

# Well log analysis for lithology and fluid contacts in Rovuma Basin – Mozambique: Application of cluster and discriminant analyses

Medja, Jone Lucas<sup>1</sup> & Bassrei, Amin<sup>2</sup>. <sup>1</sup> Universidade Eduardo Mondlane (UEM), <sup>2</sup> Universidade Federal da Bahia (UFBA)

Copyright 2021, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 17<sup>th</sup> International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 August 2021.

Contents of this paper were reviewed by the Technical Committee of the 17<sup>th</sup> International Congress of the Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

## Abstract

This study applies the cluster and discriminant analyses in geophysical well log data from the Rovuma Basin - Mozambique. The main objective was to determine the lithological profile and fluid contacts in reservoirs. Five wells drilled on the same basin were evaluated. For the discrimination, a reference well was chosen for training, and the obtained functions from it were then applied to the remaining wells. The classification process comprehended three main phases. The two methods were applied in parallel and the results are compared in each classification phase. Reservoirs quality was also assessed by applying cutoffs in relation to shale content and effective porosity to delineate net reservoirs. Overall, both methods converged to the same lithological model and fluidtypes in reservoirs. Gas has been indicated as the most predominant hydrocarbon in the basin.

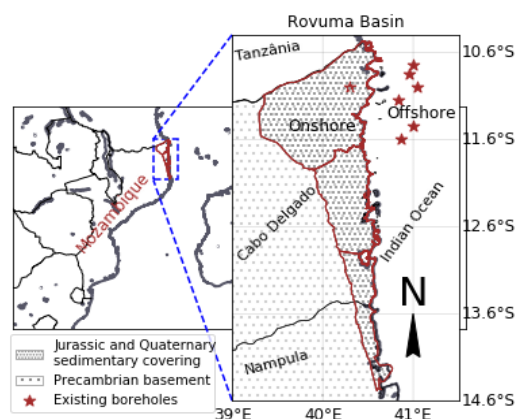
## Introduction

In general, well log analysis comprises a set of methods that aim to extract information from the logs, in order to derive a geological model. The cluster and discriminant analyses have shown good applicability in the processing and interpretation of well log data. In particular, many researchers have applied these techniques to identify and separate different litho-types and fluids in formations (Flexa et al., 2004; Rosa et al., 2008).

Cluster analysis is an unsupervised classification process, what makes the method quite useful, especially in the exploratory phase when there are no prior hypotheses about the classifying model in the field. On the other hand, discriminant analysis is a supervised process and it may be applied for many purposes, from the cases in which it is intended to perform the statistical separation between groups of data, in cases when it is intended to predict the occurrence of a certain event and even when it is intended to explain the relationship between variables.

In this work, we evaluate well log data from five boreholes over the Rovuma Basin - Mozambique. We have applied the cluster and discriminant analyses in parallel, to determine the lithological profile and fluid contacts in reservoirs. In addition, cutoffs in relation to shale content and effective porosity have been applied to gross reservoirs in order to delineate net reservoirs.

The Rovuma Basin is one of the most important sedimentary Basins in Mozambique, both in terms of the volume of accumulated sediments and in terms of the occurrence of hydrocarbons. It is located in the north-east of Mozambique. The onshore area is about 17,000 km<sup>2</sup> comprising the entire eastern part of Cabo Delgado and partially the eastern part of Nampula provinces. The offshore area is about 12,500 km<sup>2</sup>. Figure 1 shows the onshore and offshore parts of the Rovuma Basin. Several wells drilled in the region have confirmed the occurrence of hydrocarbons both onshore and offshore. Currently, only natural gas is being explored in the Basin. However, future oil exploration has been confirmed. Sediments of Rovuma Basin were deposited between the Jurassic and Quaternary periods (Hancox et al., 2002). The maximum thickness of sediments in the entire Basin is about 10 km based on geophysical surveys (Key et al., 2008).



**Figure 1** - Study area, the Rovuma Basin in Northern Mozambique.

## Data and Methods

### Well log Data

Five wells drilled over the Rovuma Basin were evaluated, namely Well-1 (2243 - 3243 m), Well-2 (2455 - 4393 m), Well-3 (2454 - 3414 m), Well-4 (596 - 3104 m) and Well-5 (1052 - 3842 m). All well log data have a sampling interval of 0.1524 meters. Table 1 shows the list of logs used for this study.

**Table 1** - Description of the well log data.

MNEMONIC	CURVE TYPE	UNIT
HCAL	Caliper	in
GR	Natural Gamma Ray	gAPI
RHOZ	Bulk Density	g/cm <sup>3</sup>
TNPH	Thermal Neutron Porosity	V / V
DTCO	Sonic (Transit Time)	μ sec
AO10	Shallow Resistivity	Ω.m
RT	Deep Resistivity	Ω.m

### Cluster analysis (CA)

Cluster analysis is a process that basically consists of arranging a set of data or objects with common characteristics into groups. Objects belonging to the same group have more similarities to each other which differentiate them from those belonging to other groups.

In real life, data sets are so complex that different groups may not be clearly separated. The clustering algorithms face the same challenge, so we need to define some criteria of dissimilarity, which in some cases can be a challenging task. Here we make use of the centroid-based method, known as K-Means Clustering. The K-Means clustering technique partitions data into  $K$  groups or categories so that the data points in the same group are closer and the data points in different groups are more distant. The algorithm tries to minimize the distances within group and maximize the distance between groups (Everitt et al., 2011). There are many methods for measuring distance. Usually, the metric adopted for many situations is the Euclidean Distance.

To describe the K-Means algorithm, we will consider  $\mathbf{X}$  as a  $N \times M$  matrix which columns are the log vectors  $\mathbf{x}_j$ , so that  $x_{ij}$  will be the value of a log  $j$  recorded at a depth  $i$ . Each row of  $\mathbf{X}$  is a vector  $\mathbf{p}_i$  defining a point in the  $M$ -dimensional space. In this case, the Euclidean Distance ( $ED$ ) between two arbitrary points  $\mathbf{p}_1$  and  $\mathbf{p}_2$  in the data space, is expressed as follow:

$$ED = \sqrt{(\mathbf{p}_1 - \mathbf{p}_2)(\mathbf{p}_1 - \mathbf{p}_2)^T}. \quad (1)$$

The log vectors  $\mathbf{x}_j$  are standardized by subtracting from each value the log mean and then divide the differences by the log standard deviation, according to Davis (2002).

K-Means is an iterative process. The algorithm performs the following steps: (1) Initially, it randomly selects  $K$  points in the data space as centroids of the  $K$  groups to be separated; (2) Calculates the distances between each data point and the centroids, associating them to the nearest centroid; (3) Finds new centroids by averaging the data points of the same group; (4) Repeats the steps 2 and 3 until the convergence. In the convergence all group centroids become unalterable.

### Discriminant analysis (DA)

Discriminant analysis is one of the most widely used multivariate procedures in Earth science. Unlike cluster analysis, in discriminant analysis sample data from the groups to be classified must be defined prior. This information is used to generate the discriminant function. Depending on the number of groups to be classified or the problem complexity, the discriminant function can be linear or quadratic. In this work, we make use of the linear discriminant analysis, which is more convenient when only two groups are defined in each discrimination phase.

Considering the case of two datasets (A and B) in a multivariate space, the method seeks an orientation in which the datasets present the maximum separation and

simultaneously a minimal variance. The linear discriminant function is mathematically expressed as:

$$Z_i = \lambda_1 x_{i1} + \lambda_2 x_{i2} + \dots + \lambda_M x_{iM}, \quad (2)$$

where  $Z_i$  is the discriminant index at a certain depth  $i$ ,  $\lambda_j$  ( $j = 1, 2, \dots, M$ ) are the coefficients of the discriminant function and  $x_{ij}$  represents the independent variables.

The discriminant function coefficients are determined using multiple regression according to Davis (2002), and it is expressed by the following matrix equation:

$$\mathbf{S} \boldsymbol{\lambda} = \mathbf{d}, \quad (3)$$

where  $\mathbf{S}$  is a  $M \times M$  matrix of pooled variances and covariances of  $M$  variables,  $\boldsymbol{\lambda}$  is a column vector of the discriminant function coefficients and  $\mathbf{d}$  is a column vector formed by the differences between the multivariate means of the two groups. The equation (3) can be solved using the basic concepts of inverse problems. Since  $\mathbf{S}$  is a square matrix and ensuring that it is non-singular, then the inverse  $\mathbf{S}^{-1}$  exists, and we have  $\boldsymbol{\lambda} = \mathbf{S}^{-1} \mathbf{d}$ .

The discriminant function generates a single number for each observation point, called discriminant index or score.

The substitution of the multivariate averages  $\bar{\mathbf{x}}^A$  and  $\bar{\mathbf{x}}^B$  in the discriminant function generates the groups centroids  $Z_A$  and  $Z_B$  respectively. The difference

$Z_A - Z_B$  is known as the Mahalanobis distance ( $D^2$ ) (Davis, 2002). It helps to calculate the relative contribution ( $e_j$ ) of each variable for the discrimination:

$$e_j = \frac{\lambda_j d_j}{D^2} \times 100\%. \quad (4)$$

The separation of the two groups is defined by a cutting score  $Z_C$ . The optimal cutting score depends on the dimensions of the groups. According to Ramayah et al. (2010), for equal dimensions, it is simply the average of the two group centroids:

$$Z_C = \frac{Z_A + Z_B}{2}, \quad (5)$$

and for different dimensions, it is found by the formula:

$$Z_C = \frac{N_B Z_A + N_A Z_B}{N_A + N_B}. \quad (6)$$

### Lithology and fluid contacts

The lithological profile and fluid contacts were determined using the cluster analysis and the linear discriminant analysis. The process consisted of three main phases: (i) the separation of shale and non-shale (potential reservoirs), (ii) the separation of water and hydrocarbon inside reservoirs and (iii) the separation of oil and gas in the hydrocarbon bearing zones.

As mentioned previously, for the discriminant analysis, it is necessary that group samples are defined prior. These sample data are used to determine the discriminant

function coefficients and the scores. Given that we did not have any information from cores and cuttings which could be helpful for the sampling, well log interpretation techniques were applied prior to select the sample data of the different groups of interest. The Well-3 was chosen as a reference well to obtain the sample data. The coefficients and scores obtained from this well are then applied to the remaining wells. According to Flexa et al. (2004), considering the absence of important diagenetic effects, these discriminant functions can be applied to the neighbouring wells. In this way, the interpretation carried out in the reference well is extended throughout the basin. Figure 4 shows the flowchart of the study methodology.

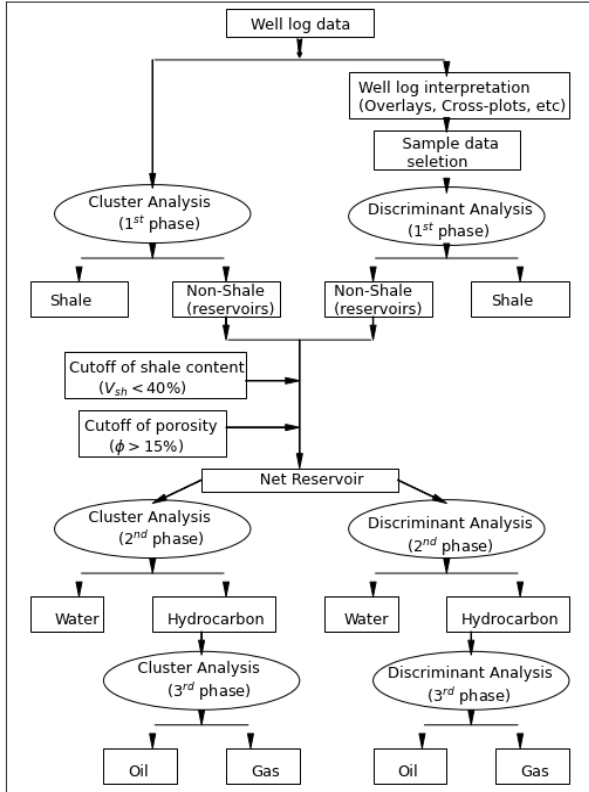


Figure 2 - Flowchart of the study methodology.

**Shale – non shale separation:** This was the first problem to be addressed. Four logs were used in this process: HCAL, GR, TNPH and DTCO. The entire logged section of the well was considered. Thus, the discriminant function was defined by the following expression:

$$Z = \lambda_{HCAL} \times HCAL + \lambda_{GR} \times GR + \lambda_{TNPH} \times TNPH + \lambda_{DTCO} \times DTCO \quad (7)$$

**Net reservoir:** Before proceeding to the next phase, which is the fluid classification, it was necessary to determine the net reservoir. The term net reservoir can be understood as being the fraction of the total thickness of gross reservoirs capable of storing fluids. So, for net reservoir determination, cutoffs in relation to shale content and porosity are applied. Cutoffs can be defined as limiting values imposed to outline the region of interest. There is no a defined criterion to determine the cutoff

values, what implies that reasonable values are arbitrarily chosen depending on the region and purpose (Worthington & Consentino, 2005). Crain (2019) have presented typical intervals for the net pay delineation as follow: maximum shale volume of 0.25 to 0.45, minimum porosity of 0.03 to 0.16, maximum water saturation of 0.30 to 0.70 and minimum permeability of 0.01 to 5.0. In this study we have applied two cutoffs. The first cutoff ( $V_{Sh} \leq 0.4$ ) was applied to the gross reservoir to eliminate parts of the formation with high shale content obtaining the net sand. And the second cutoff ( $\phi \geq 0.15$ ) is applied to the net sand to remove the portions with low porosity (e.g. tight sand), outlining the net reservoir.

**Water – hydrocarbon separation:** So far, shale formations have been discarded, and having the net reservoir, followed the second phase. For this problem, well logs considered sensitive to fluids were used: RT, RHOZ and TNPH. The discriminant function is expressed as:

$$Z = \lambda_{RT} \times RT + \lambda_{RHOZ} \times RHOZ + \lambda_{TNPH} \times TNPH \quad (8)$$

**Oil – gas separation:** This is the third and last phase. Here, all water saturated zones have been excluded, and only hydrocarbon bearing zones remained. Two logs were used in this classification: RHOZ and TNPH. The expression for the discriminant function is:

$$Z = \lambda_{RHOZ} \times RHOZ + \lambda_{TNPH} \times TNPH \quad (9)$$

**Results**

*Analysis in the reference well*

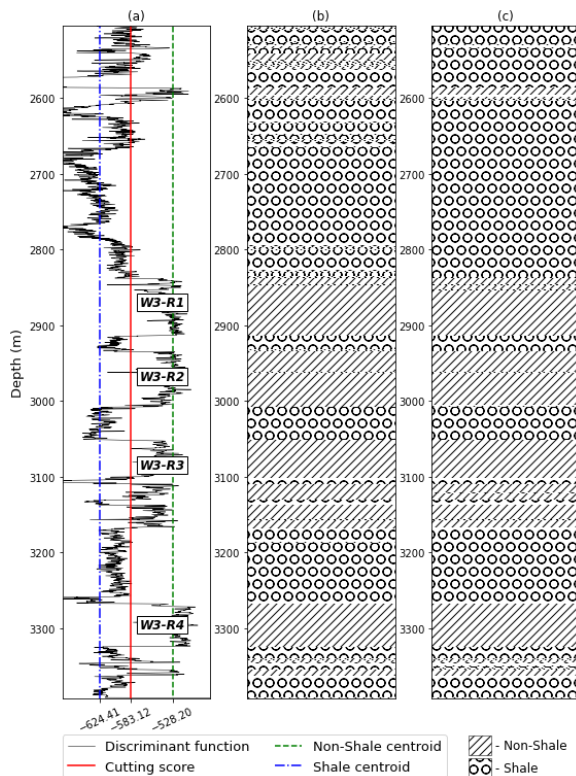
The discriminant functions coefficients and the relative contribution of each log for the three discriminatory problems are shown in Table 2. The coefficient signal indicates if the variable is contributing positively or negatively to the discriminant index. The absolute values are influenced by the variance of the respective discriminant variable. A variable with a greater variance tends to have less weight and thus contributes less to the discrimination.

In the first problem that consisted of shale and non-shale classification, the GR log had the greatest relative contribution (81.24%) as expected and the DTCO had presented the lowest contribution (0.073%) so that it could be discarded to simplify the discriminant function. The separation of groups in the DA is made by the cutting score  $Z_C$ . For each depth, a  $Z_i$  index is calculated, and is then compared to the  $Z_C$ . In this particular case, shale formations are assigned for  $Z_i < Z_C$ , and non-shale or potential reservoirs are assigned for  $Z_i > Z_C$ .

Figure 3 shows the shale and non-shale zones in the Well-3 from both methods, CA and DA. Four potential reservoirs were identified in this well, designated as W3-R1 (2845 - 2905 m), W3-R2 (2935 - 3010 m), W3-R3 (3050 - 3170 m) and W3-R4 (3270 - 3330 m). The total thickness of the four potential reservoirs constitutes the gross reservoir. The cutoffs of shale content and porosity were applied over this region to define the net reservoirs before proceeding to the second phase.

**Table 2** - Discriminant functions coefficients and the relative contribution of each variable in the three classification phases. There are three log combinations for the three different purposes. The coefficients signal indicates if the variable is contributing positively or negatively to the discriminant index.

Variables	Coefficients	Relative contribution (%)
<i>Shale – Non Shale</i>		
HCAL	-34.78	16.14
GR	-2.05	81.24
DTCO	-0.01	0.07
TNPH	-8.96	2.55
<i>Water – Hydrocarbon</i>		
RT	-0.07	15.07
RHOZ	74.23	39.66
TNPH	56.33	45.27
<i>Oil – Gas</i>		
RHOZ	-7.31	0.36
TNPH	232.95	99.64

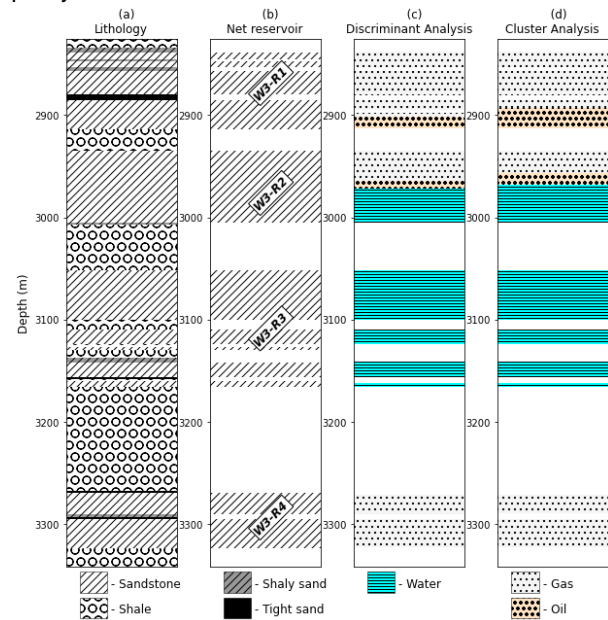


**Figure 3** - Lithological classification in the reference well (Well-3), (a) the discriminant function and scores, (b) the lithological definition by the discriminant analysis and (c) the lithological definition by the cluster analysis.

In the second phase which consisted of water - hydrocarbon classification, a new discriminant function was determined from the second database. The TNPH log presented the most significant contribution (45.27%) followed by the RHOZ (39.66%). In fact, the most significant contribution was expected to be of the RT log, as it is theoretically known that the resistivity discrepancy between water and hydrocarbon is significant. Therefore, two possibilities can be considered to explain this result. One is that the most predominant hydrocarbon is gas,

what maximizes both the TNPH and RHOZ differences between the two groups. And the other is that, the water is more likely to be fresh so that resistivity varies little from water to hydrocarbon bearing zones. Analogously to the previous procedure, an  $Z_i$  index is calculated for each depth. Hydrocarbon saturated zones are assigned for  $Z_i < Z_C$ , and water saturated zones for  $Z_i > Z_C$ .

In the third phase which is the oil - gas separation, again, a new discriminant function was determined. The most significant contribution had been of the TNPH log (99.64%) as expected. The gas saturated zones are assigned for  $Z_i < Z_C$ , and the oil saturated zones for  $Z_i > Z_C$ . Now the water-oil-gas contacts have been defined inside reservoirs. Figure 4 shows sequentially from column (a) to (d), the lithology in column (a), the net reservoirs in column (b), and fluid contacts by the DA in column (c) and by the CA in column (d). It is observed that the two methods have identified the same groups in most depths. The blank spaces presented in the columns (c) and (d) for fluid contacts reflect the cuts made in the process of determining the net reservoir. At those depths, reservoirs are considered not capable of storing fluids, due to both or one of the factors, high clay content or low porosity. Thus, this process was important to assess the quality of reservoirs.



**Figure 4** - Lithology and fluid contacts in the Well-3: (a) lithological profile, (b) net reservoirs, (c) fluid contacts by the discriminant analysis and (d) fluid contacts by the cluster analysis.

The reservoir W3-R1 is mostly filled with gas; however both methods indicated a contact with a small oil layer at the depth of 2901 m by the DA and at the depth of 2894 m by the CA. In the reservoir W3-R2 immediately below, the three fluids were identified by both methods, with water-oil and oil-gas contacts respectively at 2970 and 2960 m by the DA and at 2968 and 2957 m by the CA. In the reservoir W3-R3, only water saturation was detected and in the last reservoir W3-R4 that appears after a

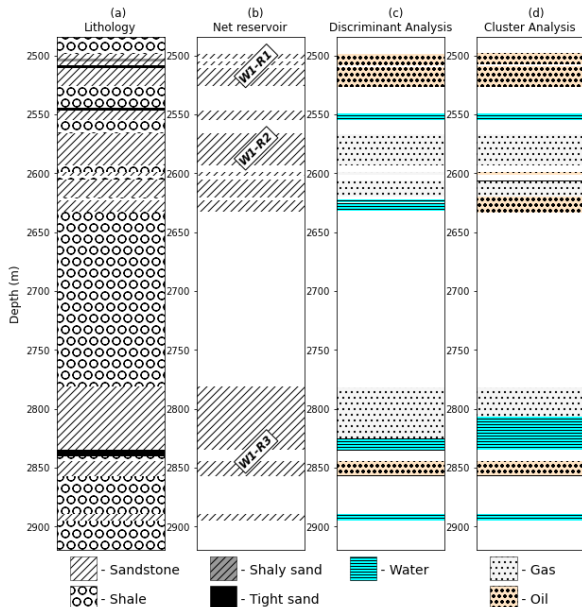


relatively extensive shale layer, only gas saturation was identified.

*Analysis in the remaining wells*

The application of the discriminant functions from the reference well to the remaining four wells extends the interpretation throughout the basin. Besides, it is also an assessment of the performance of these functions for the classification of lithotypes and fluidtypes in the field.

In the Well-1, three potential reservoirs were defined as can be seen in Figure 5. In the first reservoir W1-R1, both DA and CA indicate the presence of oil and in the second W1-R2 gas saturation is identified, with a small layer of water beneath, after a tiny shaly sand at the depth of 2622 m. In the third reservoir W1-R3, gas-water contact is indicated by both methods but at different depths. While the DA indicates at the depth of 2825 m, the CA indicates at 2807 m. Still, in the same reservoir an oil saturated layer was also detected after a small tight sand layer.



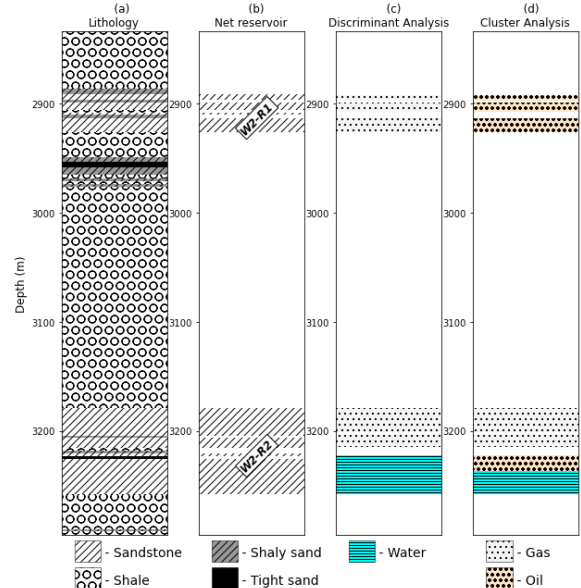
**Figure 5 - Lithology and fluid contacts in the Well-1: (a) lithological profile, (b) net reservoir, (c) fluid contact by the discriminant analysis and (d) fluid contact by the cluster analysis.**

In the Well-2 illustrated in Figure 6, two reservoirs were defined. The first W2-R1 is of very low quality (presence of high clay content and low porosity), so that most of it was discarded after being subjected to the cutoffs of clay content and porosity. The two methods did not converge on identifying fluids into the first reservoir W2-R1, while the DA indicates gas saturation, the CA suggests oil saturation, possibly due to the effect of remnant shale content. In the second reservoir W2-R2, a water-oil contact was identified by the CA at the depth of 3240 m.

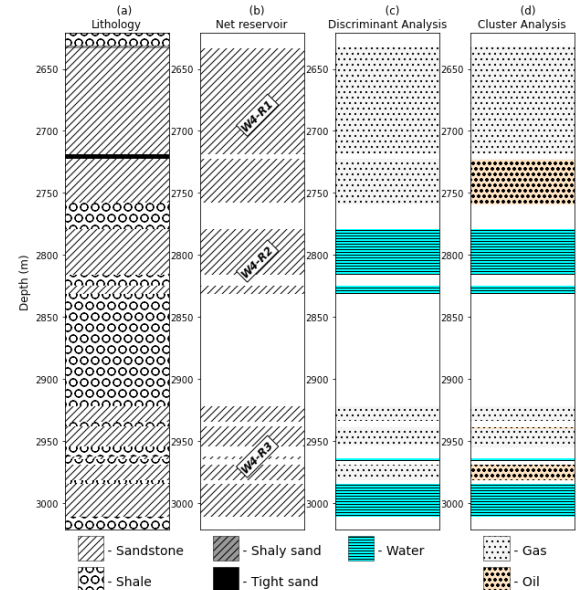
Figure 7 shows the classification for the Well-4 and three reservoirs are defined. While the DA identified only gas saturation in the first reservoir W4-R1, the CA indicates the presence of gas and oil separated with tiny tight sand between 2718 - 2723 m. The second reservoir W4-R2 is

completely saturated with water. And in the third reservoir W4-R3, both methods present gas, oil and water separated by small shaly sand layers.

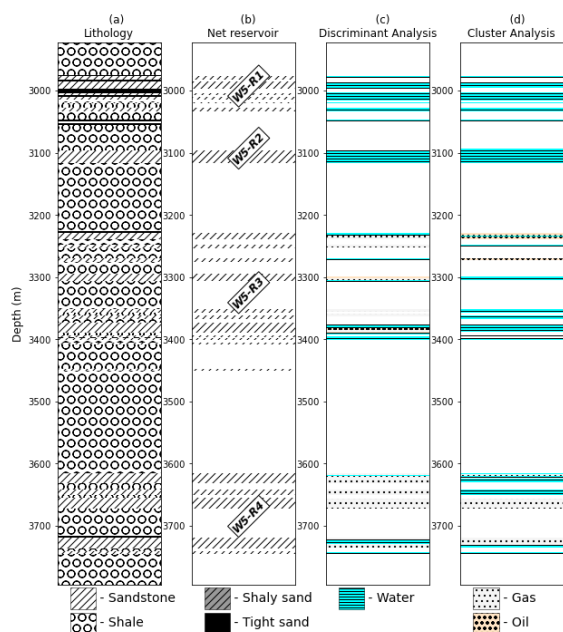
Although the Well-5 has the longest logged section among all wells, it has poor reservoirs as can be observed in Figure 8. Most reservoirs presented water saturation, except the last, W5-R4, where tiny gas saturated layers are evident in both methods. Therefore, this well has no potential for hydrocarbon production.



**Figure 6 - Lithology and fluid contacts in the Well-2: (a) lithological profile, (b) net reservoir, (c) fluid contact by the discriminant analysis and (d) fluid contact by the cluster analysis.**



**Figure 7 - Lithology and fluid contacts in the Well-4: (a) lithological profile, (b) net reservoir, (c) fluid contact by the discriminant analysis and (d) fluid contact by the cluster analysis.**



**Figure 8** - Lithology and fluid contacts in the Well-5: (a) lithological profile, (b) net reservoir, (c) fluid contact by the discriminant analysis and (d) fluid contact by the cluster analysis.

### Conclusions

Although no prior information from cores and cuttings were available in the study region, the application of discriminant and cluster analyses together has led to consistent results. The application of cutoffs for outlining the net reservoir was essential to assess the quality of reservoirs in the basin. Reservoirs with great potential for hydrocarbon exploration were found in Well-1, Well-3 and Well-4. Overall, it is observed that the three fluidtypes (water, oil and gas) are present in the Basin, being gas the most predominant hydrocarbon. The two methods converged, both in the lithology and fluid classification in most cases, although some differences occurred in indicating the fluid contacts, what is a reasonable situation since a contact between fluids in reservoirs is actually a transition zone and not a linear separation as it may seem. The discriminant functions and scores obtained from the reference well showed good performance on their application both for lithology and for fluid classification in the remaining wells of the same stratigraphic unit, what validates their efficiency in the discriminatory process. Besides, this also shows that the set of variables selected for the construction of the discriminant functions in each classification process were sufficiently good. Regarding the cluster analysis, it is important to keep in mind the existence of ambiguities because it is an unsupervised method. The K-Means always divides the input data into the predefined number of possible groups, even if erroneously the input data consists of only one cluster. Thus, the combination of the two methods is important.

### Acknowledgments

This work was carried out with the support of CAPES – Financing Code 001. Jone Medja thanks the Universidade

Eduardo Mondlane and the program GCUB/ProAfri for sponsoring his post-graduate studies at UFBA.

### References

- Crain, E. R., (2019). Crain's Petrophysical Handbook. Chapter 16 - Reserves and Productivity. Available online at <https://www.spec2000.net/16-netpay.htm>.
- Davis, J. C., (2002). Statistics and Data Analysis in Geology. New York, John Wiley & Sons. 3<sup>rd</sup> edition. 651p.
- Everitt, B. S., Landau, S., Leese, M., Stahl, D., (2011). Cluster Analysis. 5<sup>th</sup> edition. Wiley Series in Probability and Statistics. UK. 321p.
- Flexa, R. T., Andrade, A., and Carrasquilla, A., (2004). Identificação de litotipos nos perfis de poço do Campo de Namorado (Bacia de Campos, Brasil) e do Lago Maracaibo (Venezuela) aplicando estatística multivariada. Revista Brasileira de Geociências. 34(4):571–578.
- Hancox, J., Brandt, D. Edwards, H., (2002). Sequence stratigraphic analysis of the Early Cretaceous Maconde Formation (Rovuma Basin), northern Mozambique. Journal of African Earth Sciences. 34:291–297.
- Key, R. M., Smith, R. A., Smelror, M., Powell, J. H., Thorsnes, T. Njange, F., Saether, O. M., and Zandamela, E. B. (2008). Revised lithostratigraphy of the Mesozoic-Cenozoic succession of the onshore Rovuma Basin, northern coastal Mozambique. South African Journal of Geology. 111(1):89–108.
- Ramayah, T., Ahmad, N. H., Halim, H. A., Zainal, S. R. M., and Lo, M., (2010). Discriminant analysis: An illustrated example. African Journal of Business Management, 4(9):1654–1667.
- Rosa, H., Suslick, S. B., Vidal, A. C., e Sakai, G. K., (2008). Caracterização de eletrofácies por meio de ferramentas estatísticas multivariadas. Revista Escola de Minas. 61(4):415–422.
- Worthington, P. F. and Cosentino, L. (2005). The Role of Cut-Offs in Integrated Reservoir Studies. SPE Reservoir Evaluation & Engineering, 8(4):276–290.