

ELECTROFACIES CLASSIFICATION OF PONTA GROSSA FORMATION BY MULTI-RESOLUTION GRAPH-BASED CLUSTERING (MRGC) AND SELF-ORGANIZING MAPS (SOM) METHODS

Felipe Vasconcelos dos Passos^{1,2}, Marco Antonio Braga²,
Thiago Gonçalves Carelli³ and Josiane Branco Plantz⁴

ABSTRACT. In Ponta Grossa Formation, devonian interval of Paraná Basin, Brazil, sampling restrictions are frequent, and lithological interpretations from gamma ray logs are common. However, no single log can discriminate lithology unambiguously. An alternative to reduce the uncertainty of these assessments is to perform multivariate analysis of well logs using data clustering methods. In this sense, this study aims to apply two different clustering algorithms, trained with gamma ray, sonic and resistivity logs. Five electrofacies were differentiated and validated by core data. It was found that one of the electrofacies identified by the model was not distinguished by macroscopic descriptions. However, the model developed is sufficiently accurate for lithological predictions.

Keywords: geophysical well logging; lithology prediction; Paraná Basin.

RESUMO. Na Formação Ponta Grossa, intervalo devoniano da Bacia do Paraná, Brasil, restrições de amostragem são frequentes e interpretações litológicas dos registros de raios gama são comuns. No entanto, nenhum perfil geofísico único pode discriminar litologias sem ambiguidade. Uma alternativa para reduzir a incerteza dessas avaliações é executar uma análise multivariada combinando vários perfis geofísicos de poços por meio de métodos de agrupamento de dados. Nesse sentido, este estudo tem como objetivo aplicar dois algoritmos de agrupamento aos registros de raios gama, sônico e resistividade para fins de predição litológica. Cinco eletrofácies foram diferenciadas e validadas por dados de testemunhos. Verificou-se que uma classe identificada pelo modelo não foi identificada por descrições macroscópicas. Porém, o modelo é suficientemente preciso para predições litológicas.

Palavras-chave: geofísica de poços, predição litológica, Bacia do Paraná.

Corresponding author: Felipe Vasconcelos dos Passos

¹PETROBRAS, Centro de Pesquisas e Desenvolvimento Leopoldo Americo Miguez de Mello – CENPES/PDEP/GSE, Rio de Janeiro, RJ, Brazil – E-mail: felipevp@petrobras.com.br

²Universidade Federal do Rio de Janeiro - UFRJ, Programa de Pós-Graduação em Geologia, Rio de Janeiro, RJ, Brazil – E-mails: marcobraga@geologia.ufrj.br

³Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Instituto de Biociências, Departamento de Ciências Naturais, Rio de Janeiro, RJ, Brazil – E-mail: tgcarelli@gmail.com

⁴Universidade Federal do Rio de Janeiro (UFRJ), CCMN, IGEO, Departamento de Geologia, Laboratório de Geologia Sedimentar (Lagesed), Rio de Janeiro, RJ, Brazil – E-mail: josianeplantz@gmail.com

INTRODUCTION

Traditional studies on sedimentology and stratigraphy integrate several sources of information in the development of a depositional model, including surface information (outcrops and surface geophysical measurements) and subsurface information (cuttings, cores and wireline log data).

The cuttings obtained during drilling are the most frequent samples in exploratory wells, but they are also the worst quality samples for lithological analysis. It is difficult to visualize sedimentary structures, as well as performing porosity and permeability measurements. In addition, some samples are lost because of their grain size (silts) or by dissolving in the drilling mud (salts). Moreover, the samples are often contaminated due to wall collapses (Serra & Abbott, 1982).

The cores, on the other hand, are an excellent source of information since their depth can be located more accurately and are large enough for several analysis. However, due to cost and time constraints, it is not always possible to obtain core samples from the whole well and it is usual to acquire only a few meters of core in a limited interest zone. The remaining intervals are usually sampled by cuttings or sidewall cores.

In the Ponta Grossa Formation, Devonian interval of Paraná Basin, Brazil, these constraints on sampled data were found. Despite having a proven thickness of up to 654 m (Assine, 1996), most of the cores recovered are between 5 and 20 meters thick. Thus, the identification of lithology at non-cored intervals has been performed from cuttings aided by the interpretation of the gamma ray log, which is also the most widely used log for well correlations and sequence

stratigraphy approaches in this basin (Assine, 1996; Bergamaschi & Pereira, 2001; Ferreira et al., 2010).

The identification of lithology from well logs is based on the principle that well log responses are related to changes in lithology along its path. However, no single log is able to discriminate lithology unambiguously. Bentonite clays, for example, show radioactivity as low as the sandstone (Dypvik & Eriksenf, 1983).

Therefore, the main objective of this work is the use of information from multiple logs through clustering methods to reduce ambiguity and define lithology more precisely on non-cored intervals of the Ponta Grossa Formation.

Geological settings

The Paraná basin is a large intracratonic basin that extends over an area of about 1,400,000 km², distributed between Brazil, Paraguay, Uruguay and Argentina (Fig. 1). Two-thirds of its surface expression is represented by predominantly basaltic lava flows of Jurassic-Cretaceous age, which can reach a thickness of up to 1700 m and probably had a huge influence on oil generation in the basin (Thomaz Filho et al., 2008). The remaining third of the basin consists of a belt of outcrops of older sedimentary rocks surrounding the lava flows (Zalan et al., 1990).

The strata sampled in this study are from Silurian-Devonian sequence of Paraná basin, named Paraná Group, that is composed, in ascending order, by the Furnas and the Ponta Grossa formations.

The Ponta Grossa Formation, which is the focus of this study, is represented by three members ordered from the bottom to the top: Jaguariaíva (claystone-dominated), Tibagi (sandstone-dominated) and São Domingos (claystone-dominated) (Lange & Petri 1967). These rocks

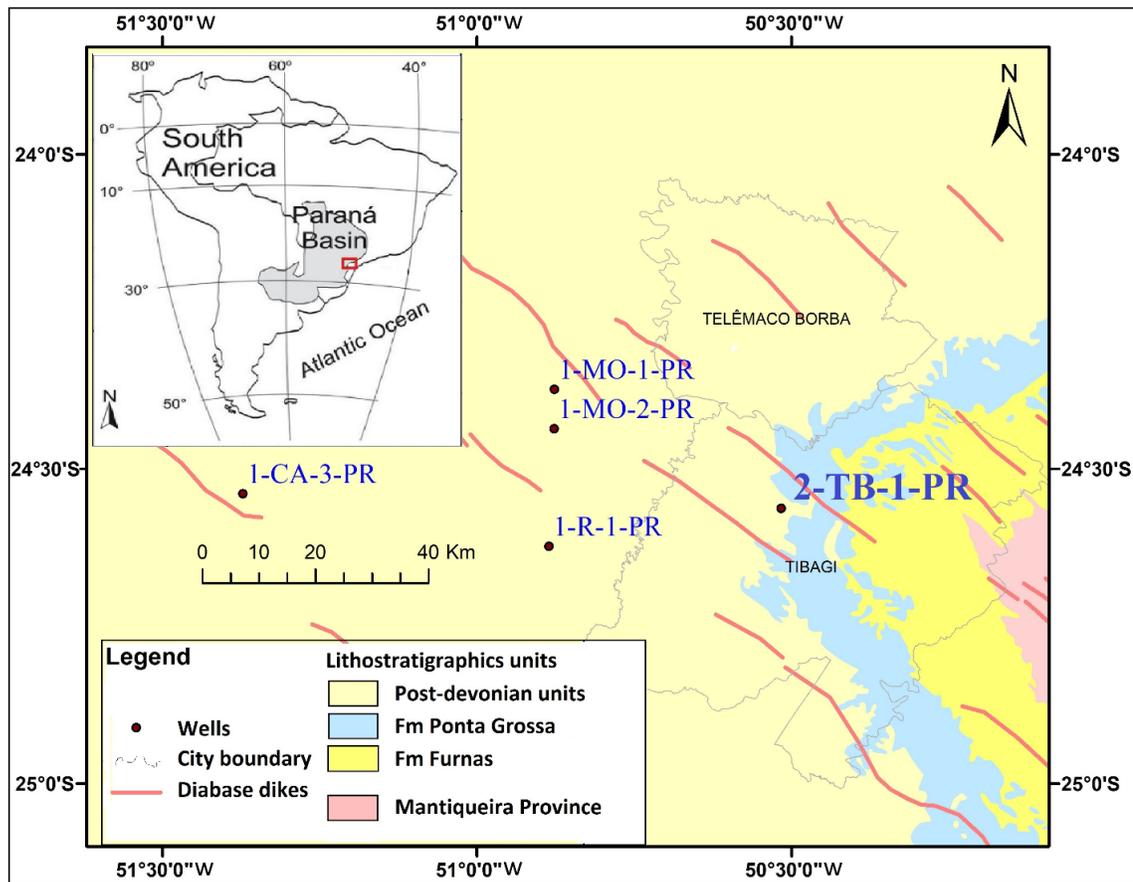


Figure 1 - Geographical location of Paraná basin and the well of study (2-TB-1-PR) in Tibagi-PR, Brazil.

has been widely studied for over a century, especially due to the rich fossil community, and partly owing to its potential as a hydrocarbon source (Zalan et al., 1990).

An important event was the discovery, in 1996, of the first commercial gas accumulation sourced from shales of the Ponta Grossa Formation (Barra Bonita gas field), associated with the Ponta Grossa-Itararé system (Campos et al., 1998).

Other studies attempted to integrate the extensive knowledge acquired over the years with the study of the fossil, sedimentary and gamma ray records in the context of sequence stratigraphy or sedimentary and paleoenvironmental processes, with recently published contributions (Bergamaschi & Pereira, 2001; Ferreira et al., 2010; Carelli

& Borghi, 2011; Grahn et al., 2013; Engelke et al., 2019; Sedorko et al., 2019).

METHODOLOGY

The data used in this study were obtained from well 2-TB-1-PR, located on the eastern edge of the Paraná Basin, in the city of Tibagi, Paraná state, Brazil. This well was completely cored and recovered about 437 meters of well-preserved cores, as well as data from conventional wireline logs, including Gamma Ray (GR), Sonic (DT), Long Normal Resistivity (RLN) and Laterolog Resistivity (RLAT).

The recovered core samples were gamma scanned with hand gamma spectrometer to obtain a log of core radioactivity (core gamma). In addition, the core provides a

lithological reference to calibrate the electrofacies (EF) model and evaluate the accuracy on lithology recognition.

The following steps were applied to the data in order to standardize the vertical resolution and reduce noise before EF modeling:

1. all logs were previously smoothed by a moving average filter, ensuring an approximately common vertical resolution for all analyzed logs. It avoids data incompatibility that generates error and bias (Doveton, 1994);
2. filling of null data points with the average value of the neighborhood points;
3. core depth adjustment: this step involves the correlation between core gamma and GR log;
4. selection of data in the interval from 46.2 meters to 441 meters for data training. Since the data outside this range have poor quality or some log are missing.

The model was built using the Geolog® Facimage software, which is a Geolog tool dedicated to EF modeling, allowing clustering analysis.

Clustering analysis involves the use of clustering algorithms to partitioning a set of log data into electrofacies units, defined as a “set of log responses that characterizes a sediment and permits the sediment to be distinguished from other” (Serra & Abbott, 1982). They can be presented as inferred lithological profiles to provide assistance in performing sequence stratigraphy and correlations (Ye & Rabiller, 2005).

In this study two unsupervised clustering algorithms were applied to predict lithology from well log data: Multi Resolution Graph-Based Clustering (MRGC) and Self-Organizing Maps (SOM). Information on the calculations performed by these methods can be found in Ye & Rabiller (2000) and Kohonen (2013), respectively. The models were trained with GR, DT and RLN logs from well 2-TB-1-PR. The clusters obtained were compared to the core lithology and merged into EF units

to form a one-to-one correspondence. So, the best match between the core lithology and EF model was analyzed within graphs to visualize the distribution of the petrophysical characteristics. A flowchart with the process performed to construct the EF models is illustrated in Figure 2.

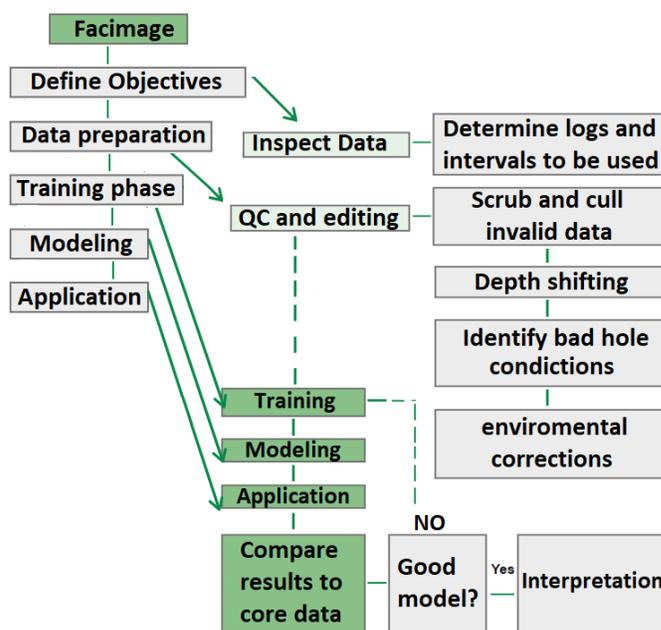


Figure 2 - Method applied to elaborate an electrofacies model on Facimage module of Geolog software.

RESULTS

The MRGC modeling resulted in five optimized models with 7, 9, 11, 13 and 18 optimal clusters. The same cluster partition configurations were used to set the SOM modeling one-dimensional grids (1x7, 1x9, 1x11, 1x13 and 1x18). The choice of one-dimensional grids was made after several experiments that demonstrated better results with this configuration. Moreover, the one-dimensional topological architecture of the SOM is consistent with the geologists' lithofacies knowledge, in which the geologist's expertise is embedded in the structure (Chang et al., 2002).

The clusters generated by the models were merged into five EF units. Four of them can be correlated to one lithology: EF 1 (for sandstones), EF 2 (for siltstones), EF 3

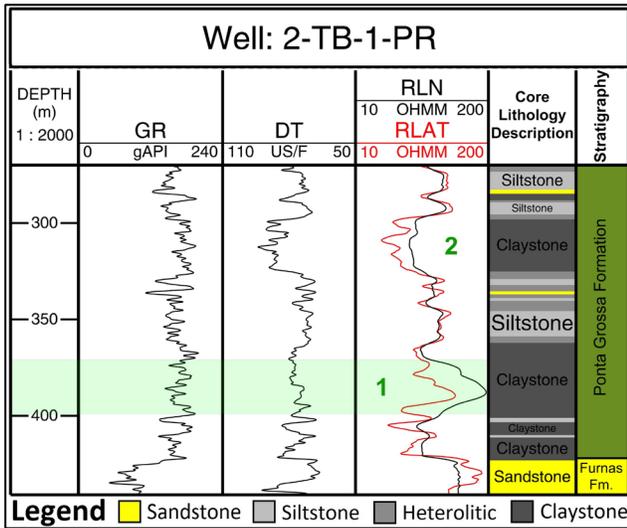


Figure 3 - Depth interval from 371m to 399m (1) in well 2-TB-1-PR. Note the significant deflection to the right of RLN log and the lower transit time (DT) compared to the claystone interval just above (2), possibly due to a decrease in the porosity by compaction or cementation.

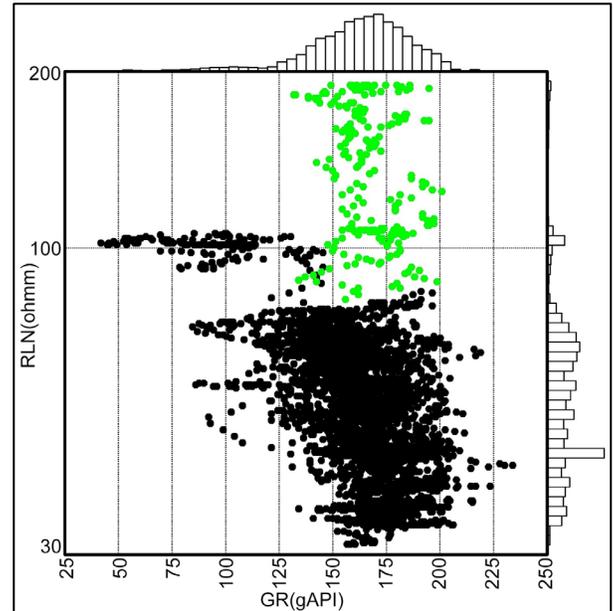


Figure 4 - High resistivity interval observed in Figure 3 highlighted in green on the GR vs RLN crossplot showing that the points on this claystone differ from the overall trend.

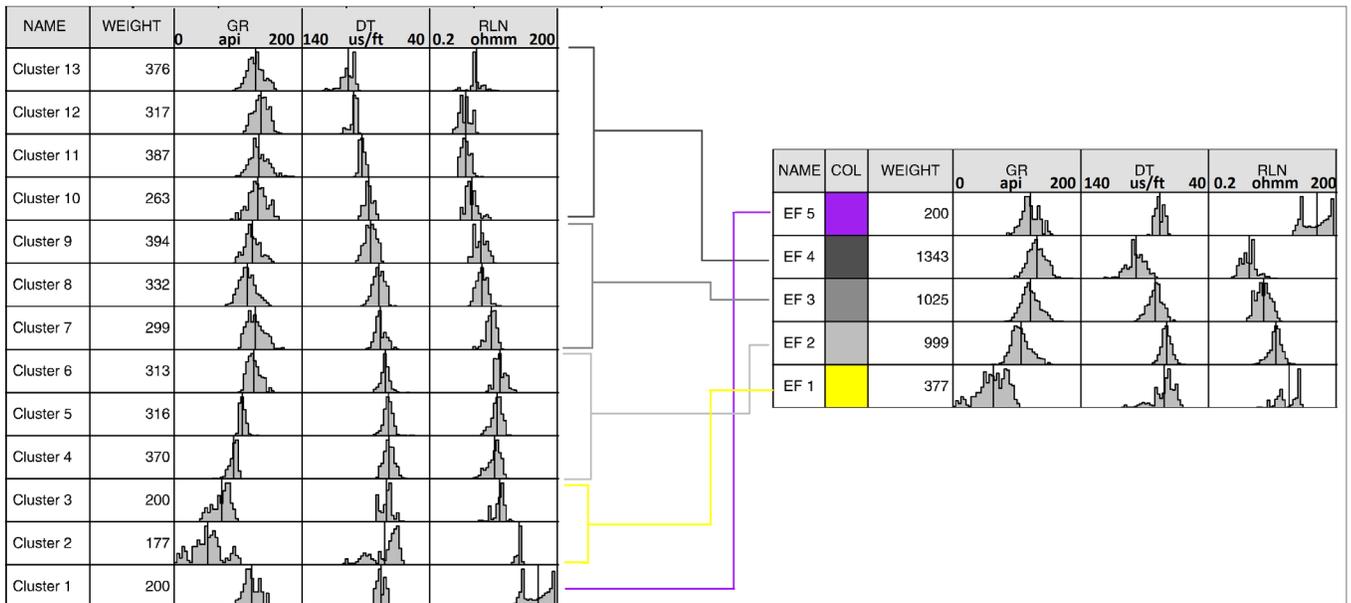


Figure 5 - SOM_13 model and the merging to form the five final electrofacies. The criterion used to merge the clusters was Similarity and correspondence with the same core lithology.

(for heterolithics - claystone intervals intercalated by thin layers of siltstone or sandstone) and EF 4 (for claystones). Additionally, the EF 5 unit was defined by very distinct characteristics from all lithologies found in the well, particularly

its resistivity values (Fig. 4). This group was easily identified by all tested algorithms and corresponds to the range of high resistivity claystones from 373m to 399m (Fig. 3), which was not differentiated from the others by macroscopic description.

Table 1 - Accuracy of MRGC models. EF=Electrofacies; SS=Sandstone; ST=Siltstone; HT=Heterolithics; CL=Claystone and CTL= % of coverage on target lithology on core.

| Model | Cluster names | EF | Recognition rates | | | | CTL |
|---------|---------------|----|-------------------|------------|------------|------------|------------|
| | | | SS | ST | HT | CL | |
| MRGC 7 | 1 | 1 | 100% | 0 | 0 | 0 | 42% |
| | 3 | 2 | 23% | 62% | 13% | 2% | 67% |
| | 4 and 5 | 3 | 1% | 16% | 58% | 24% | 87% |
| | 6 and 7 | 4 | 0% | 0% | 2% | 98% | 78% |
| | 2 | 5 | 0% | 0% | 0% | 100% | None |
| MRGC 9 | 9 | 1 | 100% | 0% | 0% | 0% | 42% |
| | 6 | 2 | 23% | 62% | 13% | 2% | 67% |
| | 4 and 5 | 3 | 1% | 16% | 58% | 24% | 88% |
| | 1,2 and 3 | 4 | 0% | 0% | 2% | 98% | 78% |
| | 7 and 8 | 5 | 0% | 0% | 0% | 100% | None |
| MRGC 11 | 11 | 1 | 100% | 0% | 0% | 0% | 42% |
| | 7 and 8 | 2 | 19% | 62% | 15% | 5% | 90% |
| | 4,5 and 6 | 3 | 0% | 6% | 67% | 27% | 82% |
| | 1,2 and 3 | 4 | 0% | 0% | 2% | 98% | 78% |
| | 9 and 1 | 5 | 0% | 0% | 0% | 100% | None |
| MRGC 13 | 13 | 1 | 100% | 0% | 0% | 0% | 42% |
| | 9 and 1 | 2 | 19% | 62% | 15% | 5% | 90% |
| | 6,7 and 8 | 3 | 0% | 6% | 67% | 27% | 82% |
| | 1,2,3,4 and 5 | 4 | 0% | 0% | 2% | 98% | 78% |
| | 11 and 12 | 5 | 0% | 0% | 0% | 100% | None |
| MRGC 18 | 16,17,18 | 1 | 100% | 0% | 0% | 0% | 52% |
| | 12 and 13 | 2 | 16% | 64% | 15% | 5% | 90% |
| | 8,9,10 and 11 | 3 | 0% | 6% | 67% | 27% | 82% |
| | 1,2,3,4,5,6,7 | 4 | 0% | 0% | 2% | 98% | 78% |
| | 14 and 15 | 5 | 0% | 0% | 0% | 100% | None |

It is important to explain that due to the sensitivity of the resistivity tool to the formation fluid, two or more EF may correspond to one geologic facies, as changes in the formation fluid salinity cause changes in resistivity measurements. However, even considering that there are no changes in the formation fluids, a diagenetic event, such as compaction or cementation, could decrease the rock porosity and, consequently, increase your resistivity.

For each resulting EF, the recognition rates and coverage in the core lithology were evaluated. The former one shows the distribution of all samples classified in the EF unit by lithology at the same intervals and the latter one shows the percentage of lithology described in the well that was recognized by the EF unit.

The evaluation aims to find the models that have the highest values in both indices, but in a balanced way. Since an EF with a high recognition rate that covers little of the target lithology is as bad as an EF that covers all examples of the target lithology throughout the well, but without distinguishing it from the others.

From the analysis of all MRGC models (Table 1) it is clear that they all have excellent recognition rates for sandstones (100%) but with little coverage in core lithology (CTL), from 42% to 52%. In other words, this means that the EF 1, correlated to sandstones, is defined by a very small range of radioactivity (GR), delay time (DT) and long resistivity (RLN) log values that always correspond to sandstones, but that represents only 42% to 52% of all sandstones described in the core.

SOM models (Table 2) have the best balance between recognition rates and core coverage, particularly for 13 clusters. Therefore, the SOM model with 13 clusters (SOM_13) was chosen to compose the final model. Similar clusters were merged (Fig. 5) and the final model was propagated in the well (Fig. 6). Additionally, an overview of the geophysical signature of each EF are shown with the help of a Box and Whisker Plot (Fig. 7).

Table 2 - Accuracy of SOM models. EF=Electrofacies; SS=Sandstone; ST=Siltstone; HT=Heterolithics; CL=Claystone and CTL= % of coverage on target lithology on core.

| Model | Cluster names | EF | Recognition rates | | | | CTL |
|--------|-----------------|----|-------------------|------------|------------|------------|------------|
| | | | SS | ST | HT | CL | |
| SOM 7 | 7 | 1 | 46% | 0% | 0% | 54% | 49% |
| | 5 and 6 | 2 | 16% | 63% | 15% | 6% | 94% |
| | 3 and 4 | 3 | 0% | 4% | 64% | 32% | 70% |
| | 1 and 2 | 4 | 0% | 0% | 10% | 90% | 65% |
| | | 5 | Not clustered | | | | |
| SOM 9 | 9 | 1 | 45% | 0% | 0% | 55% | 45% |
| | 6,7 and 8 | 2 | 17% | 60% | 17% | 6% | 95% |
| | 4 and 5 | 3 | 0% | 4% | 68% | 28% | 63% |
| | 1,2 and 3 | 4 | 0% | 0% | 12% | 88% | 70% |
| | | 5 | Not clustered | | | | |
| SOM 11 | 2 | 1 | 97% | 0% | 0% | 3% | 49% |
| | 3,4 and 5 | 2 | 16% | 63% | 15% | 6% | 95% |
| | 6 and 7 | 3 | 0% | 5% | 72% | 23% | 58% |
| | 8,9,10 and 11 | 4 | 0% | 0% | 16% | 84% | 85% |
| | 1 | 5 | 0% | 0% | 2% | 98% | none |
| SOM 13 | 2 and 3 | 1 | 77% | 21% | 1% | 2% | 72% |
| | 4,5 and 6 | 2 | 10% | 66% | 16% | 7% | 87% |
| | 7,8 and 9 | 3 | 0% | 3% | 69% | 28% | 72% |
| | 10,11,12 and 13 | 4 | 0% | 0% | 9% | 91% | 79% |
| | 1 | 5 | 0% | 0% | 0% | 100% | none |
| SOM 18 | 2,3 and 4 | 1 | 80% | 15% | 0% | 4% | 71% |
| | 5,6,7 and 8 | 2 | 11% | 65% | 17% | 7% | 84% |
| | 9,10,11,12,13 | 3 | 0% | 6% | 65% | 29% | 74% |
| | 14,15,16,17,18 | 4 | 0% | 0% | 8% | 92% | 77% |
| | 1 | 5 | 0% | 0% | 0% | 100% | none |

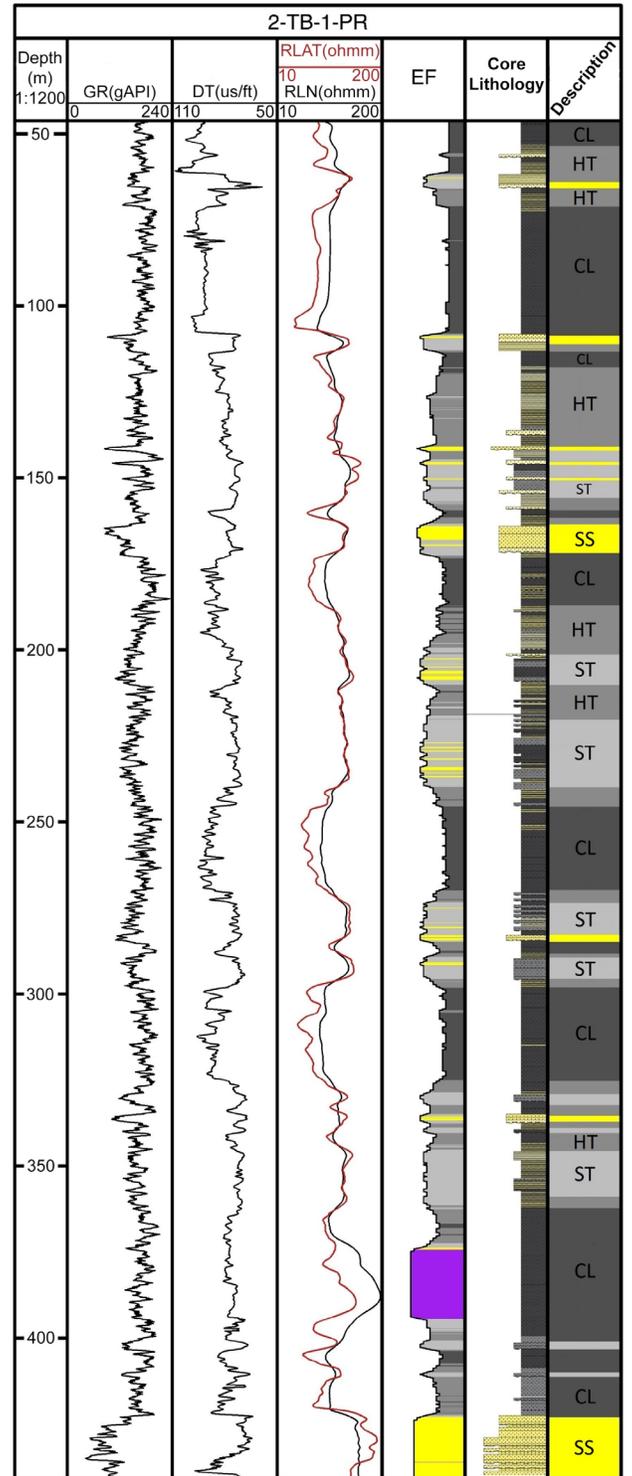


Figure 6 - Electrofacies model propagated to well 2-TB-1-PR. Note the similarity between the EF model and the Core Lithology. (SS=Sandstone, ST=Siltstone, HT=Heterolithics, CL=Claystone).

CONCLUSIONS

In this study, five models from Multi-Resolution Graph-based clustering analysis and five models from Self-organizing maps analysis were constructed. Gamma ray (GR), sonic (DT) and long resistivity (RLN) logs were the input data and core simplified lithology are associated data used to estimate the accuracy and validate the model.

The method used in this study was applied to fill the current lack of an electrofacies model for the Ponta Grossa Formation. This model proved to have good accuracy and can be used for lithological prediction in adjacent wells at least for data in the same range of study.

The main conclusions and usefulness of the results obtained in this study as:

1. Integrating conventional log information with cluster analysis improves facies classification by grouping common signatures that refer to the same lithology. This helps to standardize the lithological profile of different wells in order to correlate these lithofacies in geological sections.
2. The model allows refinement of lithological classification by providing additional details on petrophysical characteristics that are not recognizable in the macro descriptions by geologists, such as the high resistivity claystone range detected in this study.
3. Considering a succession of sandstone, siltstone and claystone, the lithofacies based on conventional log data gives satisfactory results when compared to core observed lithofacies.
4. Identification of unexpected values in data analysis before performing clustering prevents from contaminating the signature of the main electrofacies, reducing uncertainty in the prediction of lithologies. These values should be excluded from training data or clustered into electrofacies units that could be related to diagenetic or formation fluid changes. These hypotheses should be investigated with additional laboratory analysis.

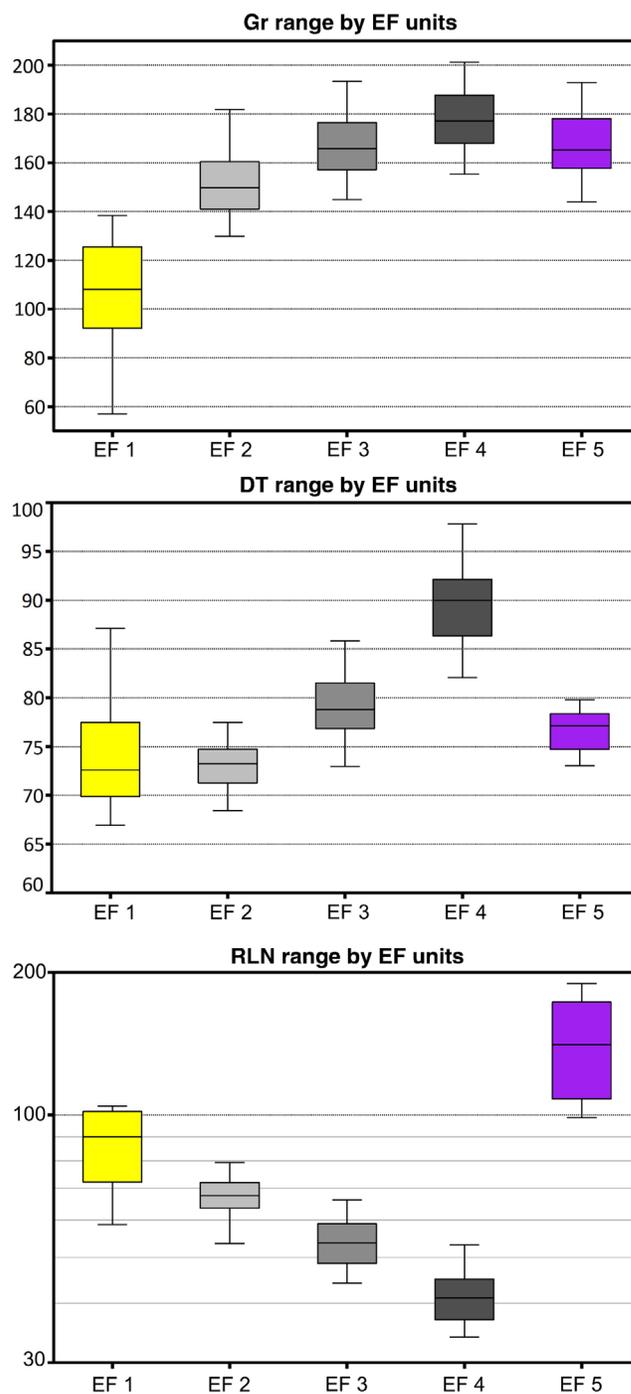


Figure 7 - Overview of the geophysical signature of Electrofacies. Two main points to note are: (i) the increasing tendency of radioactivity and transit time and the decreasing tendency of resistivity as the rocks become more shaliness (from EF 1 to EF 5); and (ii) the distinction of EF 5 from the others mainly due to high resistivity.

5. Clustering methods provide a lot of information in a short time and at a low cost, just by using conventional log information that is widely acquired in oil and gas exploration projects. A simplified model based only on the largest petrophysical contrasts can be used as a guide for visualizing the subsurface stratigraphic framework. However, it can also be refined and calibrated with core data, increasing geological meaning to each electrofacies.

ACKNOWLEDGEMENTS

The authors are grateful to anonymous reviewers who considerably improved the quality of manuscript.

REFERENCES

- ASSINE ML. 1996. Aspectos da estratigrafia das seqüências pré-carboníferas da Bacia do Paraná no Brasil. PhD Thesis on Geology - Programa de Pós-Graduação em Geologia Sedimentar, Universidade de São Paulo, Brazil. 1996. 207 pp.
- BERGAMASCHI S & PEREIRA E. 2001. Caracterização de Seqüências deposicionais de 3ª ordem para o Siluro-Devoniano na Sub-bacia de Apucarana, Bacia do Paraná, Brasil. *Ciência Técnica do Petróleo*, 20: 63–72.
- CAMPOS L DE, MILANI E, TOLEDO M, QUEIROZ R, CATTO A & SELKE S. 1998. Barra Bonita: a primeira acumulação comercial de hidrocarboneto da Bacia do Paraná. In: Rio Oil & Gas Conference. Rio de Janeiro. Brazil. pp 5-8.
- CARELLI T & BORGHI L. 2011. Caracterização de Microfácies Sedimentares em Folhelhos da Formação Ponta Grossa (Devoniano) na Borda Leste da Bacia do Paraná. *Anuário do Instituto de Geociências - UFRJ*, 34: 84–104.
- CHANG HC, KOPASKA-MERKEL DC & CHEN HC. 2002. Identification of lithofacies using Kohonen self-organizing maps. *Computers and Geosciences*, 28(2): 223–229.
- DOVETON JH. 1994. *Geologic Log Analysis Using Computer Methods*. Tulsa, Okla.: American Association of Petroleum Geologists. AAPG Computer Applications in Geology, no. 2, 169 pp. 169 pp.
- DYPVIK H & ERIKSENF D. 1983. Natural radioactivity of clastic sediments and the contributions of U, Th and K. *Journal of Petroleum Geology*, 5(4): 409–416.
- ENGELKE V, RODRIGUES DE VARGAS M, SANTOS DA SILVEIRA A, ROBERTO MOURA DE CARVALHO P, AGNESA TRENTIN F & EDUARDO FACCION J. 2019. Stochastic modeling from Ponta Grossa Formation: Integrating outcropping and subsurface data. In: 4th EAGE Conference on Petroleum Geostatistics. Florence, Italy, 2019.
- FERREIRA FJF, CANDIDO AG & ROSTIROLLA SP. 2010. Correlação gamaespectrométrica de afloramentos e poços: estudo de caso na Formação Ponta Grossa (Bacia do Paraná, Brasil). *Revista Brasileira de Geofísica*, 28(3): 371–396.
- GRAHN Y, MAULLER PM, BERGAMASCHI S & BOSETTI EP. 2013. Palynology and sequence stratigraphy of three Devonian rock units in the Apucarana Sub-basin (Paraná Basin, south Brazil): Additional data and correlation. *Review of Palaeobotany and Palynology*, 198: 27–44.
- KOHONEN T. 2013. Essentials of the self-organizing map. *Neural Networks*, 37: 52–65.
- LANGE FW & PETRI S. 1967. The Devonian of the Paraná

Basin. *Boletim Paranaense de Geociências*, 21(22): 5–55.

SEDORKO D, BOSETTI EP, GHILARDI RP, MYSZYNSKI JÚNIOR LJ, SILVA RC & SCHEFFLER SM. 2019. Paleoenvironments of a regressive Devonian section from Paraná Basin (Mato Grosso do Sul state) by integration of ichnologic, taphonomic and sedimentologic analyses. *Brazilian Journal of Geology*, 48(4): 805–820.

SERRA O & ABBOTT HT. 1982. The Contribution of Logging Data to Sedimentology and Stratigraphy. *Society of Petroleum Engineers Journal*, 22(01): 117–131.

THOMAZ FILHO A, MIZUSAKI AMP & ANTONIOLI L. 2008. Magmatismo nas bacias sedimentares brasileiras e sua influência na geologia do petróleo. *Revista Brasileira de*

Geociências, 38(2): 128–137.

YE S & RABILLER P. 2000. A New Tool for Electro-Facies Analysis: Multi-Resolution Graph-Based Clustering. In: SPWLA 41st Annual Logging Symposium. Dallas, Texas, p. 1–14.

YE S & RABILLER P. 2005. Automated Electrofacies Ordering. *Petrophysics*, 46(6): 409–423.

ZALAN PV, WOLFF S, ASTOLFI MAM, VIEIRA IS, CONCEIÇÃO JCJ, APPI VT, SANTOS NETO EV, CERQUEIRA JR & MARQUES A. 1990. The Paraná Basin, Brazil. In: LEIGHTON MW, KOLATA DR, OLTZ DF & EIDEL JJ (Eds.). *Interior Cratonic Basins (AAPG Memoir 51)* AAPG, Tulsa, p. 681–708.

Recebido em 31 de março de 2020 / Aceito em 9 de junho de 2020
Received on March 31, 2020 / Accepted on June 9, 2020