Lithology Mapping by Affinity Propagation Algorithm

Nayara Safira Caldas (IEG, UFOPA), Carlos Guerra (IEG, UFOPA), Cleudilene Rufino (IEG, UFOPA)*

Abstract

This work presents a methodology for solving the problem of the identification of lithologies at depth, directly from wireline logs. For this purpose, the *M-N* cross-plot is used as the formation evaluation technique for identifying the lithologies in the logged borehole, in terms of the physical properties of the main mineral. However, the visual interpretation of this graphic is limited by the large spread of points in the graph. The proposed method is algorithm for a computational interpretation of the *M-N* data by means of the Affinity Propagation (AP) clustering technic. A minimum distance criterion, with respect to *M-N* fixed points, is applied to the set of exemplary points, acquired from AP, in order to associate the lithologies. Finally, the methodology was applied on a real *M-N* data, proving its effectiveness, both in reducing the number of groupings obtained by the clustering algorithm as well as in the lithological identification over highly-spread *M-N* data set.

Introduction

One of the geophysical problems is the extraction of petrophysical and geological information of rocks in the subsurface. Such information should be obtained accurately with analysis of well core, taken along the entire depth of the well. Three factors hinder the daily performance of this procedure. The first refers to the technical infeasibility of achieving the coring in horizontal wells. The second factor is the high cost of coring and the last treats about the recovery factor , which is the inability to recover 100% of rock retained in the sampler at a coring operation . In the absence of such information a lot of techniques were developed for the identification of lithologies [1-3]. A quantitative interpretation of these well logs is the interpretation of the M-N Lithology Plot that, in principle, would provide a quick visual interpretation of lithology of any point along the depth of a well. However, the presence of noise in the porosity well log and simplified rock model which disregards the shaliness, makes the visual interpretation of M-N Plot complex and ambiguous.

In this paper, we present a method that aims to alleviate the limitations of interpretation of M-N Plot, due to the rock model adopted (free shaliness) and by the large spread of points in the graph. For this, is used the shaliness values (Vsh) to solve problem in ambiguous classification and a optimization in the preference vector set as a input to the clustering algorithm Affinity Propagation which will process the data. The result is the simplification of data to be interpreted by setting examples for the resulting groups and, subsequently, the classification of lithology at depth.

Theory

The M-N Plot

The M-N Plot [1] is a particular combination of the three porosity logs (sonic, density and neutron porosity) to produce lithological identification of a well logged section in terms of the physical properties and main mineral of each lithology. The M-N Plot is built with ordered pairs defined by M and N parameters, which are formulated in order to make them relatively independent of the effect of porosity over porosity logs records. The rock model considered for the parameters M and N definition disregards the presence of clay $(V_{sh} = 0)$ in the reservoir rock Constitution. Burke [1] defines the *N* and *M* parameter as:

$$N = \frac{\phi_{Nmf} - \phi_{Nb}}{\rho_b - \rho_{ma}}, \quad (1)$$

$$M = \frac{\Delta t_{mf} - \Delta t_b}{\rho_b - \rho_{ma}} \times 0,01, \quad (2)$$

In **Eq. (2)** a scale matching factor of 0.01 is introduced.

The M-N Plot interpretation is based on fact this plot is built by points formed by ordered pairs calculated for the main constituent minerals of sedimentary rocks. Each mineral produces a single point (**Figure 1**), referred to as fixed points of the plot. It is assumed that the fixed points values of parameters M and N are different enough such that each fixed point occupies a position relatively far from each other, in order to allow the classification of lithology in function of the physical properties of its main mineral.

Cluster analysis

Well log cross plots are always used for mapping facies and lithology and frequently exhibit a tendency to form groups [3, 4]. These groups can be identified and labeled by a classification process which uses well log formation evaluation techniques and clustering analysis.

Cluster analysis is a kind of unsupervised process or a technique that divides a set of objects into homogeneous groups [4-6]. This technique can be used to reduce the size of a data set, reducing a wide range of objects to the

information center of a set [7]. These grouping techniques consist of gathering objects with common characteristics in groups (clusters), with features that differ from each other. There are several algorithms that aim to divide a sample data space into groups whose elements have some similar characteristics. Most algorithms use a pre-established criterion of dissimilarity. Usually, the metric used is the Euclidean distance, but other metrics can be used, preferably ones that better reflect some common data feature.

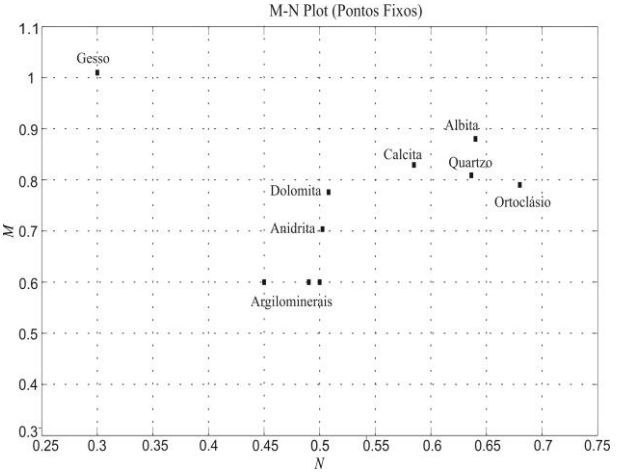


Figure 1. M-N Plot. Black squares represent the most common minerals with their respective M-N values.

Affinity propagation

Most clustering algorithms use as input parameter, a pre-determined number k of clusters to partition the sample data space. The Affinity Propagation (AP) [8, 9] adopts the principle that all data points can be elected as a "cluster exemplar". The data set forms a network configuration where the points represent the nodes and message transmissions occur between the edges of the network [8]. The goal is to optimize the convergence of a similarity function for the election of a set of exemplars that will be considered as representatives of each group to which they are linked.

AP has basically one input, the values of similarities $s(i,k)$ between each point's pair $\{x_i, x_k\}$, where these similarities indicate how the point x_k would fit as an exemplar for the point x_i. When the goal is to minimize the square of the error, every similarity is defined as the negative of the square of the error (Euclidean norm):

$$s(i,k) = -\|x_i - x_k\|^2 . \quad (3)$$

In addition to the similarities, AP also has as input parameter, values called "preferences" $s(k,k)$ for each point k. By **Eq. (3)** the main diagonal of the similarity matrix is null, which indicates that the greatest similarity occurs from a point to itself. Substituting this value for a $s(k,k) \neq 0$ will inform the preferences for these points to be exemplars. These preference values interfere on the results (number of clusters) found by the AP, high and low $s(k,k)$ values result in large and low number of clusters found, respectively. The messages exchanging between points can be of two types, transmission of responsibilities and availability. The responsibilities $r(i,k)$ are messages sent from one point i for a possible exemplar k which indicate how suitable would be the point k to be exemplar for the point i. The availability $a(i,k)$, sent from an exemplar candidate k for a point i, indicate evidence of how appropriate it would be to the point i to choose the point k as an exemplar. The responsibilities are defined by the following rule:

$$r(i,k) \leftarrow s(i,k) - \max_{k' \neq k} \{a(i,k') + s(i,k')\}. \quad (4)$$

The availabilities define if an exemplar is a good one and are defined such as:

$$\begin{cases} a(i,k) \leftarrow \min\left\{0, r(k,k) + \sum_{i' \notin \{i,k\}} \max\{0, r(i',k)\}\right\}, \\ a(k,k) \leftarrow \sum_{i' \neq k} \max\{0, r(i',k)\}. \end{cases} \quad (5)$$

Thus, the definition of an exemplar occurs when the combination of responsibilities and availabilities, according to **Eq. (6)**, has maximum value:

$$AP = \max\{a(i,k) + r(i,k)\}. \quad (6)$$

Methodology

In the formation evaluation, due to the rarefaction of wells core, the geophysical well log codified in M-N plot is widely used for the identification of lithology in depth [10-12].

Due to the simplifications adopted in the rock model of the M-N Plot, such as free shaliness and the occurrence of a single mineral in the matrix composition, which in most practical cases do not represent the real rocks, it becomes the lithology identification ambiguous. Thus, the points defined by ordered pairs (N, M) calculated for an interval of well log data have a large scatter, which distances data points from the reference points(fixed points) and makes difficult the visual interpretation of the M-N plot. One way to attenuate the distortions is compute the shaliness (Vsh) and to perform the lithological identification in terms of the physical properties of the main mineral and clay in the constitution of the reservoir rock.

The problem of computational interpretation of the M-N Plot can be established in the form of a clustering problem in order to determine a point on the graph that is

representative of a group or set of points with similar characteristics. This approach seeks to reduce the number of points to be classified as a function of the fixed points in M-N plot. Different from a general clustering problem, the computational interpretation of the M-N plot requires the establishment of the geological coherence and the depth continuity of the same layer.

In practice the interpretation of M-N plot must be guided by local geological information in order to eliminate reference points that represent absent minerals or appear only as traces in the constitution of the present rocks.

The methodology consists of the following steps. In the first step, data containing information relating to the porosity well logs (density, sonic and neutron), Gamma Ray and the depth associated to each measure are selected. The second step is to calculate the vectors of lithology parameters ***N*** and ***M*** from ***rhob, dt*** and ***phin*** vectors who represent the log values over the zone of study. **Eq. (1)** and **(2)** are used for calculating N_i and M_i in function of each depth z_i.

In the next step, the preference vector ***P*** is defined as follows: The fixed points are included in the set of data points, to be processed by the AP algorithm, and receive slightly a higher preference value (i vector position) and, for the others points, the preference values will remain the average value of the similarity matrix. Based on the similarity matrix and vector preference, the AP makes the dataset (*N, M*) division in clusters and saves its respective exemplars. The classification follows by the association of exemplars with fixed minerals (**Figure 1**) that make up the main lithologies in sedimentary rocks. The association occurs by the smallest Euclidean distance between the exemplars defined by the AP and Minerals fixed points. The set of points in a cluster will be classified as belonging to the lithology with which its exemplar was associated. Finally, the classified lithologies are shown in depth, allowing a posterior comparison with the results already obtained from core samples.

Results

Firstly, the methodology was applied on synthetic data. Afterwards, it was applied on real data to infer different lithology's in depth, even in the absence core analysis. The real data selected are referred to two well logs from Namorado Field, in Campos Basin (Brazil) [13, 14], whose cores are available.

Real Data

Figure 2 shows a real well log data recorded in the Campos Basin. The core samples are described in **Table 1** for the depth intervals selected (highlighted in green/gray colors on the log). **Figure 3** shows the *M-N* plot calculated from logs. The circles in black represents (*N*, *M*) points and the black squares the main minerals fixed points. The visual interpretation of *M-N* Plot suggests the existence of three compact and well defined clusters. However, the visual interpretation does not necessarily reflect the number of different lithologies in situ. A common problem in formation evaluation is the displacement of (*N, M*) points, due to shaliness toward clay mineral and noises, preventing a more reliable classification.

Table 1
Geological Description

Layer	Lithology	Matrix Composition
Green	Sandstone	Quartz
Gray	Radioactive Shale	Clay

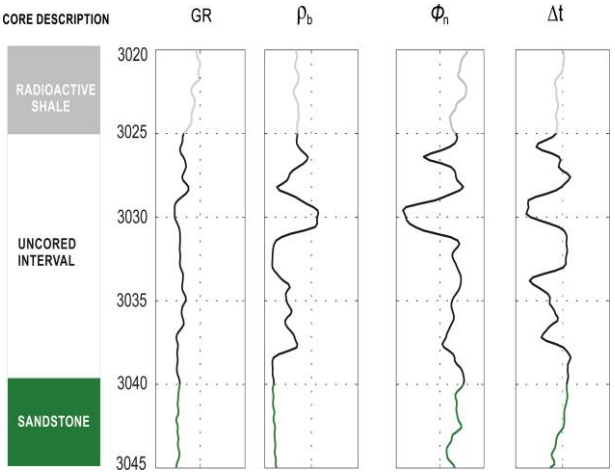


Figure 2. Core model, the Gamma ray and Porosities logs. Chosen intervals are highlighted in gray and green colors.

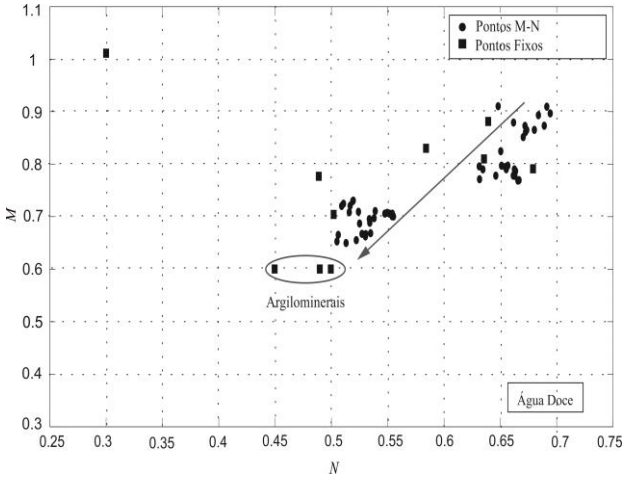


Figure 3.M-N Plot with three groups of points suggested by visual interpretation.

In **Figure 4** the AP algorithm results can be seen over *M-N* Plot data. It is remarkable the presence of five clusters resulting from the clustering when it uses only the AP algorithm without changing of preference vetor. The

visual interpretation suggests two reservoir layers (light gray / violet) and the ambiguity referring to the three exemplars of the sealing rocks which could be interpreted as Dolomite / Anidrite in the M-N plot analysis.This problem was quickly solved by computing the clay content (Vsh) resulting in high values not common in this kind of rocks.

By applying the AP algorithm with the modifield preferences vector (above described), the number of clusters is reduced to three (**Figure 5**).

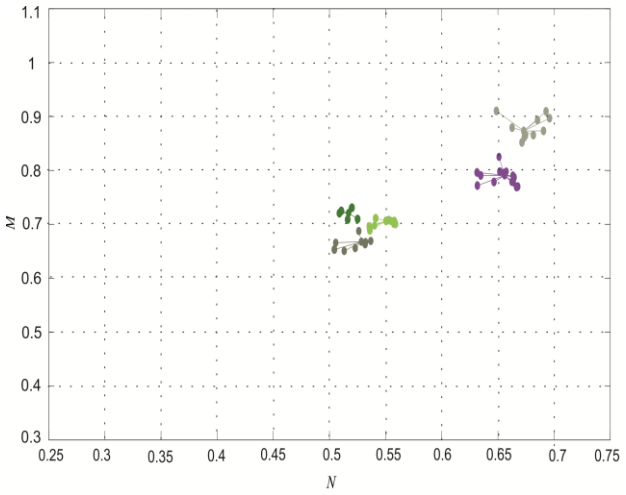


Figure 4. AP resulting clusters with usual preference vector

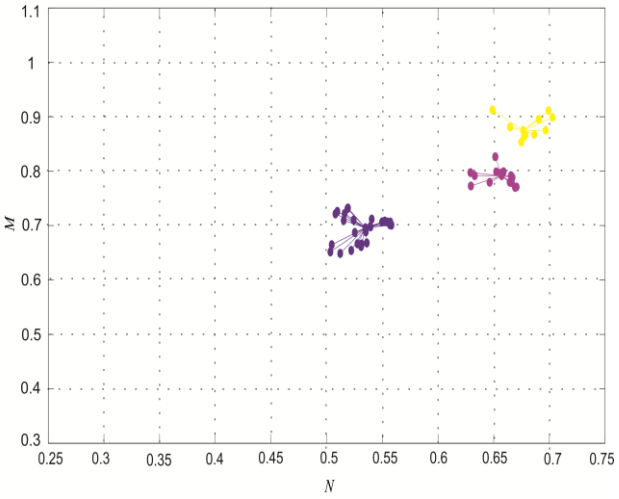
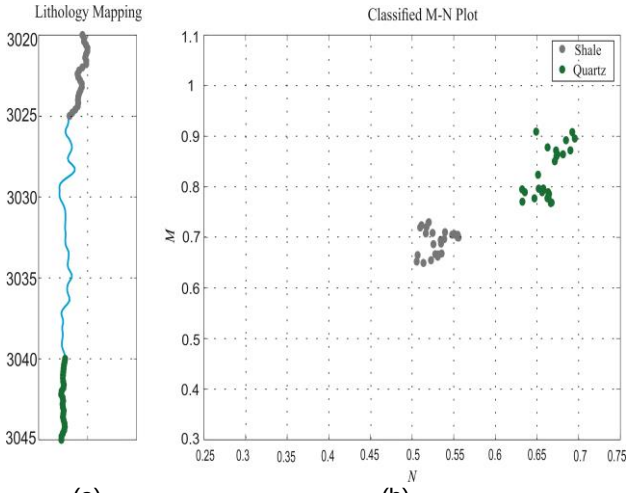


Figure 5. AP clusters resulting from the change in preferences vector.

The final classification is shown in **Figure 6**. The first break in depth is associated with mineral components of shale(clay), the second interval is associated with the fixed point of mineral representative of quartz(sandstone).



(a) (b)

Figure 6. (a) Lithology mapping in depth. (b) Resulting classification by exemplar association.

Conclusions

The identification of lithology, in the strict sense of well logging geophysics, is a very complex problem to be solved only with the analysis of logs, since it is not completely valid the assumption correspondence that variation in lithology implies variation of physical properties. A quantitative interpretation of these logs is taken from the *M-N* Lithology Plot that, in principle, would provide a quick visual interpretation of lithology of any point along the depth of a well. However, the presence of noise in the porosity well log and simplified rock model which disregards the shaliness, makes the visual interpretation of *M-N* Plot complex and ambiguous. A solution for this kind of classification problem was presented here like an unsupervised method that, regardless of the availability of well core, is able to perform the identification and distinction of lithology of rocks.

The representation of a cluster by its exemplar produced by the Affinity Propagation algorithm, in the interpretation of *M-N* Plot, results in a significant reduction of the data set content to be processed, and mitigate ambiguity regarding shaliness present in *M-N* Plot or scattering due to noise in the well log data.

The benefit obtained by using AP is the fact that it is not necessary to establish the number of clusters, as an input parameter. Another advantage of using Affinity Propagation algorithm, compared with other classification algorithm, is the substitution of centroid by exemplary, where centroid points are only representatives of a group in the graphic, but are not necessarily points of the processed data. The exemplars, however, are representatives of groups and still relevant data.

References

[1] J.A.C.J. Burke, R.L.; Schimidt, A.W., The Litho-porosity cross plot: a method of determining rock characteristics for computation of log data., in: Society of Petroleum Engineers (Ed.) SPE ILLINOIS BASIN REGIONAL MEETING, Society of Petroleum Engineers, Illinois Basin.Evansville, Indiana:, 1969, pp. p. 187- 198.

[2] C. Clavier, D. Rust, MID-plot: A new lithology technique, The log analyst, 17 (1976) 16-24.

[3] G.B. Asquith, C.R. Gibson, Basic well log analysis for geologists, American Association of Petroleum Geologists, 1982.

[4] G. Gan, C. Ma, J. Wu, Data clustering: theory, algorithms, and applications, Society for Industrial and Applied Mathematics, 2007.

[5] M. Steinbach, M., Karypis, G., & Kumar, V. 2000. A comparison of document clustering techniques.

[6] P. Rai, S. Singh, A survey of clustering techniques, International Journal of Computer Applications, 7 (2010) 1-5.

[7] H. Spath, Cluster analysis algorithms for data reduction and classification of objects, Ellis Horwood, Ltd., 1980.

[8] B.J. Frey, D. Dueck, Clustering by passing messages between data points, Science, 315 (2007) 972-976.

[9] B.a.D. Frey, D., Affnity Propagation | Frey Lab, in, 2007.

[10] W.C. Savre, J. Burke, Determination of true porosity and mineral composition in complex lithologies with the use of Sonic, Neutron and Density Survey, The Log Analyst, 4 (1963) 439-474.

[11] W. Savre, Determination of a more accurate porosity and mineral composition in complex lithologies with the use of the sonic, neutron and density surveys, Journal of Petroleum Technology, 15 (1963) 945-959.

[12] P. Delfiner, O. Peyret, O. Serra, Automatic determination of lithology from well logs, SPE Formation Evaluation, 2 (1987) 303-310.

[13] G. Bacoccoli, R.G. Morales, O.A. Campos, The Namorado oil field: a major oil discovery in the Campos basin, Brazil, AAPG Bulletin, 63 (1980) 329-338.

[14] H. Rangel, F. Martins, F. Esteves, F. Feijó, Bacia de campos, Boletim de Geociências da PETROBRAS, 8 (1994) 203-217.