SPE 98013

Developing Synthetic Well Logs for the Upper Devonian Units in Southern Pennsylvania

Rolon, L. F., Chevron, Mohaghegh, S.D., Ameri, S., Gaskari, R. West Virginia University and McDaniel B. A., Dominion E&P, Inc.

ABSTRACT

A methodology to generate synthetic wireline logs is presented. Synthetic logs can help 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 and 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. As an important conclusion, it was demonstrated that quality of the data plays a very important role in developing a neural network model.

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 posses all the logs that are required to determine reservoir characteristics. This paper presents a methodology that can 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. 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 approach presented involves the use of artificial neural networks, as the main tool, in conjunction with data obtained from conventional wireline logs.

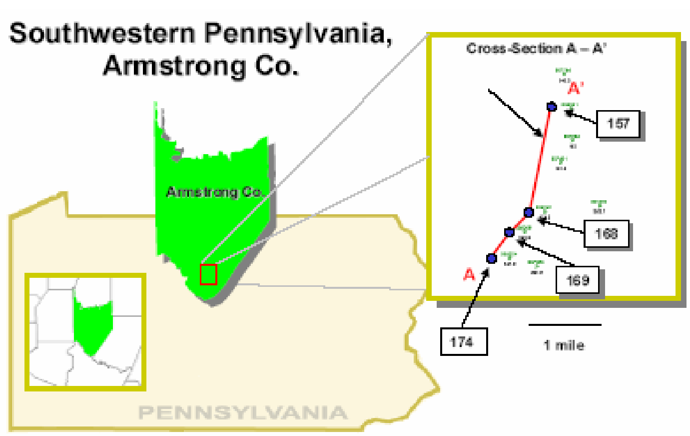


Figure 1. Location of the study area and wells analyzed (157, 168, 169, 174). A-A' represents the line of the cross-section.

The study area is located in southern Pennsylvania (figure 1). The sequence of rocks recorded in the set of logs used for this work belong to the Upper Devonian of southern Pennsylvania (the Venango and Bradford plays), which have 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 (Figures 2 and 3 at the end of the paper).

The 2nd Bradford and the Speechley sands 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 (Figure 3).

The Gordon sandstone is composed of a series of conglomerate, sandstone, 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 [1]. The conglomerate and coarse-grained sandstone of the upper upper shoreface and the foreshore tend to have more cement and therefore a lower porosity and permeability [1]. Gamma ray, density, and resistivity values are more or less constant, although changes in thickness occur from location to location (Figure 2).

METHODOLOGY

The methodology applied in this work is based on the approaches presented by Bhuiyan[2], and Mohaghegh et al.[3]. Bhuiyan[2], developed a neural network to generate synthetic magnetic resonance imaging logs (MRI). Mohaghegh et al. [3], used an intelligent software to build, learn, and reproduce the analyzing capabilities of the engineer on the remaining wells and also generate those logs that were missed and that were necessary for analysis. 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:

$$R^2 = 1 - \frac{SSE}{SS_{yy}}, \text{ where}$$

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

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

$y =$ 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.

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 Speechely, and the 2nd Bradford formations. Taken in account the sedimentary characteristics of these formations, the studied interval can be divided in 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. Figure 4 is an example of the arrangement used to prepare the matrices.

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

Figure 4. Segment of the matrix prepared for well 157.

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 would provide the best results.

The network 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:

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 Figure 5.

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.

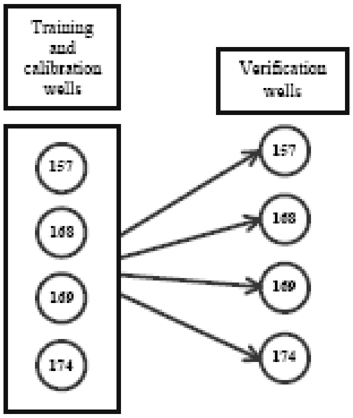


Figure 5. Schematic diagram showing distribution of wells used for training, calibration and verification datasets through exercise 1.

Combinations of different inputs/outputs pairs were chosen to train the network (Figure 6); 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.

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, never used during training and calibration, was selected to generate synthetic logs out of the other three wells (verification). Since the verification data set in this exercise consisted of a data set never used during training, the model was developed only with training and calibration data sets. Therefore, wells 157, 168, and 169 where combined to generate logs in well 174; wells 157, 168, and 174 where combined to generate logs in well 169; wells 157, 174, and 169 where combined to generate logs in well 168; and wells 174, 169, and 168 where combined to generate logs in well 157. The percentages used were distributed 85% for training and 15% for calibration. Figure 7 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.

RESULTS

The values of R^2 obtained for the training, calibration and verification dataset, 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 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.

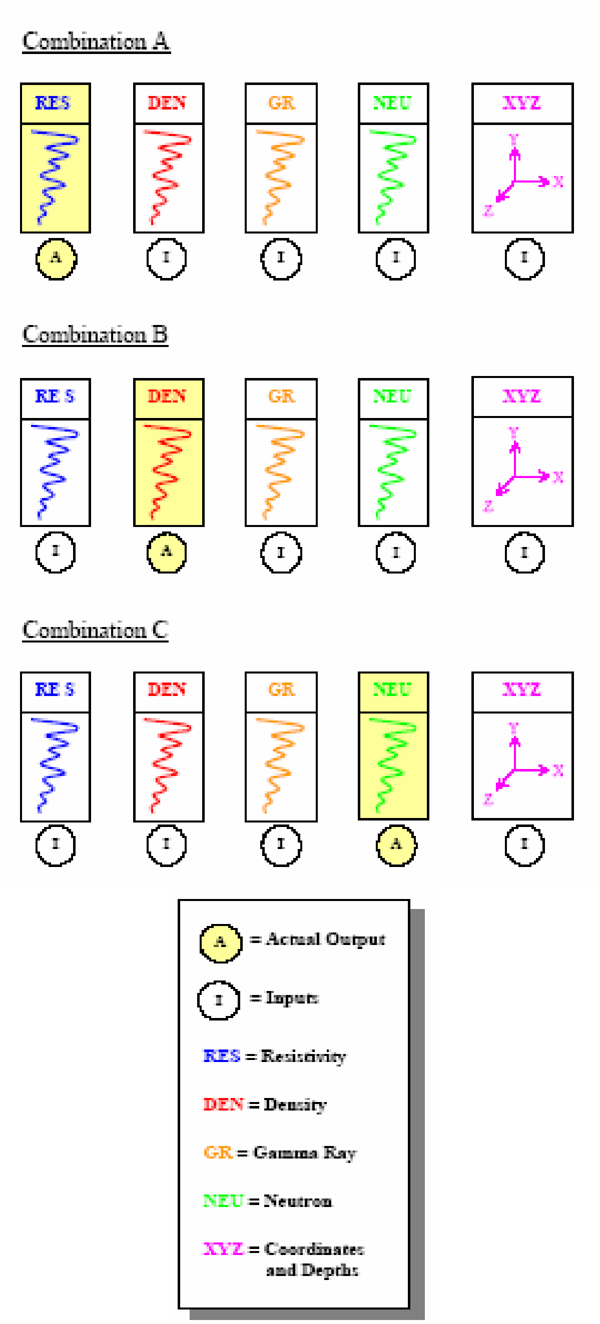


Figure 6. Different combinations of inputs/ outputs used for development of neural network model.

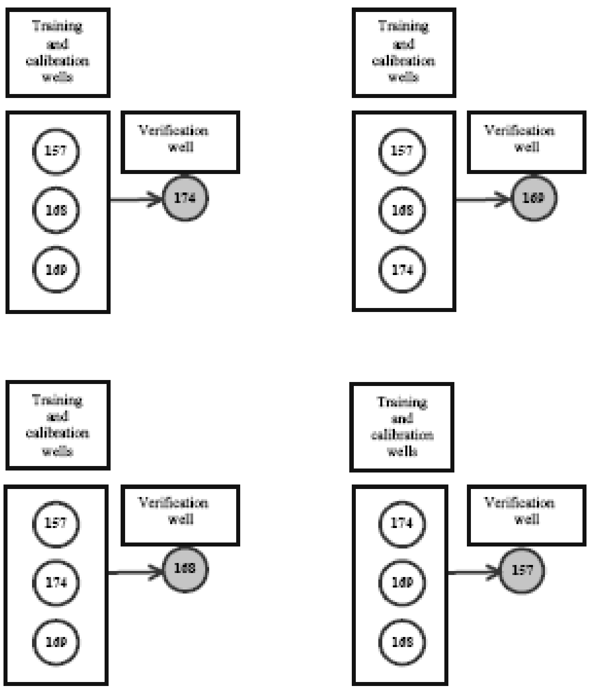


Figure 7. Combinations of wells for training, testing and verification used in exercise 2.

First Attempt: Data set from Buffalo Valley Field (New Mexico)

A first attempt to generate synthetic well-logs was done using a set of logs obtained from the website of the New Mexico Energy, Minerals and Natural Resources Department.

The log data set analyzed consisted of resistivity, gamma ray, density, and neutron porosity logs. Available logs were image files (tif) of the hard copies, therefore they had to be digitized in order to be converted to a format compatible with excel (las format). Figure 8 (at the end of this paper) shows the quality of the original tif files and the resulting curves after digitalization.

R^2 values obtained for the verification data set, out of the Buffalo Valley field data, were fairly good, when the neural network model was developed using four wells for training and calibration, and one of the wells used during training was used again for verification (exercise 1). Conversely, when the network was trained using three wells and was verified using a fourth well that has not been used during training (exercise 2) extremely poor R^2 values were observed.

Southern Pennsylvania Logs

Exercise 1

R^2 values and correlation coefficients were obtained through exercise 1 for the dataset 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 logs generated by the neural network model during verification.

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 had 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 (Figures 9 to 11 at the end of this paper). It is important to mention that the first results for the verification dataset at this interval were not successful. R^2 values obtained for this first effort were relatively low (about 0.45 to 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.

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 (Figures 11 and 12).

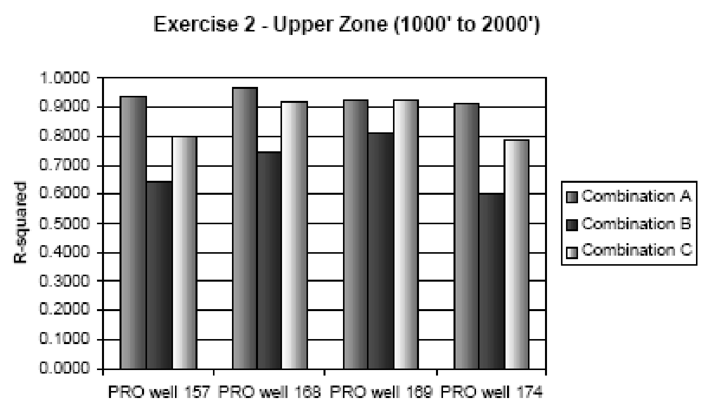


Figure 11. R2 values obtained for the upper zone of the Southern Pennsylvania data set, through exercise 1.

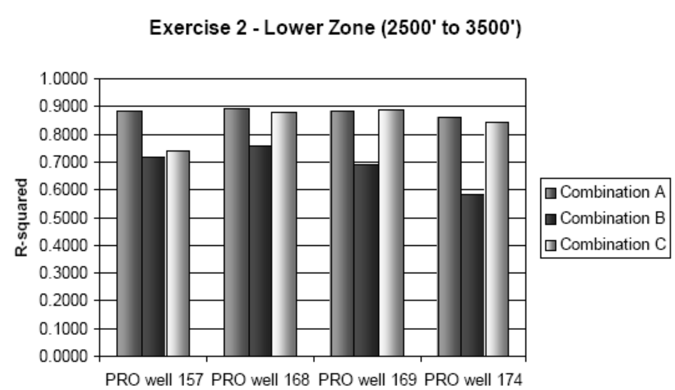


Figure 12. R2 values obtained for the lower zone of the Southern Pennsylvania data set, through exercise 1.

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 (Figures 13 and 14). 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 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.

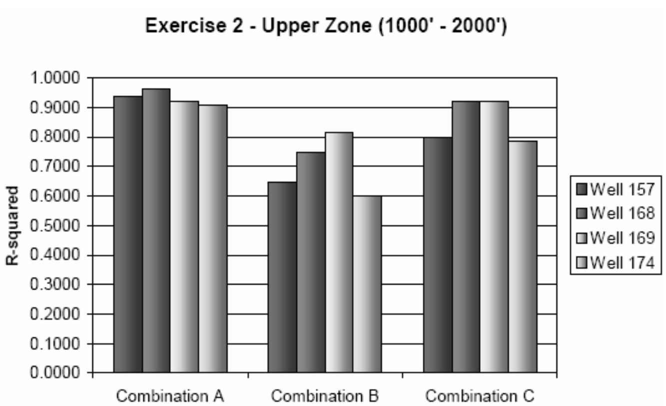


Figure 13. R2 values obtained for the upper zone of the Southern Pennsylvania data set, through exercise 2.

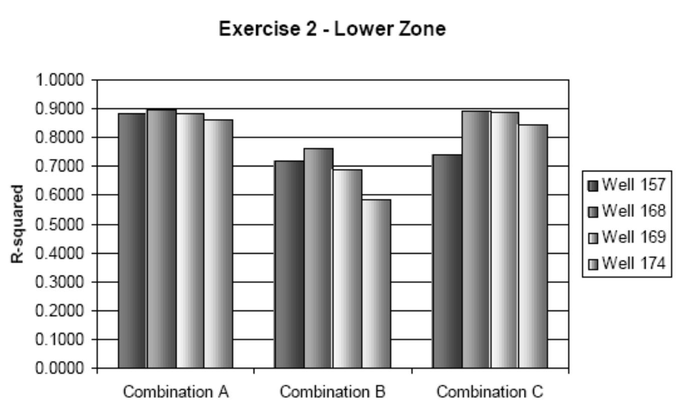


Figure 14. R2 values obtained for the lower zone of the Southern ennsylvania data set, through exercise 2.

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, 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 R2 values obtained for both zones were very similar to each other.

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.

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.

REFERENCES

1. McBride, P., 2004, Facies Analysis of the Devonian Gordon Stray Sandstone in West Virginia. M.S. Thesis, West Virginia University, Morgantown, West Virginia.
2. Bhuiyan, M., 2001. An Intelligent System's Approach to Reservoir Characterization in Cotton Valley. M.S. Thesis, West Virginia University, Morgantown, West Virgini.
3. Mohaghegh, S., Popa, A., Koperna, G., 1999 Reducing the Cost of Field-Scale Log Analysis Using Virtual Intelligence Techniques. SPE 57454, Proceedings, SPE Eastern Regional Conference Oct. 21-22, Charleston, WV.

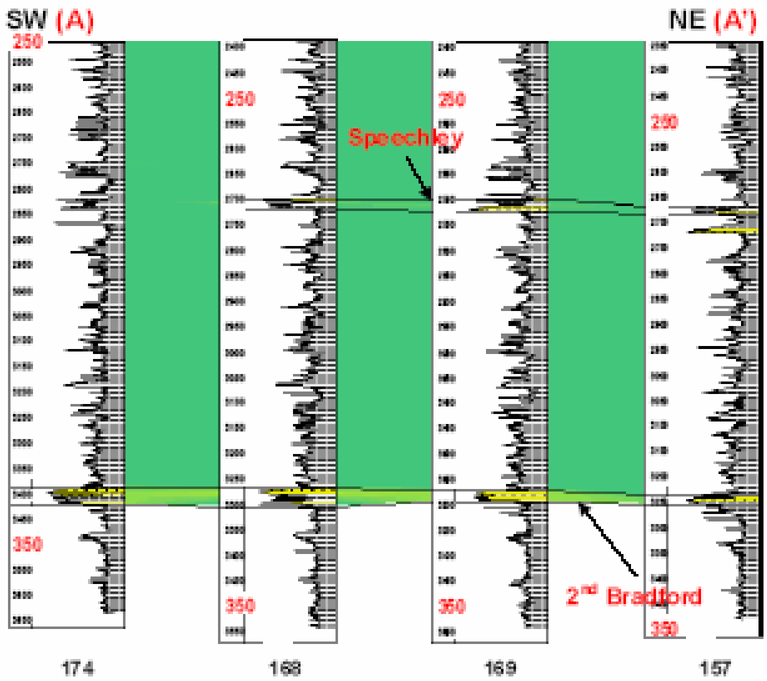


Figure 2. Gamma Ray correlation of 2nd Bradford (distributary channel – deep).

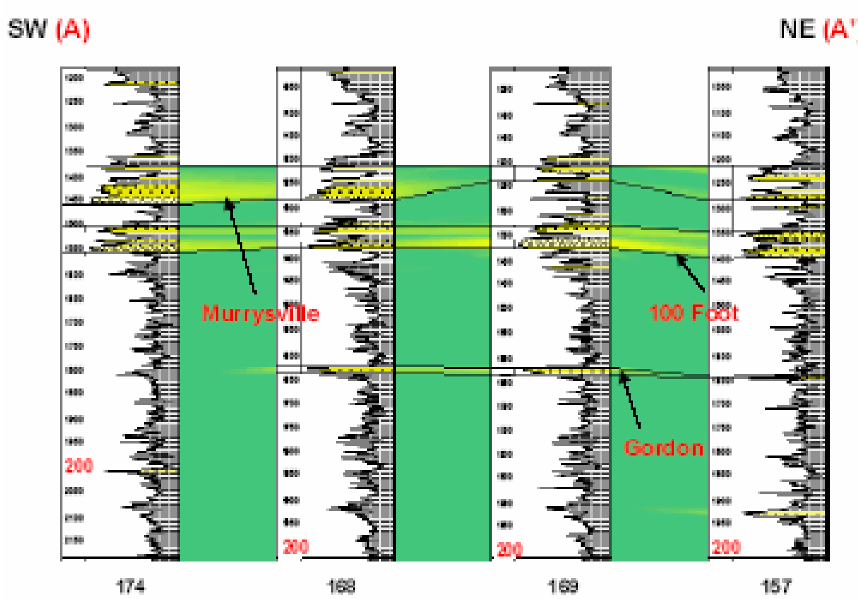


Figure 3. Gamma Ray correlation of Murrysville / 100 Foot Sands (Braided Stream – Shallow)

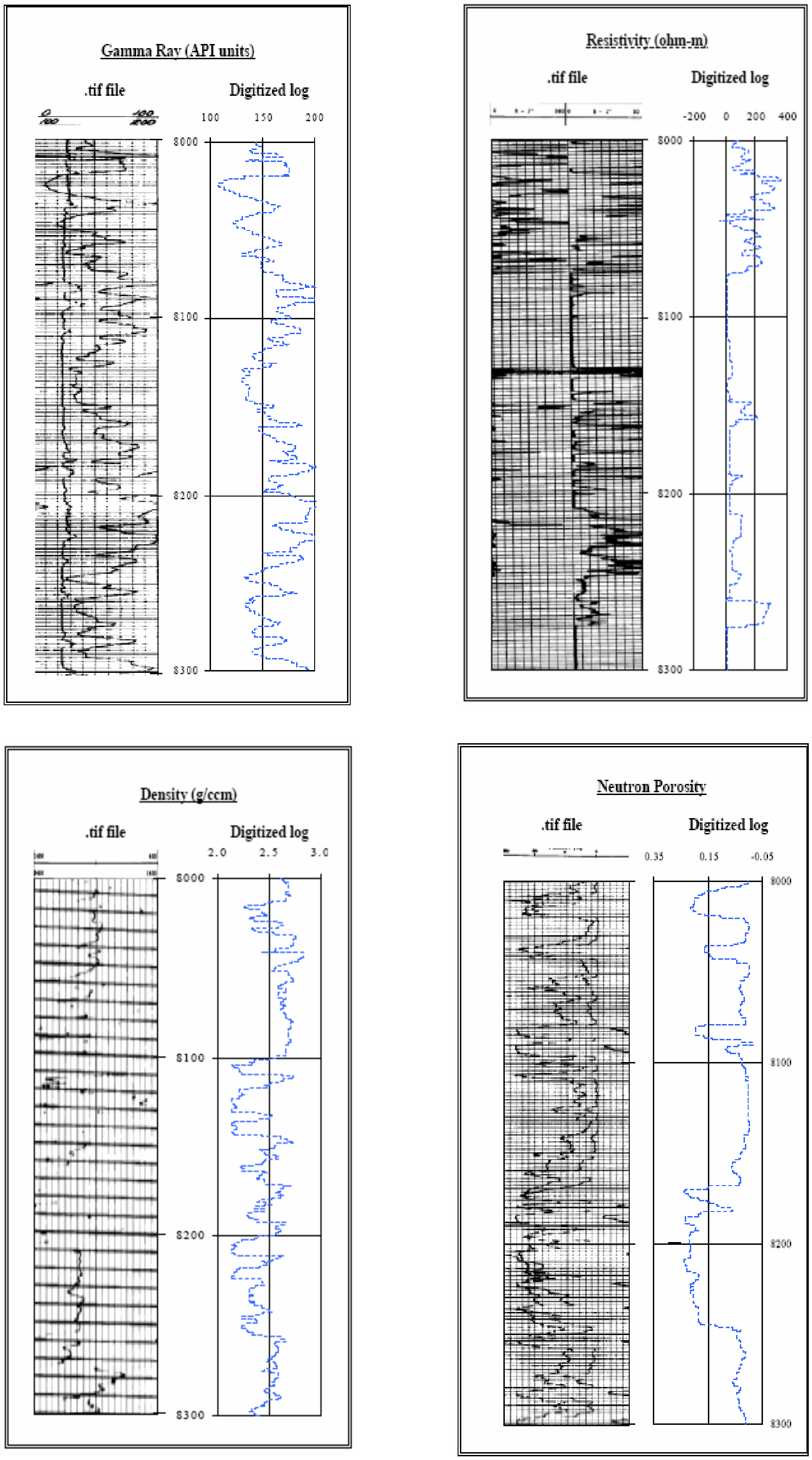
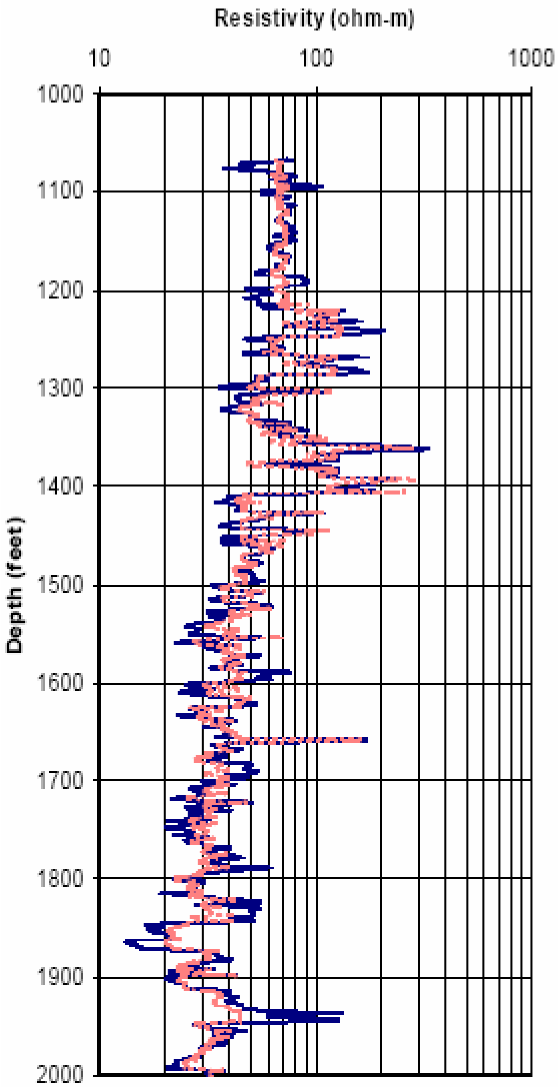


Figure 8. Original logs in .tif format and resulting curve after digitalization. Buffalo Valley Field.



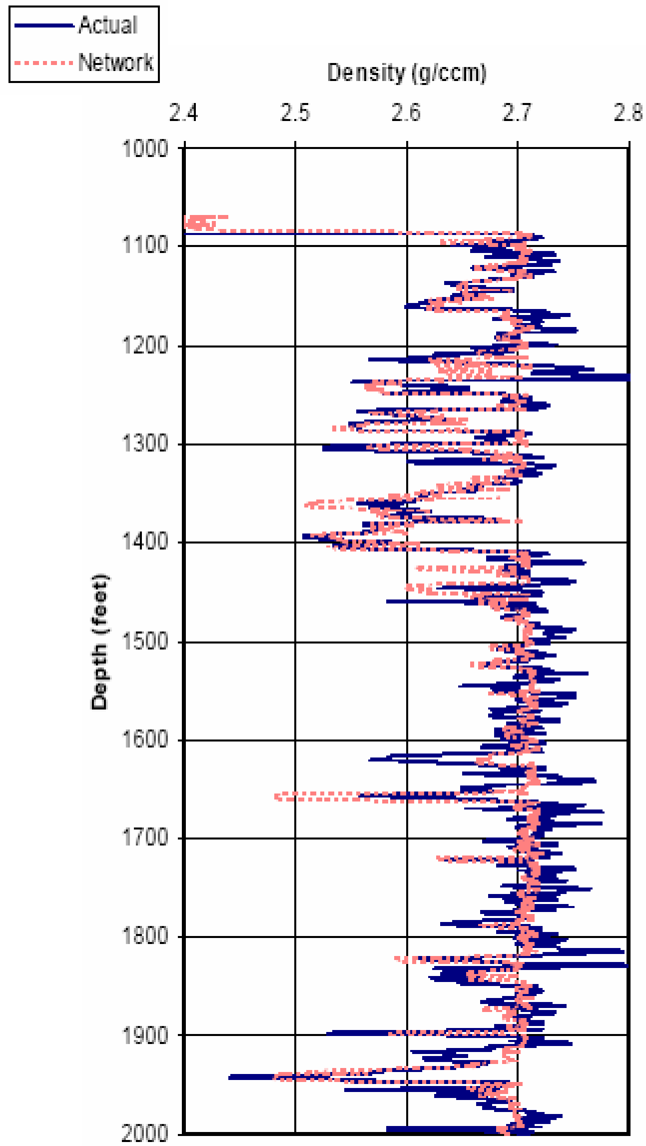


Figure 9. Synthetic resistivity log generated during verification through exercise 2 and combination A of inputs and outputs. Well 157, southern Pennsylvania area (upper zone).

Figure 10. Synthetic density log generated during verification through exercise 2 andcombination B of inputs and outputs. Well 157, southern Pennsylvania area (upper zone).

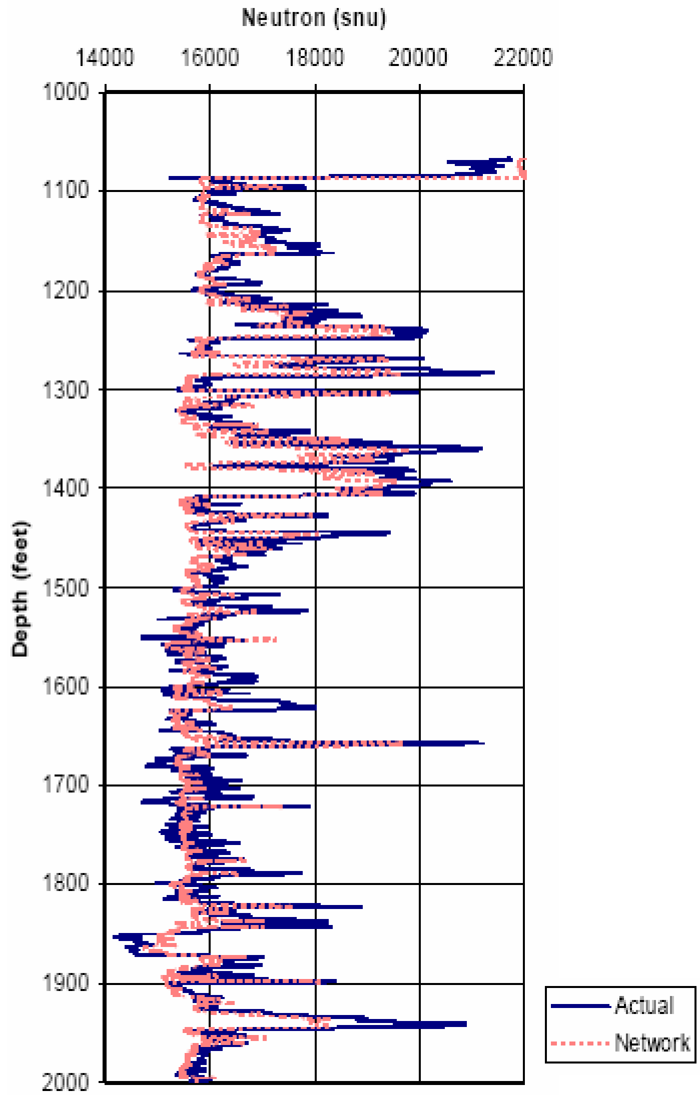


Figure 11. Synthetic neutron log generated during verification through exercise 2 and combination C of inputs and outputs. Well 157, southern Pennsylvania area (upper zone).