Brazilian Journal of Geophysics (2021) 39(2): 259-274

ISSN 0102-261X
DOI: 10.22564/rbgf.v39i3.2097

COMPARISON BETWEEN PERMEABILITY ESTIMATES IN A CARBONATE RESERVOIR OF CAMPOS BASIN USING EMPIRICAL AND MATHEMATICAL APPROACHES

Abel Gonzalez Carrasquilla * and Pedro Mureb

ABSTRACT. The permeability estimate, using geophysical well logs, is an important and difficult task in a reservoir characterization, which is usually done with empirical models or statistical regression. Thus, logs were used to achieve this goal in two wells of a carbonate reservoir in Campos Basin, together with geological information and laboratory measurements on rock samples. The test Well A10 was zoned in electrofacies and models were built for each one. Next, the hydraulic flow units were defined and correlated to the logs and to the electrofacies. The same was done in a blind test in Well A3. The Principal Component Analysis, Cluster and Discriminant techniques, Multiple Linear Regression, Alternate Conditional Expectation and Hydraulic Flow Unit were used. The quality of the estimates was calculated using the Mean Absolute Error and the Coefficient of Determination. At Well A10, the Hydraulic Flow Unit was the most promising approach. The Alternate Conditional Expectation, without zoning, was the closest to experimental laboratory data in Well A3. These results indicate that all methods are feasible in inferring permeability, however, an inadequate classification of zones can lead to erroneous estimates.

Keywords: carbonate reservoir, permeability estimates, multivariate statistics, well logs, laboratory measurements.

RESUMO. A estimativa da permeabilidade, usando perfis de poços, é uma tarefa importante e difícil na caracterização de reservatórios, a qual é normalmente feita com modelos empíricos ou regressão estatística. Com esse fim, os perfis de poços juntamente com informações geológicas e medições de laboratório em amostras de rocha foram usados em dois poços de um reservatório carbonático da Bacia de Campos. O poço de teste A10 foi zoneado em eletrofácies e modelos foram construídos para cada uma. As unidades de fluxo hidráulico foram definidas e correlacionadas aos perfis a às eletrofácies. Isso foi feito num teste cego no poço A3. As técnicas Análise de Componentes Principais, de Cluster e Discriminante, Regressão Linear Múltipla, Expectativa Condicional Alternada e Unidade de Fluxo Hidráulico foram utilizadas. A qualidade das estimativas foi calculada com o Erro Médio Absoluto e o Coeficiente de Determinação. No Poço A10, a Unidade de Fluxo Hidráulico foi a abordagem mais promissora. A Expectativa Condicional Alternada, sem zoneamento, foi a técnica mais próxima dos dados experimentais de laboratório no poço A3. Esses resultados indicam que todos os métodos são viáveis na inferência da permeabilidade, porém, uma classificação inadequada das zonas pode levar a estimativas errôneas.

Palavras-chave: reservatório carbonático, estimativas de permeabilidade, estatísticas multivariadas, perfis de poços, medições laboratoriais.

Corresponding author: Abel Carrasquilla

Universidade Estadual do Norte Fluminense Darcy Ribeiro - UENF, Laboratory of Engineering and Exploration of Petroleum (LENEP), Amaral Peixoto Highway, Km 163, Brennand Avenue, S/N, Imboacica, Macaé, RJ, Brazil. 27925310. Phone: +55222765-6564 ¬– Emails: abel@lenep.uenf.br, pedro.mureb@gmail.com

INTRODUCTION

The characterization of a reservoir is an especially important part of the management and development of an oil field. This involves many processes, such as analyzing experimental data, interpreting well logs, integrating geological and geochemical information, and so on. With that, it is possible to do a better quantitative distribution of the properties and prediction of the flow behavior in the reservoir. The estimation of the permeability of a rock, from well logs, within these activities, is an important and difficult task, in the characterization of a reservoir (Abbaszadeh et al.,1996; Martin et al., 1997).

Empirical models need adjustments to be applied in different depositional environments. (Wendt et al., 1986). The statistical regression, on the other hand, has been used more often because it is more flexible to do this kind of estimate, but involves *a priori* assumptions regarding the parameters. The conventional statistical regression has, generally, been done parametrically, using multiple linear or nonlinear models, that require *a priori* assumption about the data (Jensen & Lake, 1985). In late years, non-parametric regression techniques, such as Alternating Conditional Expectations (ACE), have been presented to overcome the restrictions of conventional multiple regression methods (Datta-Gupta et al., 1999; Breiman & Friedman, 1985; Breiman, L. 1993).

Another technique that is employed, to facilitate this characterization process, is the reservoir zoning, which consists in the division into zones, with geological, petrophysical and/or depositional similar characteristics. The zoning technique was proposed by Amaefule et al. (1993), in which the reservoir is divided into Hydraulic Flow Units. These units have characteristics that describe the fluid flow internally, with similar or different properties from others.

The applications, in complex carbonate reservoirs, have shown outstanding promise in the management of many patterns of heterogeneity in rock properties (Barman et al., 1998; Lee et al., 2002). Nevertheless, it remains significantly difficult to identify sharp local variations in a reservoir property caused by abrupt changes in the depositional environment (Huang et al., 1996). Another distinctive feature, in carbonate reservoirs, is the inconsistency of the porosity – permeability relationship, i.e., low permeability, in regions that exhibit high porosity, and vice-versa. All these characteristics are extremely important from the point of view of predicted behavior of the flow in the reservoir (Lee & Datta-Gupta, 1999; Xue et al., 1997; Mathisen et al., 2003).

Therefore, the real purpose of this study is to compare the Multiple Linear Regression, Alternating Conditional Expectations and Hydraulic Flow Units techniques in the permeability estimation of a carbonate reservoir in Oilfield A in Campos Basin, Southeastern Brazil. The lithological information, the results of laboratory tests on rock samples and, the geophysical logs from two wells were used to do this.

Thus, all of the aforementioned approaches were tested, and their performances were evaluated throughout the study. These results lead towards the search for techniques with better performance than those achieved in the research and, in the application of the knowledge acquired, in these estimates, in more complex reservoirs, such as those found in the pre-salt layer.

Geological Context

The Campos Basin is in the coast of Rio de Janeiro, in the southeastern region of Brazil. It occupies an area approximately of 120.000 km^2 and has a bathymetry of 3.500 m. The basin is limited to the north by the Espírito Santo Basin, at High of Vitória and, to the south, by the Santos Basin, at Cabo Frio High. Its origin is related to the separation of the South American and African plates, followed by the opening of the Atlantic Ocean (Bizzi et al., 2003). The Oilfield A is composed by Albian carbonate reservoirs of Quissamã Formation and it is located in the southwest part of the Campos Basin (Fig. 1).

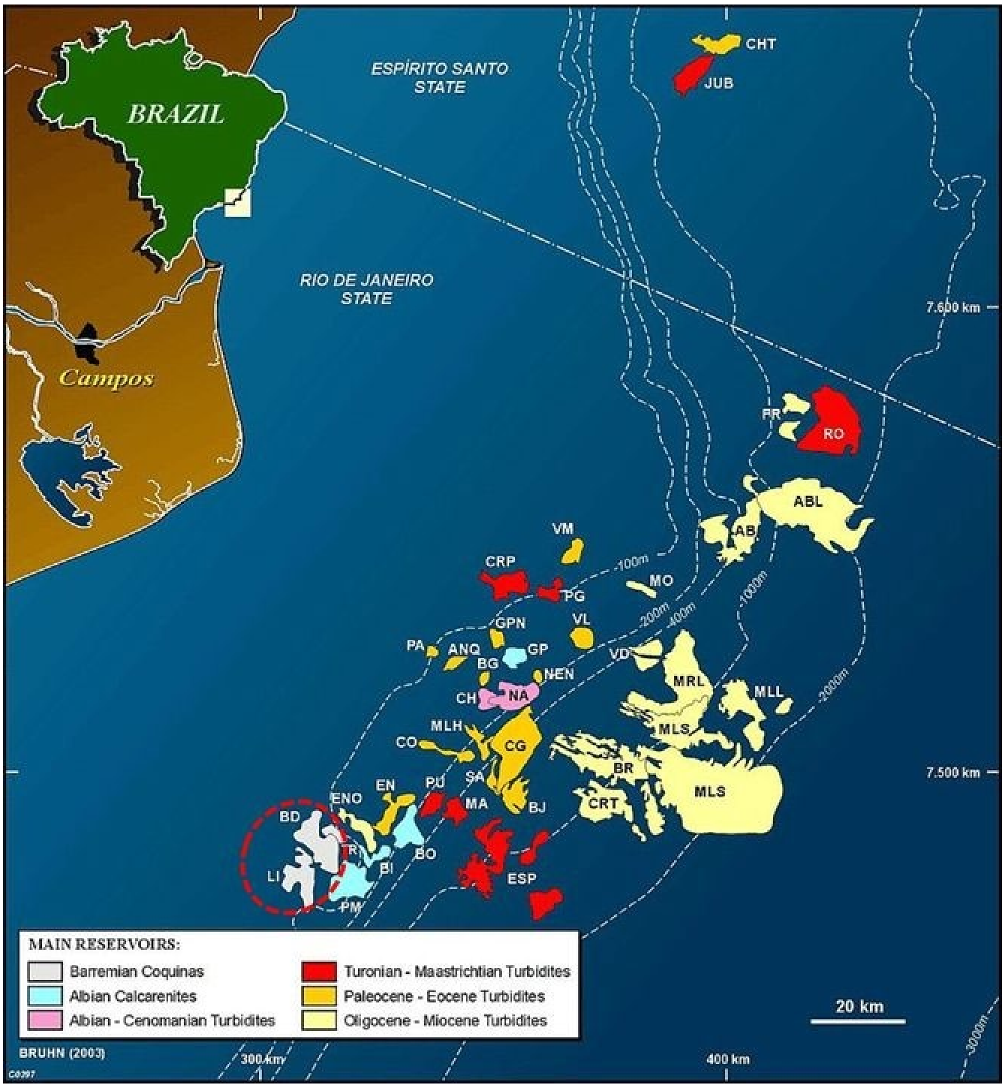

Figure 1 - Albian carbonate reservoirs in the Campos Basin showing the approximate location through a yellow dashed ellipse (modified from Bruhn et al., 2003).

According to Winter et al. (2007), the sedimentation of this carbonate starts in the drift phase, at the beginning of Albian with marine deposition (Fig. 2).

The depositional model of this oilfield corresponds to a carbonate platform (Spadini et al., 1988). This depositional model characterizes, according to Okubo et al. (2015), the sedimentation in high energy environment (oolitic and oncolitic granstones), moderate energy environment (oolitic peloidal grainstones and oncolitic bioclastic packstones) and low energy environment (peloidal bioclastic packstones and wackestone) (Fig. 3).

The Oilfield A has a total of 27 wells and, from this total, we selected two wells to carry out this study. The selected wells (A3 and A10) were chosen because they have a more complete dataset, with photos and sequential analysis of cores, geophysical logs, composite logs, and laboratory tests (water saturation with retort and membrane, mercury injection etc.). The Well A10 is on the carbonate ramp in the high energy zone and the Well A3 is in the moderate energy zone. They are spaced 1.36 km apart, with Well A3 32° east from Well A10 (Fig. 4).

MATERIALS AND METHODS

To develop this work, different mathematical approaches were used, such as Principal Component Analysis (PCA), The Cluster Analysis (CA), Discriminant Function Analysis (DAF), Multiple Linear Regression (MLR), Alternating Conditional Expectations (ACE). Along with these techniques, some petrophysical methodologies have also been used, such as Hydraulic Flow Units (HFU) and Winland Approach (WA).

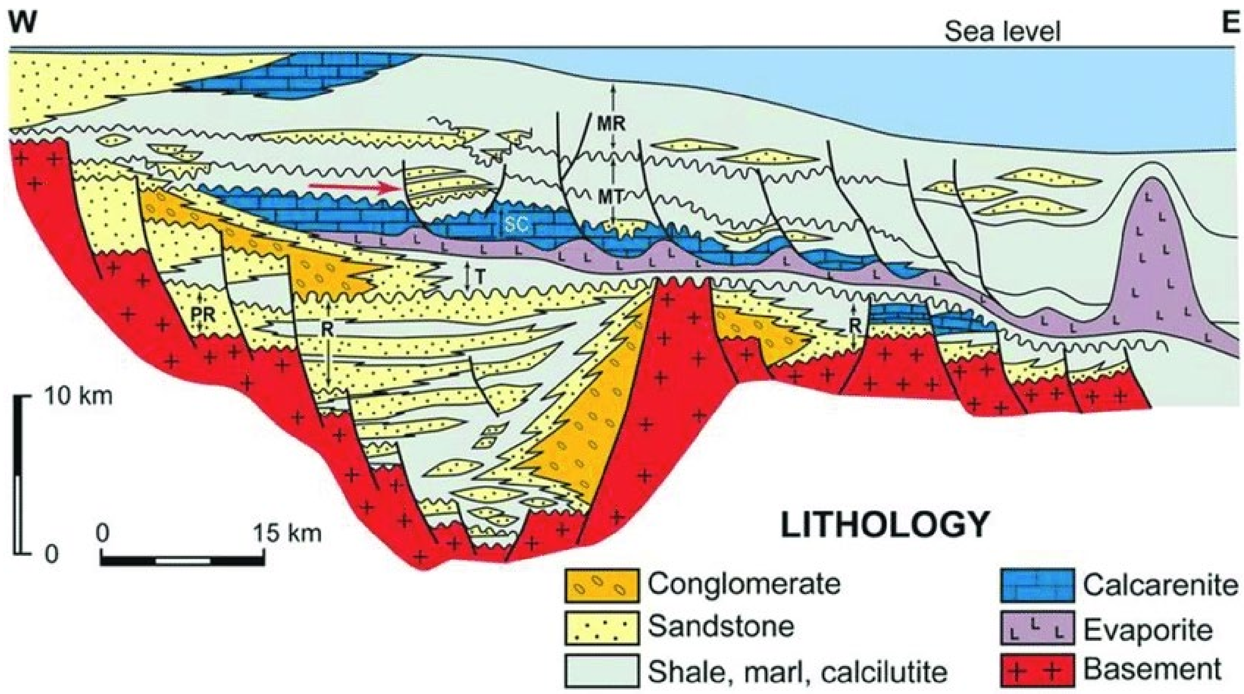

Figure 2 - Generalized geological section for the eastern Brazilian marginal basins. Shallow carbonate platform mega sequence (early to middle Albian) is identified by SC (modified from Bruhn et al., 2003).

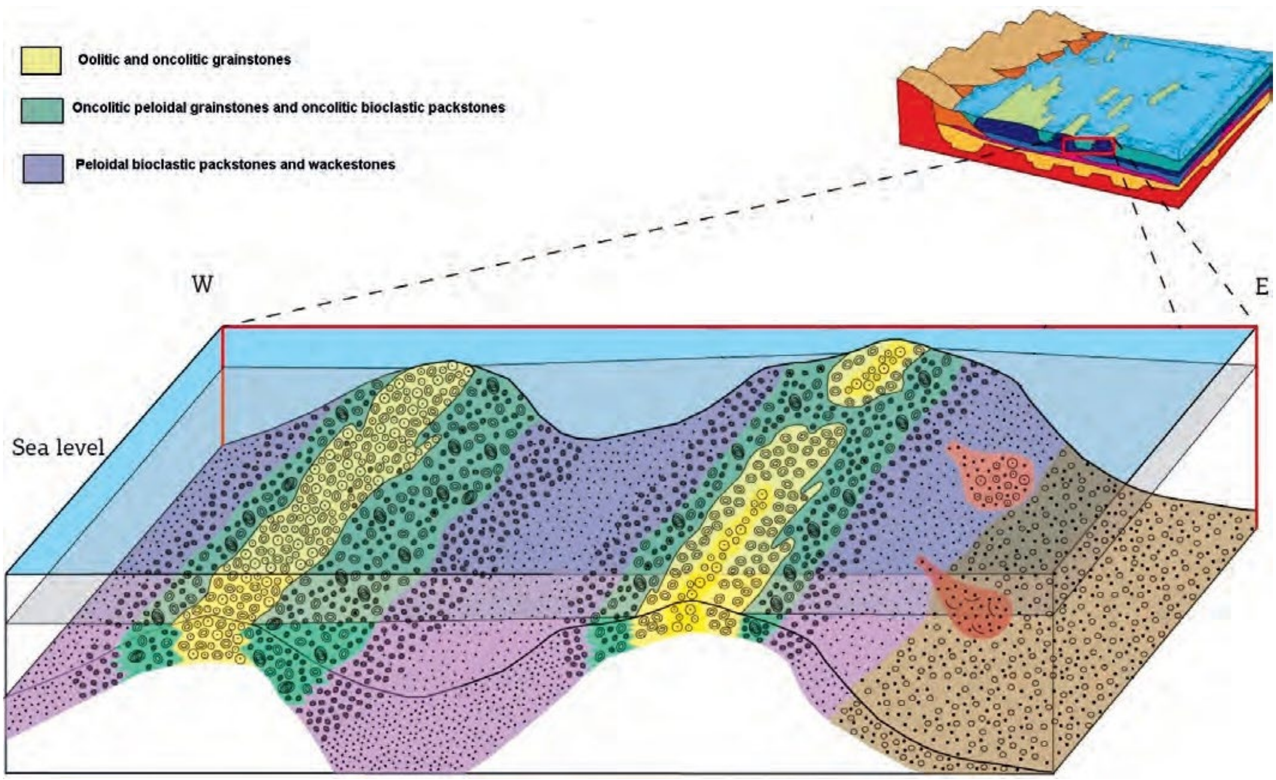

Figure 3 - Depositional model of the carbonate ramp in the Campos Basin during the Albian (modified from Okubo et al., 2015).

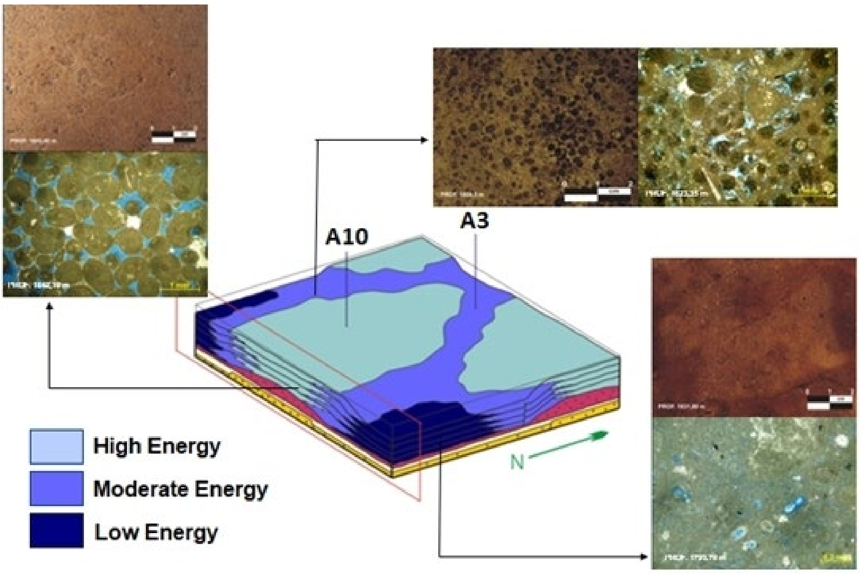

Figure 4 - Location of the Wells A3 and A10 on the carbonate ramp. In detail, the plug photos and laminae of the facies associated with energy zones: in the high energy zone we find oncolitic peloidal grainstone, in the zone of moderate energy oncolitic (modified from Okubo et al., 2015)

PCA transforms an original set of variables into another set, called the principal components (PC) of equivalent dimensions. This transformation occurs with the smallest possible loss of information, which also seeks to eliminate some original data that have little information. This transformation is only possible if the initial variables are not independent and have non-zero correlation coefficients (Vicini & Souza, 2005; Jensen & Lake, 1985).

CA classifies the data into groups, classes, or categories, in a situation where no prior information on groups is available, known as unsupervised classification (Kriegel et al., 2012; Agrawal et al., 2005). The objective of the CA is that the units within the groups should be as similar as possible and the formed groups should be as different as possible. This work uses k-means clustering, which randomly divides the data in k initial groups. K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. This clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. Then, the centroid of initial groups is computed, and each data point is relocated to the nearest centroid. The centroids are recalculated and, after that, each point in the data set is reallocated to the nearest centroid. This is made until there is no more reallocation. Once the k-means clustering is complete, several larger groups, which can be geologically justified, are reached. The results of these groupings can be presented by graphs dendrograms. The dendrograms present the elements and the melting points or divide the groups formed at each stage (Sancevero et al., 2008; Konate et al., 2017; Puskarczyk, 2020).

The next step is to use a hierarchical clustering technique, which consists in the fusion of the groups, obtained by the k-means clustering, by similarity until all data forms only one group (Steinbach et al., 2000; Vattani, 2011). With the hierarchical clustering, the initial groups that are not geologically explained converge into groups that can be geologically explained: the electrofacies (EF). With the data from well logs grouped in the EF, a function was used to classify them into new groups (Ferreira, 1996). DAF was used to classify and distribute existing groups into new classes or categories. It is a method used in statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification. This technique is utilized when the groups are known beforehand, whatever goes in the opposite direction to CA. has the following form (Hair et al., 2009; Esbensen et al., 2002):

$$Z_{ft} = \propto + w_1 x_{1t} + w_2 x_{2t} + \cdots + w_p x_{pt} \quad (1)$$

where Z_{ft} is the discriminant score of the *f* discriminant function to the *t* observation; *α* is the y-axis intercept; w_p is the discriminant weight to the independent variable x_p; and x_{pt} is the *t* observation in the x_p independent variable (Büyüköztürk & Çoklukbökeoglu, 2008).

MLR is used to study the relationship between a dependent variable with multiple independent variables. The MLR has the following form (Montgomery et al., 2015):

$$y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_u X_u + \varepsilon \quad (2)$$

where *y* is the dependent variable; X_u is the independent variable; α is the linear coefficient; β_u is the angular coefficient; and ε is the error.

ACE is one of the methods to find nonlinear transformations of variables that produce the best fitting additive model in multiple regression problems. This produces the maximum linear effect between the transformed independent variables and the transformed response variable (Breiman & Friedman, 1985). The knowledge of such transformations aids in the interpretation and understanding of the relationship between the response and predictors, to minimize the fraction of variance not explained. The transformation is obtained from data in an iterative way (Wang & Murphy, 2004):

$$\theta(Y) = \sum_{n=1}^{m} \varphi_n(V_i) + e \quad (3)$$

where θ is the transform of the dependent variable; *Y*, φ_n are the transforms of independent variables V_p; and *e* is the error.

To quantify the HFU, a graphic method like the Modified Lorenz Plot (MLP) was used, which, according to Gunter et al. (1997), offers a guide on how many HFU are necessary to represent the geological structure. In accord with these authors, the MLP shows, in percentage, the accumulated flux capacity (k_z) versus the accumulated storage capacity (ϕ_z). To characterize the HFU, it is applied the equations proposed by Amaefule et al. (1993) to calculate the Flow Zone Indicator (FZI). The FZI is an exclusive indicator for each HFU and it is based on its porosity and its permeability data, which are measured in plugs:

$$FZI = \frac{RQI}{\phi_z} \quad (4)$$

where RQI (μm) is the Reservoir Quality Index and the ϕ_z (fraction) is the pore by grain volume ratio, which can be computed by the following equations:

$$RQI = 0.0314\sqrt{\frac{k}{\phi_e}} \quad (5)$$

$$\phi_z = \frac{\phi_e}{1 - \phi_e} \quad (6)$$

where ϕ_e is the effective porosity and *k* the permeability (md). Using the FZI, it is possible to estimate the permeability (md) with the following equation:

$$k = 1014(FZI_n)^2 \frac{\phi_e{}^3}{(1 - \phi_e)^2} \quad (7)$$

where FZI_n is the *n-th FZI* mean of the *n-th* HFU. To determine the petrofacies, the pore throat radii were calculated by Winland (1972) for the accumulated saturation of 35 % of mercury injection capillary pressure (R35), to represent the curves capable of distinguishing different energies in the depositional environments. The proposed equation considers the pore throat radius (r), the laboratory permeability (k) and the laboratory porosity (ϕ):

$$\log(r) = A * \log(k) + B * \log(\phi) + C, \quad (8)$$

where *r* is the radius of the pore throat; A = 0.59 and B=-0.86 are respectively the permeability and porosity coefficients; and C (-0.73) is the intercept with the y-axis. He obtained an accumulated saturation of the Pearson´s correlation coefficient R² > 0.90.

To reach the objectives of this work, along with these previously described approaches, it was utilized data from gamma ray (GR), sonic (DT), neutron porosity (PHIN), density (RHOC), deep resistivity (RT) and micro resistivity (RXO) logs and plug experimental laboratory data (porosity and permeability) from Wells A10 and A3. All models were built in Well A10, that has more laboratory data, and to validate the results, a blind test was performed in Well A3. In the building model, firstly, the Well A10 was zoned in EF using PCA and CA. Then, MLR and ACE models were built for each EF and for the whole well. The next step was to make MLR to identify and characterize the HFU. After that, the HFUs were calculated and the k models for each HFU were built. Finally, a Discriminant Analysis (DA) was made, linking well logs to the EF and HFU classifications. In the blind test for Well A3, the DA was used to classify the well in EF and HFU. Then, MLR and ACE models were applied on each EF and the FZI model in each HFU. MLR and ACE models for the whole well were utilized too.

In this work, the Microsoft Excel software (Microsoft, 2021) was used to organize, filter, and convert the output data from Interactive Petrophysics - IP software to be used in Python and vice versa. In Python, discriminant function analysis, error calculation (Sklearn library) and regression with ACE (The ACE Package) were performed. In addition to these libraries, NumPy, Matplotlib and Pandas were used (McKinney, 2010; Hunter, 2007; Walt et al., 2011; Pedregosa et al., 2011; Touran, 2012). The IP was used for other analyses, such as principal component analysis, cluster analysis, multiple linear regression, hydraulic flow zones and also for interpretation and graphing of the results (LR Senergy, 2021).

To verify the quality of the estimates or the goodness of fit, the Mean Absolut Error (MAE) and the Pearson´s coefficient of determination (R^2) were utilized. MAE measures the absolute differences between prediction and observation, where all individual differences have equal weight. R^2, on the other hand, is a measurement used to explain how much variability of an estimate can be caused by its relationship to observations. The more MAE and R^2 approach 0 and 1, respectively, the better the estimate is considered. (Harrel, 2015; Williams, 1978; Lindley, 1987).

DISCUSSIONS AND RESULTS

According to Carrasquilla & Silva (2019), from the data in the logs it is possible to correlate the wells. The correlation is shown in Figure 5, considering the logs in the interest zone. Presenting the records at the same scale, it is possible to conclude that the interest zone in Well A10 is around 10 m deeper than in Well A3. In this figure, four of the main intervals with similar characteristics between the logs are identified:

1) high RT, RHOB log to the left of the NPHI curve and higher DT values. These features point to the presence of hydrocarbons;
2) the values of RT log and the falling of DT log, indicating a transition zone;
3) RT log and the distance between RHOB and NPHI logs, indicating a water zone;
4) a small RT peak indicating a second zone with hydrocarbons, but other logs indicating low porosity and permeability

Still following Carrasquilla & Silva (2019), the laboratory data are shown in Figure 6, having a strong linear dependence between permeability (k_{LAB}) and porosity (ϕ_{LAB}), with R^2 = 0.81 (Pearson´s determination coefficient) and a linear relationship equation in the form log (k_{LAB}) = -1.4142 + 7.2875ϕ_{LAB} between these two parameters. The light and dark blue dots indicate oil and water, respectively. ϕ_{LAB} range between 0.7% and 35%, with values concentrating between 20% and 24%. In the identified oil section, ϕ_{LAB} is found between 24% and 28%, while in the identified piece of water ϕ_{LAB} is localized between 12% to 16% and 20% to 24%. k_{LAB} values are concentrated between 1.6 mD to 1.8 mD and range between 0.1 mD and 40 mD. In the identified hydrocarbon portion, permeability is focused between 1.6 and 1.8 mD.

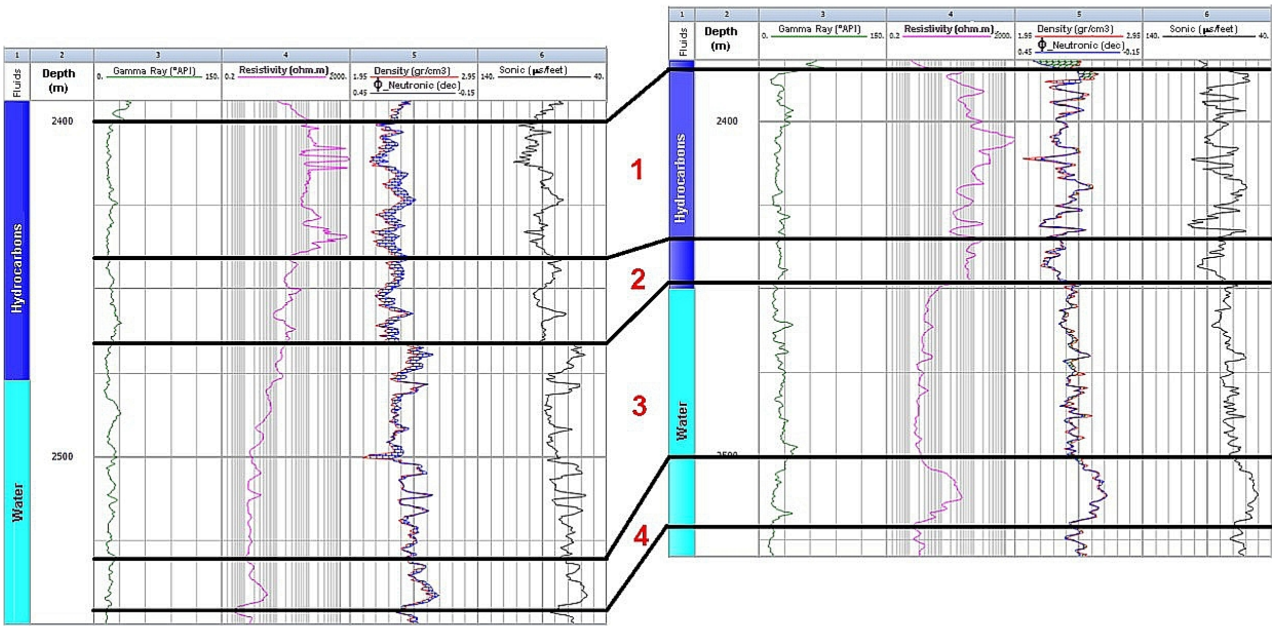

Figure 5 - Correlation between GR, RT, RHOB, NPHI and DT logs to the Wells A10 (left) and A3 (right) of Oilfield A in Campos Basin. The red numbers 1 to 4 indicate areas with similar characteristics (after Carrasquilla & Silva, 2019).

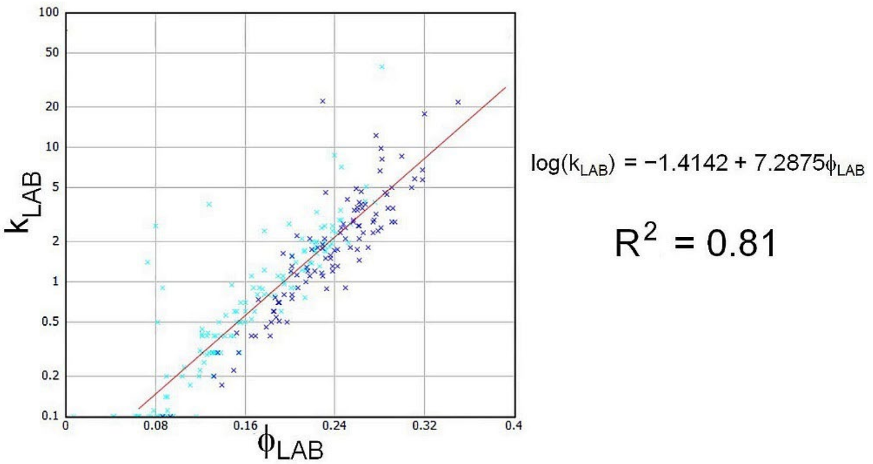

Figure 6 - Porosity (phiLAB) and permeability (kLAB) laboratory data crossplot for Well A10 showing a strong direct relationship, with dark blue indicating the hydrocarbon zone and light blue the aquifer (after Carrasquilla & Silva, 2019)..

In the case of Well A10, the model was built first utilizing the PCA, aiming to eliminate noise from the data. The technique was applied in all logs, so it was obtained four principal components with 60.24 %, 25.06 %, 9.98 % and 4.73 % of the variance. Then, the last PC was discarded because the first three explain 95.28 % of the information.

Reservoirs in Wells A10 and A3 were classified in five lithofacies (Petrobras, 2012), using a supervised classification in the plug analysis. They are: grainstone, cemented grainstone, packstone, cemented packstone, and wackestone. Therefore, for this reason, it was decided to make the CA with five EF. Figure 7 shows the cross plot of PC 1 and PC 2 (Principal Components), with the lithofacies in the color bar, showing that where it is clear there is an overlap between EF 1 and EF 2, like in EF 3 and EF 4. Thus, as shown in this figure, the algorithm fails to distinguish between high energy (grainstone and cemented grainstone) and moderate energy (packstone and cemented packstone) lithologies.

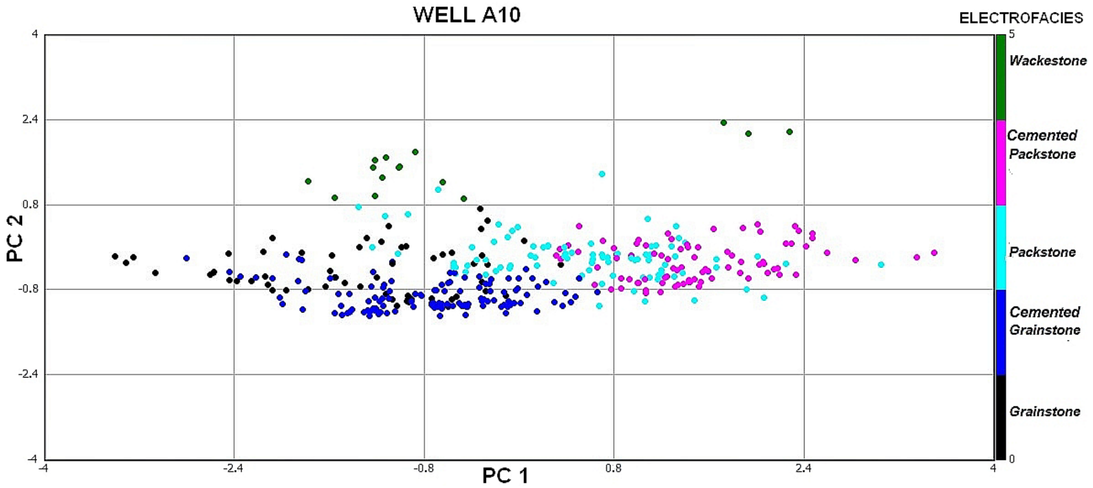

Figure 7 - Cross plot of CPs 1 and 2 with the lithotypes classification in the color bar

Similar results are shown in Figure 8 (track 6), where there is an overlap between the red dots (estimate) and, the black dots (Petrobras classification). So, it can be said that the CA, in Well A10, did not perform well, using the Petrobras (2013) lithofacies classification of EF as model. The explanation for this, as shown above, is because the algorithm does not distinguish the different EF with the same level in the energy.

Then, the MLP was plotted to zone the Well A10 in HFU and to characterize them (Fig. 9). As seen in this figure, the HFU 5, 6, 7, 8 and 9, where the curve has the highest slope, are considered as speed zones. The HFU 1, 2, 3 and 4, where the curve is more horizontal, are barrier zones. After that, FZI was calculated, and the average FZI for each HFU was obtained (Table 1). An upward trend is observed between the means of FZI 1 (0.25) and FZI 9 (17.62). On the other hand, Figure 8 shows the Winland (1972) porosity vs permeability cross plot, and Figure 9 shows the ϕ_Z (Phi Z) vs RQI cross plot. Both figures show a good separation of the HFU, being clear that there is no overlap. Next, the permeability was then estimated, using the estimated FZI.

Thus, Figure 10 highlights the barrier zones in pink, and the speed zones in blue. This figure shows, at the same time, the GR (track 2), DT (track 3), RT/ RXO (track 4) and PHIN/RHOC (track 5) logs, the HFU permeability model (track 6), the joint graph between the FZI permeability model and the laboratory permeability (track 7), and the laboratory porosity (track 8). To perform the blind test in Well A3, the DA was performed to make the EF and HFU classifications. Thus, the EF discriminant function was estimated using the logs, with mean accurate of 94 %, five EF (track 6, Fig. 11). Unlike Well A10, this classification worked well. For the HFU, the DA was resulted in an average accurate of 59 % and DA classified five HFU (1, 2, 3, 6 and 9), less than the nine existing in Well A10 (track 3, Fig. 12).

Finally, the models of permeability were applied with EF for both wells (MLR EF and ACE EF) and without EF (MLR, ACE and HFU) for the whole well (Fig. 13). Based on the lowest MAE values, Table 2 shows that the best zoning techniques to Well A10 are: HFU, MLR EF, ACE, MLR and ACE EF, in order. In other words, there is no clear distinction between the use or not of zoning. For Well A3, meanwhile, the best zoning techniques were those without EF zoning: ACE, MLR and HFU, in order. The worst were those that used zoning: MLR EF and ACE MLR, in order. When analyzing without zoning, the ACE was the best estimate for the Well A3 and the third for the Well A10 (Table 2). The explanation for this is because, in this condition, the number of data used to construct the permeability estimate is larger with 372 data points. Thus, there is only one EF with more than 100 observations to do the zoning. In accord to Breiman (1993) and Harrell (2015), dataset with less than 100 observations hinders the estimate of all transformations in ACE.

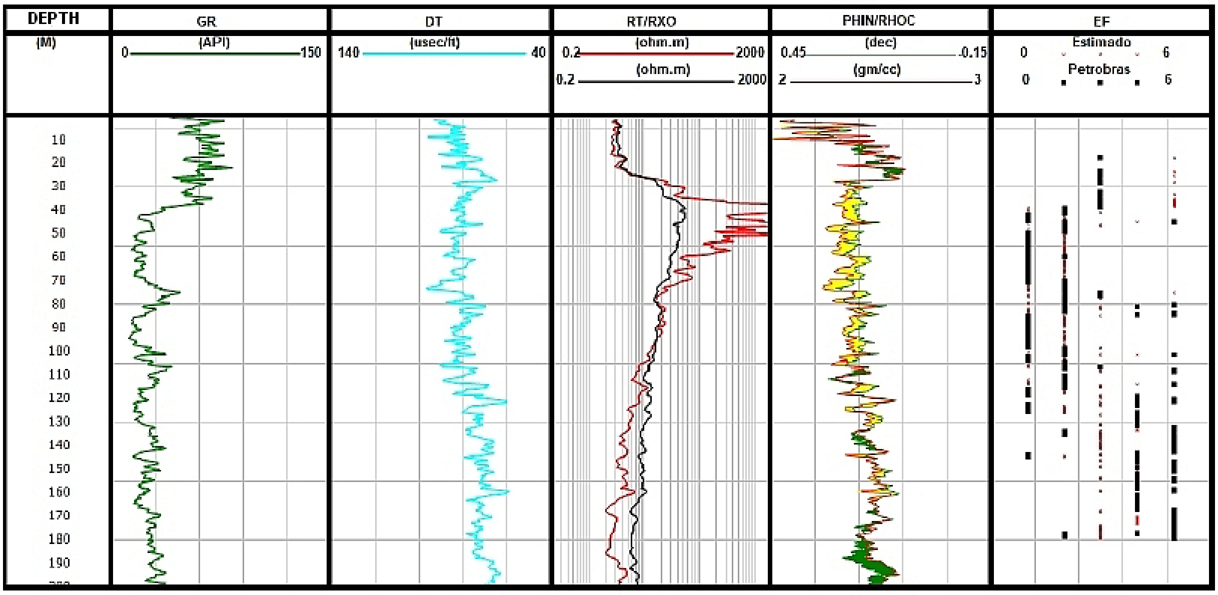

Figure 8 - CA classification compared with Petrobras classification of EF.

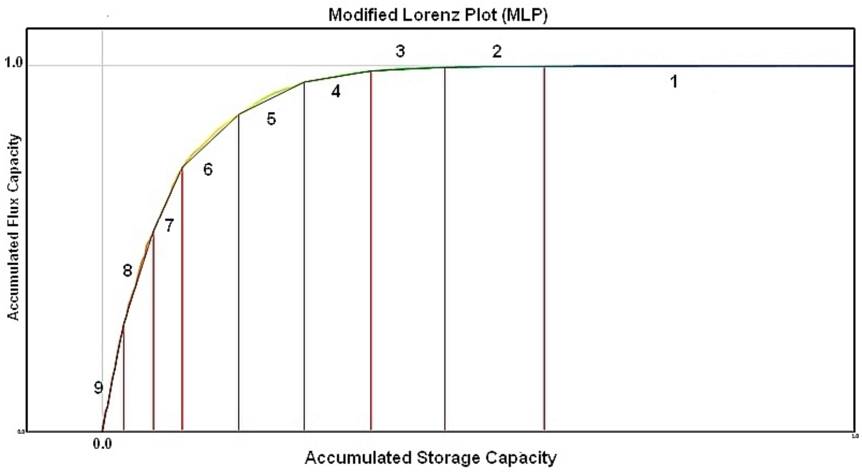

Figure 9 - MLP showing the nine HFUs.

Table 1 - HFU statistical characterization.

HFU	FZI range (μm)	Mean FZI (μm)	RMSE (mD)	MAE (mD)	Number of data
1	0.081 - 0.470	0.25	0.48	0.39	170
2	0.470 - 1.249	0.73	0.56	0.42	49
3	1.249 – 2.339	1.65	0.33	0.26	34
4	2.339 – 4.131	3.00	0.30	0.23	30
5	4.131 – 6.263	4.96	0.26	0.22	28
6	6.263 – 9.104	7.25	0.27	0.22	25
7	9.104 – 11.664	10.27	0.21	0.17	12
8	11.664 – 15.574	13.30	0.20	0.20	13
9	15.574 – 40.907	17.62	0.29	0.24	11

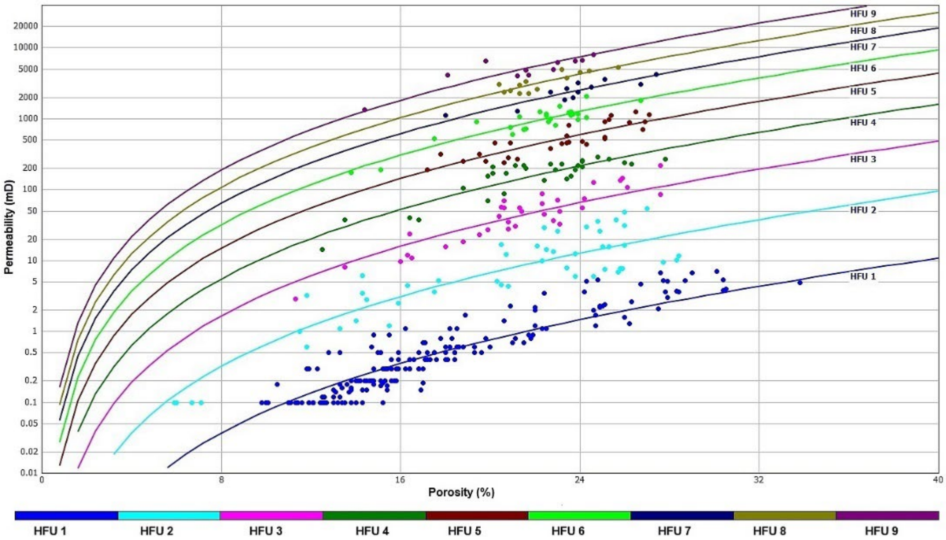

Figure 10 - Cross plot of plug permeability and porosity with the HFUs in the color bar.

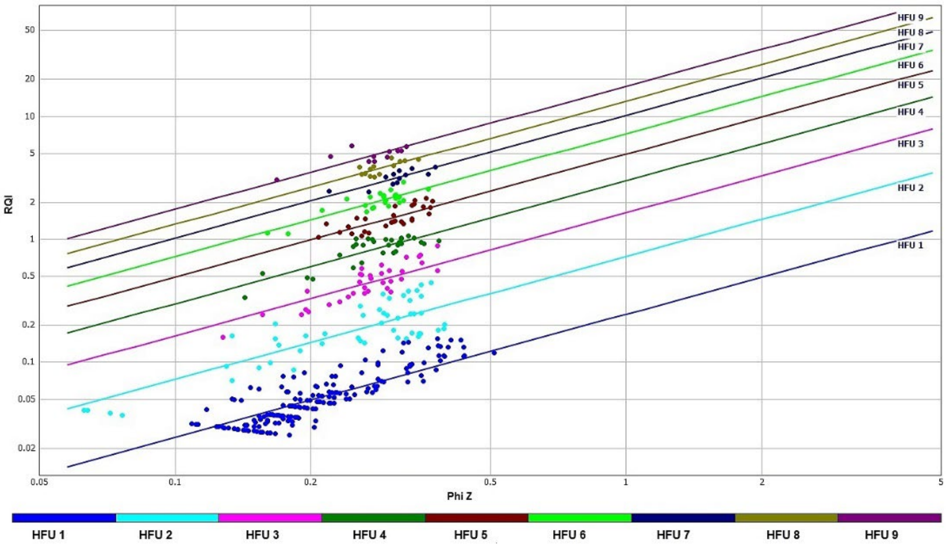

Figure 11 - Cross plot of RQI and ϕZ with the HFUs in the color bar.

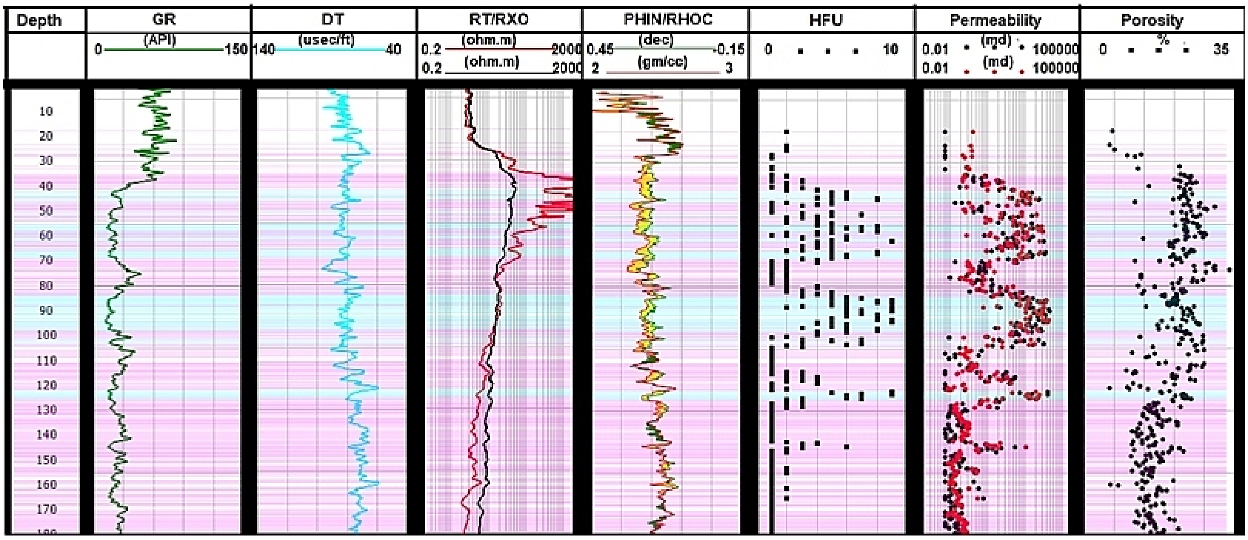

Figure 12 - Graph of the UFH, the basic well logs, the permeability and porosity of the plugs and the estimated permeability. In blue and pink, the velocity and barrier zones, respectively.

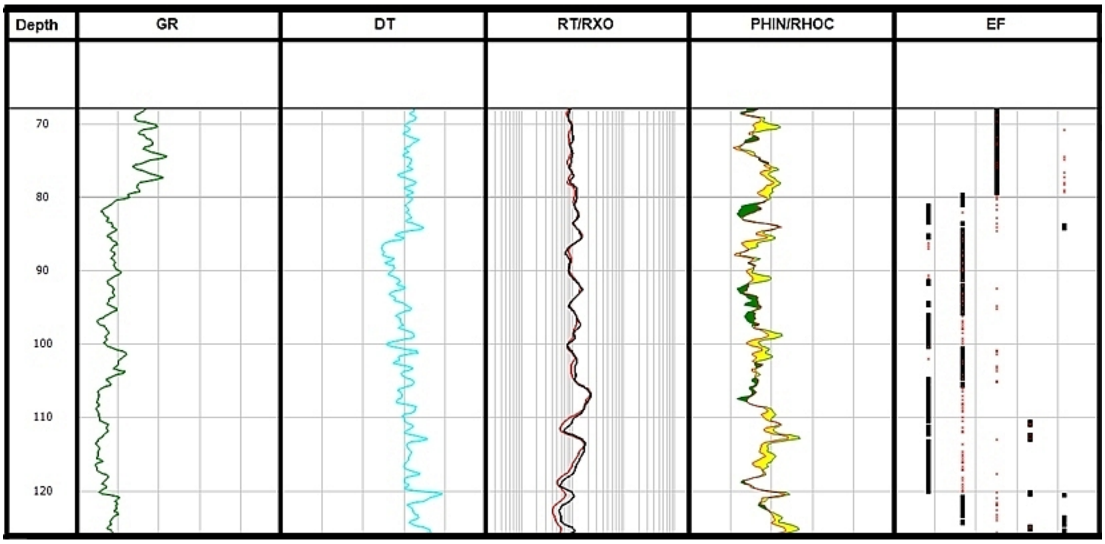

Figure 13 - EF classification through DA.

Table 2 - Errors of all used permeability estimation methods.

METHOD FOR K ESTIMATE	WELL A10 (reference)		WELL A3 (blind test)	
	MAE (mD)	Ranking	MAE (mD)	Ranking
MLR EF	0.59	2	0.92	4
ACE EF	0.99	5	0.99	5
MLR	0.77	4	0.81	2
ACE	0.61	3	0.79	1
HFU	0.33	1	0.95	3

For Well A3, the best estimate for permeability is presented by the high energy region, that can be explained by the predominance of grainstone, which is a more homogeneous facies, when compared to the other lithologies found in the reservoir (red dots in Fig. 14). The best correlations in this high energy region were obtained by ACE without zoning (R2 = 0.47), which can be considered low (red dots in Fig. 15).

The HFU technique was a promising model, placing first on the permeability estimate for Well A10 and third for Well A3. This can be explained by the fact that Well A10 has a greater permeability range than Well A3 (Almeida, 2015). In accord with Lichotti (2016), Well A10 has FZI values ranging from 0.081 to 40.907 µm, while Well A3 ranges from 0.1405 to 21.1688 µm. Thus, Well A3 may not have the same number of HFU obtained in Well A10, or, the HFU has a quite different FZI range. This produced a wrong HFU classification of Well A3 and, therefore, provided a bad permeability estimation.

CONCLUSIONS

In this work, geological information, geophysical well logs and experimental laboratory data in samples were used to make a comparison between different permeability estimates. The data came from Wells A3 and A10 of a carbonate reservoir in Oilfield A, Campos Basin, Southeastern Brazil. The multivariate statistical techniques Multiple Linear Regression (MLR) and Alternating Conditional Expectations (ACE) were used, besides reservoir zoning in Hydraulic Flow Units (HFU) and Electrofacies (EF). MLR and ACE models were built with and without EF zoning, and together with HFU methodology, were all initially applied in the reference Well A10. After this, a blind test in Well A3 was made, and the errors were calculated for comparison. The Component Analysis (CA) in both wells did not perform well, when Petrobras lithofacies classification was used as comparison, because the algorithm failed to distinguish between EF, in the same energy zone. This is reflected in a poor permeability estimate,

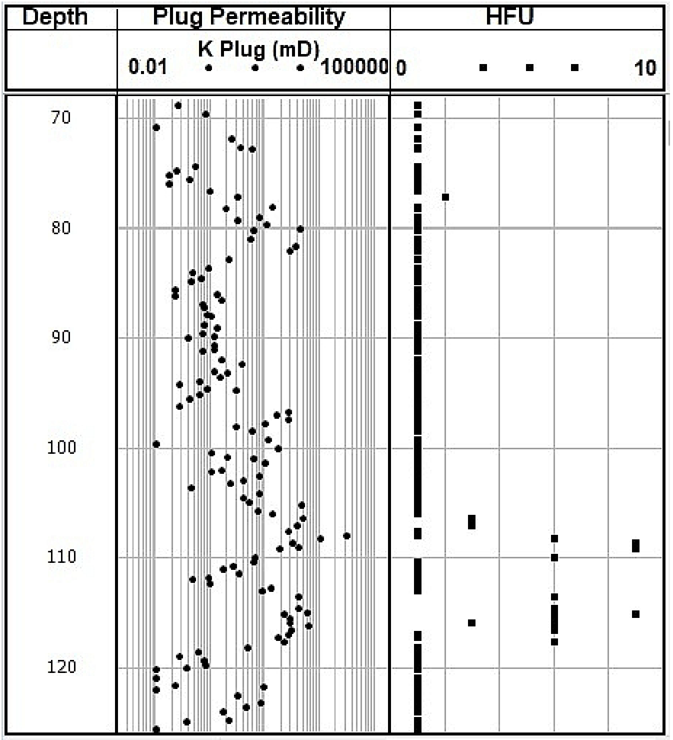

Figure 14 - Well A3 classification in HFUs.

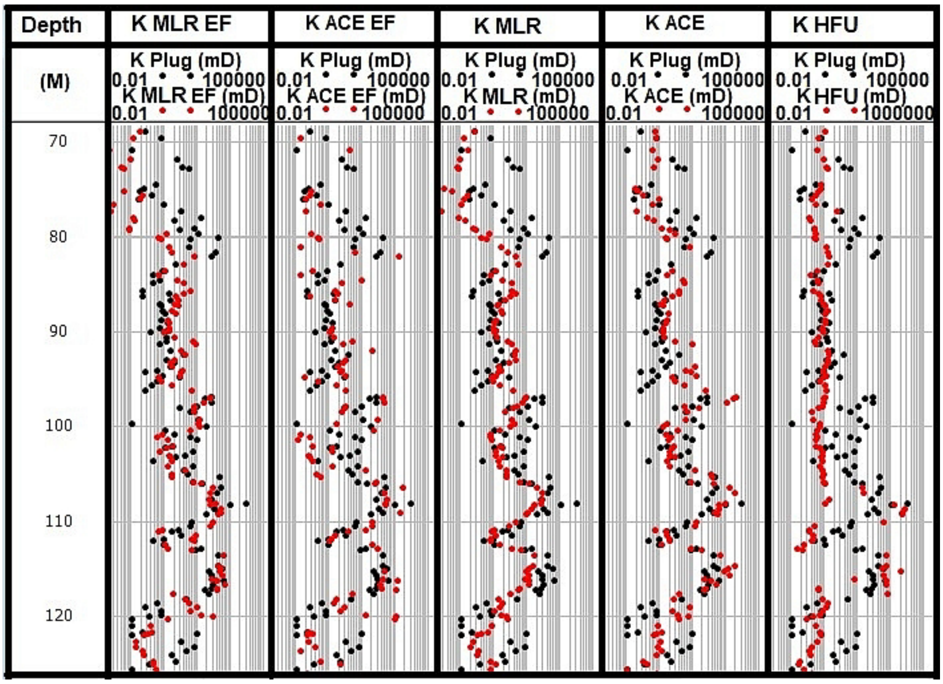

Figure 15 - In red all the estimated permeability and in black the plug permeability.

with MLR and ACE. When no zoning is used, MLR and ACE performed better, with ACE being the best technique. Besides the fact that there was no bad classification impairing this technique, the ACE method had a bigger number of data to perform the regression, 372 data points against 122 data points of the EF. The dataset with less than 100 observations hinders the estimative of all transformation of the ACE. When Well A10 was zoning in HFU, it was obtained nine units, with the 5, 6, 7, 8 and 9 HFU characterized as speed zones, or flux zones, and the 1, 2, 3 and 4 HFU characterized as barrier zones. The cross plots approach showed a good performance in zoning, because it does not result in data overlap. The HFU technique was the promising model on the permeability estimate, placing first in Well A10 and third in Well A3. This is explained by the larger range of permeability in Well A10 than Well A3. This resulted in a wrong HFU classification of Well A3 and, therefore, it provided a bad permeability estimate. It is concluded that the three utilized methods are useful in the permeability estimate. The zoning process, when applied, should be used with caution, because a poor classification may lead to an imprecise permeability estimate. The results of this article were good to estimate permeability, using statistics and deterministic approaches. Further research in petrophysical parameter estimates indicates that time should be invested in a better selection of input parameters for these estimates, using data mining and pattern recognition. The use of artificial intelligence techniques and stochastic methods can also improve the petrophysical assessments.

ACKNOWLEDGMENTS

The authors would like to thank LENEP for the computational infrastructure, CNPq by the research grant, LR Senergy for software academic license and Petrobras/ANP/UENF covenant for grant, dataset, and fees.

REFERENCES

ABBASZADEH M, FUJII H & FUJIMOTO F. 1996. Permeability prediction by hydraulic flow units: theory, and applications. SPE Formation Evaluate, 11: 263–271.

AGRAWAL R, GEHRKE J, GUNOPULOS D & RAGHAVAN P. 2005. Automatic subspace clustering of high dimensional data. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11: 5–33.

ALMEIDA P. 2015. Integration of geological attributes and well logs in the permeability estimation of a carbonatic reservoir in Campos Basin. Master Thesis, UENF/CCT/LENEP, Macaé, RJ, Brazil. 101 pp. (In Portuguese).

AMAEFULE J, ALTUNBAY M, TIAB D, KERSEY D & KEELAN D. 1993. Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/wells. SPE Annual Technical Conference and Exhibition, Houston, Texas, 3 - 6 October, p. 205 - 220.

BARMAN I, SHARMA A, WALKER R, DATTA-GUPTA A. 1998. Permeability predictions in carbonate reservoirs using optimal non-parametric transformations: an application at the salt creek field unit, Kent County, TX. In: Permian Basin Oil & Gas Recovery Conference, Midland, Texas, USA, p. 113 -125.

BIZZI L, SCHOBBENHAUS C, VIDOTTI R & GONÇALVES J. 2003. Geology, Tectonics and Mineral Resources of Brazil: text, maps, and GIS. Brasília: CPRM, 2003. 673 pp. (In Portuguese).

BREIMAN L. 1993. Fitting additive models to regression data: diagnostics and alternative views. Computational Statistics & Data Analysis, Elsevier, v. 15, n. 1, p. 13 - 46.

BREIMAN L & FRIEDMAN J. 1985. Estimating optimal transformations for multiple regression and correlation. Journal of the American Statistical Association, Taylor & Francis Group, v. 80, n. 391, p. 580 - 598.

BRUHN C, GOMES J, LUCCHESE JR C & JOHANN P. 2003. Campos Basin: reservoir characterization and management - historical overview and future challenges. Offshore Technology Conference, Houston TX, USA. OTC 15220.

BÜYÜKÖZTÜRK Ş & ÇOKLUK-BÖKEOĞLU Ö. 2008. Discriminant function analysis: concept and application. Egitim Arastirmalari - Eurasian Journal of Educational Research, 33: 73–92.

CARRASQUILLA A & SILVA R. 2019. A petrophysical multivariate approach using geophysical well logs and laboratory measurements to characterize an Albian carbonate reservoir in Campos Basin, Southeast Brazil. In: Carbonate Pore Systems: New Developments and Case Studies, Mountjoy Symposium 2, Special Publication No. 112, Society for Sedimentary Geology (SEPM). Edited by Donald McNeill, Paul Harris, Eugene Rankey & Jean Hsieh. Tulsa, OK, USA, 18(2): 61–72.

DATTA-GUPTA A, XUE G & LEE S. 1999. Nonparametric transformations for data correlation and integration: from theory to practice. AAPG Special Volume, Memoir 71, Chapter 27, pp. 381–396.

ESBENSEN K, GUYOT D, WESTAD F & HOUMOLLER L. 2002. Multivariate data analysis in practice: an introduction to multivariate data analysis and experimental design, Multivariate Data Analysis, 598 pp.

FERREIRA D. 1996. Multivariate analysis. Ed. UFLA, Lavras, MG, Brazil. 662 pp. (In Portuguese).

GUNTER G, FINNERAN J, HARTMANN D & MILLER J. 1997. Early determination of reservoir flow units using an integrated petrophysical method. SPE Annual Technical Conference and Exhibition. San Antonio, TX, 5 - 8 October, Paper 38679.

HAIR J, BLACK W, BABIN B, ANDERSON R & TATHAM R. 2009. Multivariate data analysis. Ed. Bookman, Porto Alegre – RS, 688 pp. (In Portuguese).

HARRELL F. 2015. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. Springer International Publishing, Switzerland, 582 pp.

HUANG Z, SHIMELD J, WILLIAMSON M & KATSUBE J. 1996. Permeability prediction with artificial neural network modeling in the Venture Gas Field, offshore Eastern Canada. Geophysics, Tulsa - OK, 61(2): 422–436.

HUNTER J. 2007. Matplotlib: A 2d graphics environment. Computing In Science & Engineering, IEEE, 9(3): 90-95.

JENSEN J & LAKE L. 1985. Optimization of regression-based porosity-permeability predictions. In: CWLS 10th Symposium, Calgary, Alberta, Canada, September.

JIANG H & ESKRIDGE K. 2000. Bias in principal components analysis due to correlated observations. Proceedings, Annual Conference on Applied Statistics in Agriculture, Department of Statistics, Kansas State University, pp. 148 -160.

KONATE A, MA H, PAN H, QIN Z, AHMED H & DEMBELE N. 2017. Lithology and mineralogy recognition from geochemical logging tool data using multivariate statistical analysis. Applied Radiation and Isotopes: Including Data, Instrumentation and Methods for Use in Agriculture, Industry and Medicine, 128: 55-67.

KRIEGEL H; KRÖGER P & ZIMEK A. 2012. Subspace clustering. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2 (4), pp. 351–364.

LEE S & DATTA-GUPTA A. 1999. EF characterization and permeability predictions in carbonate reservoirs: role of multivariate analysis and nonparametric regression. SPE Annual Technical Conference and Exhibition, Houston, TX, SPE56658, 13 pp.

LEE S, KHARGHORIA A & DATTA-GUPTA A. 2002. Electrofacies characterization and permeability predictions in complex reservoirs. SPE Res Eval & Eng., 5(03): 237–248. Paper Number: SPE-78662-PA.

LR SENERGY. 2021. Interactive Petrophysics - IP User Manual.

LICHOTTI H. 2016. Characterization of a carbonate reservoir in the Campos Basin from the integration of geological attributes, petrophysical properties and well logs studies. Master Thesis, UENF/CCT/LENEP, Macaé – RJ, 122 p. (In Portuguese).

LINDLEY D. 1987. Regression and correlation analysis. New Palgrave: A Dictionary of Economics, v. 4, pp. 120–123.

MARTIN A, SOLOMON S & HARTMAM D. 1997. Characterization of petrophysical flow units in carbonate reservoir. American Association of Petroleum Geologists Bulletin, 81: 734–759.

MATHISEN T, LEE SH & DATTA-GUPTA A. 2003. Improved permeability estimates in carbonate reservoirs using electrofacies characterization: a case study of the North Robertson Unit, West Texas. SPE Reservoir Evaluation & Engineering, 6(3): 176–184.

MCKINNEY W. 2010. Data structures for statistical computing in Python. In: SCIPYAUSTIN, TX. Proceedings of the 9th Python in Science Conference, 445: 51–56.

MICROSOFT. 2021. User's Manual Excel, Microsoft 365 (https://www.office.com).

MONTGOMERY D, PECK E & VINING G. 2015. Introduction to linear regression analysis. Wiley - Interscience, John Wiley and Sons, New Jersey, 672 pp.

OKUBO J, LYKAWKA R, WARREN L, FAVORETO J & DIAS-BRITO D. 2015.

Depositional, diagenetic, and stratigraphic aspects of Macae group carbonates (Albian): example from an oilfield from Campos Basin. Brazilian Journal of Geology, 45(2): 243–258.

PEDREGOSA F, VAROQUAUX G, GRAMFORT A, MICHEL V, THIRION B, GRISEL O, BLONDEL M, PRETTENHOFER P, WEISS R, DUBOURG V, VANDERPLAS J, PASSOS A, COURNAPEAU D, BRUCHER M, PERROT M & DUCHESNAY E. 2011. Scikit - learn: machine learning in Python. Journal of Machine Learning Research, 12: 2825–2830.

PETROBRAS 2013. Data and explanatory material of Oilfields A and B provided by Petrobras to SCTC. (In Portuguese). Brazil.

PUSKARCZYK E. 2020. Application of multivariate statistical methods and artificial neural network for facies analysis from well logs data: an example of Miocene deposits. Energies, 13(7): 1548–1566.

SANCEVERO S, REMACRE A, VIDAL A & PORTUGAL R. 2008. Application of multivariate statistical techniques in the definition of lithology from geophysical well logs. Brazilian Journal of Geology, 38(1): 61–74. (In Portuguese).

SPADINI A, ESTEVES F, DIAS-BRITO D, AZEVEDO R & RODRIGUES R. 1988. The Macae Formation, Campos Basin, Brazil: its evolution in the context of the initial history of the South Atlantic. Brazilian Journal of Geology, 18(3): 261–272. (In Portuguese).

STEINBACH M, KARYPIS G, KUMAR V. 2000. A comparison of document clustering techniques. In: KDD Workshop on Text Mining, 400, 1, pp. 525–526.

TOURAN N. 2012. A modal expansion equilibrium cycle perturbation method for optimizing high burn up fast reactors. PhD Thesis, University of Michigan, 199 pp.

VATTANI A. 2011. k-means requires exponentially many iterations even in the plane. Discrete and Computational Geometry, 45, 4, pp. 596–616.

VICINI L & SOUZA A. 2005. Multivariate analysis of theory to practice. UFSM, CCNE, Santa Maria - RS, 215 pp. (In Portuguese).

WALT S, COLBERT S & VAROQUAUX G. 2011. The Numpy array: a structure for efficient numerical computation. Computing in Science & Engineering, IEEE, 13(2): 22–30.

WANG D & MURPHY M. 2004. Estimating optimal transformations for multiple regression using the ACE algorithm. Journal of Data Science, Tingmao Publish Company, 2(4): 329–346.

WENDT W, SAKURAI S & NELSON P. 1986. Permeability prediction from well logs using multiple regression. In: LAKE L & CARMIL Jr H, (Eds.). Reservoir characterization, Orlando Academic Press, Orlando, FL, 659 pp.

WILLIAMS E. 1978. I. Regression. In: KRUSKAL W & TANUR J. Ed., Linear Hypotheses, International Encyclopedia of Statistics. Free Press, 1, p. 523–541.

WINLAND H. 1972. Oil accumulation in response to pore size charges, Weyburn field, Saskatchewan. Amoco Production Company Research Department, F72-G-25.

WINTER W, JAHNERT R & FRANÇA A. 2007. Campos Basin. Petrobras Geosciences Bulletin, Rio de Janeiro, 15(2): 511–529. (In Portuguese).

XUE G, DATTA-GUPTA A, VALKO P & BLASINGAME T. 1997. Optimal transformations for multiple regression: application to permeability estimation from well logs. SPE Reservoir Evaluation & Engineering, 12(2): 85–94.

A.G.C.: proposed the work, worked on the interpretation of data, wrote the text of the article and modified the final figures.
P.M.: worked on the master's thesis and on data modeling.

Received on April 6, 2021 / Accepted on December 29, 2021