Predicting porosity logs from seismic attributes using geostatistics

Natalia Soubotcheva and Robert R. Stewart

ABSTRACT

At Pikes Peak, the Lower Cretaceous Waseca Formation, about 450 m below the surface, is the producing reservoir. The Waseca is filled with a homogeneous sand unit, an interbedded sand and shale unit, and a capping shale unit. These sands exhibit lithological variation, which affects the porosity distribution. After a proper well-to-seismic tie, we are able to analyze sample-based seismic attributes and select the most reliable ones using cross-validation. Effective porosity logs and various seismic attributes from PP and PS seismic data were used as inputs for porosity prediction. We have found that a probabilistic neural network showed the highest crosscorrelation (86%) between actual and predicted porosity logs at seven wells in the study area. After validation, the predicted volume of porosity along a 2-D seismic line was displayed. This final section provides a geologically realistic porosity distribution and helps in understanding the subsurface image.

INTRODUCTION

Mapping the physical properties of the reservoir is very important for assessment and development of the hydrocarbon therein. The Pikes Peak heavy-oil field is a heterogeneous reservoir, so we employ geostatistical methods to predict the rock properties between the drilled wells. Emerge (from Hampson-Russell) is a good tool to merge well logs and seismic data for predicting well log properties along the seismic line.

In this paper, we examine the prediction of density porosity logs from seismic attributes using neural networks.

Several wells were added to the project, since Emerge requires at least 6 control points (wells in our case) to conduct the statistical analysis (Figure 1). A typical set of well logs contains: P-wave sonic, density, gamma-ray, resistivity, neutron porosity, spontaneous potential, and caliper logs.

Some progress has been made in the rock property estimation using the integration of 3C-3D seismic data and well logs. Todorov (2000) has already investigated some geostatistical methods for reservoir analysis, such as variograms, kriging, cokriging, and stochastic simulation.

POROSITY LOGS

There are three types of measurements that are used to estimate porosity: density, neutron and acoustic. Although none of these logs actually measures pore volume directly, they can detect the contrast between the physical characteristics of water and rock-forming minerals. (Doventon, 1994). For porosity prediction, the density porosity log is recommended, as it shows the actual porosity of the rocks. If we compare two types of porosity logs in one window we can see the difference. (Figure 2).

The neutron porosity log (red line in the third column) measures hydrogen concentration in the rock matrix (Schlumberger oilfield glossary). The log was calibrated to read the correct porosity for a sandstone matrix. That means it is still correct for our productive sand formation. (Note where the red and blue lines overlap.) However, it shows "unrealistic" porosity values for the shale, about 60%. The reason that shales show such high neutron "porosities" is the bound water content and hydrogen in the clay minerals, rather than effective porosity.

To solve this problem, we used the density log (second column) to calculate the density porosity (third column, blue line) using the "Transforms" menu (Figure 3). The values for the fluid density and matrix density were taken to be the default.

Analogous transforms have been applied to the remaining wells, which are close to the seismic line and chosen for statistical analysis (Figure 1, blue line). Since Emerge will be correlating the target logs with seismic data, the proper depth-to-time conversion is critical. Manual correlation of well logs and seismic was performed in the eLog application (Hampson-Russell Software).

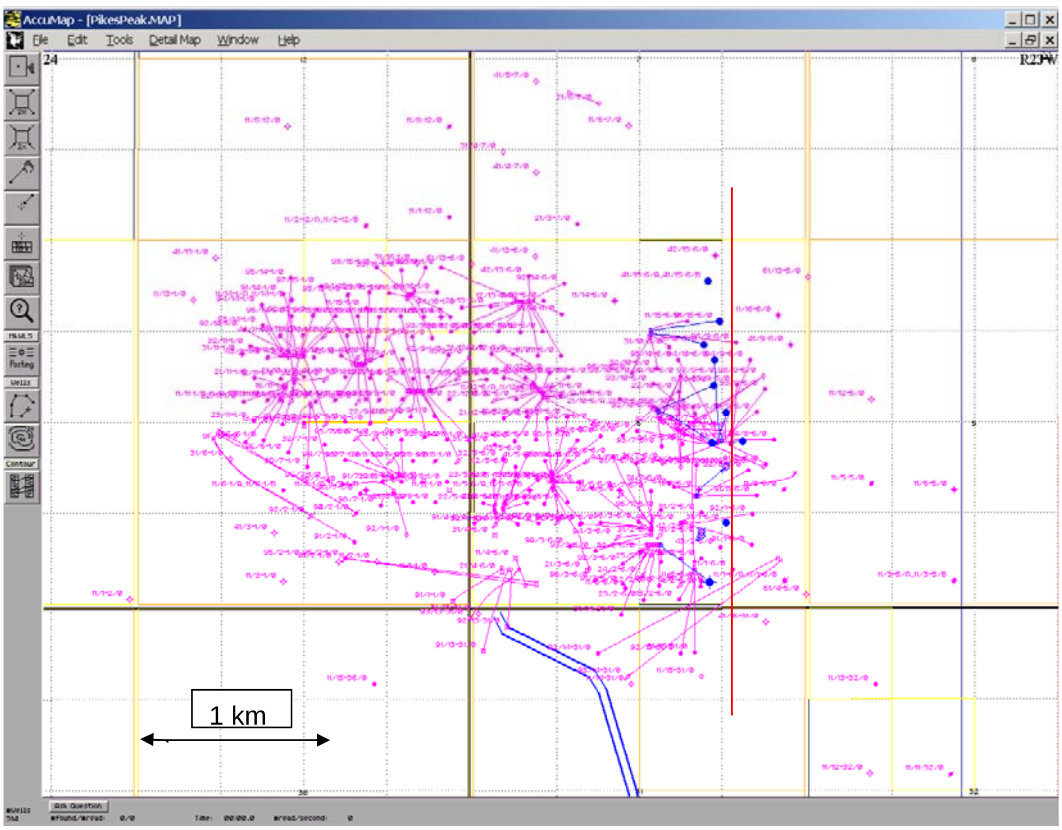


FIG. 1. Pikes Peak area with indicated wells (purple and blue) and seismic line (red) from the AccuMap system.

GEOSTATISTICAL METHODS

Any geostatistical method typically goes through the following steps:

- Quantify the spatial continuity of the well log data using variograms;
- Find and quantify a relationship (statistically) between the log and seismic data at this well location using cross-validation plots. This relationship could be linear (multivariate regression) or non-linear (neural network);
- Use this relationship to "predict" or estimate the volume of the log property at all locations of the seismic volume via kriging;
- Estimate the reliability of the result.

The result is validated by "hiding" wells and predicting them from other wells or seismic. (Yarus and Chambers, 1994)

We are going to analyze not the seismic data by itself, but attributes of the seismic data (Figure 4). A seismic attribute is any mathematical transform of seismic trace data which may or may not include other data sources.

First of all, we are going to use the wells for which density porosity logs exist, and which are located close to seismic line H00-131. Emerge requires sufficient well control — at least six. So, seven calculated porosity logs are loaded into Emerge (Figure 5). The same window contains an extracted seismic trace (from a PP dataset) at the well location. To improve the prediction, we can also use another seismic dataset as an external attribute. In this case, we converted PS data into the PP time domain with the help of ProMC and loaded it into the same Emerge project (Figure 5, blue line).

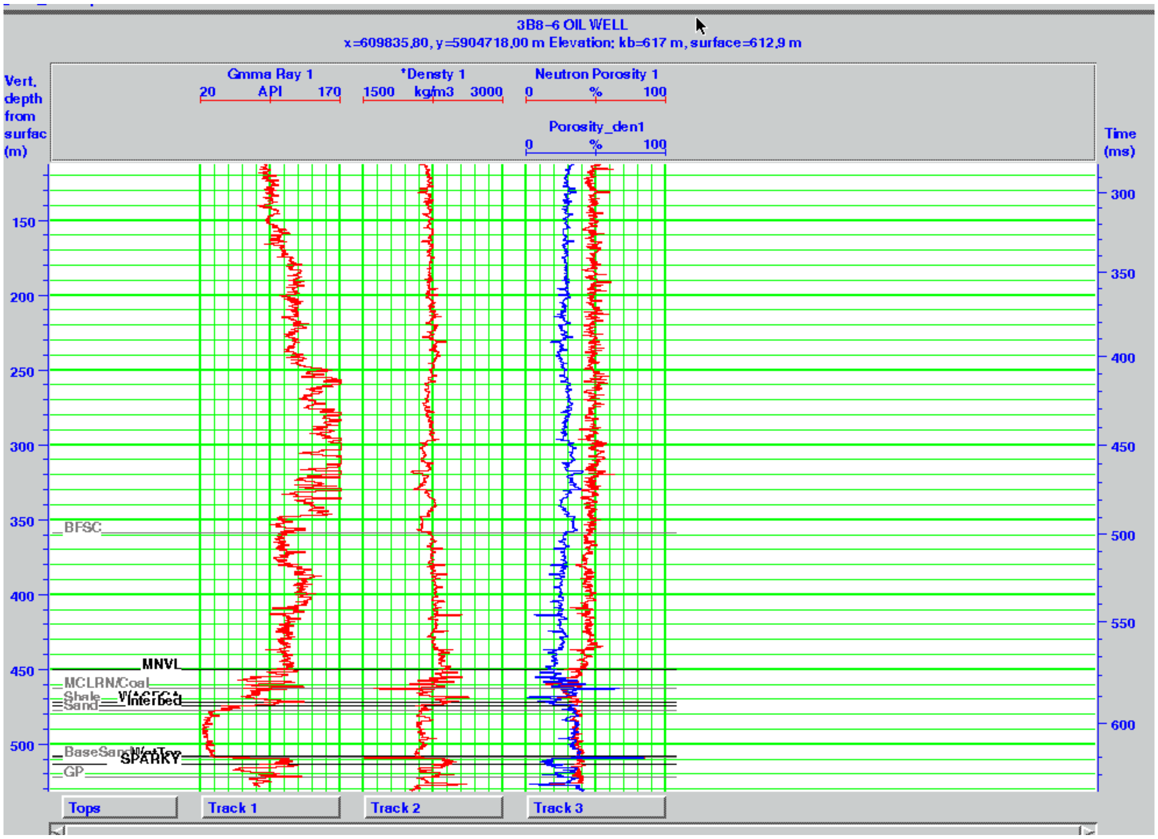


FIG. 2. Gamma-ray, density and porosity logs for well 3B8-6.

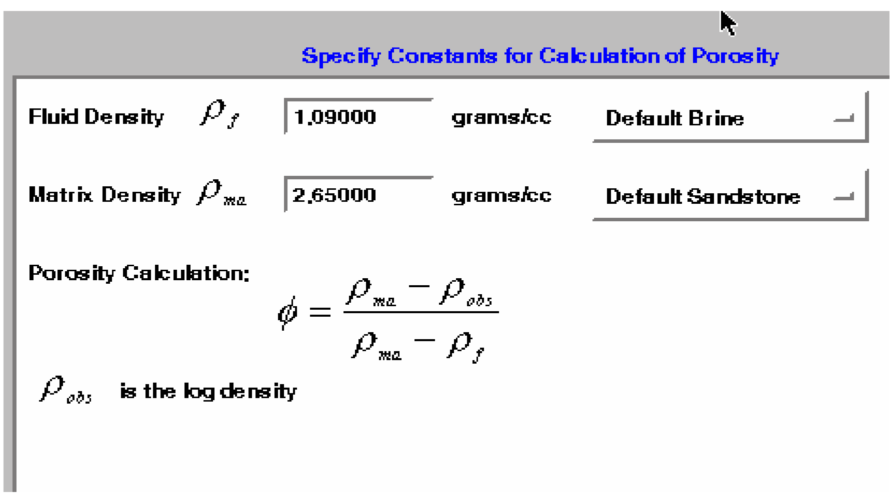


FIG. 3. “Transforms” menu window.

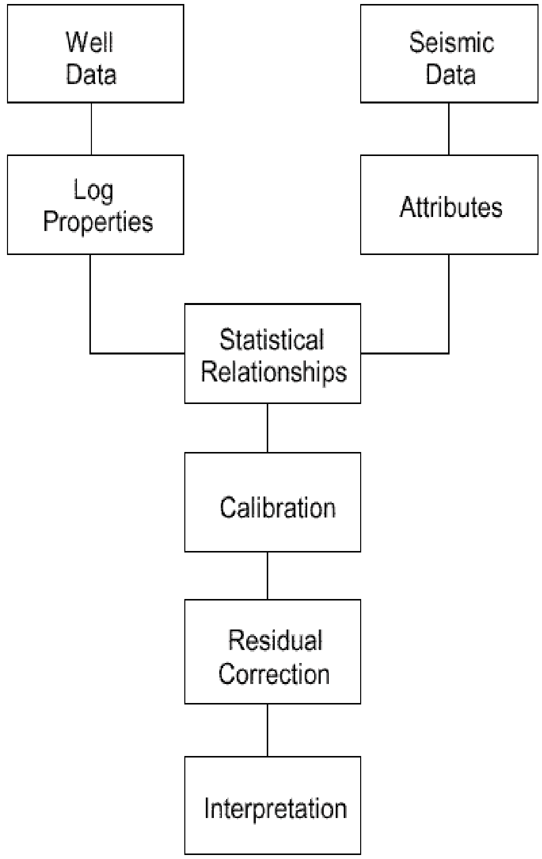


FIG. 4. The data-driven statistical interpretation (from Schultz et al.).

As was mentioned before, Emerge will be analyzing the seismic attributes, not the seismic traces. In Emerge we can plot any attribute (or attributes) for the particular well. The most important for log prediction are:

- Integrated trace;
- Integrated absolute amplitude of the trace;
- Near-angle stack;
- AVO intercept/gradient;
- Frequency and absorption estimates;
- Seismic velocity.

Each target log sample (Figure 6) is modelled as a linear combination of attribute samples at the same time:

$$\phi(t) = w_0 + w_1 A_1(t) + w_2 A_2(t) + w_3 A_3(t), \qquad (1)$$

where A - is a seismic attribute and w- its weight.

This equation can be written as a series of linear equations or in matrix form. The solution is optimal in a least-square sense (Emerge theory, on-line guide, Hampson-Russell Software).

The data are now ready for analysis. The first step is to examine the single-attribute transforms: Attribute/Create single attribute list.

The comparison will be performed with all attributes in the project (Figure 7): raw seismic, PS data in PP time, amplitude envelope, amplitude-weighted cosine phase, amplitude-weighted frequency, amplitude-weighted phase, average frequency, apparent polarity, cosine instantaneous phase, derivative, derivative instantaneous amplitude, dominant frequency, filters (different), integrate, integrated absolute amplitude and second derivative, and time.

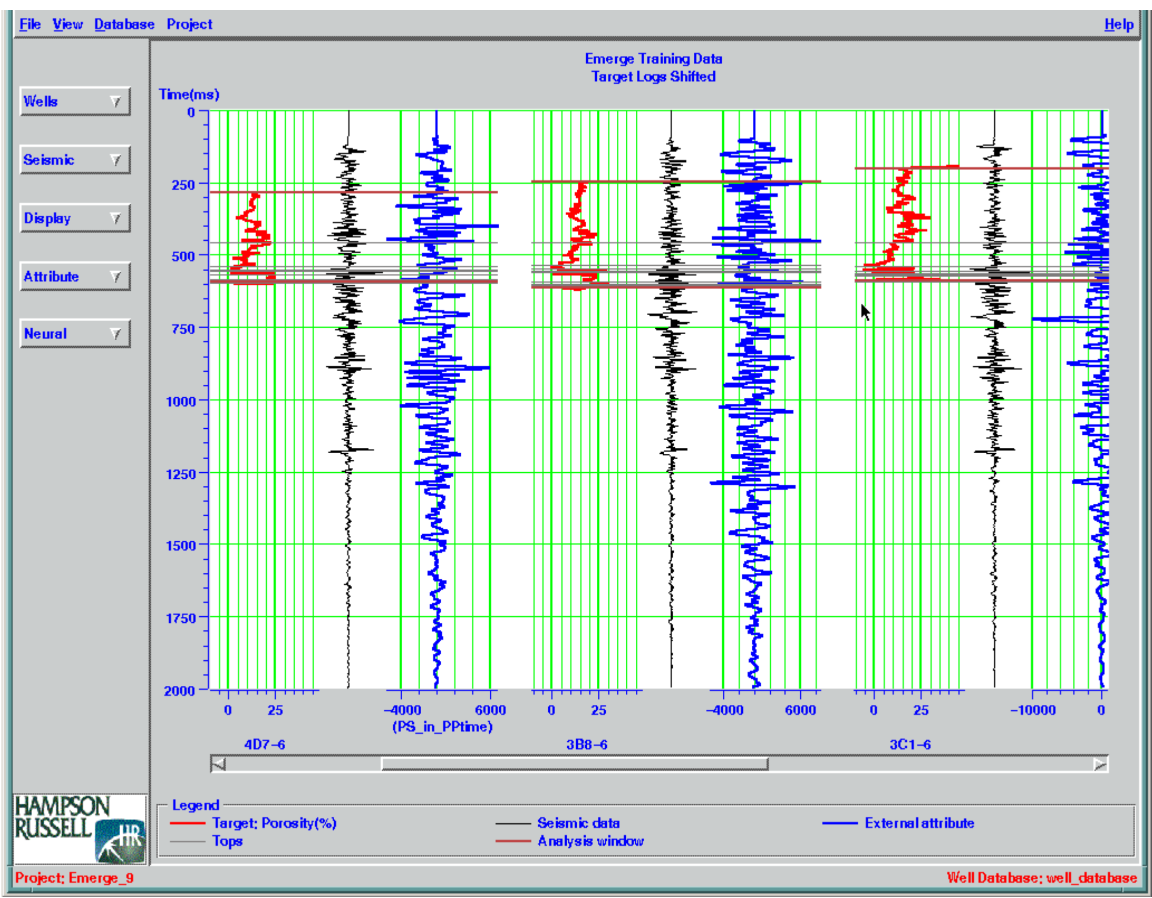


FIG. 5. Emerge data window with loaded porosity logs (red), extracted seismic traces at the well location (black — from PP dataset; blue — from PS dataset).

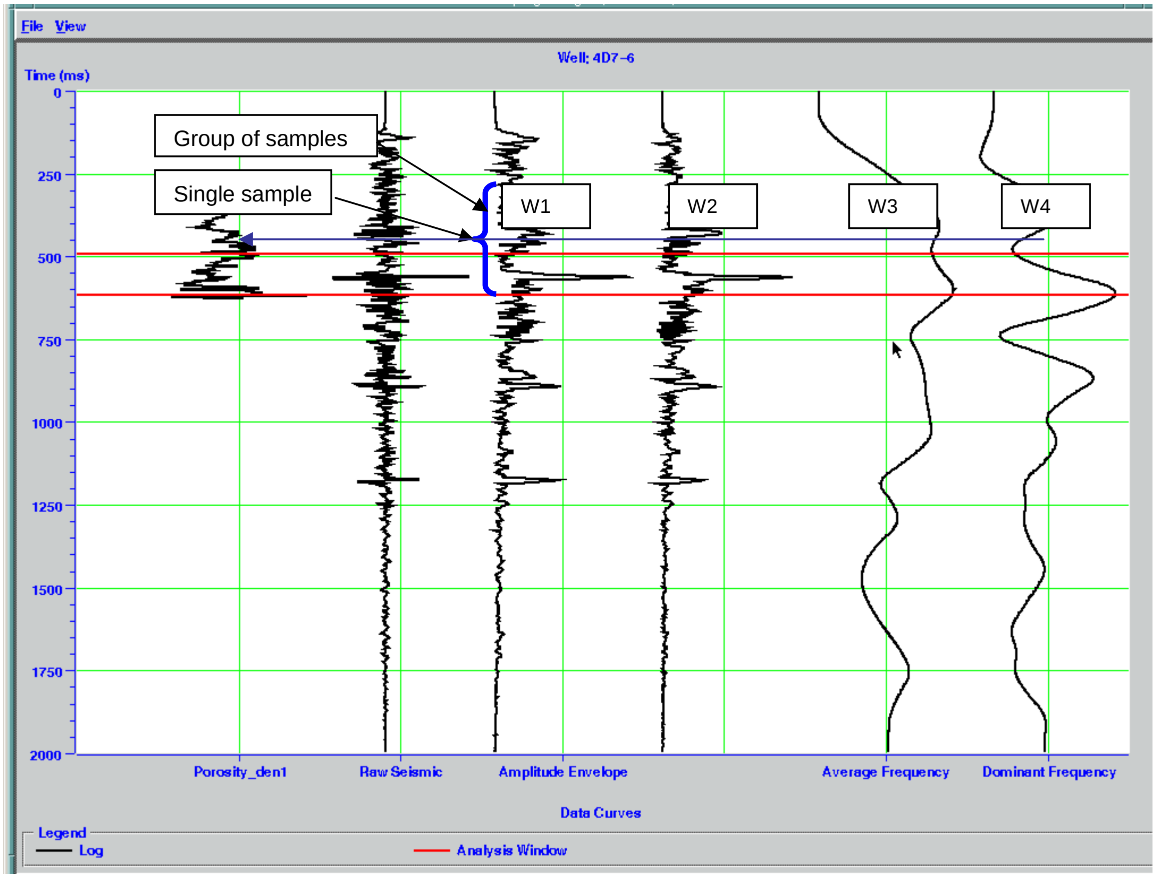


FIG. 6. Seismic attributes for well 4D7-6.

In Figure 8, we can see the list of internal seismic attributes, with corresponding prediction errors and correlation coefficients. The best correlation in this case looks rather poor – 38%. However, we can improve the correlation by applying the residual time-shift between the target porosity logs and the seismic data.

The problem is that the frequency content of the target log is much higher than that of the seismic attribute (Figure 6), therefore the convolution operator is recommended to resolve the difference. (Emerge theory, on-line guide, Hampson-Russell Software). In this case, each target sample is predicted using a weighted average of a group of samples on each attribute (Figure 6). This can be done by multi-attribute analysis, where we have chosen the operator length equal to 7 and the number of attributes to use as 8.

Now we can check how the program predicts the logs (in our case porosity logs) from the seismic data (Figure 9). For example, we have chosen the 5th (integrated absolute amplitude) and the 7th (cosine instantaneous phase) attributes for analysis.

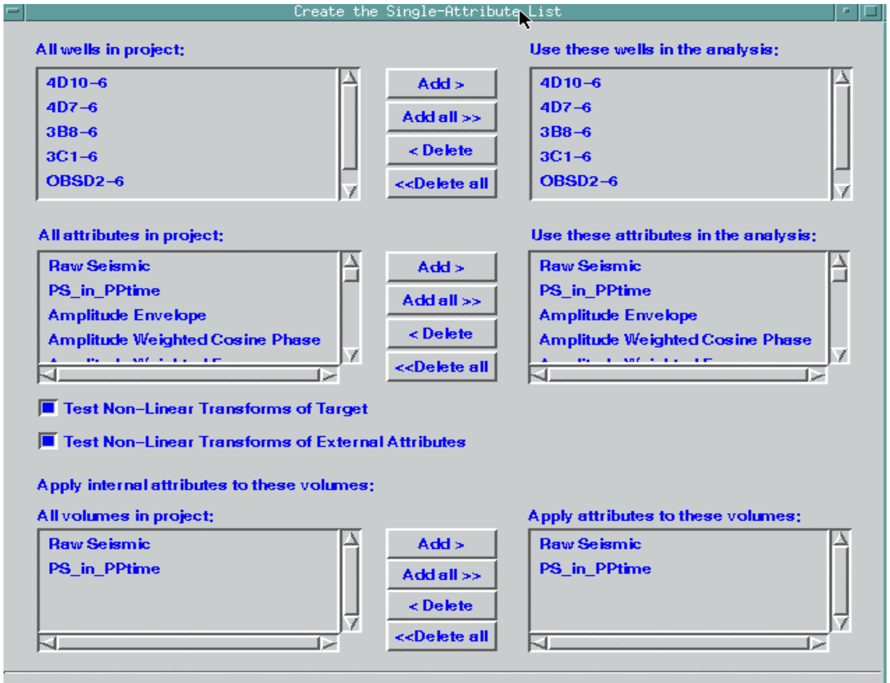
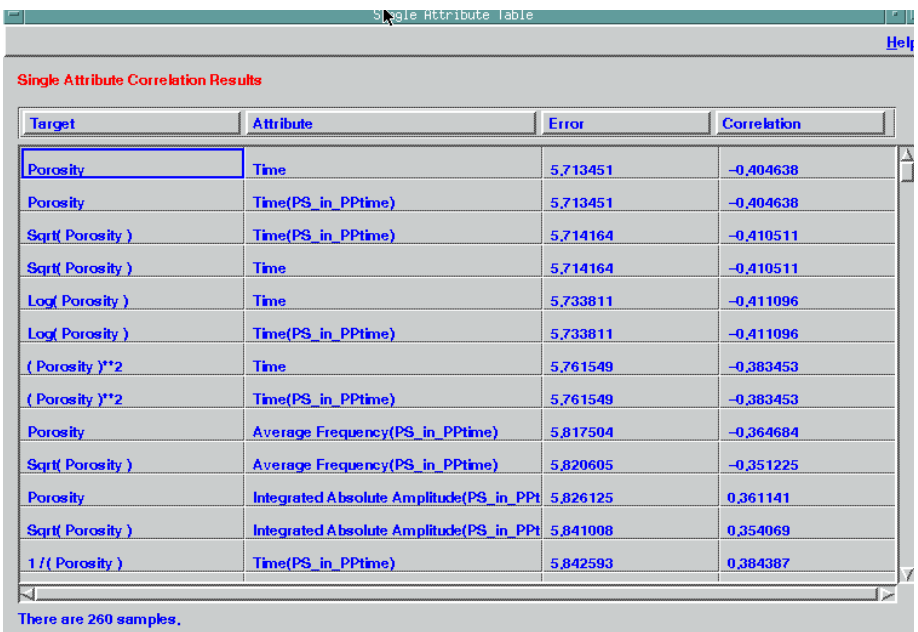
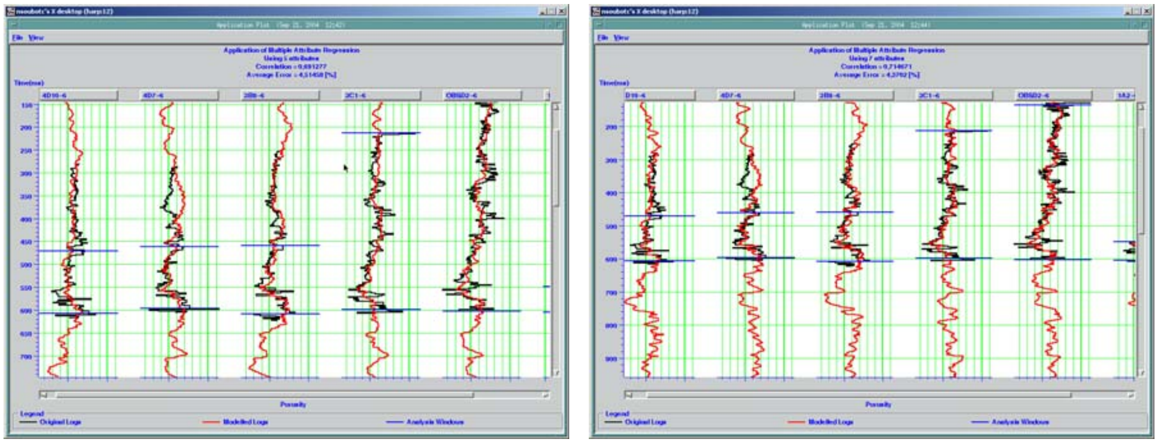


FIG. 7. Window to create the single attribute list.



Target	Attribute	Error	Correlation
Porosity	Time	5.713451	-0.404638
Porosity	Time(PS_in_PPtime)	5.713451	-0.404638
Sqrt(Porosity)	Time(PS_in_PPtime)	5.714164	-0.410511
Sqrt(Porosity)	Time	5.714164	-0.410511
Log(Porosity)	Time	5.733811	-0.411096
Log(Porosity)	Time(PS_in_PPtime)	5.733811	-0.411096
(Porosity)**2	Time	5.761549	-0.383453
(Porosity)**2	Time(PS_in_PPtime)	5.761549	-0.383453
Porosity	Average Frequency(PS_in_PPtime)	5.817504	-0.364684
Sqrt(Porosity)	Average Frequency(PS_in_PPtime)	5.820605	-0.351225
Porosity	Integrated Absolute Amplitude(PS_in_PPt	5.826125	0.361141
Sqrt(Porosity)	Integrated Absolute Amplitude(PS_in_PPt	5.841008	0.354069
1 /(Porosity)	Time(PS_in_PPtime)	5.842593	0.384387

FIG. 8. Single attribute table.



a) b)

FIG. 9. Comparison target logs (black) with predicted logs (red) using 5th (a) and 7th (b) attributes.

For the 5th attribute we can observe a 69% correlation between the predicted logs and target logs, and an RMS error of 4.5% for porosity. For the 7th attribute, correlation between the predicted logs and target logs is 71% and the RMS error is 4.4% for porosity.

In other words, each predicted log has used an operator calculated from the other well, and we can see how well the process will work on a new well.

How can we know when to stop adding attributes?

Emerge divides the entire dataset into two groups (Figure 10): a training dataset (original wells, in black) and a validation dataset (predicted data, in red).

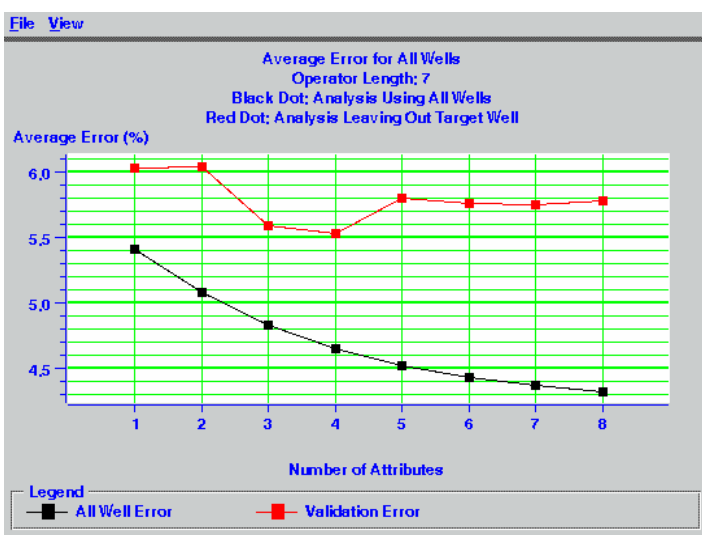


FIG. 10. Prediction error plot.

The horizontal axis shows the number of attributes used in the prediction. The vertical axis is the root-mean-Square prediction error for that number of attributes. I have chosen the maximum number of attributes to use, 8, and an operator length of 7 samples. From this picture, we can see that it is best to use only 4 attributes to avoid a greater error in prediction.

To determine which seismic attribute is more reliable in predicting the target log, we can cross-plot the particular attribute versus the density porosity log (Figure 11) and look for the normalized correlation value. For the integrated absolute amplitude this value is equal 0.69, for the cosine instantaneous phase, 0.71. Thus the internal attribute, cosine instantaneous phase, gives the better fit with the line of perfect correlation (in red).

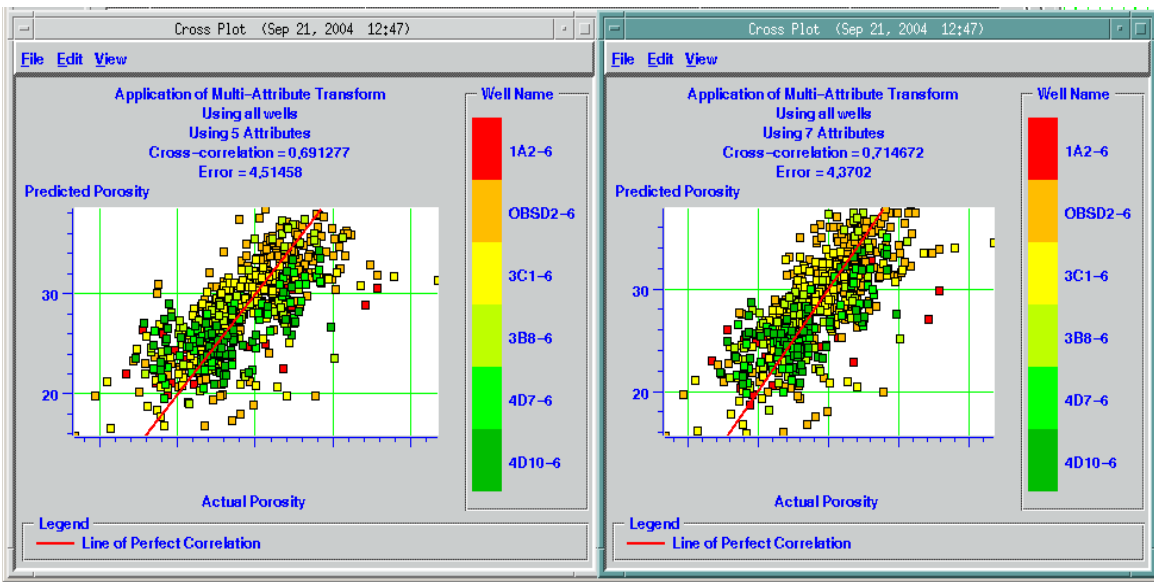


FIG. 11. Cross-validation of predicted and actual porosity using the 5th (left) and 7th (right) attributes.

NEURAL NETWORKS IN EMERGE

The analysis so far has been linear. The neural network in Emerge represents the non-linear approach to the problem. It can increase both the predictive power and resolution of derived porosity volumes. For this project, we have chosen to use a probabilistic neural network, which is similar to the kriging interpolating technique.

Let us assume that we know the exact values of porosity for three seismic attributes,

$$\begin{bmatrix}\phi_1\\ X_1\\ Y_1\\ Z_1\end{bmatrix} \quad \begin{bmatrix}\phi_2\\ X_2\\ Y_2\\ Z_2\end{bmatrix} \quad \begin{bmatrix}\phi_3\\ X_3\\ Y_3\\ Z_3\end{bmatrix} \quad \begin{array}{l} \longleftarrow \quad \textbf{Log value} \\ \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \longleftarrow \quad \textbf{Attributes} \end{array},$$

and wish to get a new output point:

$$\begin{bmatrix}?\\ X_0\\ Y_0\\ Z_0\end{bmatrix}.$$

We solve the problem by comparing the new attribute with the known attributes. In a probabilistic neural network, the estimated value is a linear combination of the known training values:

$$\phi_0 = w_1 * \phi_1 + w_2 * \phi_2 + w_3 * \phi_3, \qquad (2)$$

where * is the convolution, ϕ is the porosity value, and w_i are the weights. The weights depend on the distance from the desired point to the training point.

The main disadvantage of PNN is that the application time is slow (about 50 minutes to calculate the colour porosity volume for a seismic line with 761 CDPs), because it carries all training data and compares each output sample with each training sample.

Figure 12 shows the result of neural network training.

We notice that the correlation coefficient is much higher in this (86%) than for multi-attribute analysis (71%). After validation of this result we can apply it to the seismic data.

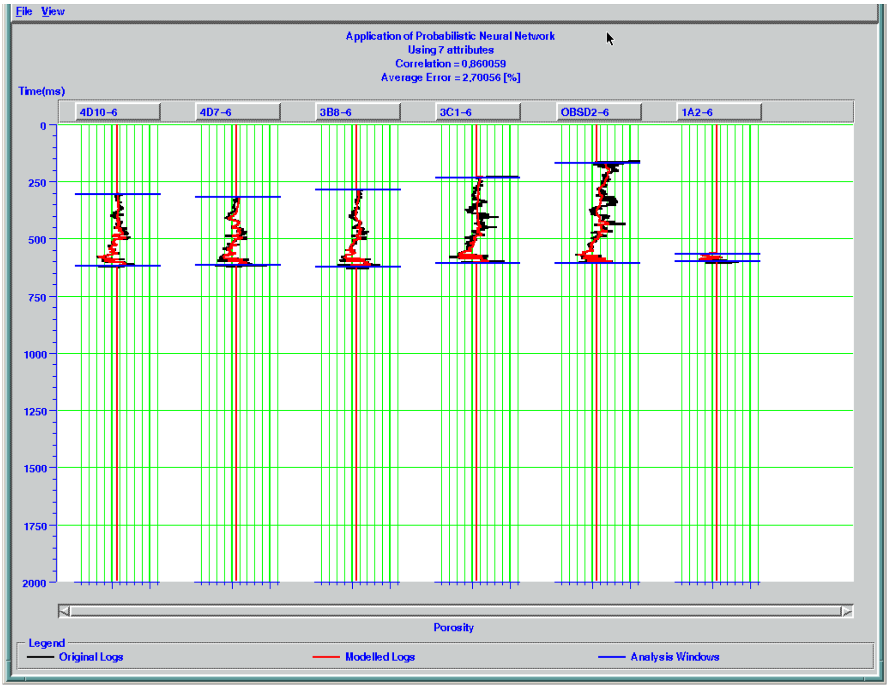


FIG. 12. Correlation of target logs (black) and predicted logs (red) after neural network training

Neural Networks: Cosine instantaneous phase

In this case, Emerge is acting similar to the kriging technique: it interpolates the well log values using the seismic data as a guide. The predicted porosity along the seismic line looks as shown in Figure 13.

According to van Hulten (1984), the porosity of homogeneous sand is about 30%. Since the productive formation is located at about 600 ms, we can see that it agrees with the colour porosity volume fairly well. Figure 13 displays only the wells which were chosen for statistical analysis. However, we should keep in mind that there are some other wells close to the seismic line, for example: 4D7-6 (CDP 339), 4D2-6 (CDP 399), D2-6 (CDP 419). It is clear that the productive wells are concentrated in the high porosity zones. For heterogeneous reservoirs these zones have limited stratigraphic and geographic extent: coloured purple on the picture (Yarus and Chambers, 1994). Thus, for the successful development of the existing oil fields, it is desirable to know how these zones (with high porosity and permeability) spread under the surface.

Figure 14 demonstrates the southern part of the seismic line. Note the high porosity zone (about 30%) corresponding to CDPs 725–735. Recall that we consider the time interval where the productive formation lies to be 580–600 ms. No wells have been drilled at this location (Figure 1). Geologically, this site could be good for oil accumulation as soon as porosity here is noticeably higher than that for the surrounding area. However, this result should be confirmed using the other methods, such as AVO or seismic inversion.

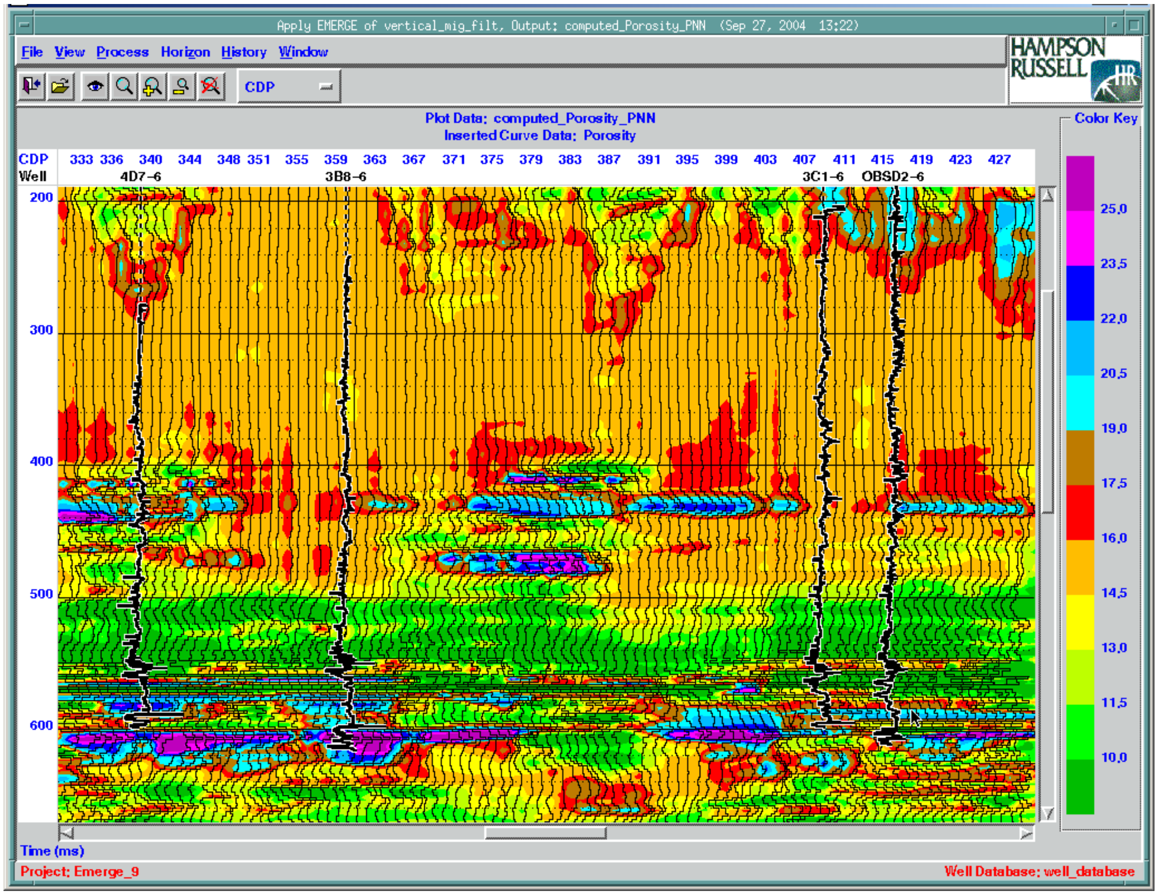


FIG. 13. Colour porosity volume along the seismic line for CDPs 330–430.

CONCLUSIONS

1. In this paper, geostatistical methods were used to predict density porosity (%) using seismic attribute as a guide.

2. Emerge Software (Hampson-Russell) was used to learn about the data, find relationships between seismic and well data, and use what had been learned to predict reservoir properties along the seismic line.

3. Geostatistical methods can lead to greater understanding of the data and the successful development of existing fields.

FUTURE WORK

1. Calculate porosity volumes using the inversion result as an external attribute.

2. Asses the accuracy of the results.

3. Estimate the original oil in place (OOIP) at the well locations.

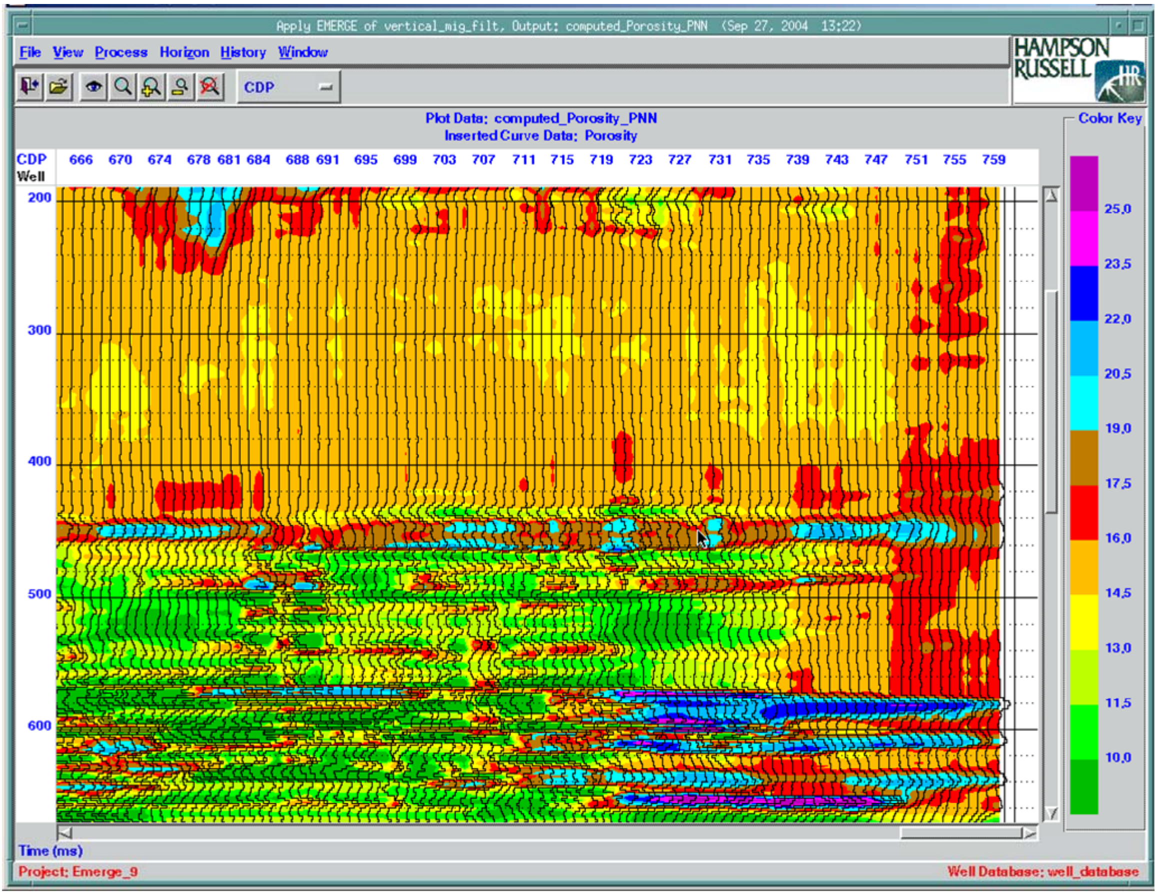


FIG. 14. Colour porosity volume along the seismic line for CDPs 660–761.

REFERENCES

Doventon, J. H., 1994, Geologic log interpretation: SEPM Short Course #29.

Hampson, D., Using multi-attribute transforms to predict log properties from seismic data: On-line Guide for Emerge Software.

van Hulten, F. F. N., 1984, Petroleum geology of Pikes Peak heavy-oil field, Waseca formation, Saskatchewan: Canadian Society of Petroleum Geologists, Memoir 9, 441–454.

Luthi, S. M., 2001, Geological Well Logs: Springer-Verlag.

Todorov, I. T., 2000, Integration of 3C-3D seismic data and well logs for rock property estimation: M.Sc. Thesis, University of Calgary.

Yarus, J. M., and Chambers, R., 1994, Stochastic Modeling and Geostatistics: AAPG.