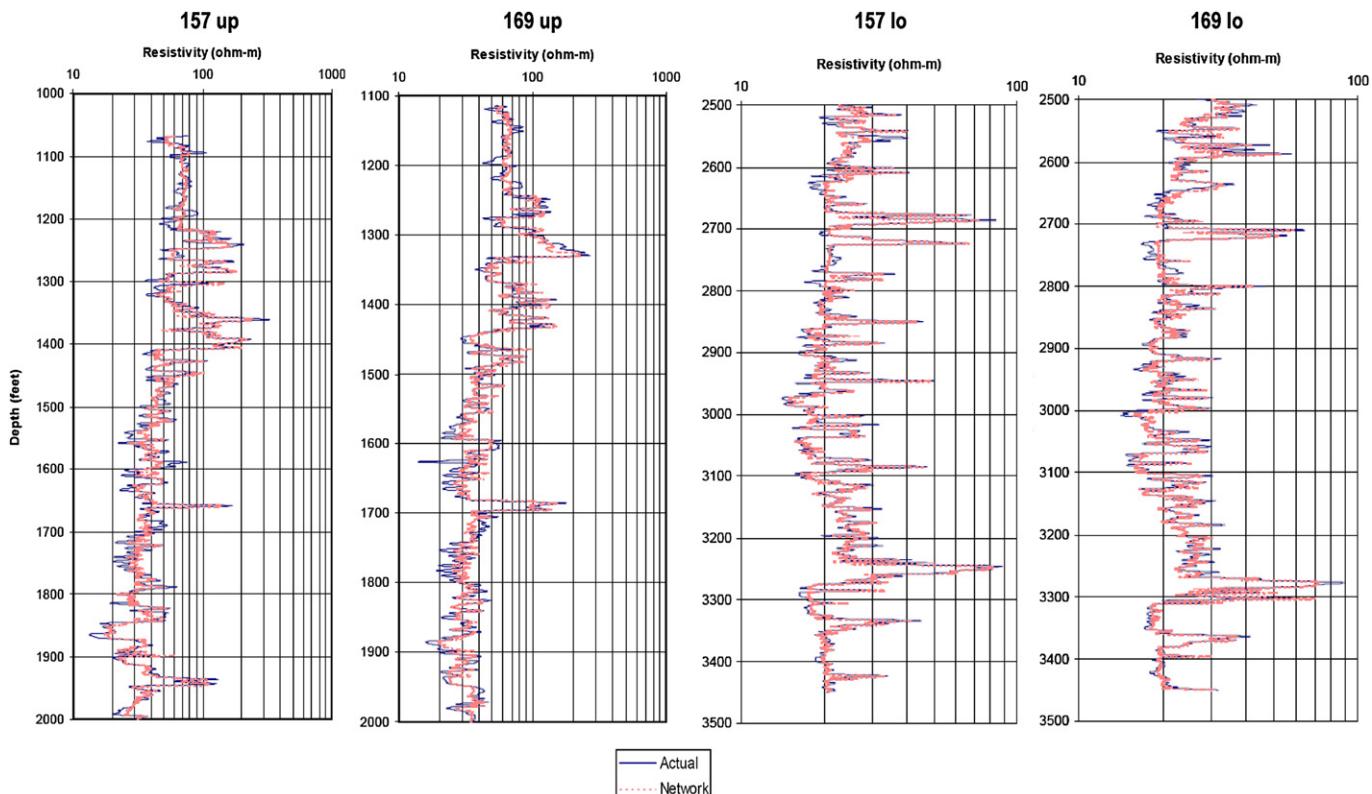


Fig. 15. Synthetic resistivity log generated during verification through exercise 1 and combination A of inputs and outputs. Wells 157 and 169 in upper and lower zones, Southern Pennsylvania.

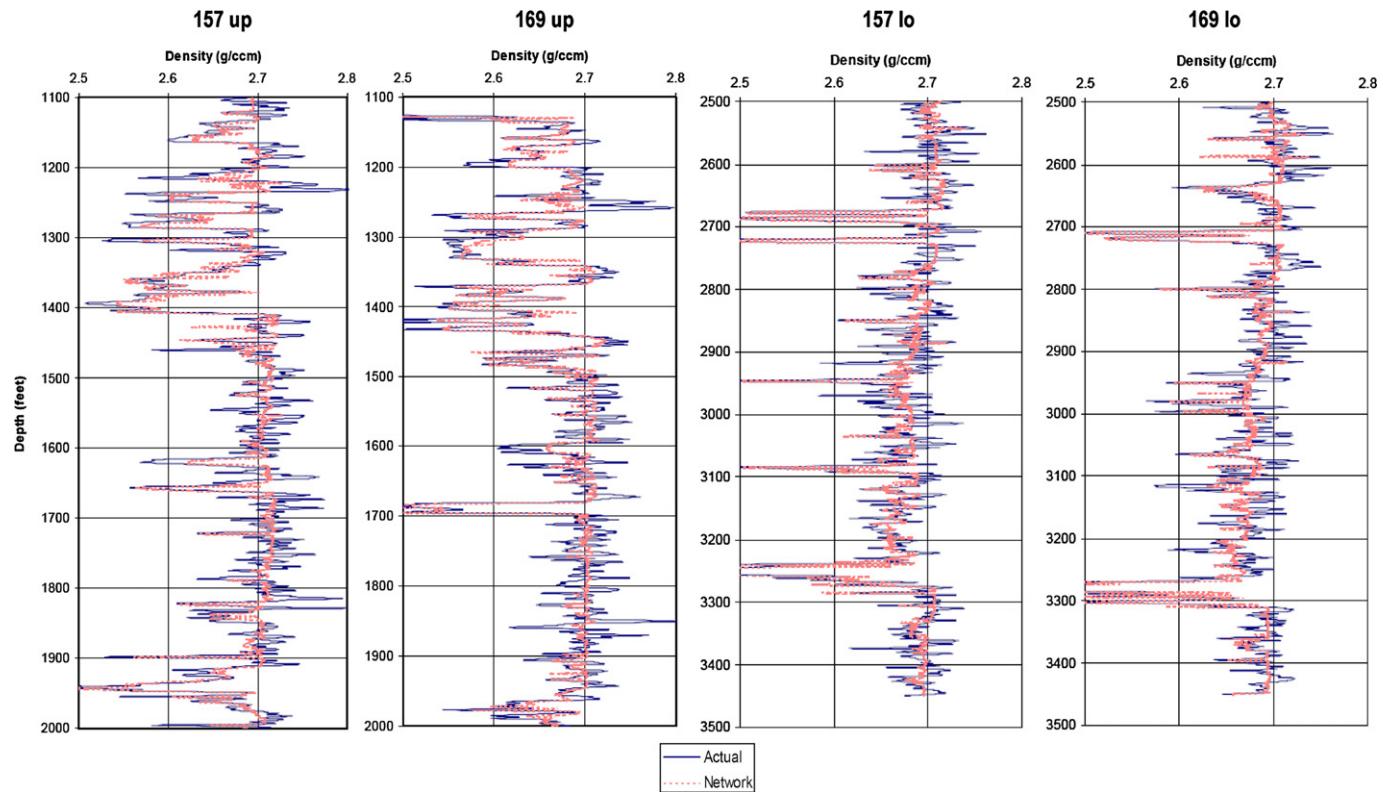


Fig. 16. Synthetic density log generated during verification through exercise 1 and combination A of inputs and outputs. Wells 157 and 169 in upper and lower zones, Southern Pennsylvania.

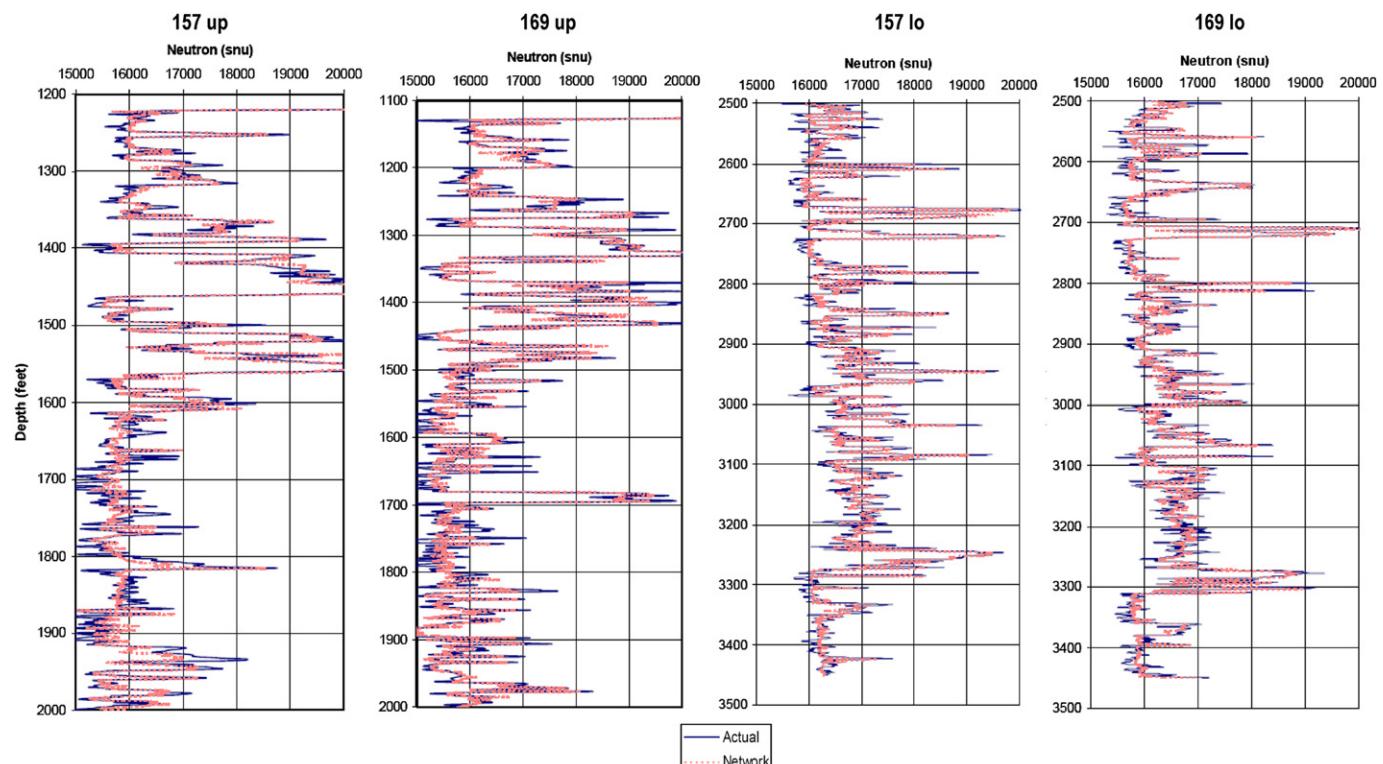


Fig. 17. Synthetic neutron log generated during verification through exercise 1 and combination A of inputs and outputs. Wells 157 and 169 in upper and lower zones, Southern Pennsylvania.

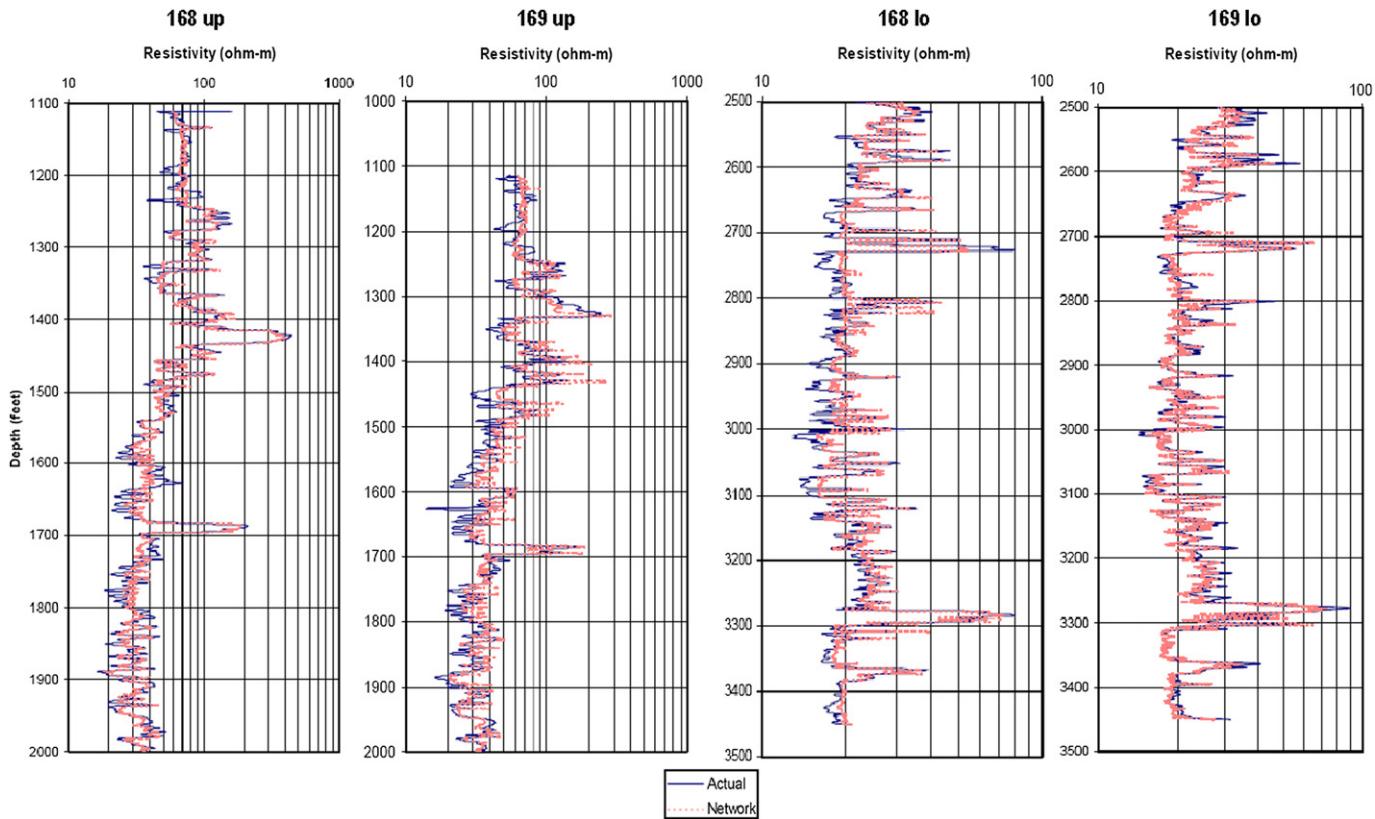


Fig. 18. Synthetic resistivity log generated during verification through exercise 2 and combination A of inputs and outputs. Wells 168 and 169 in upper and lower zones, Southern Pennsylvania.

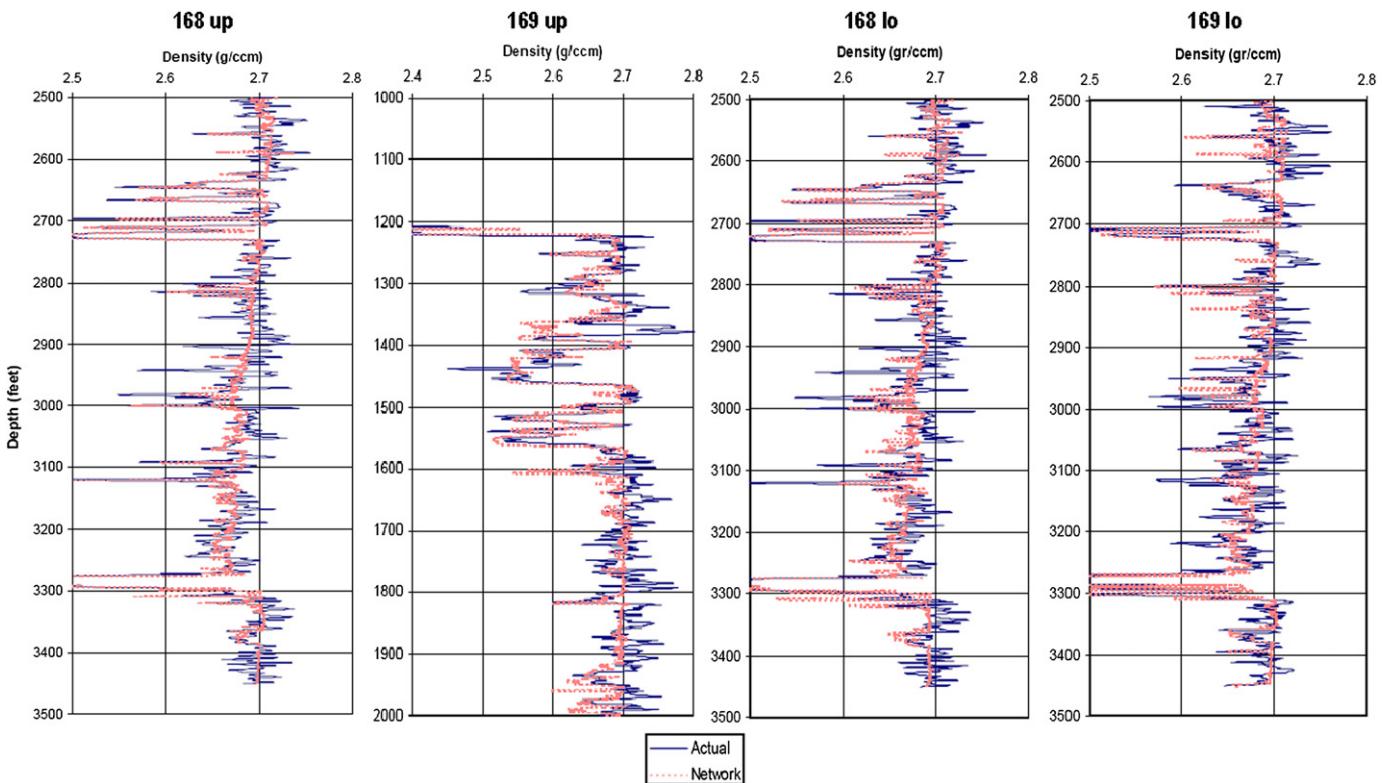


Fig. 19. Synthetic density log generated during verification through exercise 2 and combination A of inputs and outputs. Wells 168 and 169 in upper and lower zones, Southern Pennsylvania.

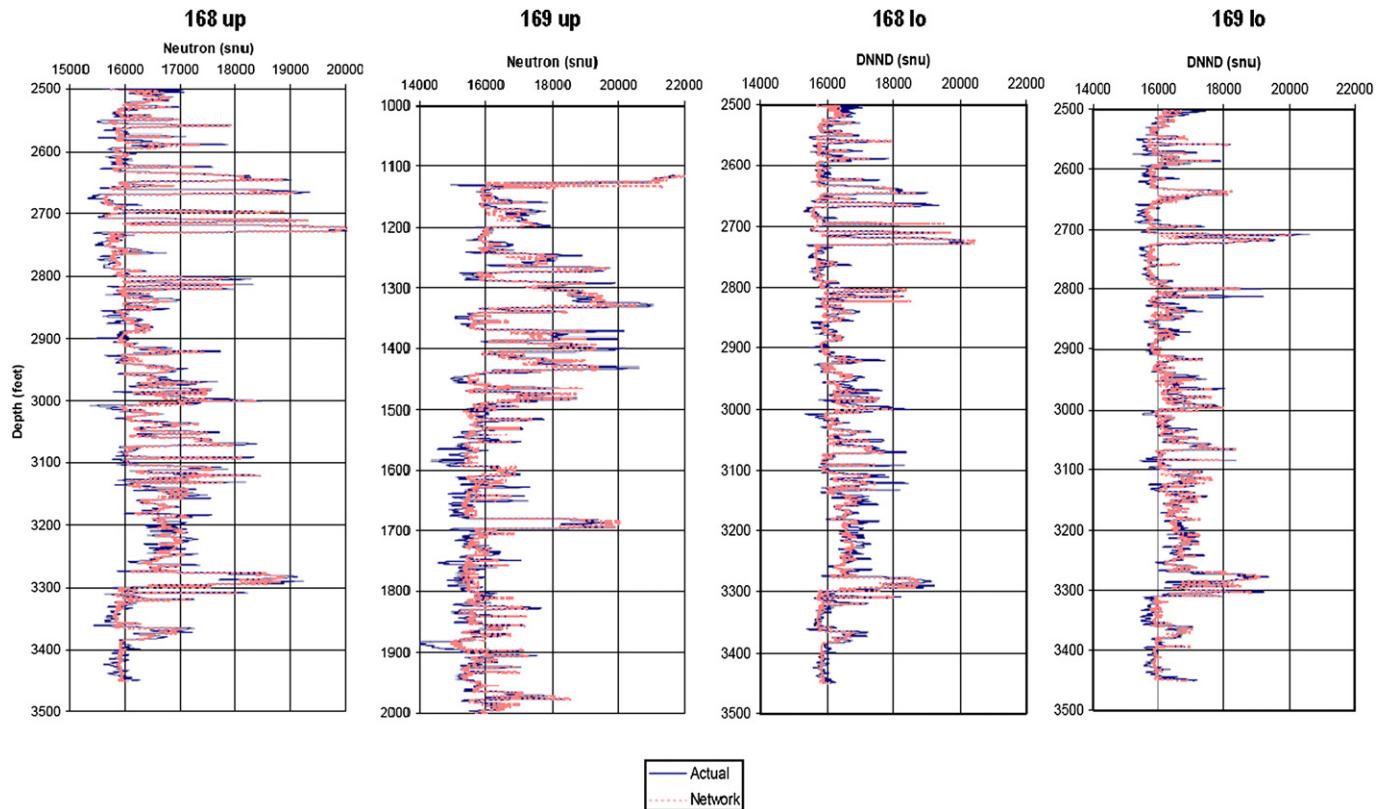


Fig. 20. Synthetic neutron log generated during verification through exercise 2 and combination A of inputs and outputs. Wells 168 and 169 in upper and lower zones, Southern Pennsylvania.

logs generated by the neural network model during verification. Figs. 15–17 show the relation between the actual logs and the logs generated by the neural network model during verification.

9.2. Exercise 2

Results obtained through exercise 1 for the upper and lower zone of the logs of Southern Pennsylvania area, substantially improved in comparison with the results obtained from Buffalo Valley data set. However, improvements were more significant when exercise 2 was performed. R^2 values rose from negative values to values over 0.8.

Despite some wells having R^2 values between 0.6 and 0.7, the synthetic logs generated by the network during verification still showed a high degree of accuracy (Figs. 18–20 at the end of this paper). It is important to mention that the first results for the verification data set at this interval were not successful. R^2 values obtained for this first effort were relatively low (about 0.45–0.58). The reason for these poor results was due to a portion of the data that initially included a log interval run at a cased segment. Hence, the values recorded were highly anomalous and consequently they did inject a significant error into the network. Once the data were cleaned of this error, results improved.

10. Discussion

Results discussed above for exercises 1 and 2, indicate that the best neural network model performance was obtained in general for combination "A". Combination "C" was ranked second and combination "B" represents the lowest performance (Figs. 21 and 22).

The inferior response of the neural network model to combination "B" is reflected in the way that the network captures the

deflections of the log. The high peaks of the log where high contrasts of density are present are captured with high accuracy, on the other hand, small changes cannot be captured accurately; in these cases the network averaged the values. An explanation for the lower degree of predictability of the density log is due to radioactive fluctuations (relative to the cesium source) during the logging operation. Hence, radiations can take different ways at each time and log response can vary from place to place.

For exercise 2 the best results, were obtained, for wells number 168 and 169 (Figs. 23 and 24 and 18–20 at the end). The reason is probably because these wells are located at the middle of the cross-section A–A', so the neural network model can interpolate

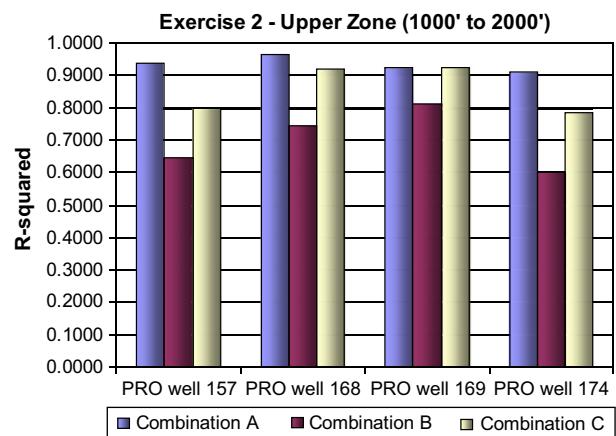


Fig. 21. R^2 values obtained for the upper zone of the Southern Pennsylvania data set, through exercise 1.

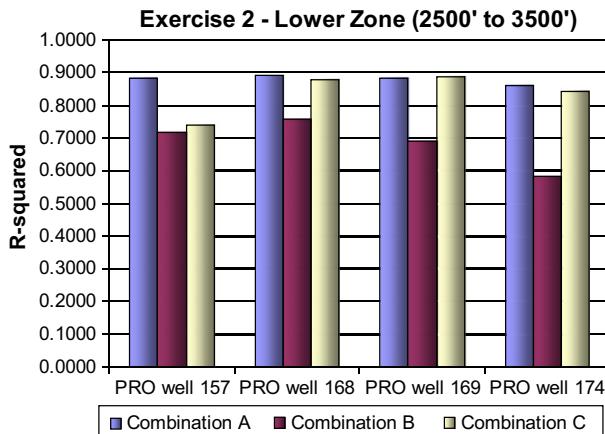


Fig. 22. R^2 values obtained for the lower zone of the Southern Pennsylvania data set, through exercise 1.

information from adjacent wells. An explanation for anomalous results obtained for combination "A" at the upper zone, could not be resolved by this study. However, it is important to be mentioned that, generally speaking, geology in the study area is simple, therefore, it is possible that for areas with more complex geology, this condition could change, and the location of the wells in terms of geometry, could not have any relation with the performance of the neural network.

A very interesting point to discuss in this study is the fact that poor results were obtained because of the quality of the data, as a consequence of digitalization of the logs in original tif format. These poor results are indicative of the high degree of sensitivity that neural network models have to the quality of the data. It is highly recommended for future studies involving generation of intelligent synthetic logs, to perform a strict quality control of data prior to building the neural network model, especially if the logs that will be used as inputs in the model are not directly obtained from the borehole, but have been digitized.

It was demonstrated that once it was realized that data from Buffalo Valley had low quality as a result of the imprecise digitizing, and they were replaced by data from Southern Pennsylvania, the effectiveness of the neural network models in predicting logs for exercises 1 and 2 improved considerably. Furthermore, before data from the upper zone of the Pennsylvanian data set were cleaned,

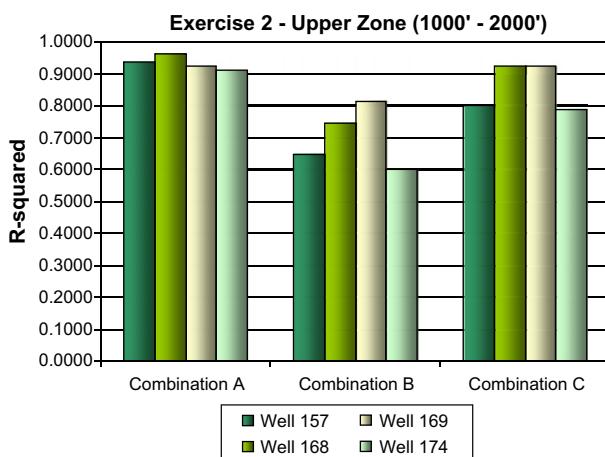


Fig. 23. R^2 values obtained for the upper zone of the Southern Pennsylvania data set, through exercise 2.

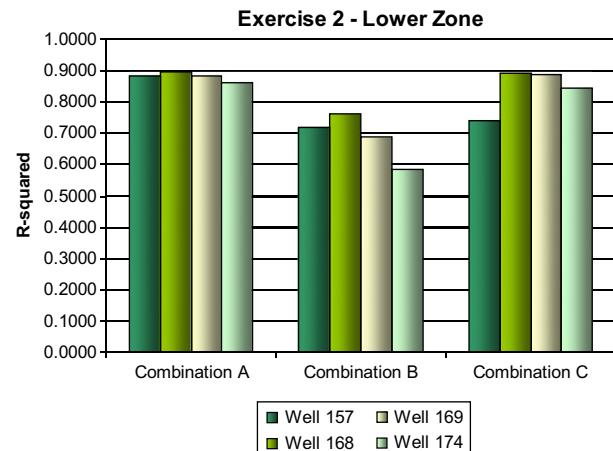


Fig. 24. R^2 values obtained for the lower zone of the Southern Pennsylvania data set, through exercise 2.

the results were poor, once the anomalous data were detected and cleaned, results improved noticeably.

Finally, it is important to mention that the main reason why logs of the Pennsylvanian area were split into two intervals was because it was desired to determine if heterogeneity of the rocks influenced the performance of the network model. It was demonstrated for all models built in this study that heterogeneities in lithology of the reservoir have low influence for the networks since R^2 values obtained for both zones were very similar to each other.

11. Conclusions

This work demonstrates that generation of synthetic logs with a reasonable degree of accuracy is possible by using a neural network model and following the methodology described herein. Furthermore it was demonstrated that using neural network to generate synthetic well logs is far superior to using conventional methodologies such as multiple-regression.

Three neural network models to predict resistivity, density, and neutron logs were built through exercises 1 and 2, as well as using different combinations of inputs and outputs, namely combination "A" to predict resistivity logs, combination "B" to predict density logs, and combination "C" to predict neutron logs.

Results indicate that the best performance was obtained for combination "A" of inputs and outputs, then for combination "C", and finally for combination "B". Therefore performance for combination "A" indicates that the resistivity log was the most predictable log. On the other hand performance in combination "B" demonstrates that density log was the least predictable as a consequence of radioactive fluctuations of the cesium source during the logging operation.

Results also demonstrate that in areas where geology is simple, as is the case of the study of this work, accuracy of synthetic logs may be favored by interpolation of data. Therefore, the best results were obtained for wells located in the central part of the cross-section studied. This condition could change in areas where geology presents more complexities.

It was demonstrated that in studies that involve generation of intelligent synthetic logs quality of data plays a very important role during the development of the neural network model. Quality of logs could be defective when logs are not available in digital format and have to be digitized. Therefore it is highly recommended for future works to perform a very careful quality control of the data before a neural network model is built.

It was also demonstrated that lithologic heterogeneities in the reservoir do not significantly affect performance of a neural network model in the generation of synthetic logs.

References

- Amizadeh, F., 2001. Soft Computing Information Applications for the Fusion and Analysis of Large Data Sets: Fact Inc./dGB-USA. University of Berkeley, California.
- Banchs, R., Michelena, R., 2002. From 3D seismic attributes to pseudo-well-log volumes using neural networks: practical considerations. *The Leading Edge* 21 (10), 996–1001.
- Bassiouni, Z., 1994. Theory, measurement, and interpretation of well logs. In: SPE Textbook Series, vol. 4, 372 pp.
- Bhuiyan, M., 2001. An Intelligent System's Approach to Reservoir Characterization in Cotton Valley. M.S. thesis. West Virginia University, Morgantown, West Virginia.
- Cant, D., 1992. Subsurface facies analysis. In: Walker, R., James, N. (Eds.), *Facies Models: Response to Sea Level Change*. Geological Association of Canada, 454 pp.
- Faucett, L., 1994. *Fundamentals of Neural Networks. Architectures, Algorithms, and Applications*. Prentice Hall, Englewood Cliffs, NJ.
- Jaeger, H., September/October 2002. Supervised Training of Recurrent Neural Networks. An INDY Math Booster Tutorial. AIS.
- Lawrence, S., Giles, C.L., Tsoi, A., 1997. Lessons in neural network training: training may be harder than expected. In: Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI-97. AAAI Press, Menlo Park, California, pp. 540–545.
- McBride, P., 2004. Facies Analysis of the Devonian Gordon Stray Sandstone in West Virginia. M.S. thesis. West Virginia University, Morgantown, West Virginia.
- Mohaghegh, S., Richardson, M., Ameri, S., 1998. Virtual magnetic imaging logs: generation of synthetic MRI logs from conventional well logs. SPE 51075. In: Proceedings, SPE Eastern Regional Conference, Nov. 9–11, Pittsburgh, PA.
- Mohaghegh, S., Popa, A., Koperna, G., 1999. Reducing the cost of field-scale log analysis using virtual intelligence techniques. SPE 57454. In: Proceedings, SPE Eastern Regional Conference, Oct. 21–22, Charleston, WV.
- Mohaghegh, S., Wilson T., Toro J., 2002. An intelligent system approach to reservoir characterization. Technical application. West Virginia University. Norwegian Formation Evaluation Society, 2002/2003 Season Stavanger May Meeting.
- Pofahl, W., Walczak, S., Rhone, E., Izenberg, S., Sep 1998. Use of an artificial neural network to predict length of stay in acute pancreatitis. *American Surgeon* 64 (9), 868–872.
- Poulton, M., 2002. Neural networks as an intelligence amplification tool: a review of applications. *Geophysics* 67 (3), 979–993.
- Prechelt, L., 1998. Early Stopping—but When? Neural Networks: Tricks of the Trade. <http://wwwipd.ira.uka.de/~prechelt/Biblio/>, pp. 55–69. Retrieved March 28, 2002 from World Wide Web:
- Sarle, W., 1997. Neural Network FAQ, Part 1 of 7: Introduction. Periodic Posting to the Usenet Newsgroup comp.ai.neural-nets. <http://ftp.sas.com/pub/neural/FAQ.html> URL:
- Specht, D., 1991. A general regression neural network. *IEEE Transactions on Neural Networks* 2 (6).
- Tonn, T., 2002. Neural network seismic reservoir characterization in a heavy oil reservoir. *The Leading Edge* 21 (3), 309–312.
- White, A., Molnar, D., Aminian, K., Mohaghegh, S., The Application of artificial neural networks to zone identification in a complex reservoir. SPE 30977, Proceedings, SPE Eastern Regional Conference and Exhibition September 19–21, 1995, Morgantown, West Virginia.