The use of well logs in logfacies modeling – example in the Namorado Field, Campos Basin, Brazil

Camila Faria de Albuquerque*, José Agnelo Soares, Claudio Bettini. PRH-ANP-18/DEGEO/IGEO/UFRJ, RJ, Brasil.

Abstract

One of the most important stages for oilfield development is the reservoir modeling, which can be performed in many ways. In this study, we decided to use facies modeling based on well logs, owing to their availability in the petroleum industry. We have used the statistical software Enterprise Guide® 2.0, with a specific tool developed for logfacies modeling. The logfacies modeling used here includes a sequence of stages: calibration, validation and application of a discriminant function, which can be linear or quadratic. The data were previously treated with discriminant and cluster analyzes. This study describes an example of the application of this technique to the Namorado Field, Campos Basin, Brazil, showing the modeling stages and the results for two wells.

1. Introduction

Reservoir modeling is fundamental for many stages in reservoir development, including new well location. The geological model can be constructed in different ways. Each model depends on the data used in its construction and, consequently, shows various levels of uncertainty.

Therefore, a model generated using core description, for example, has desirable features such as high resolution and reliability. Nevertheless, obtaining well cores is too expensive and, sometimes, recovery is poor. Thus, core availability is usually too limited.

On the other hand, geophysical data are generally abudant and available. Such data are indirectly obtained and show variable uncertainty and resolution. Seismic data, for example, have resolution that, normally, is sufficient for delimitating reservoir's top and base, but can't discriminate its internal stratigraphic units. Nevertheless, well logs have higher resolution than seismic data, and can be used for generating stratigraphic models. This application is widely known as "logfacies modeling".

Lithofacies definition consists of direct rock observation in well cores, outcrops and cuttings, and analysis of lithology, cementation, fluids, etc. Logfacies consists of indirect rock recognition using well logs, by distinguishing well logs' behavior, such as: high resistivity, low radioactivity, etc.

This study aimed at the application of the logfacies modeling strategy, described by Soares (2005), to the Namorado Field, Campos Basin, Brazil.

2. Data

We used the public data released by the Brazilian National Petroleum Agency (ANP), relative to Namorado field's core descriptions and well logs.

This package consists of 56 wells that, basically, contain an assembly of five well logs, loaded as a LAS archive: Gamma Ray (GR), Density (RHOB), Neutron (NPHI), Sonic (DT) and Resistivity (ILD). However, only 19 wells have the Sonic Data, and, other 19 have the core description, called as ANASETE by PETROBRAS. The intersection between data with core description and sonic log represents only 13 wells.

3. Statistical techniques used in this paper

The main statistical techniques used for logfacies determination based on core descriptions and well logs are: principal component analysis, cluster analysis, discriminant analysis and regression analysis. In addiction to these, there are numerical techniques, such as: neural network and fuzzy logic. In this study, we adopted the discriminant and cluster analysis as a method to settle the logfacies (Souza Jr., 1992).

According to Bucheb & Evans (1992), the study of multivariate statistical techniques have great acceptance between well log interpreters since the Serra and Abbott (1980, *apud* Bucheb & Evans, 1992) pioneer studies. The former authors quote many studies that apply these techniques in the petroleum industries, using well logs and core descriptions, and also using bioestratigraphy and geochemical data.

3.1. Cluster Analysis

Cluster analysis is a multivariate statistical technique that uses similarity to hierarchically classify individuals into groups, approximately heterogeneous, simultaneously regarding all variables (Moura, 1985).

The measures of similarity rank normally used are the Euclidian distance and the Pearson correlation coefficient. The first one is used when one wants to establish the similarity rank between objects, also known as "Q" mode in cluster analysis. The second one is used to measure the similarity rank between variables, also known as "R" mode.

According to Silva & Silva (1990), the sample classification based on similarity criteria allows definition of clusters with homogeneous features. The origin and the geographic localization of the samples can be analyzed and related to clusters, in order to detect affinities and causalities that would remain unnoticed.

Souza Jr. (1992) explains that cluster analysis can be used in logfacies determination, gathering in the same group those lithofacies, preferentially with any genetic association, that have the same characteristics when compared with the well logs. In the same way, this technique can be used in lithofacies determination where the variables could be, for example, the texture and the sedimentary structures.

In this study, the cluster analysis was used in the "non-supervised" classification, where we do not have the core description. Therefore, the sample classification was carried out using well log properties, such as "High Resistivity Logfacies", "Low Radioactivity Logfacies".

3.2. Discriminant Analysis

Discriminant analysis is a multivariate technique that statistically distinguishes two or more previously defined groups in a particular research situation, linearly combining the discriminant variables, trying to maximize the differences between the groups (Moura, 1985; Bucheb & Evans, 1992). To check the distinction between the groups, a collection of discriminant variables that measure the groups' features must be selected.

According to Moura (1985), discriminant analysis weighs and linearly combines the discriminant variables so as to maximize statistical differences between several case groups or populations. Functions that discriminate groups between themselves, termed "discriminant functions", are mathematically defined in conformity with Equation 1, where $\boldsymbol{D_i}$ is the *I-th* function discriminant score; $\boldsymbol{d_{ik}}$ is the discriminant function weight coefficient; $\boldsymbol{Z_k}$ is the standardized value of the variable; and $\boldsymbol{p}$ is the number of discriminant variable used in the analysis.

$$D_i = \sum_{k=1}^{p} d_{ik} \cdot Z_k \qquad (1)$$

Once these functions are defined, it is possible to reach the two objectives of this technique, namely, classification and analysis. The first one allows combining unknown cases within a previously established cluster. The other one provides several tools for data interpretation, such as: (a) statistical tests to measure the importance of a discriminant variable when it is inserted in the function; and (b) coefficient weight interpretation, which helps to identify variables with higher discriminant contribution in a given function (dimension).

In a representative well section of a study area, the facies are previously identified, either using core descriptions, or through automatic classification methods (such as cluster analysis), to obtain the coefficients to be applied uncored wells (Bucheb & Evans, 1992).

If the discriminant function, effectively, distinguishes the considered groups, the facies recognition based on the answers of the well logs in the whole investigated area will be feasible.

In the discriminant analysis it is possible to determine a linear function that discriminates groups (previously defined, using cluster analysis or not), in a way that the misclassification probability of an element in any group will be minimized. This objective is reached using a linear combination of the discriminant variables that maximizes the differences between the groups and minimizes the internal variability of each group (Souza Jr., 1992).

In this study, the discriminant analysis was applied in wells that have their core descriptions, generating a function that relates lithofacies with physical properties measured using well logs. It is used in facies classification where we do not have the well descriptions, expecting to find in these wells the same lithologies found in the cored wells.

4. Method

In this study, in the logfacies modeling, we used the statistical techniques described above – cluster analysis and discriminant analysis. These techniques are inserted in tools specially developed from the software Enterprise Guide® 2.0. The construction of these tools was a result of a project funded by CTPETRO/PETROBRAS/FINEP, called "Projeto Perfil".

In the cluster analysis, these tools have the following methods for data analysis:

1. Average Linkage: the mean distance is calculated between each sample in a cluster and with samples of other clusters. The groups with smaller distances are gathered in a same cluster.
2. Centroid Method: the distance between two groups is defined by the Euclidian distance between two centroids or means. Outliers do not affect this method.
3. K-means Algorithm: used when the K number of groups is known. First, K clusters are randomized. Then, this method reorganizes the samples, trying to minimize clusters' internal variability and maximize the variability between clusters.
4. Ward's Minimum Variance: this method tends to bring together clusters with fewer samples. It has a strong tendency to produce clusters with approximately the same number of samples. So, outliers affect this method.

In the discriminant analysis technique, these tools have the following methods:

1. Linear Discriminant Rule: generates a linear function between each facies and well logs, as the example below.

Facies $x = x_0 + x_1$*GR + x_2*RHOB + x_3*ILD + x_4*NPHI+ x_5*DT

2. Quadratic Discriminant Rule: generates a quadratic function between each facies and well logs, as the example below.

Facies $y = y_0 + y_1$*GR + y_2*RHOB + y_3*GR^2 + y_4RHOB^2 + y_5*GR*RHOB

3. Covariance Matrix Equality Test: automatically decides which of the above rules is more applicable in a study, based on covariance matrix homogeneity test.

4. KNN (K nearest neighbor): assigns a given sample the classification that prevails among its neighbor's classifications.
5. Canonical: using linear combination of original variables, it obtains a variable Y_1 that maximizes the discrimination power between facies. Then it obtains a second variable Y_2, not correlated with Y_1. At a certain moment, new variables Y_n do not contribute to facies discrimination.
6. Stepwise Discriminant Analysis: selects among all well logs those variables that contribute the most for facies discrimination and excludes those which are irrelevant. It uses Wilks' Lambda statistics.

5. Logfacies modeling in the Namorado Field, Campos Basin, Brazil

Logfacies modeling consists of three different stages: calibration, validation and application of the discriminant rule.

Out of the 56 wells available for this study, 14 were selected. They are all vertical, containing at least five well logs (GR, ILD, DT, RHOB and NPHI) and only three wells do not have well core description. Out of these 14 wells, six were selected for the calibration, five for the validation and three – without core description – for the application stage (Table 1).

Table 1. Namorado Field wells used in the logfacies modeling.

Well	Stage
NA-01A	Calibration of the discriminant rule
NA-04	Calibration of the discriminant rule
NA-07	Calibration of the discriminant rule
NA-12	Calibration of the discriminant rule
RJS-19	Calibration of the discriminant rule
RJS-234	Calibration of the discriminant rule
NA-02	Validation of the discriminant rule
NA-05	Validation of the discriminant rule
NA-11A	Validation of the discriminant rule
NA-21B	Validation of the discriminant rule
RJS-42	Validation of the discriminant rule
NA-13A	Application of the discriminant rule
NA-17A	Application of the discriminant rule
RJS-214	Application of the discriminant rule

According to Soares (2005), the resistivity log changes its recorded values quickly, which generally demands a logarithm scale for its presentation. In such cases, the author recommends applying some variable transformation to linearize the scale. In this paper, the resistivity logs were transformed according to Equation 2, where Ω is the original value of the resistivity log and Ω' is the modified value of this log.

$$\Omega' = \log_{10}(\Omega) \tag{2}$$

To minimize data uncertanties, we opted to standardize the well logs that would be used, because the calibration of the data acquisition tools could not be the same, as well as the fluids also could be distinct. Both aspects interfere in the geophysical survey.

Standard scores also allows that different features to be correlated. To standardize the well logs, we used Equation 3, where Z_i' is the standard score, x_i is the *I-th* well log sample value, $\bar{x}$ is the log curve mean and s is the log curve standard deviation.

$$Z_i' = \frac{x_i - \bar{x}}{s} \tag{3}$$

To select well logs among STD GR (Standardized Gamma Ray), STD LOG ILD (Standardized Resistivity Logarithm), STD NPHI (Standardized Neutron), STD RHOB (Standardized Density) and STD DT (Standardized Sonic), those that would be used in the logfacies modeling, we used the Stepwise Discriminant Analysis method, treating well logs as analysis variables and core description as the classification variable.

To assess the discriminant power of a well log, its R^2 and Wilks' Lambda were calculated. As result, we have the following discriminant power: STD RHOB > STD GR > STD NPHI > STD LOG ILD > STD DT (Table 2).

Table 2. Stepwise Discriminant Analysis Method Report.

Step	Entered	Partial R^2	Wilks' Lambda
1	STD RHOB	0.5931	0.4069
2	STD GR	0.4335	0.2305
3	STD NPHI	0.3616	0.1471
4	STD LOG ILD	0.1859	0.1198
5	STD DT	0.1409	0.0647

Based on Table 2, we noticed that the sonic log does not have good discriminant power. Therefore it was excluded. This operation allowed the inclusion of another well (NA-22). And, since it has its core description, it was included in the validation step. According to Bucheb (1991), the selected logs – ILD, NPHI, GR and RHOB – have good discriminant ability, as shown in Table 3.

Table 3. Logs' discriminant ability (Modified from Bucheb, 1991).

Log	Composition	Texture	Sedimentary Structures	Fluid Presence
ILD	Average	Strong	Average	Strong
GR	Average	Weak	Weak	Weak
NPHI	Average	Average	Weak	Average
RHOB	Strong	Average	Weak	Average

Since we know what well logs we would use and owing to the facies great diversity, we decided to adopt fewer clusters to facilitate their identification in the wells. We used cluster analysis, average linkage method, to base our decision. As result, we obtained the graphics below (Figure 1), showing that the ideal quantity of clusters would be four or five. To know which option would be the best, we performed tests with both and concluded that four would be better. The tree chart shows how a facies is classified within the logfacies.

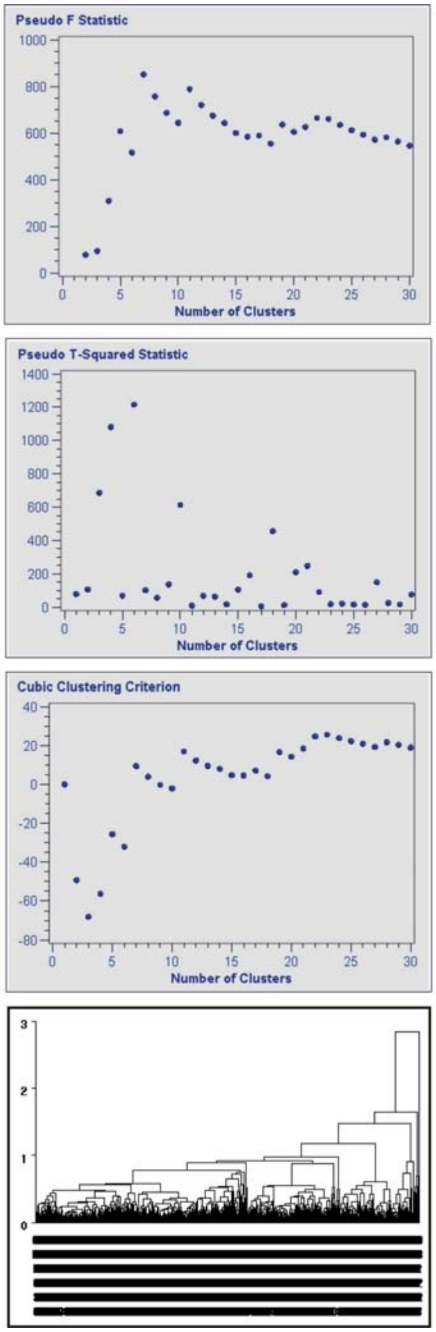

Figure 1. Average Linkage Method Answer.

Since we know the ideal number of clusters, the next stage is applying it to the K-means Algorithm to generate the composed facies. Table 4 and 5 show the results and the classification of the clusters correspondent to the clusters' means interpretation.

Table 4. Classification of the clusters.

Cluster	STD GR	STD NPHI	STD RHOB	STD LOG ILD
1	- 0.9571	- 0.9353	0.9261	0.2091
2	0.8344	0.9368	- 0.8448	- 0.0325
3	0.6518	0.1524	0.2416	- 0.5220
4	0.3280	0.7338	-1.6132	2.1252

Table 5. Standard deviation of the classification of the clusters.

Cluster	STD GR Standard deviation	STD NPHI Standard deviation	STD RHOB Standard deviation	STD LOG ILD Standard deviation
1	0.7424	0.4787	0.4382	0.8211
2	0.7256	0.3911	0.5886	0.5374
3	0.5801	0.5617	0.4260	0.5004
4	0.2348	0.5252	0.4148	1.1843

Based on Table 3 and 4, we concluded that **Cluster 1** is corresponds to a **Third Grade Reservoir**; **Cluster 2** to a **Non-Reservoir**; **Cluster 3** to a **Second Grade Reservoir**; and **Cluster 4** to a **First Grade Reservoir**.

To generate a discriminant rule to be applied in the model, a linear or a quadratic function can be used. To check which is the best option, we used the Covariance Matrix Equality Test, which indicated that a quadratic function would be better than a linear one. Table 6 displays the adjustment probability of the quadratic function, the error valuation and the *a priori* probability for each cluster.

Table 6. Adjustment probability, error valuation and *a priori* probability of the Quadratic Discriminant Rule.

Cluster	1 (%)	2 (%)	3 (%)	4 (%)	Total	Error Valuation
1	100	0.00	0.00	0.00	100%	0.00%
2	0.00	86.01	4.45	9.54	100%	13.99%
3	3.68	3.90	91.53	0.89	100%	8.47%
4	0.00	1.02	0.00	98.98	100%	1.02%
Total	22.60	31.11	30.50	15.80	100%	0.59%
Priority	25	25	25	25		

The validation stage consists of the application of the rule generated in the intervals of the wells that have the four well logs used to this study – ILD, NPHI, GR and RHOB – and the their core descriptions. Thus, the rule is applied to wells and the results are compared to their core description, as, for example, showed in RJS-42 (Figure 2A). If the results are satisfactory, the rule is validated. And, then, the rule is applied in wells that do not have core description as, for example, showed in NA-17A (Figure 2B).

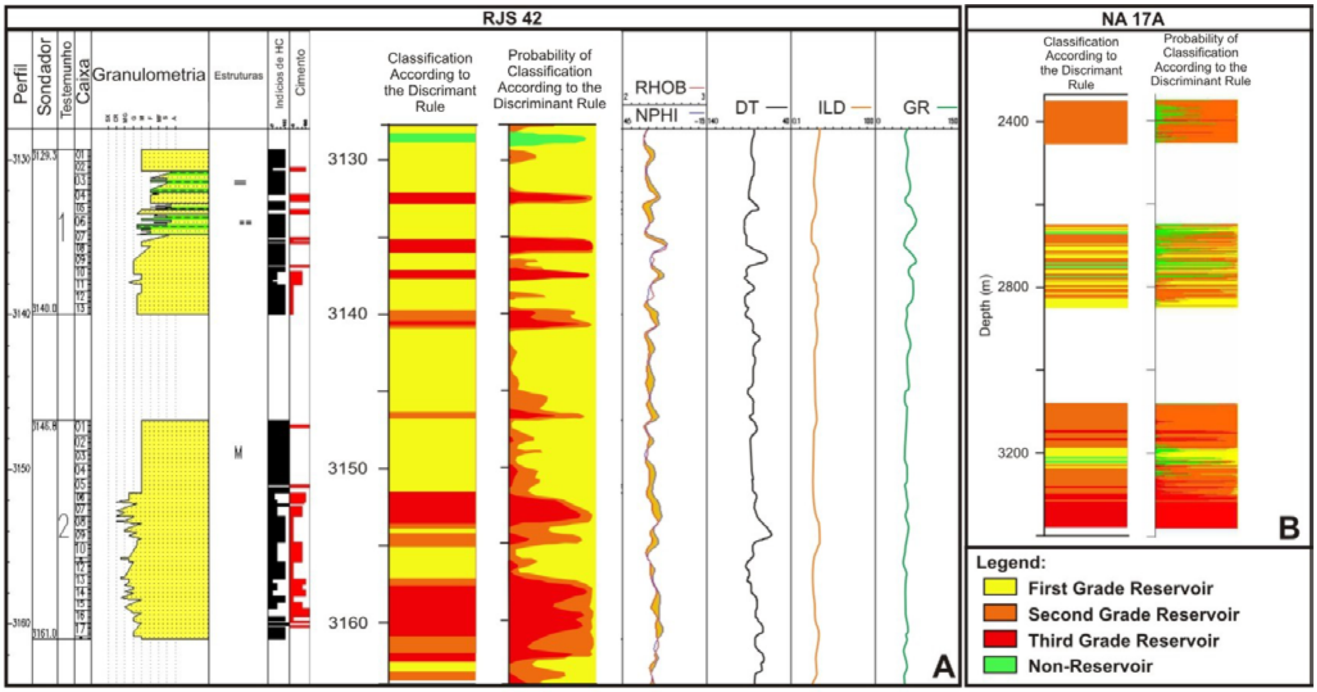

Figure 2. Example of the validation and application stages. (A) shows the RJS-42 well, where was applied the validation of the discriminant rule. (B) shows the NA-17A well, used in the application stage.

6. Conclusions

The results obtained with this study show that is possible to reach an efficient workflow for logfacies modeling using geophysical well data. The recognized facies, according to the adapted procedure, correspond to a few number of compound facies, which generally differs from the great quantity of lithofacies derived from the core descriptions. The limited number of logfacies is adequate for reservoir simulation. The application of the logfacies modeling workflow in the Namorado Field, Campos Basin, presented satisfactory results, which will be applied to the construction of a three-dimensional model of that field.

7. Acknowledgments

We would like to thank the Human Resources Program in Petroleum Geology of the Federal University of Rio de Janeiro (UFRJ – Conv. PRH-ANP/MCT No. 18) for all the support given to this project and to the Technical Data of the Brazilian National Petroleum Agency (SDT-ANP) for releasing the data used in this study.

8. References

BUCHEB, J. A. Aplicação de Tratamento Estatístico Multivariante em Dados de Perfis de Poços da Bacia de Sergipe-Alagoas. *Federal University of Pará* (UFPA). 138p. *Master's Thesis*. 1991.

BUCHEB, J. A.; EVANS, H. B. Aplicação da Análise de Componentes Principais em Dados de Perfis. *Boletim de Geociências da Petrobras*, v. 6, n. 1/2, p. 5-16, 1992.

MOURA, C. A. V. Aplicação de Tratamento Estatístico Multivariante em Dados Geoquímicos de Solo no Mapeamento Geológico na Província de Carajás (Alvo 2 – Corpo 4). *Revista Brasileira de Geociências*, v. 15, n. 3, p. 241-248, 1985.

SILVA, Z. C. G.; SILVA, M. A. G. Subsídio Estatístico à Interpretação de Dados do Complexo Gabro-Anortosítico de Angola. *Revista Brasileira de Geociências*, v. 20, n. 1-4, p. 122-132, 1990.

SOARES, J. A. Um Fluxo de Trabalho para Modelagem de Eletrofácies com Entrelaçamento de Técnicas de Classificação Supervisionada e Não-Supervisionada. *Anais do IX Congresso Internacional da Sociedade Brasileira de Geofísica*. Salvador, BA. 2005.

SOUZA JR., O. G. Análise de Dados Multivariados: Uma Eficiente Ferramenta para Descrição e Caracterização de Reservatórios. *Revista Brasileira de Geociências*, v. 6, n. 3/4, p. 149-154, 1992.