RESERVOIR CHARACTERIZATION BY CALIBRATION OF SELF-ORGANIZED MAP CLUSTERS

M.T. Taner, J.D. Walls, M. Smith, G. Taylor, M.B. Carr, D. Dumas, Rock Solid Images, 2600 S. Gessner Suite 650, Houston, TX 77003

Summary

Kohonen's Self Organizing Feature Maps (SOFM) and other unsupervised clustering methods generate groups based on the identification of various discriminating features. These methods seek an organization in the dataset and form relational organized clusters. However, these clusters may or may not have any physical analogues. A calibration method that relates SOM clusters to physical reality is desirable. This calibration method must define the relationship between the clusters and the observed physical properties; it should also provide an estimate of the validity of the relationships. With the development of a calibrated relationship, the whole dataset can be classified. The principal steps, therefore, are the Three-C's **"Clustering, Calibration and Classification".**

The clustering step reduces the multiple dimensions of the data description into logically smaller groups. Each original data point defined by multiple attributes is reduced to one or two-dimensional relational groups. This establishes some logical clustering and reduces the complexity of the classification problem. Furthermore, calibration should be more successful due to the consideration of less variability in the data.

Here in, a simple calibration method is proposed that employs Bayesian logic to provide the relationship between cluster centers and the known reservoir properties. The output will give the most probable calibration between each Self-Organized Map node and the wellbore-measured lithology. The second part of the output will give the probability of the calibration.

Method

A Bayesian decision is based on the knowledge of the probability density function of each class. The decision boundary between classes is located at the point where the probability density of adjoining classes is equal.

Figure 1 shows three different class probability densities. The Bayesian decision boundaries are located where the probabilities of different classes are equal. This is a very intuitive concept and easy to accept. Samples will be classified as belonging to the class with highest probability density.

In the method presented here, we will use Bayesian logic to establish the relationship between lithology classes and the SOM neural nodes. To establish such a relationship, we will need to compute the probability density function of each class in the SOM topology. The Euclidean distance and the scaled Gaussian function as the probability density estimator is employed.

Let $w(i,j)$ represent the SOM i'th weight of j'th neuron and , $X(i,n)$ represent the i'th attributes of the n'th lithology class. The Euclidean distance between the neural node and the input data sample is given by:

$$d(j,n) = \sqrt{\sum_{i=1}^{NI} \{ w(i,j) - X(i,n) \}^2} \qquad (1)$$

where NI is the number of attributes (number of input dimensions).

During the SOM iteration, the Euclidean distances between data points and each neural node are computed. The node with the closest distance to the data is declared the winning neuron and its weights are adjusted to be closer to the input data. Its topologically neighboring neurons are also adjusted, but in a reduced amount which is proportional to their distance to the winning neuron. Iteration continues until an acceptable convergence is reached. Because the input data is not perfectly organized, we expect the clustering around each neuron to exhibit some scatter, i.e. some variance other than zero. In the calibration stage we need to determine the degree of convergence, so our probability estimate will have some basis. The average variance of clustering will give us a measure of the distance between SOM neuron cluster centers. This average will control the shape of the Gaussian function, and be used for a control distance of 50% probability. It is considered that each data point is valid with some probability. It could belong to any one of the clusters of the SOM. However, the probability of belonging to any group is

a function of the distance to the neuron. Thus, the probability is computed as a Gaussian function of the distance.

In the first pass through the data the average Euclidean distance between the input data samples and the winning neurons is computed. This in turn determines the Gaussian shape factor α. In the second pass this shaping factor α is used in computation of the probability at each SOM neuron.

$$P(j,n) = exp\left(\alpha . d^2(j,n)\right), \qquad (2)$$

where $d(j,n)$ is the distance between n'th input data and j'th neuron.

This suggests that the closer the data point is to a node, the higher the probability of a correct calibration. The probability map is then generated, with the same topology as the SOM for each lithology class. For each lithology class data point the distance and the (Gaussian function) probability for each SOM neural node is computed. We accumulate these probabilities for data samples for that particular lithology class. Finally, we compute a scalar and divide the accumulated probabilities so the sum is equal to unity (100%). This map now represents the probability density of the particular lithology class.

Comparison of the lithology probability map with the maximum probability map (MPM) is accomplished using Bayesian logic. We update the MPM if the lithology probability map contains a higher probability than the MPM and we also update the classification number, otherwise the MPM is left unaltered. This procedure is repeated for all classes. Upon completing the computation for all of the classes, a table of classes with the highest probability for each SOM neural node and a table of corresponding probability densities, results. Because the data is given in list form containing the attribute values and corresponding lithology class, calibration could be conducted on multi-well data and for both deviated and horizontal wells, where synthetics are difficult to generate.

The Three-C procedure is analogous to the regularized Radial Basis Function Networks (RBF). The original form of RBF uses each training data sample as the center of each neuron in the Hidden Layer. This will result in an enormous amount of hidden layer neurons. The regularization process reduces the number of neurons to a level that represents the input data field by a minimum number of neurons. Following this reduction, the output layer weights are computed to linearly interpolate the desired results. Here in, a SOM clustering in the first stage is used, which is similar to a regularization of the RBF network. The calibration stage is performed via Bayesian logic rather than linear interpolators.

Example

The geologic setting for this test was a North Sea tertiary turbidite system. The seismic survey covering the area of interest was about 325 square km. There were 4 wells inside the survey area, two of which had encountered oil saturated pay sands. These wells had a full suite of high quality logs, including dipole shear wave logs in the two producing wells.

The logs were used to classify 4 different lithologies, shale, siltstone, wet sand, and oil sand. A suite of post-stack seismic attributes were computed on the 3D data volume and used as input to the Kohonen SOM with a 10x10 cluster topology. A probability field for each class was generated and scaled so the sum was equal to 1.0 representing 100 percent probability. The probability function was generated using the RMS clustering distance as the 50 percent probability value. This gives us the value of Gaussian shaping factor. Smaller RMS values will make the Gaussian curve sharper, and larger values will produce a smoother curve. The maximum probability for each cluster center is determined by comparison. The program displays final calibration and related probabilities as shown in Figure 4.

These classifications are shown as a time section from the 3D reservoir volume in Figure 5. The red colored zones indicate the oil sands. When compared to classification from other methods such as acoustic and elastic impedance inversion (Mukerji, et al., 1998), the results are comparable, even though in this case only post-stack seismic data volume was used.

Conclusions

The calibration step connects the clustering and classification steps in highly logical manner. The procedure will perform a calibration for all SOM neurons regardless of the size and topology of the network. Any wellbore configuration can be accommodated; even those highly deviated cases where the generation of synthetic for calibration purposes may be difficult.

This procedure simplifies the Probabilistic Neural Network (PNN) approach. In the PNN procedure each training data point is considered a valid point in data space and a corresponding probability function is generated in N-dimensional space. In the current implementation, the clustering at an SOM dimension where all attributes are

well organized is conducted. This reduces the dimensionality of the problem and the computation time. Since the data is clustered by the SOM, the calibration is less complicated, and most probably, more accurate.

The new method was tested on a 3D seismic data volume with 4 wells. The results appear to be in agreement with the wells and similar to results obtained by statistical methods applied to acoustic and elastic impedance volumes.

References

Duda, R. O. and Hart, P. E. , 1973, ***Pattern Classification and Scene Analysis;*** Wiley-Interscience Publication.

Haykin, S. 1994, "***Neural Networks, A Comprehensive Foundation"***, Published by Macmillan College Publishing Company, New York. (This book contains one of the most comprehensive discussions of all types of Neural Networks including the Self-Organized Maps).

Kohonen, T., 1988, "***Self-Organizing and Associative Memory"***, 3rd. ed., Published by Springer Verlag, New York.

Mukerji, Tapan, Arild Jorstad, Gary Mavko, John Reidar Granli; Applying statistical rock physics and seismic inversion to map lithofacies and pore fluid probabilities in a North Sea reservoir, SEG Annual Symposium, 1998.

Acknowledgement

We would like to thank Statoil for contribution of the seismic and log data used in this work.

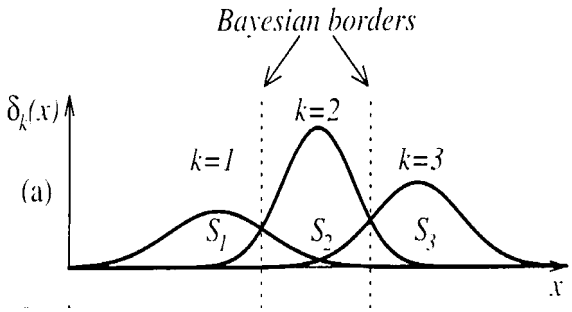

Figure 1. Bayesian Boundaries for Three Different Probability Densities

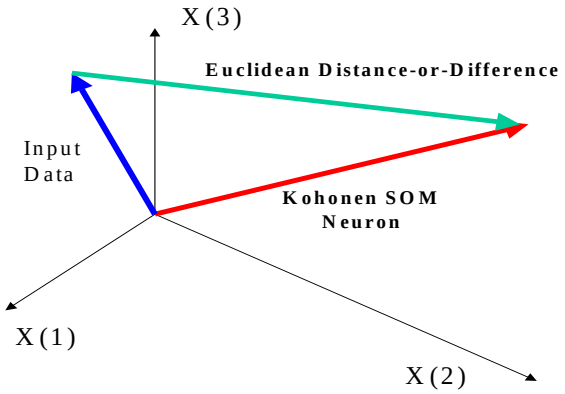

Figure 2. Input Vector, Neural Weight Vector and Euclidean Distance

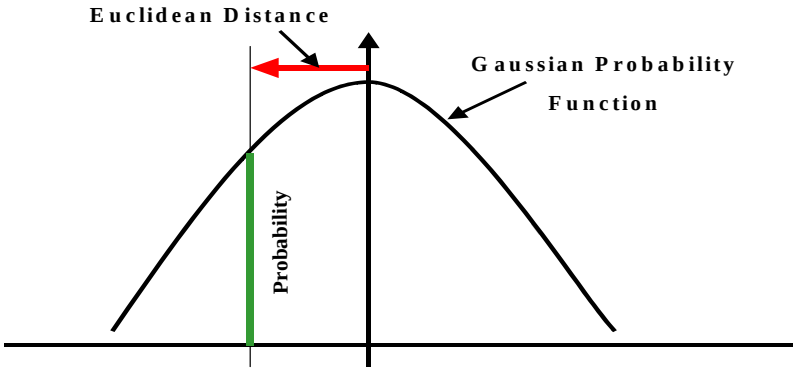

Figure 3 Computation of Probability by Euclidean Distance and Gaussian Function

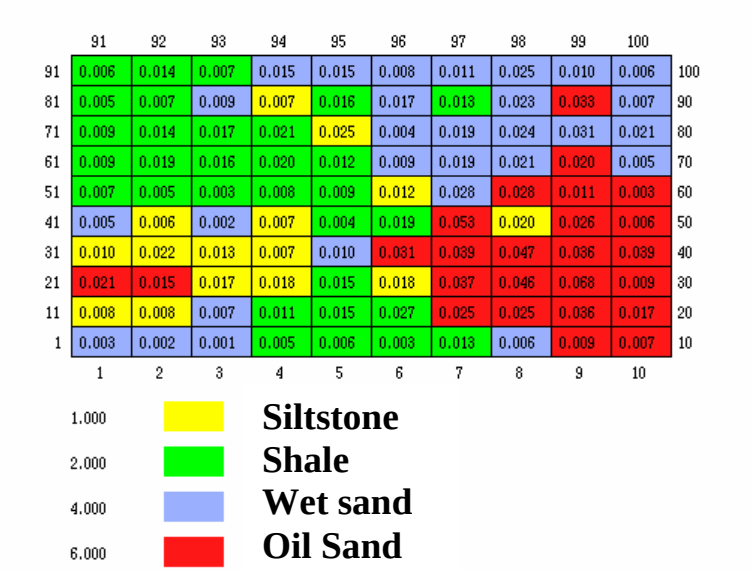

Figure 4: Kohonen SOM with 10x10 cluster map calibrated into 4 lithologic classes. Each class is shown by a specific color for ease of recognition. Number on each cluster center represent relative probability (Total adds up to 1.0)

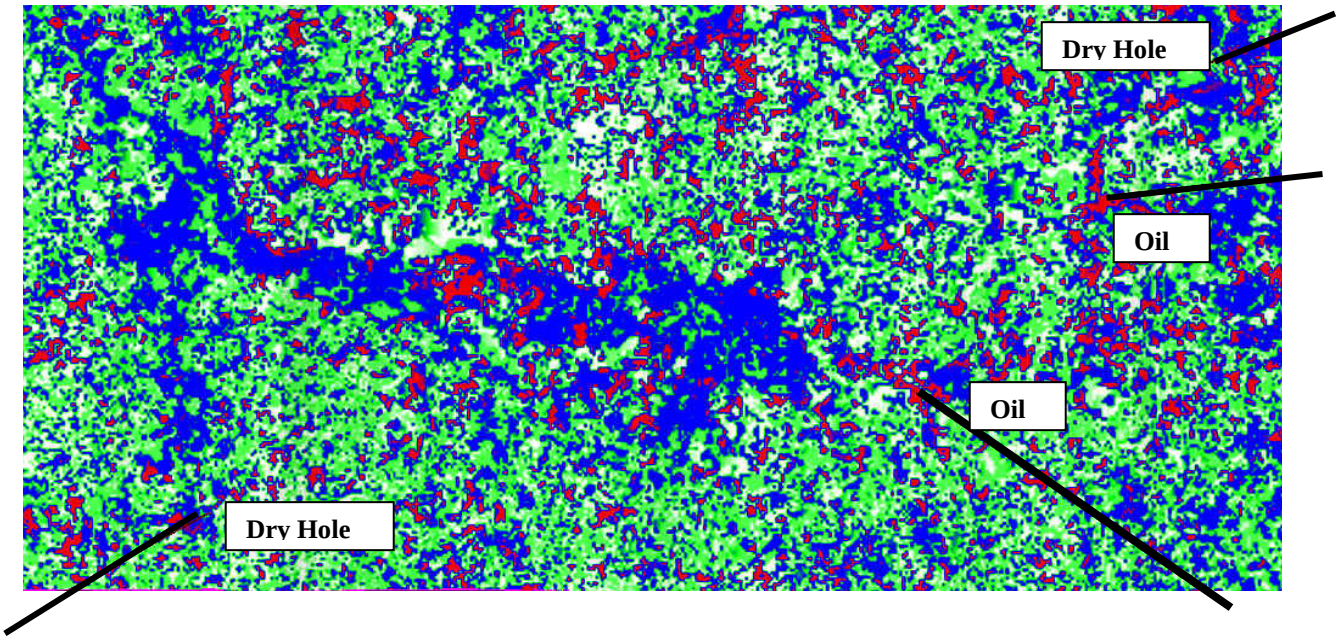

Figure 5: Time slice through 3D seismic volume showing spotty distribution of oil sand class (red) and location of 4 wells.