A Bayesian approach for lithofacies identification and classification

Igor L. S. Braga* UENF/LENEP, Luiz Geraldo Loures, UENF/LENEP

Abstract

This work introduces a methodology for lithofacies identification and classification following the probabilistic theory of pattern recognition. This methodology applies the Expectation Maximization algorithm to identify lithofacies in the learning stage. This unsupervised learning process is applied to well log data to identify lithofacies. Uncertainty analysis is performed to predict the classification success and the Bayesian decision theory is applied to classify new incoming data.

This methodology was tested in well log data. The results achieved are interesting enough to extend the methodology to seismic resolution to classify lithofacies in the inter-well zone using seismic attributes.

Introduction

Lithology, pore-fluid properties and porosity prediction from seismic data has become an important goal of petroleum exploration industry, however, the reservoir characterization process is time consuming and an uncertainty analysis challenge. The knowledge regarding these reservoir petrophysical properties is a key for hydrocarbon recovery efficiency (Mukerji, 2001).

A lithofacies mapping inside the reservoir is important for rock physics model calibration. These rock physics models are tools to extract the petrophyscal information from seismic attribute. This work presents a methodology for reservoir lithofacies classification and identification.

Recently, new methodologies for lithofacies classification have been published. Some of these methodology use neural network approach applied to textural attributes (West B. P.,2002); discrimintant analysis and bayesian classification applied to well-log and prestack seismic inversion (Avseth,2001) and acoustic and elastic impedance with a Mahalanobis distance classifier (Mukerji & Jorstad, 2001). Most of these methodologies use a pre-defined training data-set to predict lithofacies.

In this work, we aim to enhance the framework for lithofacies classification using the Expectation-Maximization algorithm (EM) on well-log data to identify clusters representing lithofacies and to perform lithofacies identification. This process is performed with a bayes classifier, which provides an associated uncertainty measurement.

The output of this methodology is:

i- the classified lithofacies;

ii- the reservoir lithofacies mapping and

iii- an associated uncertainty measurement.

The methodology was tested using well-log data from a heavy-oil sand/shale reservoir. Next section we present the theoretical background followed by test results and conclusion.

Methodology

The following steps can summarize the methodology process:

1- unsupervised learning using the EM algorithm to group the data in *k* groups (lithofacies);

2 - model selection criteria (BIC) application to define a *k* optimal (the number of lithofacies);

3 – uncertainty analysis: predict classification quality using the bayes error;

4 - new data lithofacies classification classification using the MAP classifier.

1- Expectation-Maximization algorithm:

The EM algorithm (Dampster, 1977) is a technique to clustery data using an observed part of data to estimate an unobserved part. We assume that the data is distributed according to a mixture of Gaussian density function *G*, denoted by

$$p(x) = \sum_{l=1}^{K} G(x \mid \mu_l, \Sigma_l)\omega_l \qquad (1)$$

where *k* represents the number of mixture components, μ and Σ are a mean vector and a covariance matrix and ω_l the weight of each component. Let $\theta_k = \{\mu_1,...,\mu_k, \Sigma_1,...,\Sigma_k, \omega_1,...,\omega_k\}$ be a set of parameters in the mixture. The goal of EM algorithm is to estimate the parameters θ_k based on a set of samples $\{x_1,..,x_2,...,x_n\}$ using a learning principle. We estimate these parameters by maximizing a likelihood function $L(\theta_k)$:

$$L(\theta_k) = \ln \prod_{i=1}^{n} p(x_i) = \sum_{i=1}^{n} \ln \sum_{l=1}^{K} G(x_i \mid \mu_l, \Sigma_l)\omega_l \qquad (2)$$

EM alternates between the two actions: i- calculates the expected value for the likelihood from unobserved parameters, however, we do not know this values, so an initial guess must be used; ii- the maximization, which consists in obtaining the parameters that maximize the expected value of the likelihood. The new parameters will be the guess for the next interaction. The process continues until convergence.

2- model selection criteria (*BIC*)

The problem that remains is that we do not know the number of components *k* in the mixture. To surpass this problem we define a range from k_{min} to k_{max}, which is assumed to contain the $k_{optimal}$. At each *k* we obtain a value for a penalized model selection criteria. We used BIC (Bayesian Information Criteria) (Schwarz,1978) for such selection criteria, BIC is defined as

$$BIC = -2L(\theta_k). + \ln(n)(k-1) + k(d + (d+1)/2) \tag{3}$$

where *d* is the data dimension. The optimum *k* represents the *k* that minimizes Equation (3).

3 - uncertainty analysis

This methodology uses Bayes error (or Bayes risk) to quantify the classification risk. The Bayes error represents the probability of classifying a determined lithofacies as another. The Bayes error is a measure that varies from 0 (no classification error) to 0.5 (maximum classification error), a risk analysis should be done before classification process. The following expression represents the Bayes error:

$$l^* = \int_{wj=false} P(w_j = true \mid x)P(x)dx + \int_{wj=true} P(w_j = false \mid x)P(x)dx \tag{4}$$

Takahashi (2000) empirically demonstrated that the Bayes error is invariant to non-linear coordinate transformation and is a reliable measure of uncertainty.

4 - Bayesian classification

A MAP (Maximum Posteriori Probability) classifier is applied with a Bayesian decision theory for lithofacies classification. A MAP operator is defined as

$$\delta(x) = \arg\max_k P(\omega_K)P(x \mid \omega_K) \tag{5}$$

where $P(\omega_k)$ is the a priori probability of lithofacies *k* obtained in the clustering stage. To calculate the likelihood $P(x|\omega_k)$, a *k*-nearest neighbors approach based on Euclidean distance was used.

After the Bayesian Classification**,** we analyze the petrophysical property trends in a set o cross-plots, following a rock physics diagnostic. This analysis combines theoretical models, such as Hashin-Shtrikman bounds (1963), empirical rock physics models (Dvorkin and Nur, 1996) and lab data from Han et al (1986). This analysis helps us to understand the physical properties behavior for the classified lithofacies.

Well log tests

The EM was applied on a well log data set (well 1). Using sonic logs (P-wave and S-wave velocities) the EM identified five clusters, representing five lithofacies. Figure 1 presents the *BIC* values obtained after the EM analysis from K = 2 to K_{max}. Note that *k*=5 minimize the *BIC.* Figure 2 shows in a cross-plot Vp-Vs the sonic data (Vp and Vs), with color indicating the lithofacies.

The next step is to build probability density functions (pdf) for each lithofacies. These pdfs represent the available information to classify new incoming data. Figure 3 represents the probability distribution of each lithofacies.

Figure 4 represents two pdfs. The area inside the pdfs that is not shaded represents the Bayes error. Table 1 summarizes the Bayes error associated to each one identified lithofacies.

Table 1 – Bayes error for the identified classes with different pairs of parameters.

	Vp – Vs	P-Impedance S-Impedance	Vp – Rho
Facies 1	0.049021	0.117733	0.289426
Facies 2	0.064001	0.168257	0.276442
Facies 3	0.048891	0.179122	0.322092
Facies 4	0.074666	0.115885	0.255185
Facies 5	0.069354	0.111100	0.207562

After defining the lithofacies and building the pdfs for each lithofacies, the classifier was applied on income data-set from well 2, located near well 1. Figure 4 shows Vp and Vs logs from well 2, the identified lithofacies and the probabilities to happen the classified lithofacies. Table 2 shows the mean value of Vp, Vs, density and Vp/Vs ration for the five identified classes.

Table 2 – Mean values of Vp, Vs, density and Vp/Vs ratio for the identified facies.

	Vp (m/s)	Vs (m/s)	RHO (g/cc)	Vp/Vs
Facies 1	3227	1746	2.222407	1.84
Facies 2	3427	1775	2.196168	1.93
Facies 3	3514	1841	2.266681	1.90
Facies 4	3587	1881	2.281460	1.90
Facies 5	3707	1992	2.352574	1.86

To test our results we cross-validated the samples from well 1. Half of the set was used as training data and the other half was classified with the MAP classifier. We obtained a success rate of 92% defining as 20 the numbers of nearest neighbors.

Figure 6 shows a cross-plot porosity-K with well log data from well 1 and with colorbar indicating the lithofacies. Figure 7 (from Takahashi, 2000) shows the relation of several factors that can influence rocks properties in the porosity-K domain. Figure 8 shows a crossplot the Mu-Rho x Lambda-Rho and Figure 9 (from Pelletier & Gunderson, 2004) shows the lithology variation in this domain. From analysis of these cross-plots we can observe:

i- facies 2, 3 and 4 are most likely to be sandstone, with increasing clay content. Shale has a wide range of velocities (and impedances) and increasing clay content in the rock may lead to erroneous interpretation.

ii-Facies 3 and 4 have the same Vp/Vs ratio, which is a good lithological indicator. These two facies overlap each other in the Mu-Rho X Lambda-Rho domain. They are probably the same facies and the Bayesian criteria slightly over-fitted the model.

iii-Facies 5 falls in the bounds between sandstone and tight sands, while facies 1 falls within the bounds of shale and sandstone. Unconsolidated sandstone and pure shale can present same velocity. Others parameter pair, which are sensitive to clay content, should be used to minimize this ambiguity.

Conclusion

This methodology shows good results in identifying clusters and classifying new samples. However identifying lithofacies using Vp and Vs sonic logs only can lead to erroneous interpretation.

To be able to correctly predict lithofacies, other parameters which are good lithological identifiers, as Vp/Vs ratio S-impedance, Mu*Rho, Lambda*Rho should be used.

The results sufficiently encourage us to extend this technique to inter-well region using seismic attributes after the identification of which parameters best responds to different lithofacies in this data set.

Acknowledgments

We would like to thank the PRH20/ANP program and Landmark Software Grant Program.

References

Devroye, L, Gyorfi, L. e Lugosi, G., (1996) *A Probability Theory of Patter Recognition*, Springer-Verlag Inc, New York, 636 pp

Hilterman, F., 1989, *Is AVO the seismic signature of lithology?*: 59th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 559.

Loures, L. G. C. L. & Downton, J. E., 2004**.** *Bayesian lithological inference for a sandstone reservoir*. 2004 CSEG National Convention – *Accepted for publication.*

Mavko, G., Mukerj, T. e Dvorikin, J., 1998, *The Rock Physics Handbook*, Cambridge University Press, 329 pp.

Xuelei Hu , Lei Xu , 2004, *Investigation on Several Model Selection Criteria for Determining the Number of Cluster,* Department of Computer Science and Engineering, The Chinese University of Hong Kong.

P. Avseth, T. Mukerji, A. Jørstad, G. Mavko, and T. Veggeland, *Seismic reservoir mapping from 3-D AVO in a North Sea turbidite system,,* Geophysics, Vol. 66, No. 4 (July-August 2001); P. 1157–1176.

Brian P. West, Steve R. May, John E. Eastwood, Chistine Rossen**,** *Interactive seismic facies classification using textural attribute sand neural networks,* THE LEADING EDGE ,October 2002 .

Isao Takahashi, May 2000, *Quantifying Information and Uncertainty of Rock Property Estimation From Seismic Data,* Stanford University.

Heath Pelletier, Jay Gunderson*,* 2004*, Application of Rock Physics to an Exploration Play,A Case Study from the Brazeau River 3D* **,** CSEG National Convencion.

A. Dempster, N. Laird, and D.Rubin. *Maximum likelihood estimation from incomplete data via the EM*

algorithm. *J. Royal Statistical Soc. B*, 39:1–38, 1977.

Russel S., Norvig P., 2003*, Artificial Intelligence* 2nd edition, Prenice Hall.

Joseph H. F., Rolph A., Tatham R., Black W., 1998, *Multiraviate Data Analysys* 5th Ed., Prentice Hall.

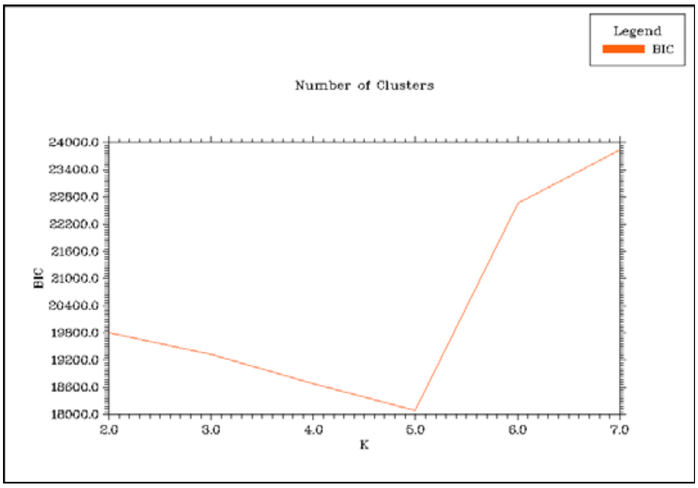

Figure 1 – Results of the BIC (Bayes Information Criteria) obtained after the execution of the EM algorithm. The minimum value indicates the number of components that best fits the model, showing five different lithofacies in the data set.

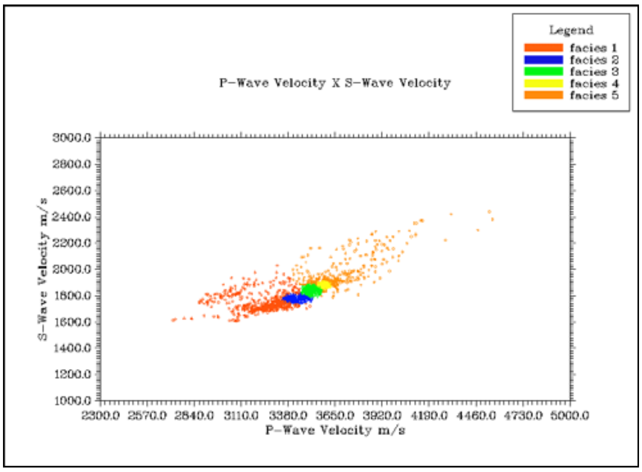

Figure 2 – The five identified facies after the execution of the EM algorithm.

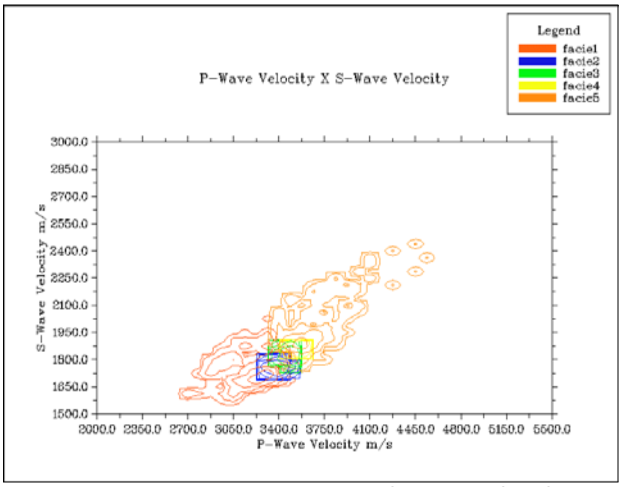

Figure 3 – Bi-variated distributions of the identified facies, used as priori information to classify well 2.

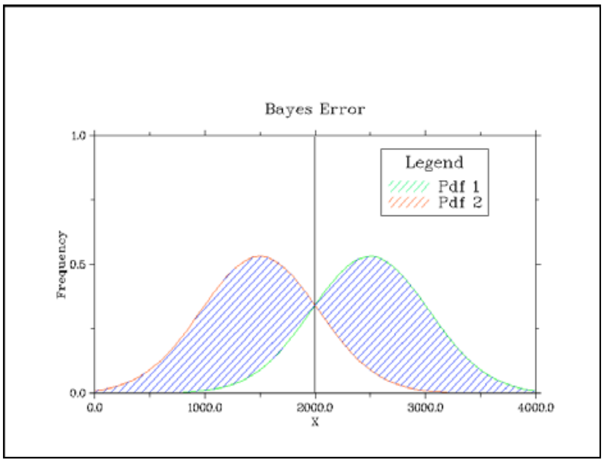

Figure 4 – The area in white under the curve represents the Bayes error of two univariated distribution.

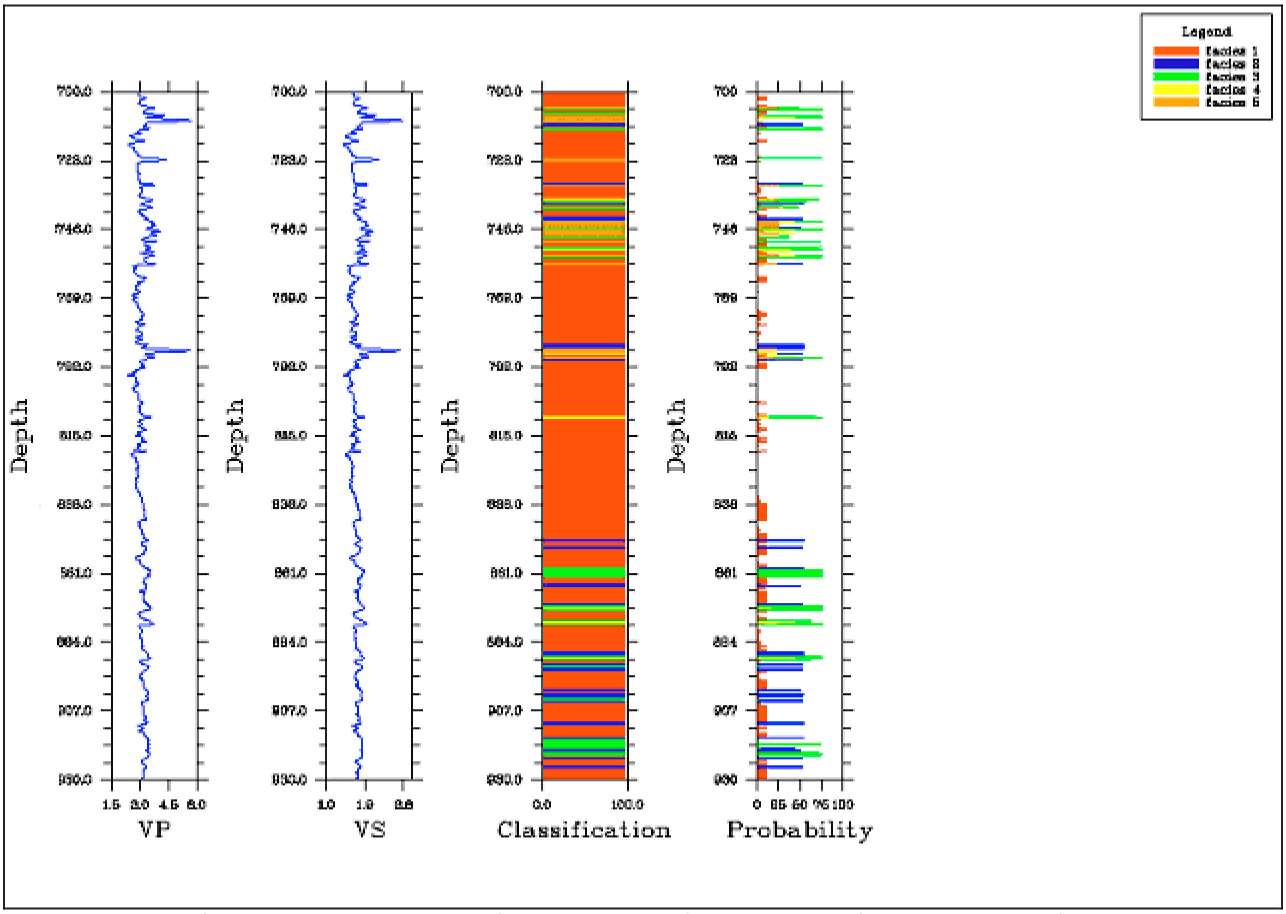

Figure 5 – From left to right , Vp and Vs logs of well 2, the classification results after the MAP classifier and the probability of each classification.

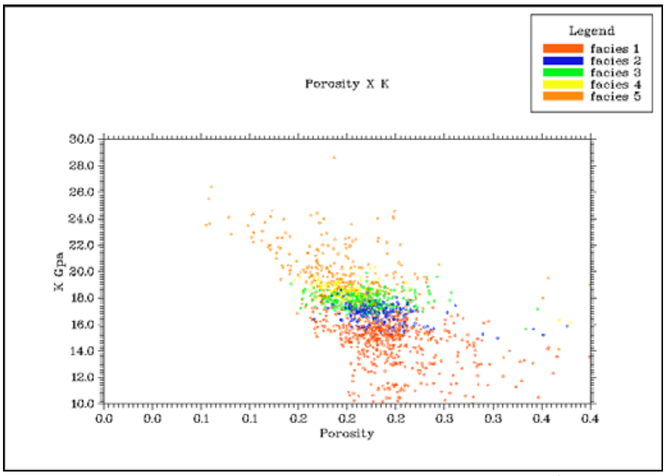

Figure 6 – K (bulk modulus) – Porosity cross plots of the identified facies.

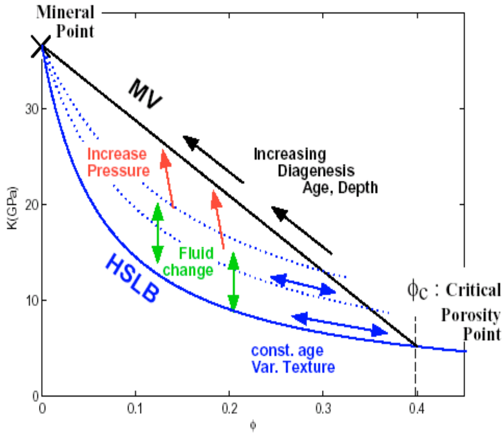

Figure 8 – Trends of sandstone in the K (bulk modulus) – Porosity domain, (Takahashi, 2000).

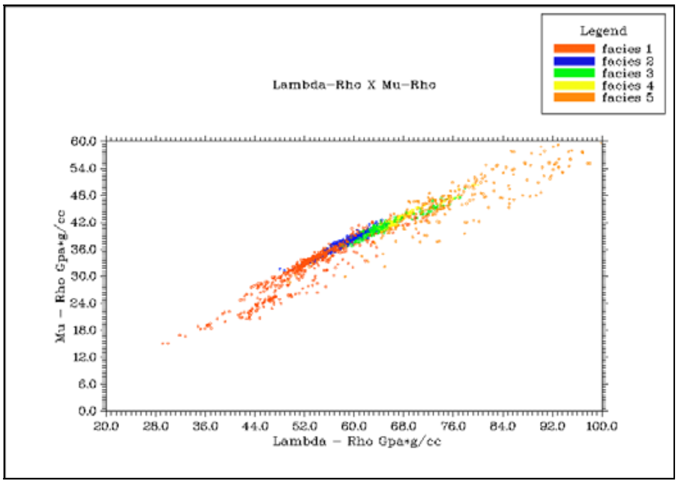

Figure 7 – Mu–RHO X Lambda-RHO crossplots of the identified classes.

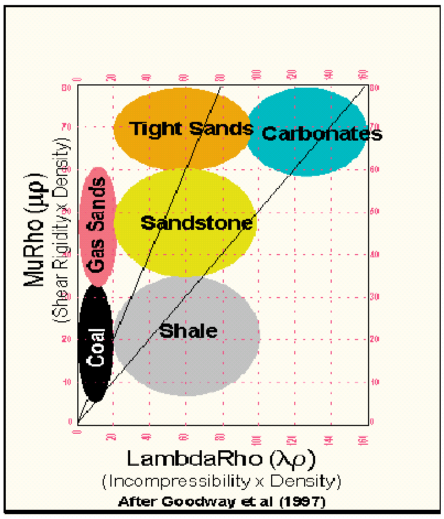

Figure 9 - Trends of different rocks in the Mu–RHO X Lambda-RHO domain, (Pelletier & Gunderson, 2004)