Petroleum Res. J. Vol. 13 (2001); P. 47-56, 15 Figs

Defining Electrofacies from Logs and Core Data: Principles of Supervised and Non-Supervised Approaches

Gaël Lecante*, Pierre-Yves Dequirez** and Pascal Lavest*

معرفة السحنات الإلكترونية بواسطة السرود وبيانات العينات اللبية: أسس التطلعات الموجهة وغير الموجهة

ج. ليكانتي وبييرا ديكويرز وبسكال لافيست

إن التعرف على السحنات الصخرية المستنبطة من تسجيلات الآبار السلكية ومعايرتها باستخدام بيانات العينات اللبية يعتبر أحد مفاتيح التقييم الناجح للمكمن، وتحديد نوع الصخر حيث يمكننا من إشتقاق البيانات المدخلة واللازمة لبناء نماذج ثلاثية الأبعاد للمكمن.

إن التحليلات البيانية متعددة المتغيرات توفر الآلية الصحيحة التي تسمح بالآتي:

- تفسير تسجيلات الآبار السلكية وبيانات العينات اللبية.
- توقع السحنات الإلكترونية لعدد كبير من الآبار باستخدام مجموعة متكاملة من السرود.
- معايرة السحنات الالكترونية المستخلصة من سرود الآبار مقابل بيانات العينات اللبية.
- توقع السحنات الالكترونية في الأعماق التي لا تتوفر بها عينات لبية وكذلك الآبار التي لم تؤخذ منها هذه العينات.
- تحديد مقدار الريبة في تعيين السحنات باستعمال السرود.

هذه الورقة والتي اعتمدت على حالة دراسة حقيقية تصف طريقتين أساسيتين لتعيين أو توقع السحنات الالكترونية المستندة على التفسير البياني متعدد المتغيرات.

- الأسس غير الموجهة وهي تستند بدرجة أساسية على التفسير البياني متعدد المتغيرات لتسجيلات الآبار السلكية بغض النظر عن بيانات العينات اللبية.
- الأسس الموجهة هي التي تربط أو تدمج بين تسجيلات الآبار السلكية وبيانات العينات اللبية.

وكخطوة ثانية، فإنها تصف كيفية الربط بين هاتين الطريقتين بشكل ناجح وتكاملي لمعرفة أو للتوقع الأمثل للسحنات باستخدام السرود في الآبار التي توجد بها عينات لبية أو التي لا توجد بها هذه العينات.

- تفسير دالة الكثافة عديدة المتغيرات لتحديد تجمعات النقاط.
- تفسير هذه التجمعات لتعريف السحنات الإلكترونية باستخدام بيانات العينات اللبية.
- توقع السحنات الإلكترونية في المواضع التي لم تؤخذ بها عينات لبية وكذلك في الآبار التي تتوفر فيها هذه العينات.

Abstract: *The definition of log facies from wireline logs and their calibration against core data is one of the keys to successful formation evaluation and rock-typing. It allows to derive the input data that is essential for building accurate 3D geocellular reservoir models.*

* Beicip – Franlab, ** IFP

- *Multi-variate statistics provide the right tools that allow:*
- *to analyze wireline logs and core data;*
- *to predict electrofacies from a large number of wells and using complete sets of logs;*
- *to calibrate the detected log facies against core data;*
- *to predict the log facies at the non-cored intervals and non-cored wells;*
- *to quantify the uncertainty of the log facies determination.*

This paper, which is based on a real case study, describes two basic approaches for determining and predicting electrofacies, based on multi-variate statistical analysis:

- *A non-supervised approach, that is purely based on multi-variate statistical analysis of the wireline logs, regardless of the core data.*
- *A supervised approach, that integrates wireline logs with core data.*

As a second step, it describes how these two basic approaches can be combined to identify and predict optimal log facies at cored and non-cored wells, in an integrated and robust workflow:

- *Multi-variate density function interpretation and cluster identification*
- *Cluster interpretation and electrofacies definition, using core data*
- *Electrofacies prediction at the non-cored intervals and non-cored wells.*

INTRODUCTION

What is an Electrofacies?

Well data are a set of z values. A set of log measurements is associated to each z value. Data can be represented as shown in figure 1.

Samples are in a p-variable hyperspace, in which each axis is linked to a tool measurement (Fig. 2).

An electrofacies is nothing but a cluster of z values in the space. Its geological meaning is a priori known or it must be interpreted (from core data for instance).

$$\begin{array}{ccc} & \text{log1} & \text{logp} \\ [X] = & \begin{bmatrix} & (x_{ij}) & \end{bmatrix} & \end{array}$$

Fig. 1. Matrix of the data. (xij) = log j measurement at the point i.

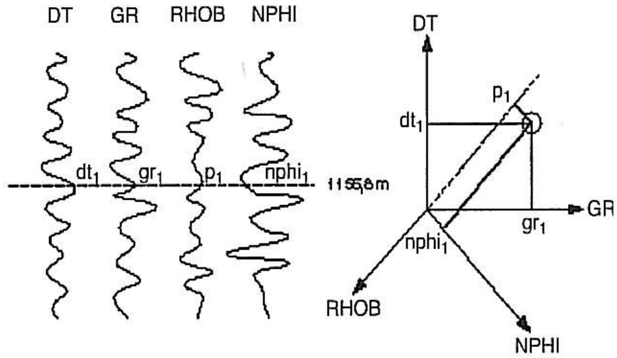

Fig. 2. Display of the samples in the log hyperspace.

The electrofacies gathering depends closely on the variables (logs) in use. The electrofacies definition and prediction presented in this paper combines a statistical analysis with the standard geological interpretation of logs (Fig.3). Logs and core data are both taken into account in order to generate facies logs that honour both pieces of information.

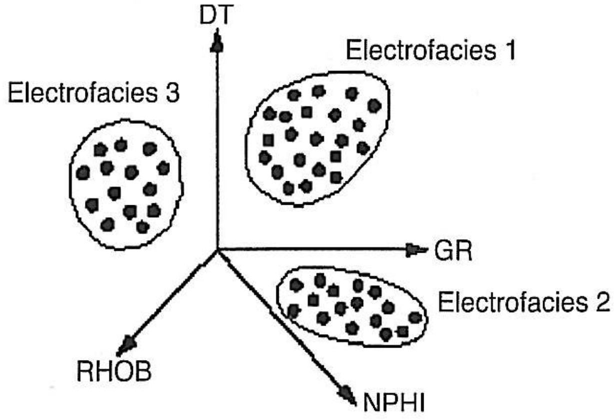

Fig. 3. Gathering into electrofacies.

Supervised or Unsupervised Approach?

Two different classification methods can be performed: a supervised and unsupervised. The supervised approach takes into account a priori data given by the geologist which constitute the "training sample". The number of classes and their log characteristics are determined by these data. When running unsupervised approach, the training samples are defined from the interpretation of the density function calculated on log data. The number of classes and the corresponding characteristic samples are decided and selected at this step. Their geological meaning will be interpreted from their location in cross-plots and/ or from core data.

In both cases, points not yet belonging to any class will be next assigned to a class according to a classification function created from the training sample.

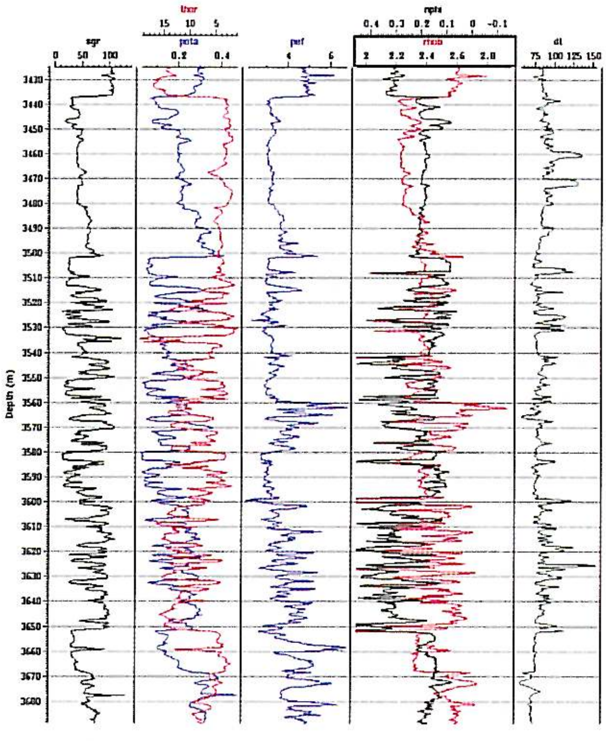

Fig. 4. Composite log

The two methods can be run independently or be linked together:

- electrofacies detected during an unsupervised approach can constitute a training population to add as a priori information to a supervised study;
- the geologist can add a priori supplementary class to these that have been interpreted during the unsupervised approach.

The study we use for showing the electrofacies definition methodology study takes place in a fluvio-deltaic environment. Figures 4 and 5 display the composite log and some classical cross-plots.

The Unsupervised Approach

The unsupervised approach can be divided into 3 important steps.

- density function estimation,
- density function interpretation: number of classes, creation of the training sample,
- data points assignment.

Density Function Estimation

An electrofacies is nothing but a gathering of points together in the log hyperspace. Electrofacies classes constitute a set of points separated by areas with low density of points. By detecting and analyzing the variation of the density in the log hyperspace, it is possible to detect these different sets. The number of density peaks defines the number of classes. Points highlighted in these zones of high density correspond to the most typical samples of each class.

The probability density function (PDF) describes the distribution of a variable and gives its associated probabilities. In a univariate case, the PDF is often described in a discrete manner with a histogram. However this histogram is not a good density function estimator. It does not give the number of samples for a particular value; it only gives the number of samples that fall into an interval.

In a p-dimension space (corresponding to the p logs which are used), the PDF is estimated using KNN (K nearest neighbors) or kernel techniques (Gaussian, Epanechnikov, *etc.*). If only one variable is used, the density function is a curve (that can be approximated by smoothing a histogram). If two variables are used, the density function is a surface with valleys and hills like a topographic surface. When more variables are

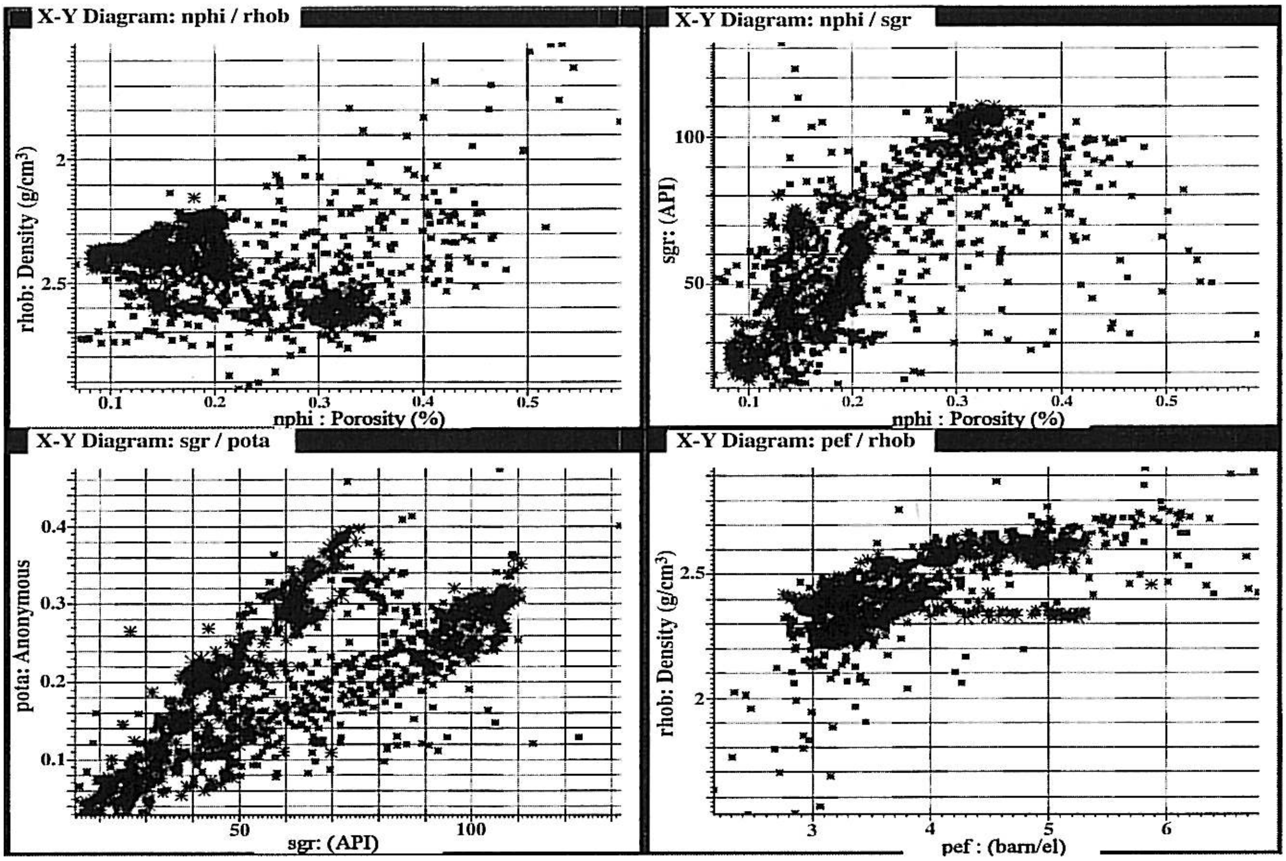

Fig. 5. Cross-plots

used, it is not possible to represent the PDF, however, we can analyze the peaks of the PDF using Kittler algorithm[1]. This technique is called the mapping of the PDF or PDF mode mapping (Fig. 6):

- A random point is chosen.
- A neighbouring is defined in order to calculate the slope in all directions.
- The next point is chosen from nearby in order to move as far uphill on the PDF as possible (*i.e.* the greatest slope is used)
- The process goes on until reaching a density peak.
- From this, the following points are selected in order to ensure that one moves as little downhill as possible until a minimum of the PDF.

This gives a path that climbs steeply uphill to a mode and then goes down slowly, visiting all points. The neighbouring, called « Mode mapping parameter », is the smoothing parameter to choose. If it is too big, you loose information about density. If it is too small, all the details are mapped, even the small ones, introducing noise in the interpretation.

Assignment of Samples to a Class: the Classification Method

The next step of the classification aims at classifying all log samples. At this point statistical techniques such as discriminant analysis or pattern recognition can be used. In the methodology presented in this paper, the discriminant analysis is used. It is fast, robust and has shown an unmatched efficiency and reliability on any type of environment (clastics and carbonates).

In a discriminant analysis, the probability of belonging to a class is computed for each sample. The point is next allocated to its most probable belonging class. This calculation uses the well-known Bayes formula. Various kinds of discriminant functions can be used for classifying the samples:

- The linear hypothesis: the covariance of each class is assumed to be the same,
- The quadratic hypothesis: for each class a distinct covariance is calculated from the training sample.
- Non-parametric hypotheses can also be made.

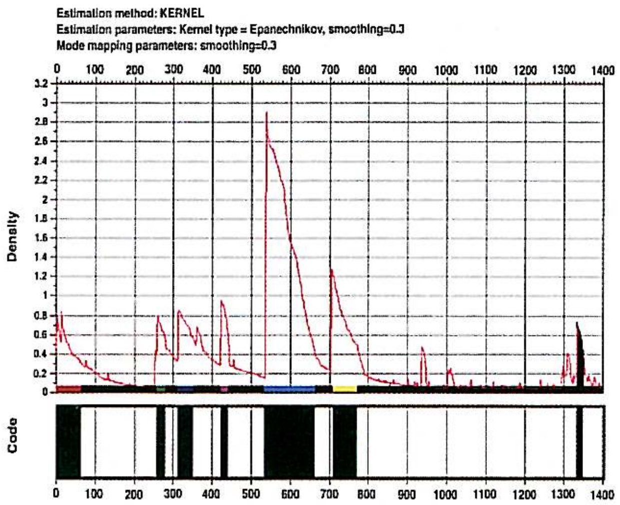

Fig. 6. Density function mapping

In such a case, the probability laws for each class have to be estimated using Kernel or K Nearest Neighbors methods. These laws are nothing but the PDF of each class.

Obviously, the less hypothesis on the model is assumed, the more closer we are to the reality. Ideally, the non parametric approach is the most promising, but with this approach, we need a lot of points in the training sample to obtain a reliable result

Once the assignment method is chosen, the training samples are stored as non allocated points (Fig. 7a & b). The method efficiency can be evaluated by computing the percentage of well-assigned samples (i.e. which are allocated to their actual class).

In the example displayed on figure 8, all samples are correctly assigned, whatever the method is chosen. However, this method is it underestimates the error rate because observations to evaluate the classification results are the same to build the classification rules.

For this reason, a second series of validations is carried out. A sample is taken out of the training sample before building the classification rules. It is re-allocated using this rule afterward. This is a more pessimistic but realistic measure of the classification efficiency (Fig. 9).

In the case displayed on figure 9, all the samples of the class 6 are not correctly assigned using the non parametric method. Quadratic and linear hypothesis give good results. Cross

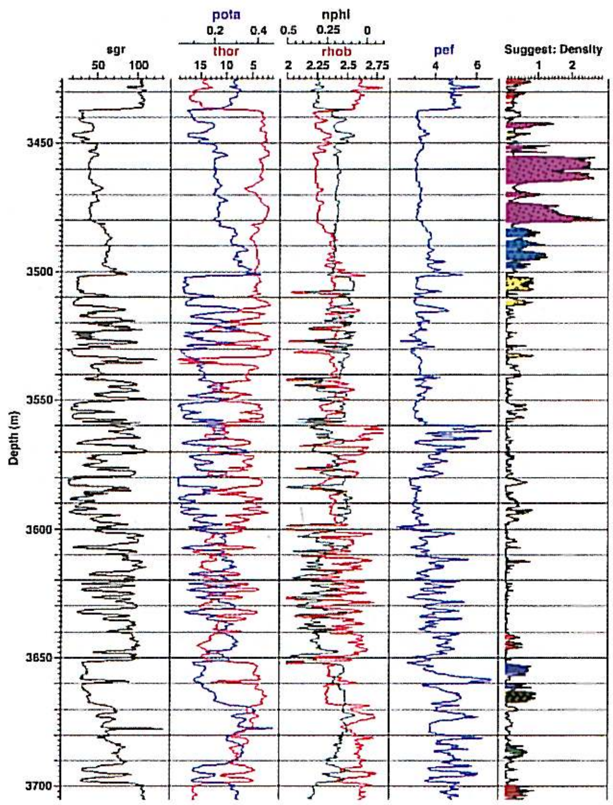

Fig. 7a. Highlighting the training samples on composite logs

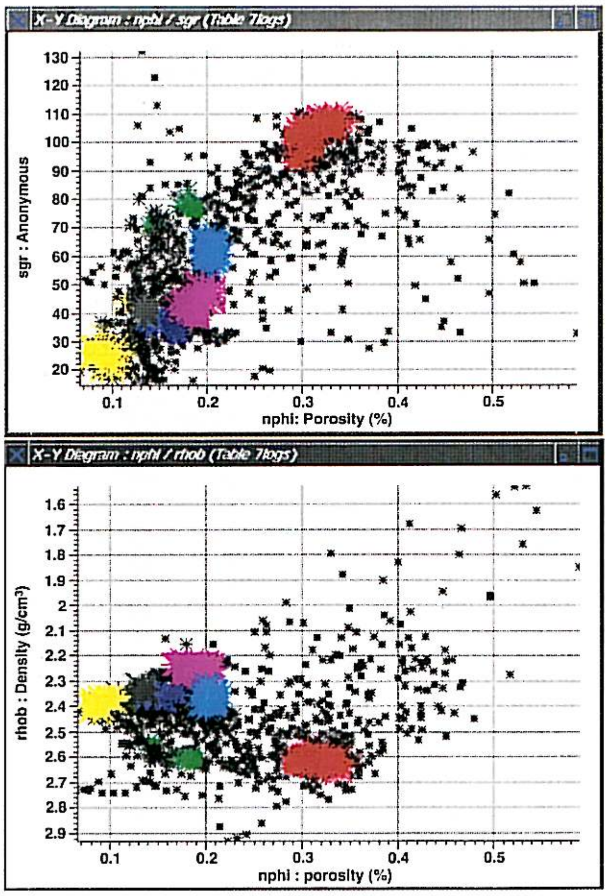

Fig. 7b. Highlighting the training samples on cross-plots

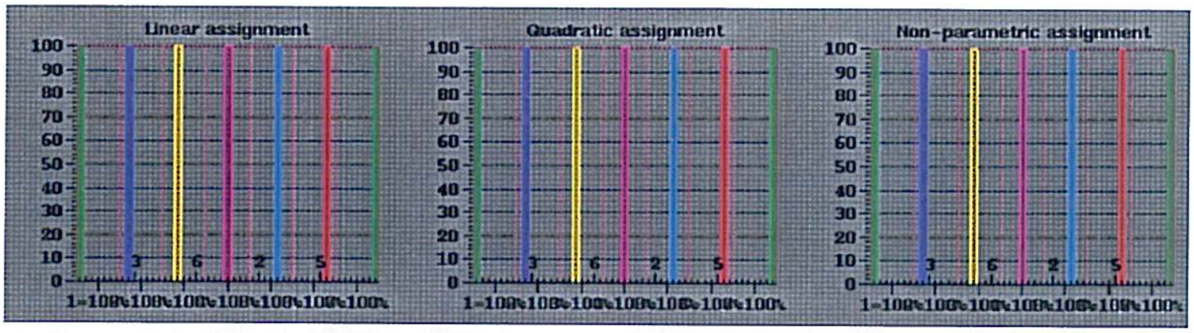

Fig. 8. Direct validation results.

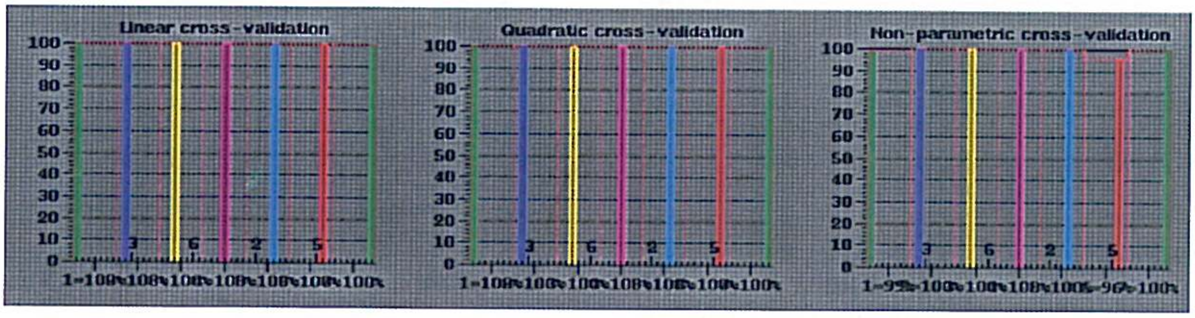

Fig. 9. Cross validation results.

validation is interesting, in the linear and quadratic case, only if the number of points is not too large. Otherwise, if you remove a sample from a lot of points, it has a little impact on statistical methods using only mean and covariance matrices. However, in a non parametric case, this diagram is always interesting to be consulted because removing even one point could affect intra-class density estimations.

The results, obtained with the linear hypothesis are displayed (Fig. 10a and b).

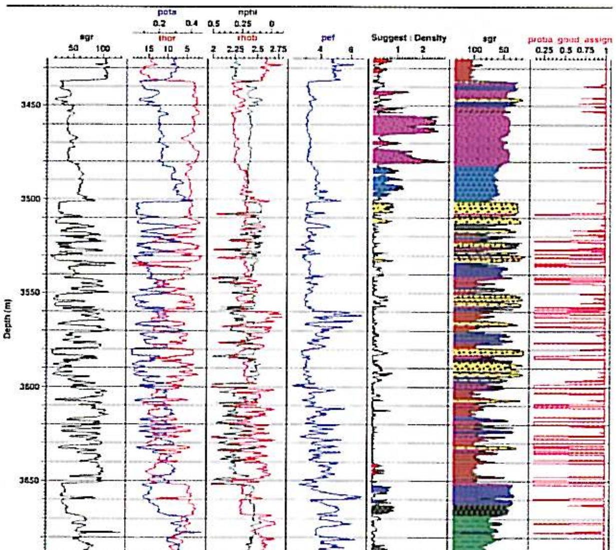

Fig. 10a. Assignments results: logs

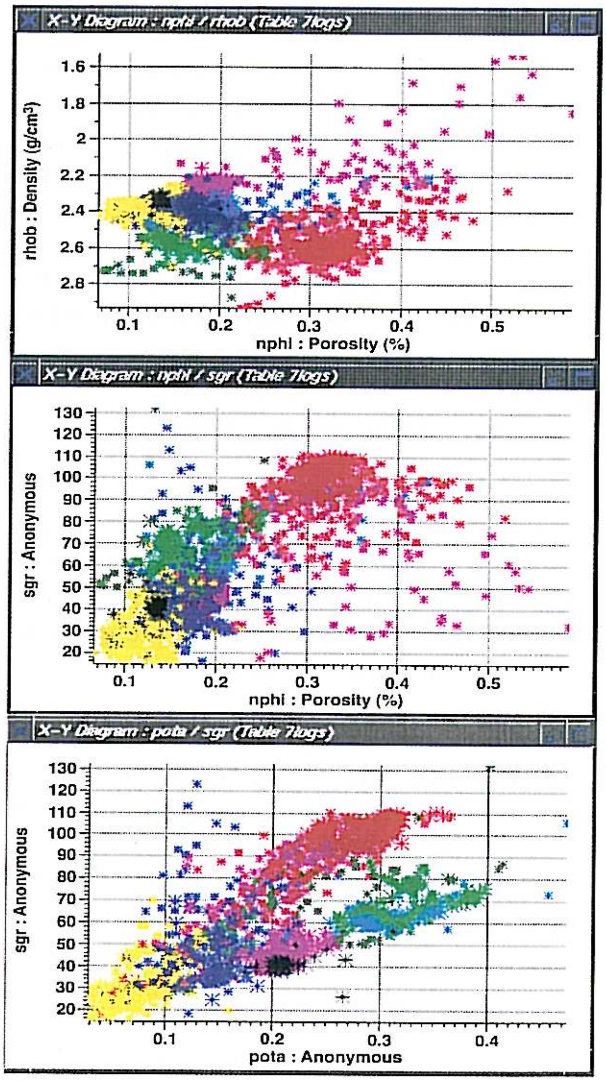

Fig. 10b. Assignments results: crossplots.

At this point, the probability of good assignment is also calculated. It quantifies the uncertainties and controls the quality of the newly built lithologic column. In figure 11a, 10% of the samples are allocated to a class with a very small probability. These samples are outlying as we can see on the cross-plots on which they are highlighted as triangles (Fig. 12).

Supervised Approach

In a supervised approach, we have a priori training sample coming from a geologist prior interpretation. The number of classes is given, it is a part of the geological a priori. The classification function is created from this a priori

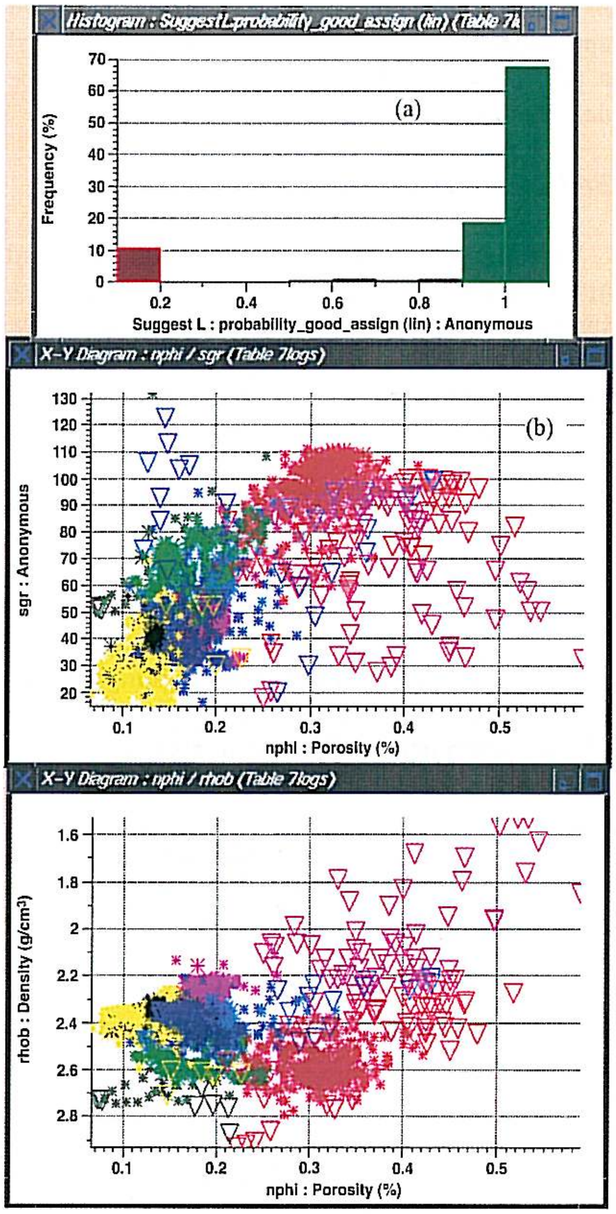

Fig. 11. Highlighting points with a not reliable assignment.

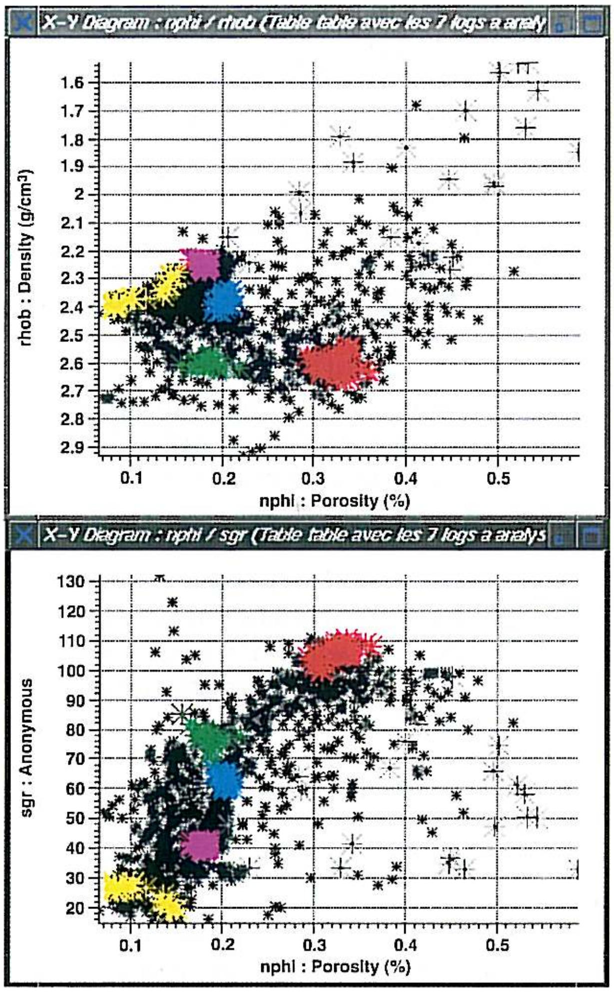

Fig. 12. Training samples displayed on cross-plots.

training sample. All the points will be assigned to a class using this classification function.

Everything is known since we have already discussed the discriminant analysis in the frame of the unsupervised interpretation: we have to choose the hypothesis (linear, quadratic or non parametric), in order to run the classification and check the results.

The Training Samples

In this case, we have to highlight a class of coal not discriminated during the unsupervised approach (Figs. 7a and b). In fact, a class can be identified if it forms a statistical population. Scattered points in the log hyperspace cannot be detected using a unsupervised approach. In the case of a supervised approach, samples with characteristics close to coal will be assigned to a coal class, which does not correspond to the real situation. As a consequence, the set of training sample is very important. The confidence in the results depends on the confidence in the training samples.

Choosing the Discriminant Method

In the case presented here, the cross validation shows that the linear method does not allow a reliable identification of the coals. We cannot use the non-parametric method either, as we only have 205 training samples (20 to 50 per class), which is not enough. At last, if we use the quadratic assumption, almost 70% of anonymous samples were classified into the class of coals. This is not good but foreseeable, as Coals have very scattered log responses.

The quadratic method computes each class covariance from the training sample. If sample characteristics are very dissimilar, the class covariance will be very high and the attraction of this class will be very important. In our particular case, we want to highlight samples with no typical features. The quadratic assumption is not reasonable.

Consequently, we have to use the linear hypothesis, even though we know that the discrimination may be not as reliable as expected. The obtained results (Fig. 13) are assigned to the different classes in a way that is consistent with the geological expectation.

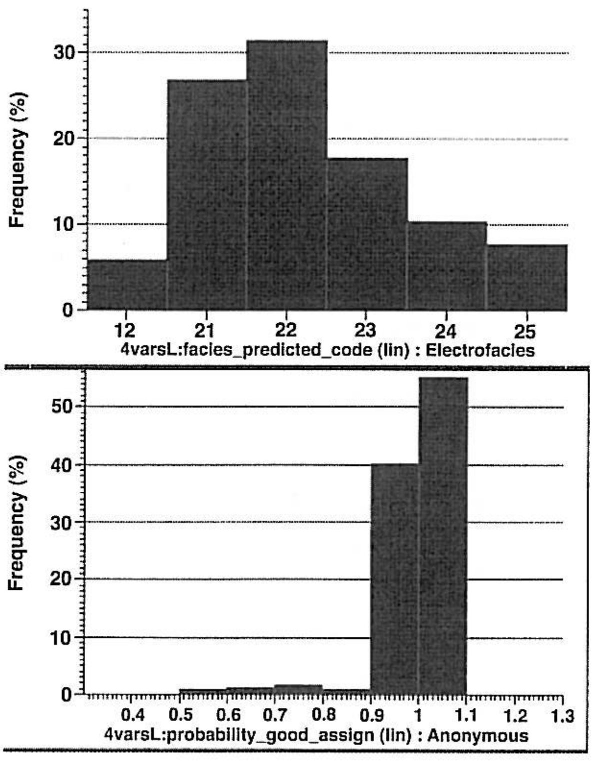

Fig. 13. Samples distribution and probability of good assignment.

As a result, an electrofacies log, calibrated to the geological model, are obtained. They have to be compared to the unsupervised log result.

Supervised or Unsupervised Approach?

Both of Them Can be Used Together

Even if you have an a priori information, both approaches have to be performed, to ensure a better control on the predicted facies. The unsupervised approach increases your knowledge about the sample distribution in the log hyperspace. The convergence in results of the two approaches is a way to confirm that the geological a priori used in the supervised method can be detected on the available logs.

Both of Them Can be Linked

In our particular case, the unsupervised approach did not identify the coals, which are very important from the geologist point of view. Conversely, the supervised approach did not take into account some classes that appear on the density function.

Samples allocated as coals can be used as constrains to unsupervised approach. The samples corresponding to coals can be picked on the PDF (Fig. 14).

The coals are now a part of our training sample, used in the unsupervised approach. The classification using a linear rule gives the results that we can see in figure 15.

The coals seem accurately identified. The seven classes discriminated during the unsupervised approach are consistent with the a priori information.

Automatic Interpretation of the Other Wells of the Field

The well that has been studied first is now used as the reference. The obtained classifier allows an automatic interpretation of all wells drilled in a similar geological environment in this reservoir.

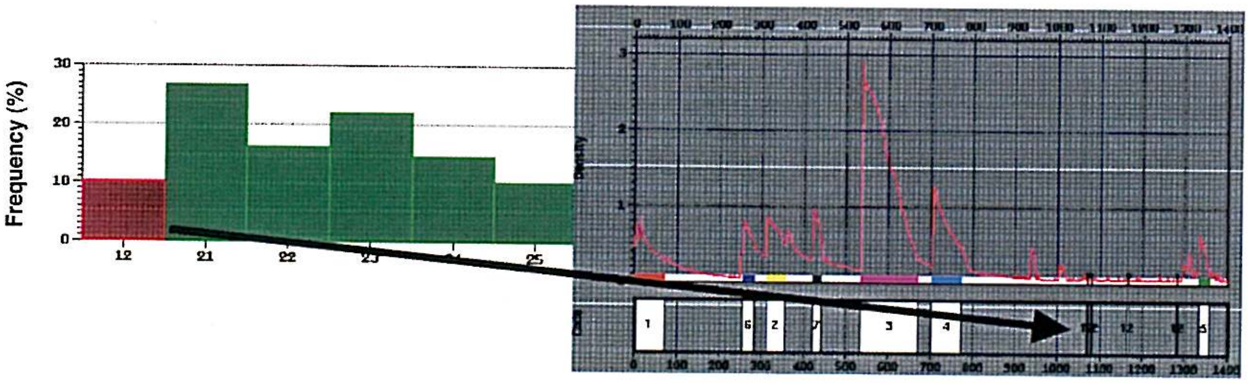

Fig. 14. The training coals are selected on the a priori histogram and a class is created.

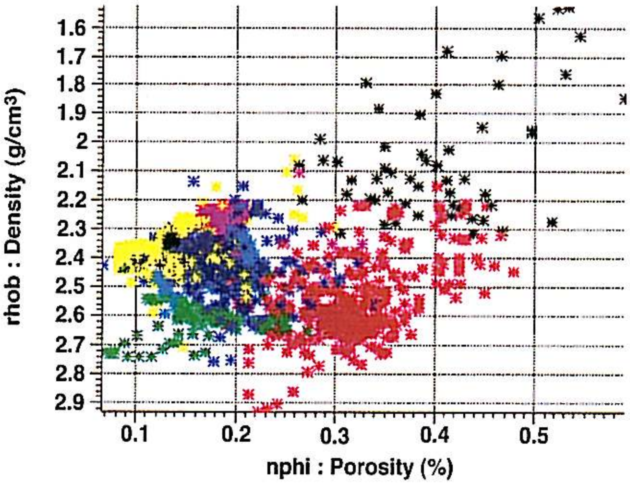

Fig. 15. The coals are identified.

If we consider a second well, where the spectral components of Gamma Ray was not acquired, the previous classifying rule cannot be applied directly. A new one has to be defined using all available logs as Nphi, Rhob and GR.

The new rule is instantly created using the unsupervised application with the training sample selected previously. The newly built classifying function allows the assignment of the samples of each well to classes.

The multi-variate statistical approach for accurate electrofacies determination at wells, within the non cored intervals and at the non cored wells is extremely robust and easy-to-use, and has been applied on a large number of studies, for both clastic and carbonate reservoirs. Thus, it is also used as a starting point of rock-typing, as it is far easier to infer rock-types from calibrated electrofacies at wells, than from raw wireline logs.

REFERENCES

[1] Kittler, J. 1976. A locally sensitive method for cluster analysis. *Pattern Recognition, 8, 23-33.* Patent SN 97/12.182 – France & SN 157/572 – USA : Méthode statistique d'évènements liés aux propriétés d'un milieu complexe tel que le sous-sol.