10º Congreso de Exploración y Desarrollo de Hidrocarburos
Simposio de Evaluación de Formaciones: Maximizando la eficiencia a través de la innovación

SPECTRAL GAMMA-RAY LOGGING OF CERRO BALLENA ANTICLINE, SANTA CRUZ PROVINCE: DATA-DRIVEN MODEL FOR FACIES CLASSIFICATION, WITH IMPLICATION FOR SUBSURFACE OILFIELD DEVELOPMENT

Francisco Castillo Gamarra[1], Marina A. Lema[1], José Matildo Paredes[2]

1: Sinopec Argentina Exploration and Production, Inc., francisco_castillo@sinopecarg.com.ar, marina_lema@sinopecarg.com.ar

2: Universidad Nacional de la Patagonia San Juan Bosco, paredesjose@yahoo.com

Keywords: **facies, spectral gamma-ray log, machine learning, classification, data driven, R, h$_2$o.ai**

RESUMEN

Perfil de rayos gamma espectral del anticlinal del Cerro Ballena, provincia de Santa Cruz: modelo Data-driven para clasificación de facies, con implicancias en el desarrollo de yacimientos petroleros. La respuesta del perfil de rayos gamma espectral en formaciones productivas aporta información clave para el análisis de subsuelo y la caracterización de reservorios. En la Formación Bajo Barreal de la cuenca del Golfo San Jorge, la aplicación de métodos comúnmente usados para clasificar litofacies en base a este perfil se ve dificultada por la complejidad mineralógica y la presencia común de componentes piroclásticos.

Este trabajo describe la construcción de un modelo experimental para clasificación de litofacies en base al perfil de rayos gamma espectral para afloramientos del Cerro Ballena, provincia de Santa Cruz. El proceso de modelado matemático se enmarcó en el contexto geológico durante todas las etapas. El modelo de clasificación de facies (litofacies) fue generado mediante un detallado estudio sedimentológico realizado previamente, y comprende categorías de depósitos clásticos gruesos, así como diferentes tipos de depósitos de planicie de inundación. El perfil de rayos gamma registrado para una sección de 380 m, con una separación vertical de 50 cm entre muestras, constituye la variable de entrada para el modelo generado. La aplicación de análisis de varianza simple (ANOVA) determinó que existe diferencia estadísticamente significativa entre las categorías de facies respecto a las mediciones de rayos gamma espectral. Esta instancia validó la factibilidad de desarrollar un algoritmo con Machine Learning para transformar la variable de entrada (mediciones de rayos gamma espectral) en facies. El objetivo del estudio fue generar un modelo empírico capaz de predecir litologías en nuevos sets de datos, evaluando su consistencia con los datos disponibles. Para lograrlo, durante el proceso de modelado evitó el sobreajuste mediante una cuidadosa combinación de bootsraping, validación cruzada y retención. Para maximizar la precisión del modelo, se empleó la métrica log loss para penalizar falsas clasificaciones. El modelo generado replica con efectividad las categorías determinadas mediante criterios sedimentológicos en el estudio de afloramiento.

Los perfiles de rayos gamma de las secciones aflorantes fueron correlacionados con el subsuelo con alta certidumbre durante el estudio de superficie, validando el carácter de análogos de los afloramientos para los yacimientos cercanos. Los resultados de este estudio habilitan expectativas respecto de la construcción de modelos de facies para el subsuelo en base a mediciones de rayos gamma.

INTRODUCTION

Outcrop analog models provide key data for subsurface reservoir characterization, including quantitative information to reconstruct 3D reservoir architecture and to reduce uncertainty regarding spatial distribution of critical parameters (Szerbiak *et al.*, 2001; Miall, 2014; Hamon *et al.*, 2015). Measurement processes for physical attributes of reservoirs using borehole logging tools result in high disparity of sampling density in different dimensions. Spatial distribution of reservoir properties can be assessed with lower uncertainty using information provided by this type of models. However, feasibility to apply outcrop parameters to subsurface and the method to accomplish this, depend largely on the type of variable and the available tools to assess it. The use of ab initio or first principle models, which describe the physical behavior of the system from basic physics laws is not always possible, due to complex and unknown relationships among variables (Paredes *et al.*, 2016). Empirical or data-driven modeling (DDM) involves mathematical equations derived not from the physical complex process of the phenomena that needs to be modeled, but from the analysis of the data that characterizes the system under study (Solomatine and Ostfeld, 2008). Therefore, a model can be defined finding connections between the system state variables (input and output variables) without the explicit knowledge of the physical behavior.

The Bajo Barreal Formation records a compositionally complex fluvial succession (Limarino and Giordano, 2016; Paredes *et al.*, 2016a; Limarino *et al.*, 2017; Paredes *et al.* this congress), including mineralogically immature siliciclastic sandstones, mudstones and fine-grained conglomerates. Compositional complexity limits the possibilities to define, based on borehole logging data, facies models that reflect a complete depiction of relevant sedimentologic categories, since log data do not allow a consistent definition of classes beyond simplified models (Ferreira *et al.*, 2011; Paredes *et al.*, 2016b). Usually, models with two or three categories are applied (e.g. sandstone-mudstone or good quality sandstone-low quality sandstone-mudstone/non-reservoir), defined according to reservoir quality criteria (Rodríguez *et al.*, 2011). Frequently, electrofacies models are calibrated with core data (Plazibat *et al.*, 2013; Schiuma *et al.*, 2013). These models are similar to borehole logs regarding sampling density disparities, but at the same time attain incorporation of sedimentological criteria in the definition of categories. Application of outcrop facies models to the subsurface in the Golfo San Jorge basin deals with scarcity of complete data sets that allow to reflect the wide range of existent sedimentary deposits (Ferreira *et al.*, 2011, 2013; Acosta, 2013).

Spectral Gamma Ray (SGR) logging of Cerro Ballena outcrops, carried out together with a detailed fluvial architecture and sedimentary facies study (Paredes *et al.*, 2017; Paredes *et al.*, this congress) provided a suitable dataset to assess the feasibility of an analog facies model for nearby oilfields. The outcrop model comprises seven lithofacies: LITH1: Conglomerates, LITH2: Sandstones, LITH3: Siliciclastic siltstones, LITH4: Siliciclastic mudstones, LITH5: Fine-grained tuffs, LITH 6: Tuffaceous siltstones, and LITH 7: Coarse-grained lahar (see Paredes *et al*., this congress). Based on paleoenvironmental criteria, five facies associations were generated in the same study: FA1: fluvial channels; FA2: proximal floodplain; FA3 distal, epiclastic floodplain; FA4 distal, volcaniclastic floodplain, and FA5: coarse-grained lahar. Facies classes were defined based on lithology, grain size, sedimentary structures, geometry and bounding surfaces (Paredes *et al*., 2017). Noticeably, a number of these parameters are feasible to be measured in subsurface, whereas others, such as geometry and bounding surfaces, cannot be completely characterized on borehole log data. Nonetheless, sedimentary facies stacking patterns provide critical information to understand the depositional history and to infer controls on the sedimentation, key aspects to increase predictability and to reduce uncertainty during oilfield development strategies (Miall, 2014).

In this contribution, we describe the construction of a data-driven model based on SGR data, with the goal to replicate facies classification of outcrop model for Cerro Ballena anticline. The use of Supervised Machine Learning (SML), particularly Distributed Random Forest (DRF) and Gradient Boosting Machines (GBM) algorithms (h2o.ai-Arno Candel et. al; h2o.ai-DRF; h2o.ai-GBM), aims at finding a tuned algorithm that defines a function which maps the input (SGR log measurements) in the target output (facies). First stage of model building consists of the analysis and exploration of input data. The learning algorithm produces an inferred function to make predictions about the output target (facies). Machine Learning algorithm is able to provide targets for any new input after sufficient calibration (training). The learning process involves a comparison of the model output with the correct intended target and finding errors in order to modify the model accordingly, maximizing the accuracy but avoiding overfitting (Hastie *et al*., 2009; James *et al*., 2013). As the final goal is to obtain a model with an adequate capability of prediction and generalization in new datasets, consistency checking is imperative. This is accomplished using the available input dataset with a very controlled training or algorithms calibration (Daumé III, 2017).

METHODOLOGY

Data exploration

Initial exploration of input data included descriptive statistics, multidimensional plots, box plots and correlation analysis among variables, aiming at visualizing clear trends or patterns, delineating possible models to build. Next, analysis of SGR log signal significance for the three components in the discrimination of facies was carried out, to define feasibility of facies discrimination based on SGR log. The following R software packages were used in this stage: car, data table, corrplot, dplyr, ggplot2 and MASS. All the software is GPLv3 (General Public License version three).

Predictor variables are K (measurements of potassium in units of %), U (measurements of uranium in units of ppm) and Th (measurements of thorium in units of ppm). Normalization of each log (A, B and C) was performed to consider only the effect of the signals and not the scale effect. The objective is to improve predictive accuracy, avoiding impact of a particular input variable on the prediction due to its large numeric value range. The selection of minimum-maximum normalization transforms each value of the K, Th and U for each well log A.B, C into a [0, 1] range. The target output or response variable is facies, either Facies Association or Lithology, derived from previous sedimentary study of the outcrops (see Paredes *et al.*, this congress). The output variable (facies categories) are shown in Fig 1.

Class	Facies category name	Type
Facies Association	1	Fluvial channel
Facies Association	2	Proximal floodplain
Facies Association	3	Distal epiclastic floodplain
Facies Association	4	Distal volcaniclastic floodplain
Facies Association	5	Coarse-grained lahar
Lithology	LITH1	Conglomerates
Lithology	LITH2	Sandstones
Lithology	LITH3	Siltstones
Lithology	LITH4	Mudstones
Lithology	LITH5	Fine-grained tuffs
Lithology	LITH6	Tuffaceous siltstones
Lithology	LITH7	Coarse-grained lahar

Figure 1. Defined names of the target variable (facies).

Parametrical one way ANOVA was performed, to find out whether mean values of the distribution of each signal, K, Th and U, show significant differences for each type of facies (Lithology or Facies Association). In this case the null hypothesis is that that there is no significant difference in the groups of Lithology respect the signal of K, Th, U of the spectral gamma-ray log. A very low value of the p-value (Pr) is enough to reject the null hypothesis, but it is necessary to check if the residuals of the ANOVA models are normally distributed and the assumptions on ANOVA are not potentially violated (Hastie *et al.*, 2009). If ANOVA test is significant, then there is statistically significant difference between the means of some facies categories, but this procedure does not inform which pairs show significant difference. Therefore, Tukey HSD (Tukey Honest Significant Differences) test for multiple pairwise-comparison between the means of groups was carried out.

The detection of important deviations from normally distributed ANOVA residuals enables the use of Kruskal-Wallis test, like a non-parametrical alternative to one-way ANOVA (Hastie *et al.*, 2009). As in the previous case, if Kruskal-Wallis test is significant, it indicates that some of the group means differ significantly, even though the differing pairings are not defined. Multiple pairwise-comparison can be performed through Dunn test, to determine if the mean difference between specific pairs of group is statistically significant. This procedure was applied for U log.

Machine Learning

The problem to solve is a multiclass classification problem. SML (Supervised Machine Learning) implies a process to find an algorithm which learns from data, supervising the learning process and testing the performance in unseen data (extracted from the original data in a random process). To secure internal consistency of data, a true learning process must be guaranteed, avoiding overfitting (Hastie *et al.*, 2009). This is accomplished through calibration and training of algorithms.

In this case-study the sequential process for training of algorithms and metrics was:

a) Initial retention of the data. Original data was split in a random way to generate two sets: training data (i.e. 70 %-80% of the data) and test data (i.e. 20 %-30% of the data). Test data are never used in the calibration or adjustment of the algorithms, it is genuine unseen data for the algorithms.

b) With training data set, calibration of the algorithms involved the tuning of the parameters of each machine learning algorithm. A new random splitting of training data was carried out, generating a new training data set and test data set to check the performance of the training process. Ten-folds cross validation over the new training data was applied. The metric selected to evaluate the performance of the classification model was Log Loss metric, which penalizes false classification. In the case of very imbalanced data, stratified cross validation could be used, changing the metric for performance evaluation or using a sequential SMOTE (Chawla *et al.*, 2002) for an initial balancing of the training data.

c) Selection of the model with best performance over the test data retained in the training process. Associated error is called “training error”. Prediction over the initial retained data and performance checking on the test data (retained from the beginning), with an associated error called “model error”. If the value of model error is acceptable and similar to training error, then it is possible to assume that that the algorithm was calibrated. If not, a new step b) is performed with extended grid or stacked SML algorithms are applied.

The reported statistics for the model calibration are: (1) Total Accuracy: total percent correct classified, (2) CI (interval of confidence) for Total Accuracy, (3) Balanced Accuracy per class (Brodersen *et al.*, 2010), average between the total true positive and negative detected, (4) Cohen’s Kappa statistic (Landis and Koch, 1977), which shows how better is performance over the performance of randomly guessing according to the frequency of each class. This last one is important in imbalanced data of multiclass problems.

Software for Machine Learning

With the use of h2o.ai (software) and R (software) it is possible to implement deep learning (neural networks), distributed random forest(DRF) and gradient boosting machines(GBM) in a simple way using the pre-built libraries. The h2o.ai (software) is an open source license Apache 2.0.

In this case, the use of deep learning in a small dataset requires a neural network with very limited flexibility (degree of freedom) limiting and making more difficult the training process in order to avoid the overfitting. Using h2o.ai (software) the training process involves the use of 2 ensemble machine learning techniques (Daumé III, 2017): Distributed Random Forest (DRF) and Gradient Boosting Machines (GBM). For detailed information refer to h2o.ai (software).

RESULTS

Data Exploration

In Figs. 2 and 3, a summary of basic statistics for predictor and target variables is shown.

K	U	Th	Facies Association
Min. :0.0000	Min. :0.00000	Min. :0.0000	1:266
1st Qu.:0.3667	1st Qu.:0.07407	1st Qu.:0.3110	2:149
Median :0.4800	Median :0.13068	Median :0.4207	3:613
Mean :0.4717	Mean :0.16280	Mean :0.4237	4:180
3rd Qu.:0.5667	3rd Qu.:0.21296	3rd Qu.:0.5366	5: 11
Max. :1.0000	Max. :1.00000	Max. :1.0000	NA

Figure 2. Summary of the data with the target variable Facies Association.

K	U	Th	Lithology
Min. :0.0000	Min. :0.00000	Min. :0.0000	LITH1: 21
1st Qu.:0.3667	1st Qu.:0.07407	1st Qu.:0.3110	LITH2:326
Median :0.4800	Median :0.13068	Median :0.4207	LITH3: 51
Mean :0.4717	Mean :0.16280	Mean :0.4237	LITH4:388
3rd Qu.:0.5667	3rd Qu.:0.21296	3rd Qu.:0.5366	LITH5: 50
Max. :1.0000	Max. :1.00000	Max. :1.0000	LITH6:364
NA	NA	NA	LITH7: 19

Figure 3. Summary of the data with the target variable Lithology.

The Fig. 4 shows the correlation matrix and distribution of the predictor variables. Visually, K and Th distributions are normal, whereas U distribution seems to deviate from normal. There is some degree of correlation between K, Th and U, but it is not enough to call for use of MANOVA instead of ANOVA one-way.

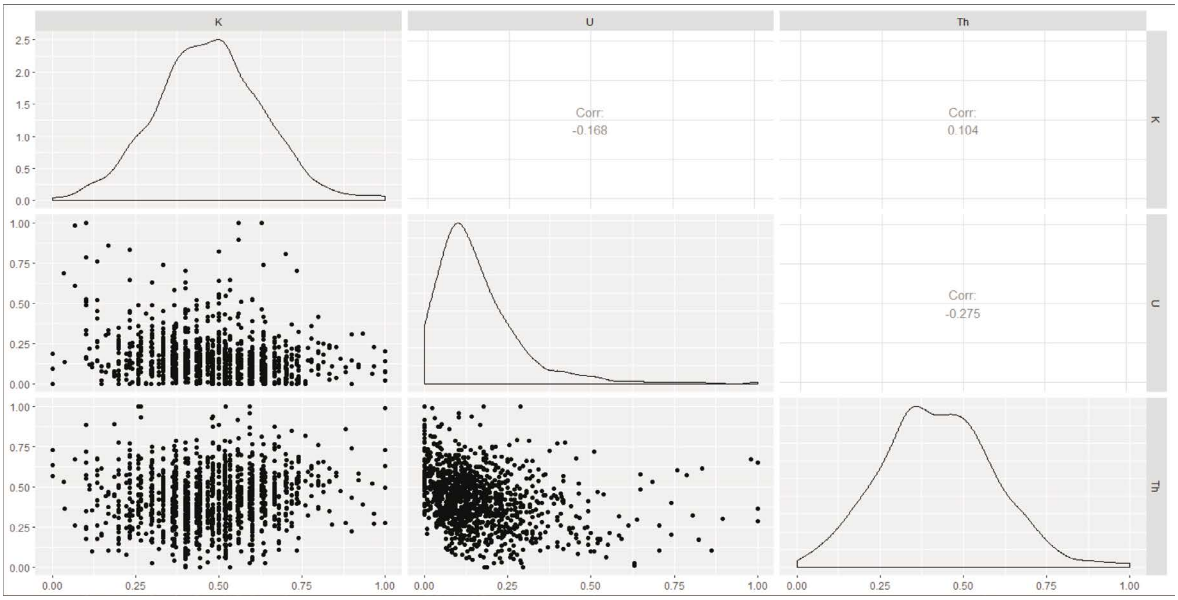

Figure 4. Correlation and distribution for the predictor variables.

The Figs. 5 and 6 show multidimensional plots that can be examined in search of trends or patterns, which are very difficult to define. A very complex relation between input and output variables is evidenced.

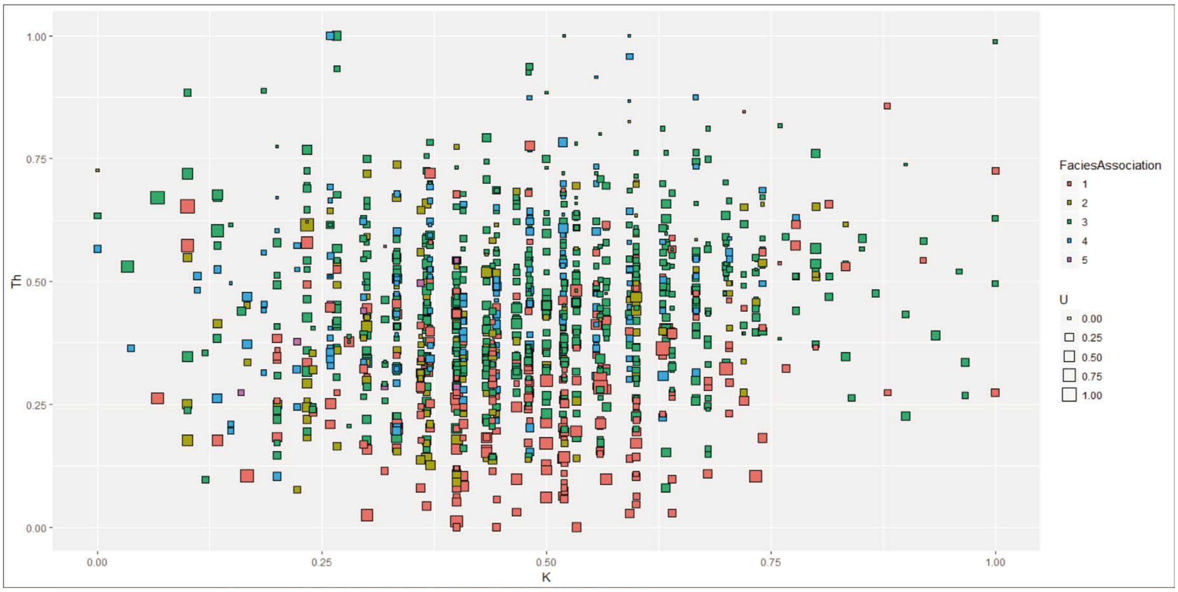

Figure 5. Multidimensional plot for the Facies Association.

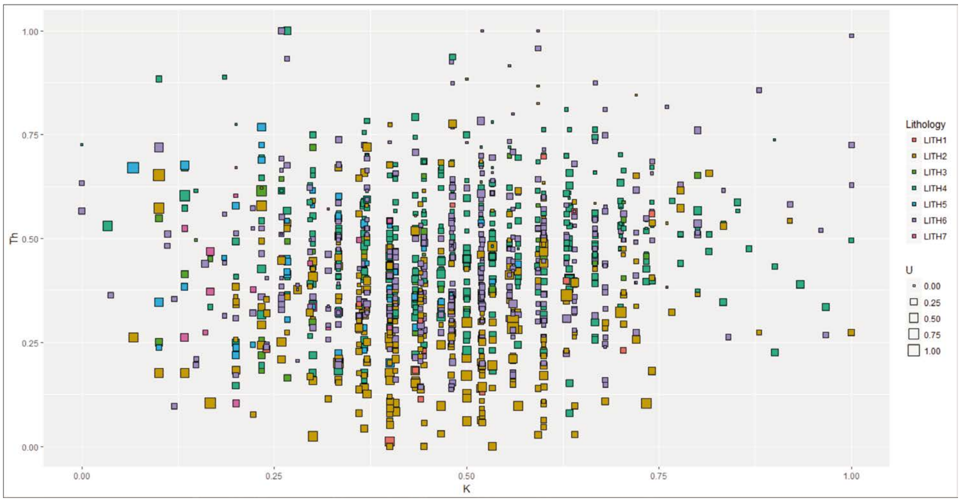

Figure 6. Multidimensional plot for the Lithology.

ANOVA one-way tests for Lithologies

For this set of tests, the null hypothesis is that there is no significant difference in the groups of Lithology respect the signal of K, Th or U from SGR log.

Results for ANOVA one-way for K~Lithology are shown in Figs. 7 and 8. The very low value of Pr is enough to reject the null hypothesis. Fig. 4 shows the plausibility of significant differences between the categories of Lithology respect the signal of K from spectral gamma-ray logs, while Fig. 8 shows that the residual distribution of the ANOVA one way is normal with no systematic pattern and is plausible the homogeneity of variance assumption.

```
Analysis of Variance Table

Response: K
            Df  Sum Sq Mean Sq F value    Pr(>F)
Lithology    6  2.5443 0.42404  17.285 < 2.2e-16 ***
Residuals 1212 29.7332 0.02453
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 7. ANOVA one way of the K signal for classes of Lithology.

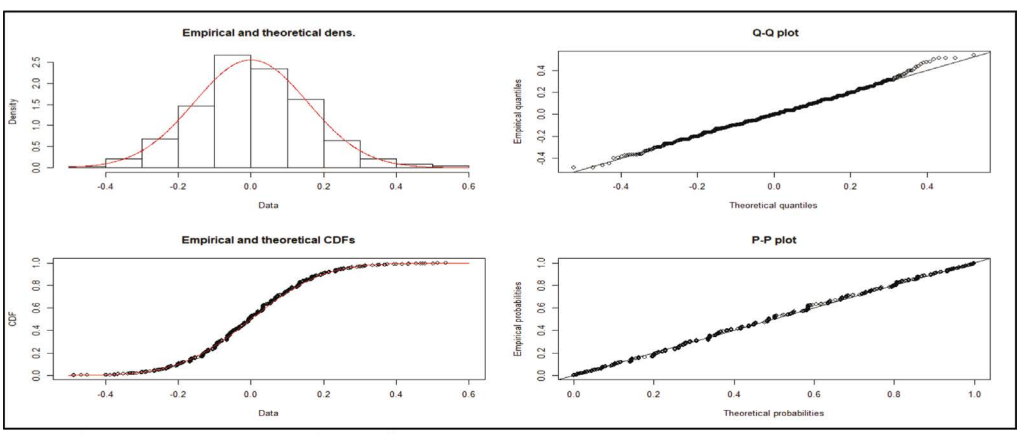

Figure 8. Residuals analysis for ANOVA one-way K ~ Lithology.

As the ANOVA test is significant, it is plausible to compute Tukey HSD (Tukey Honest Significant Differences), to perform multiple pairwise-comparisons between the means of groups of Lithology. This post-hoc test is for discriminate which groups of Lithology are different respects the signal of K of the spectral gamma-ray log. There are 21 possible combinations, between pairs of categories. The Fig. 9 shows the only 11 combinations that are significantly different respect to K signal.

Pair discrimination	p adj
LITH5-LITH2	0
LITH5-LITH4	0
LITH7-LITH4	0
LITH6-LITH5	0
LITH7-LITH6	0
LITH7-LITH2	0
LITH7-LITH1	0.0001
LITH5-LITH1	0.0006
LITH7-LITH3	0.0015
LITH5-LITH3	0.0124
LITH4-LITH3	0.0250

Figure 9. Pairwise-comparisons (only significant different) between the means of classes of Lithology respect to K signal.

Results for ANOVA one-way for Th~Lithology are shown in Figs. 10 and 11. The very low value of Pr is enough to reject the null hypothesis. The Fig. 10 shows the plausibility of significant differences between the categories of Lithology respect to the signal of Th from spectral gamma-ray log. The Fig. 11 shows that the residual distribution of the ANOVA one way is normal with no systematic pattern and is plausible the homogeneity of variance assumption.

```
Analysis of Variance Table

Response: Th
            Df  Sum Sq Mean Sq F value    Pr(>F)
Lithology    6  5.3695 0.89491  35.382 < 2.2e-16 ***
Residuals 1212 30.6549 0.02529
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 10. ANOVA one way for Th~ Lithology.

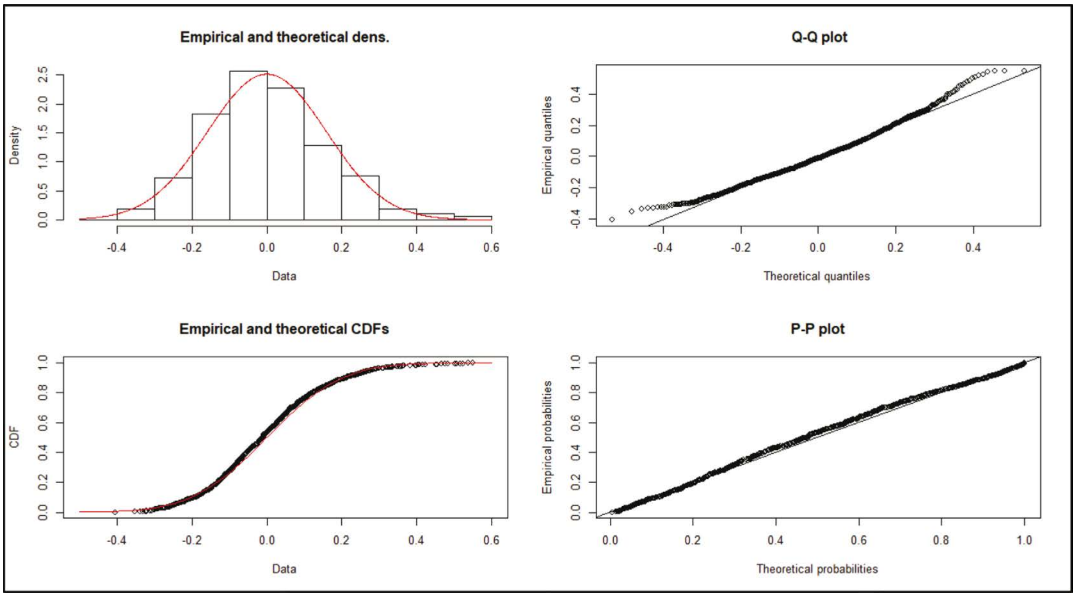

Figure 11. Residuals analysis for ANOVA one-way Th ~ Lithology.

Fig 12 shows that only 7 possible combinations are significantly different respect to Th signal.

Pair discrimination	p adj
LITH3-LITH2	0
LITH4-LITH2	0
LITH6-LITH2	0
LITH4-LITH1	0.0000397
LITH5-LITH2	0.0000508
LITH6-LITH1	0.0024886
LITH3-LITH1	0.0038981

Figure 12. Pairwise-comparisons significantly different between the means of groups of Lithology respect to Th signal.

Results for ANOVA one-way for U~Lithology are shown in Figs. 13 and 14. The very low value of Pr is enough to reject the null hypothesis. Fig. 13 shows the plausibility of significant differences between the categories of Lithology respect the signal of U from spectral gamma log. Residuals analysis for ANOVA one-way U~Lithology highlights the need to check if hypotheses of parametrical ANOVA are plausible.

```
Analysis of Variance Table

Response: U
            Df Sum Sq Mean Sq F value    Pr(>F)
Lithology    6  2.590 0.43167  25.294 < 2.2e-16 ***
Residuals 1212 20.684 0.01707
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 13. ANOVA one way for U~ Lithology.

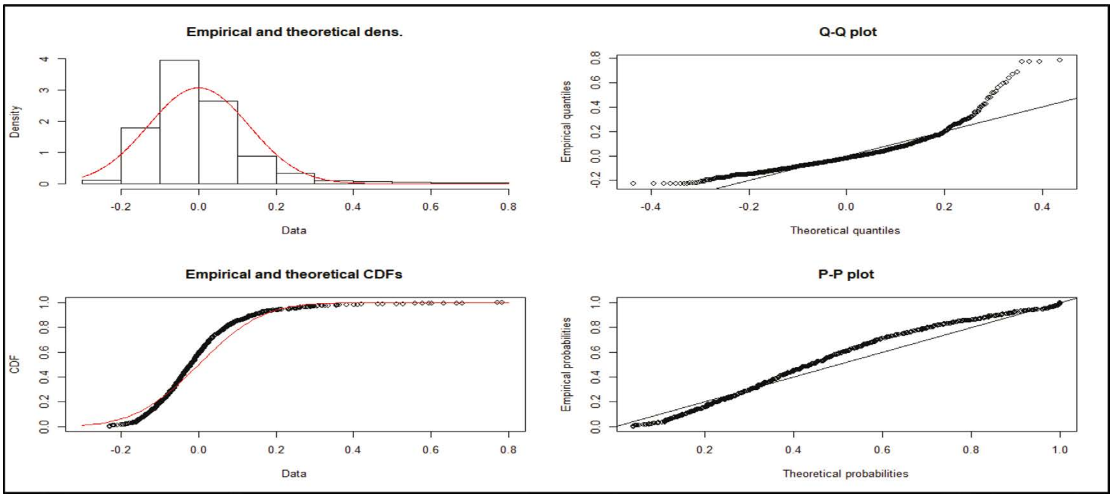

Figure 14. Residuals analysis for ANOVA one-way U ~ Lithology.

Fig. 14 shows that the residual distribution for ANOVA one-way U ~ Lithology has some deviation of the normal, therefore it is necessary to run a non-parametrical ANOVA, for example Kruskal-Wallis test and avoid a false positive. The test (Fig. 15) shows that it is plausible to reject the null hypothesis with no assumption of homogeneity of variance.

```
        Kruskal-Wallis rank sum test

data:  U by Lithology
Kruskal-Wallis chi-squared = 120.58, df = 6, p-value < 2.2e-16
```

Figure 15. Kruskal-Wallis test for U~ Lithology.

As Kruskal-Wallis test is significant, it is plausible to compute Dunne test, for performing multiple pairwise-comparisons between the means of groups. The Fig. 16 shows that only 5 of the combinations are significantly different respect to U signal.

Pair discrimination	p adj
LITH4-LITH2	0
LITH6-LITH2	0
LITH6-LITH4	0.0000029
LITH6-LITH5	0.0001258
LITH3-LITH2	0.0018034

Figure 16. Pairwise-comparisons significant different between the means of groups of Lithology respect to U signal.

ANOVA one-way tests for Facies Association

Results of ANOVA for K~Facies Association are shown in Figs. 17 and 18. The very low value of Pr is enough to reject the null hypothesis. The Fig. 17 shows the plausibility of establishing significant differences between the categories of Facies Association respect to K signal. The Fig. 18 shows that the residual distribution of ANOVA one way is normal with no systematic pattern and is plausible the homogeneity of variance assumption.

```
Analysis of Variance Table

Response: K
                    Df  Sum Sq Mean Sq F value    Pr(>F)
FaciesAssociation    4  1.2837 0.32093  12.571 5.032e-10 ***
Residuals         1214 30.9938 0.02553
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 17. ANOVA one way of the K signal for classes of Facies Association

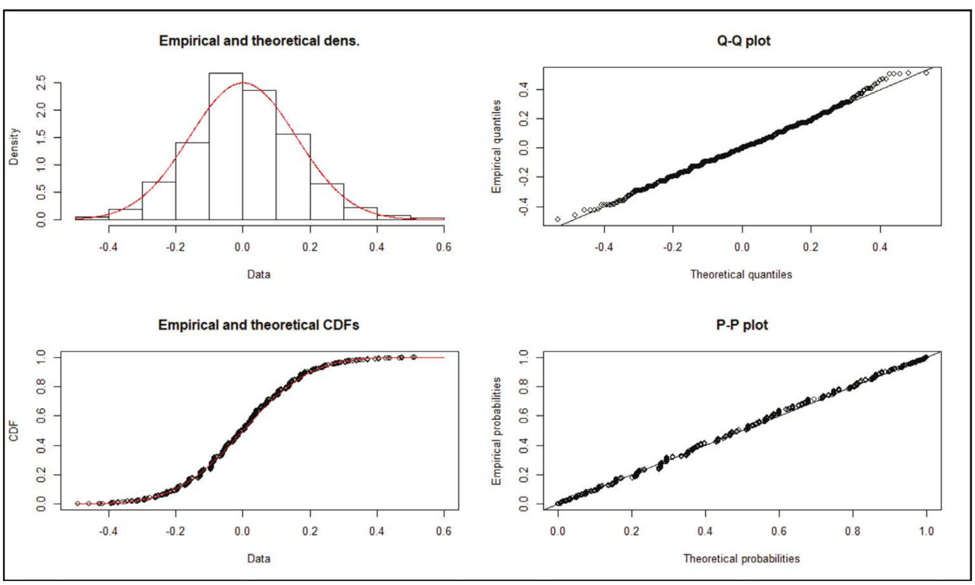

Figure 18. Residuals analysis for K ~Facies Association.

As the ANOVA test is significant, it is plausible to compute Tukey HSD (Tukey Honest Significant Differences), for performing multiple pairwise-comparisons between the means of groups of Facies Association. This is done to discriminate which groups of Facies Association are different respect to K signal. There are 10 possible combinations, between pairs of categories. The Fig. 19 shows the 6 out of all possible combinations that are significantly different respect to K signal.

Pair discrimination	p adj
*4-3	0.00001
*3-2	0.00010
*4-1	0.00100
*5-3	0.00158
*5-1	0.00331
*2-1	0.00436

Figure 19. Pairwise-comparisons significant different between the means of groups of Facies Association respect the K signal.

Results of ANOVA for Th~Facies Association are shown in Figs. 20 and 21. The very low value of Pr is enough to reject the null hypothesis. The Fig. 20 shows the plausibility of establishing significant differences between the categories of Facies Association respect to Th signal. The Fig. 21 shows that the residual distribution of ANOVA one way is normal with no systematic pattern and is plausible the homogeneity of variance assumption.

```
Analysis of Variance Table

Response: Th
                    Df Sum Sq Mean Sq F value    Pr(>F)
FaciesAssociation    4  5.253 1.31325  51.811 < 2.2e-16 ***
Residuals         1214 30.771 0.02535
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 20. ANOVA one way for Th~ Facies Association.

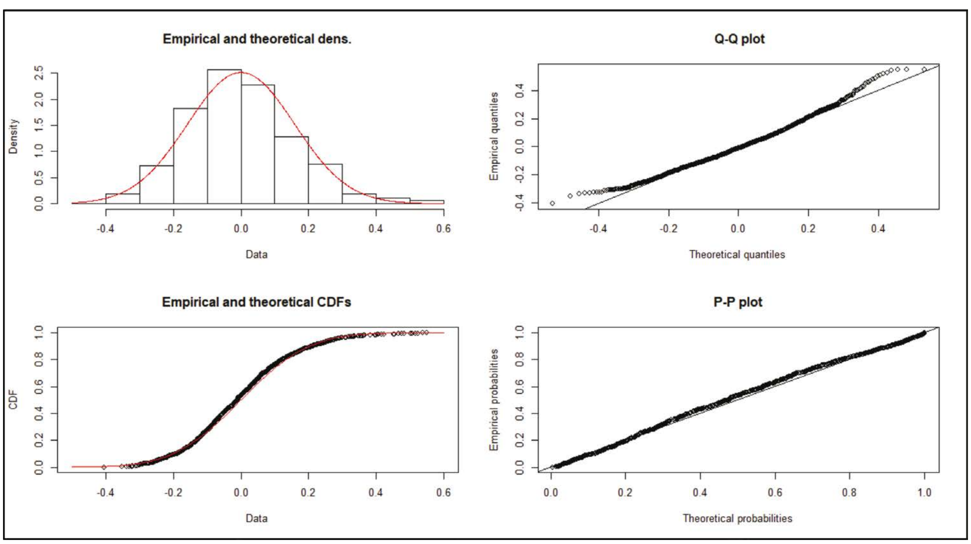

Figure 21. Residuals analysis for K ~Facies Association.

Fig. 22 shows the 5 combinations that differ significantly respect to Th signal.

Pair discrimination	p adj
*4-3	0
*3-2	0
*4-1	0
*5-1	0.0006213
*2-1	0.0008795

Figure 22. Pairwise-comparisons significant different between the means of groups of Facies Association respect the Th signal.

Results of ANOVA for U~Facies Association are shown in Figs. 23 and 24. The very low value of Pr is enough to reject the null hypothesis. The Fig. 23 shows the plausibility of establishing significant differences between the groups of Facies Association respect to U signal. Residuals analysis for ANOVA one-way U~Facies Association shows it is necessary to check if hypotheses of parametrical ANOVA are plausible. The Fig. 24 shows that the residual distribution for ANOVA one-way U~Facies Association has some deviation from the normal, then it is necessary to run a non-parametrical ANOVA, for example Kruskal-Wallis test, and avoid a false positive.

```
Analysis of Variance Table

Response: U
                      Df  Sum Sq Mean Sq F value    Pr(>F)
FaciesAssociation      4  2.1924 0.54809  31.561 < 2.2e-16 ***
Residuals           1214 21.0822 0.01737
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 23. ANOVA one way for U~ Facies Association.

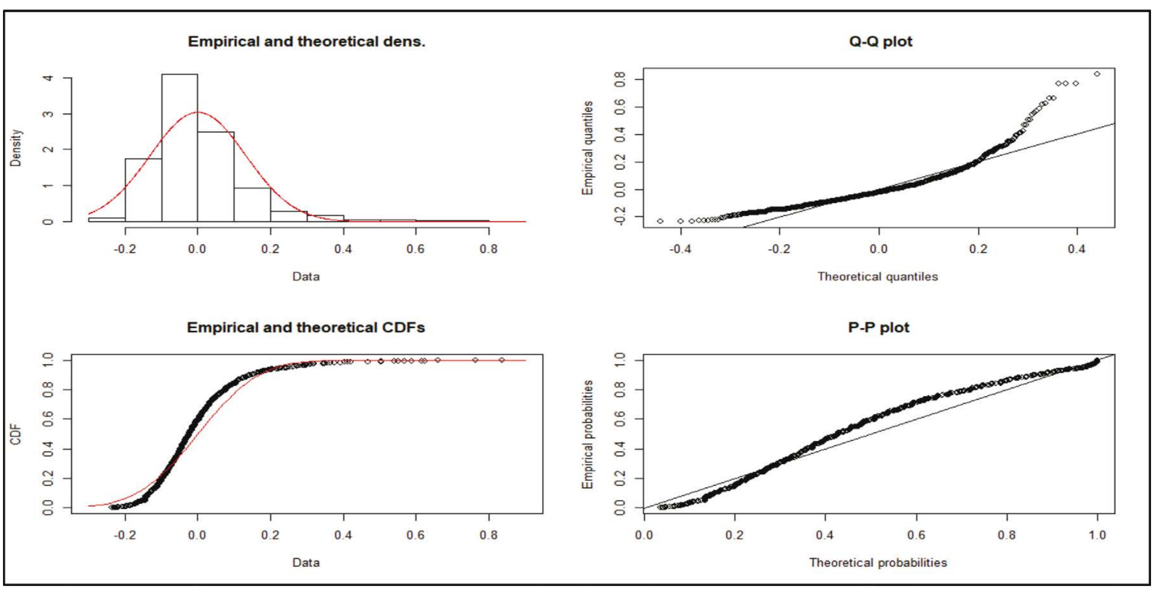

Figure 24. Residuals analysis for ANOVA one-way U ~ Facies Association.

```
        Kruskal-Wallis rank sum test

data:  U by FaciesAssociation
Kruskal-Wallis chi-squared = 86.444, df = 4, p-value < 2.2e-16
```

Figure 25. Kruskal-Wallis test for U~ Facies Association.

Kruskal-Wallis test (Fig. 25) shows that it is plausible to reject the null hypothesis with no assumption of homogeneity of variance.

As the Kruskal-Wallis test is significant, it is plausible to compute the Dunne test, to perform multiple pairwise-comparisons between the means of groups. The Fig. 26 shows the 6 combinations that significantly different respect to U signal.

Pair discrimination	p adj
*3-1	0
*4-1	0
*2-1	0.000017
*4-2	0.0002883
*5-1	0.0065832
*4-3	0.018806

Figure 26. Pairwise-comparisons significant different between the means of groups of Facies Association respect the U signal.

Data Driven Modeling

The results of the data exploration show a complex no linear relation among K, Th and U signals and the facies (Lithology or Facies Association). Although there is statistically significant difference between categories of facies and K, Th and U signals, this does not imply directly the feasibility of an accurate model of Machine Learning which can find the function f(K,Th,U) that transforms the K, Th, U signals into facies. The great complexity of the physical phenomena connecting the predictors and target variables suggests the use of Data Driven Models (DDM) is more viable, since derivation of f(K,Th,U) from basic physical laws is often unachievable.

Results from data exploration make it feasible to build 4 models of machine learning avoiding extremely imbalanced classes, considering the limitations of the predictor variables and taking into account the results of Tukey HSD and Dunne test.

Machine Learning Models for classes of Lithology

Model 1 for classes of Lithology: Discarding the class LITH7 (extremely imbalanced), the response variable will have 6 categories with imbalanced classes LITH1 and LITH3, LITH5. The predictor variables are the K, Th and U signals normalized for each well log to their respective minimum and maximum value. Using h2o.ai (software) with the previously described training, the DRF (Distributed Random Forest) shows the best performance. The Figs. 27 and 28 show the confusion matrix and error of the model for all the classes.

```
Confusion Matrix and Statistics

          Real
Predicted LITH1 LITH2 LITH3 LITH4 LITH5 LITH6
    LITH1     5     0     0     0     0     0
    LITH2     1    76     0     5     3    18
    LITH3     0     0     9     1     0     1
    LITH4     1    11     5    84     5    17
    LITH5     0     0     0     0     9     2
    LITH6     0    15     3    23     2    80

Overall Statistics

               Accuracy : 0.6995
                 95% CI : (0.6504, 0.7454)
```

Figure 27. Confusion Matrix and error model for Model 1 for classes of Lithology.

Class	Balanced Accuracy
LITH1	0.85714
LITH2	0.8233
LITH3	0.76192
LITH4	0.7975
LITH5	0.73404
LITH6	0.7556

Figure 28. Balanced Accuracy per class for Model 1 for classes of Lithology.

The Cohen's Kappa statistic is 0.5836. According to Landis and Koch (1977), this result suggests a moderate agreement. Despite the complexity of the problem, an acceptable total performance was obtained, clearly depending of the aim of application of the model. Class 1 with high imbalance shows a good balanced accuracy and suggest an accurate treatment of the imbalance in the training (using SMOTE).

Model 2 for classes of Lithology: In order to improve the accuracy of the model 1, the model has 5 categories of Lithology where, LITH1 and LITH2 are merged (LITH=LITH2) and without LITH7, with imbalanced classes LITH3, LITH5. With this change in the variables plus SMOTE balancing of the training data, a better performance was obtained. Using h2o.ai (software) with the previously descripted training, DRF (Distributed Random Forest) shows the best performance respect to GBM (Gradient Boosting Machine), although there is little difference (minor than 2%). The Figs. 29 and 30 show the confusion matrix and error of the model for all the classes.

```
Confusion Matrix and Statistics

          Real
Predicted LITH2 LITH3 LITH4 LITH5 LITH6
    LITH2    53     0     7     1     3
    LITH3     0     9     3     0     0
    LITH4     7     2    51     4    12
    LITH5     1     0     0     3     1
    LITH6     7     0     6     2    64

Overall Statistics

               Accuracy : 0.7627
                 95% CI : (0.7032, 0.8155)
```

Figure 29. Confusion Matrix and error model for Model 2 for classes of Lithology.

Class	Balanced Accuracy
LITH2	0.857
LITH3	0.9042
LITH4	0.8066
LITH5	0.64558
LITH6	0.8519

Figure 30. Balanced Accuracy per class for Model 2 for classes of Lithology.

Cohen's Kappa statistic is 0.6675. According to Landis and Koch (1977), this result suggests at substantial agreement. The performance of the Model 2 for Lithology classes is better than Model 1. Merging the class LITH1 (with high imbalance) with LITH2, improves performance and the model maintain a substantial amount of information about the classes.

Machine Learning Models for classes of Facies Association

Model 1 for classes of Facies Association: The response variable has 4 categories (excluding Facies Association 5) of Facies Association, with imbalanced classes 2 and 4. Using h2o.ai (software) with the previously descripted training, the GBM (Gradient Boosting Machines) shows the best performance respect to DRF (Distributed random forest), although with a small difference (minor than 2%). The Figs 31 and 32 show the confusion matrix and error of the model for all the classes.

```
Confusion Matrix and Statistics

          Real
Predicted   1   2   3   4
        1  53   4   6   1
        2   5  25   3   0
        3  19  19 168  10
        4   0   4  10  36

Overall Statistics

               Accuracy : 0.7769
                 95% CI : (0.7305, 0.8187)
```

Figure 31. Confusion Matrix and error model for Model 1 for Facies Associations.

Class	Balanced Accuracy
1	0.8249
2	0.7275
3	0.8128
4	0.86

Figure 32. Balanced Accuracy per class for Model 1 for Facies Associations

The predictor variables are the K, Th and U signals, normalized for each well log to their respective minimum and maximum values. Cohen Kappa's statistic is 0.6431. According to Landis and Koch (1977) this result suggests at substantial agreement. Class 2 with high imbalance has the minor balanced accuracy.

Model 2 for classes of Facies Association: To improve the accuracy of Model 1, the response variable has 3 categories, not considering Facies Association 5 and merging Facies Associations 3 and 4. The Figs. 33 and 34 show the confusion matrix and error of the model for all the classes.

```
Confusion Matrix and Statistics

          Real
Predicted   1   2   3
        1  44   2   7
        2   1  10   0
        3  15  13 153

Overall Statistics

                Accuracy : 0.8449
                  95% CI : (0.7934, 0.8878)
```

Figure 33. Confusion Matrix and error model.

Class	Balanced Accuracy
1	0.8423
2	0.69773
3	0.8134

Figure 34. Balanced Accuracy per class.

Cohen's Kappa statistic is 0.6628. According to Landis and Koch (1977) this result suggests at substantial agreement. The performance of the Model 2 of class of Facies Association is only slightly better than Model 1 and the problem of the imbalanced class Facies Association 2 is still present. Also, the model 2 losses relevant information about the classes of Facies Association 3 and 4.

DISCUSSION AND CONCLUSIONS

Both the quality of SGR outcrop measurements and the sedimentological determination of an outcrop facies model further consideration of data analytics and Data-Driven modeling using Machine Learning.

In the data exploration, the multidimensional plot shows a complex relation between the K, Th, U (predictor variables) and the classes of Lithology (response variable). ANOVA one-way and Kruskal-Wallis tests with the corresponding Tukey HSD and Dunn tests (for each signal of K, Th and U) show that it is feasible the discrimination of 18 (out of 21) significantly differing pairwise-comparisons between the means of Lithology groups, considering simultaneously the 3 signals. The 3 pairwise-comparisons that cannot be discriminated with this criterion are LITH6-LITH3 (Tuffaceous siltstones-Siltstones), LITH7-LITH5 (Coarse-grained lahar-Fine-grained tuffs), and LITH2-LITH1 (Sandstones-Conglomerates). As can be observed, some grade of sedimentological proximity exists between both categories in each of the 3 pairs.

The process of modeling using Machine Learning must take into account the limitations of the predictor variables. The use of Facies Association categories favors a higher quality model, since data exploration tests show that it is plausible the discrimination of 8 significantly differing pairwise-comparisons (from a total of 10). The concept of Facies Association increases the number of significantly differing pairwise-comparisons between these classes (using a linear analysis and only K, Th and U signals), and then simplifying the target learning of the machine learning algorithms (maximizing the future performance), compared to results for Lithology classes.

Model 1 (6 classes of Lithology) for classes of Lithology gives an acceptable performance using h2o.ai-Arno Candel and GBM (Gradient Boosting Machine) algorithm with a total accuracy of 70 %, with a 95% CI (Confidence Interval) of (65,75)%. A Cohen's Kappa statistic value of 0.5836 suggests a moderate agreement according to Landis and Koch (1977).

Model 2 (5 classes of Lithology) for classes of Lithology gives an acceptable performance using h2o.ai-Arno Candel and DRF (Distributed Random Forest) algorithm with a total accuracy of 76 %, with a 95% CI (Confidence Interval) of (70,82)%. Cohen's Kappa statistic value of 0.67 suggests a substantial agreement according to Landis and Koch (1977).

Model 1 (4 classes of Facies Association) for classes of Facies Association gives an acceptable performance using h2o.ai-Arno Candel and the GBM (Gradient Boosting Machine) algorithm with a total accuracy of 78 % with a 95% CI (Confidence Interval) of (73,82)%. Cohen's Kappa statistic value of 0.63 suggests a substantial agreement according to Landis and Koch (1977).

Model 2 (3 classes of Facies Association) for classes of Facies Association gives an acceptable performance using h2o.ai-Arno Candel and DRF (Distributed Random Forest) algorithm with a total accuracy of 85 % with a 95% CI (Confidence Interval) of (79,89)%. A Cohen's Kappa statistic value of 0.66 suggests a substantial agreement according to Landis and Koch (1977).

The accuracy of the obtained models could be satisfactory, depending on the potential intended application. The best results were obtained merging variables: Model 2 for classes of Lithology accomplishes higher accuracy by merging categories corresponding to Sandstones and Conglomerates. Model 2 for classes of Facies Association shows increased accuracy with respect to Model 1 by merging the two classes of Distal Floodplain (Volcaniclastic and Epiclastic). Although the improvement is only slightly better than Model 1, the issue of imbalanced class Facies Association 2 is still present. The advantages and drawbacks of variables merging must be carefully considered in each case, estimating the impact of associated information loss in the light of the focus and objectives of model construction. Decisions regarding the convenience of accepting these simplifications can only be made in the context of each particular modeling project. The possibility to discriminate Distal Floodplain Facies Association from Proximal Floodplain, in addition to Fluvial Channel Facies Association, can have a considerable impact in the context of stacking patterns analysis. In this case, decision must be made to favor a higher global accuracy accepting a lower accuracy for imbalanced class 2 (Model 2), or higher balanced individual accuracy with lower global accuracy (Model 1).

The limitations of the Machine Learning algorithm to learn from the input data (SGR measurements) and to discriminate corresponding facies categories, require both a careful definition of the facies, taking into account the feasibility of the discrimination of the categories with the input data, and the availability of consistent information according to the predefined objectives. In some cases, incorporating additional relevant data, such as other well logs, could be necessary.

In this study we determined, based on models with substantial agreement, the feasibility of SGR use to obtain facies categories on the outcrop (where measurements were carried out). This precedent encourages expectations regarding feasibility to generate detailed facies models in subsurface based on an easily available tool, such as SGR log. Although the models are independent of the scale (K, Th and U signals were normalized to min-max for each well log) transferring these results to the subsurface requires a thorough, sound validation process with independent data. Development of additional sets of Machine Learning models might be mandatory. Nevertheless, it is important and relevant to highlight that it is possible to obtain an acceptable accuracy in the facies classification using ensemble Machine Learning algorithms. These algorithms can learn from the input data following a careful calibration avoiding overfitting, and defining a function f(K,Th,U) that transforms the K, Th and U signals of SGR into facies defined with detailed sedimentological criteria.

ACKOWLEDGEMENTS

We would like to thank Sinopec Argentina, Sergio Giordano, Román Paulino, and Juan Techenski for their support during the developments of this study.

CITED REFERENCES

Acosta, N., 2013. Análisis comparado de reservorios del cretácico superior en subsuelo y en afloramientos. Relaciones y correspondencias entre referentes instrumentales y algunos casos de estudio en la Cuenca del Golfo San Jorge. 1° Jornadas Geológicas de la Cuenca del Golfo San Jorge, Libro de resúmenes, p. 12. Comodoro Rivadavia, Argentina.

Brodersen K.H., Ong, C.S., Stephan, K.E. and Buhmann, J.M., 2010, The balanced accuracy and its posterior distribution. 20th International Conference on Pattern Recognition. Istanbul, pp. 3121-3124. DOI: 10.1109/ICPR.2010.764

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16: 321-357.

Daumé III, Hal., 2017. A course in Machine Learning. Self-published. Available in http://ciml.info/dl/v0_99/ciml-v0_99-all.pdf

Ferreira, L., Santangelo, A. and Georgieff, S.M., 2011, Modelo geológico para la Fm. Bajo Barreal (Cretácico): Integración y optimización entre datos de superficie y subsuelo. Cuenca del Golfo San Jorge, provincia de Santa Cruz, Argentina. VIII Congreso de Exploración de Desarrollo de Hidrocarburos. Trabajos Técnicos, p. 327-352. Mar del Plata.

Ferreira, L., Santangelo, A. and Georgieff, S.M., 2013, Modelo geológico del yacimiento Cerro Guadal Norte. Integración de datos de superficie y subsuelo: geología, geofísica y petrofísica. Cuenca del Golfo San Jorge, Argentina. 1° Jornadas Geológicas de la Cuenca del Golfo San Jorge, Libro de resúmenes, p. 42-43. Comodoro Rivadavia, Argentina.

h2o.ai. (software), Arno Candel, Cliff Click and Sri-Satish Ambati. Retrieved from https://www.h2o.ai/resources/ h2o.ai-DRF, Distributed Random Forest, http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/drf.html.

h2o.ai-GBM, Gradient Boosting Machine, http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/gbm.html.

Hamon, Y., Barbier, M., Deschamps, R., Doligez, B., Joseph, P., Rohais, S., Schmitz, J. and Callot, J.-P., 2015. Application of Outcrop Analogues in Carbonate Reservoir Characterization and Modeling: Multiscale and Multidisciplinary Approaches. Search and Discovery Article #51084.

Hastie, T., Tibshirani, R. and Friedman, J., 2009. The Elements of Statistical Learning. Data Mining, Inferencen and Prediction. Second edition. Springer.

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. An Introduction of Statistical Learning. Springer-Verlag New York. DOI 10.1007/978/-1-4614-7138-7.

Landis, J.R., and Koch., G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33(1): 159-174. DOI: 10.2307/2529310.

Limarino, C.O. and Giordano, S.R., 2016. Análisis de modas detríticas de areniscas: identificando múltiples áreas de proveniencia en el flanco sur de la Cuenca del Golfo San Jorge. 2° Jornadas Geológicas de la Cuenca del Golfo San Jorge, Libro de resúmenes 60-61. Comodoro Rivadavia, Argentina.

Limarino, C.O., Giordano, S.R. and Rodríguez Albertani, R.J., 2017, Diagenetic model of the Bajo Barreal Formation (Cretaceous) in the southwestern flank of the Golfo de San Jorge Basin (Patagonia, Argentina). Marine and Petroleum Geology, 88: 907-931.

Limarino, C.O., Giordano, S.R., and Rodríguez Albertani, R.J., this congress. Petrografía y análisis de proveniencia de los reservorios del Cretácico Superior – Flanco Sur, Cuenca del Golfo San Jorge. Rocas Reservorio de las Cuencas productivas argentinas – Actualización. 10° Congreso de Exploración y Desarrollo de Hidrocarburos. Mendoza.

Miall, A., 2014. Fluvial Depositional Systems. Springer International Publishing Switzerland. ISBN 978-3-319-00666-6.

Package ‘car’, Retrieved from R Software: https://cran.r-project.org/web/packages/car/car.pdf

Package ‘corrplot’, Retrieved from R Software: https://cran.r-project.org/web/packages/corrplot/corrplot.pdf

Package ‘data.table’, Retrieved from R Software https://cran.r-project.org/web/packages/data.table/data.table.pdf

Package ‘dplyr’, Retrieved from R Software https://cran.r-project.org/web/packages/dplyr/dplyr.pdf

Package ‘ggplot2’, Retrieved from R Software: https://cran.r-project.org/web/packages/ggplot2/index.html

Package ‘MASS’, Retrieved from R Software: https://cran.r-project.org/web/packages/MASS/MASS.pdf

Package ‘caret’, Retrieved from R Software: https://cran.r-project.org/web/packages/caret/caret.pdf

Paredes, J.M., Foix, N. and Allard, J.O., 2016a. Sedimentology and alluvial architecture of the Bajo Barreal Formation (Upper Cretaceous) in the Golfo San Jorge Basin: outcrop analogues of the richest oil bearing fluvial succession in Argentina. Marine and Petroleum Geology, 72: 317-335.

Paredes, J.M., Valle, M.N., Olazábal, S.X., Foix, N., Allard, J.O. and Ocampo, S.M., 2016b. Usos y limitaciones del perfil de rayos gamma en la Cuenca del Golfo San Jorge: mediciones en afloramiento. 2° Jornadas Geológicas de la Cuenca del Golfo San Jorge, Libro de resúmenes, p. 70-71. Comodoro Rivadavia.

Paredes, J.M., Foix, N. and Allard, J.O., 2017. Depositional architecture and evolution of fluvial systems of the Bajo Barreal Formation in the Cerro Ballena anticline: implications for subsurface studies and oilfield development. Internal report. Sinopec Argentina Exploration and Production.

Paredes, J.M., Giordano, S.R., Valle, M.N., Allard, J.O. y Foix, N., this congress. Spectral Gamma-Ray logging of Cerro Ballena anticline, Santa Cruz province: an outcrop analogue to subsurface characterization and well-log correlation of fluvial sandstone reservoirs. 10° Congreso de Exploración y Desarrollo de Hidrocarburos, Sesiones Técnicas. Mendoza.

Plazibat, S., Acuña, C., Gavilán, J., Asef, M. and Cayo, E., 2013. Estudio integrado de reservorios fluviales de la Formación Comodoro Rivadavia, Yacimiento Cerro Dragón, Cuenca del Golfo San Jorge. 1° Jornadas Geológicas de la Cuenca del Golfo San Jorge, Libro de resúmenes, p. 74-75. Comodoro Rivadavia. Argentina.

R software, https://www.r-project.org/

Rodríguez, R.J., Caprioglio, P., Loss, M.L. y Aguirre, H.D., 2011. Caracterización de reservorios y modelado estático. Delineado de un yacimiento complejo. VIII Congreso de Exploración de Desarrollo de Hidrocarburos. Trabajos Técnicos, p. 443-467. Mar del Plata.

Schiuma, A., Crovetto, C., Andersen, D. and Gómez, J., 2013. Caracterización y desarrollo de reservorios fluviales en el yacimiento Piedra Clavada, empleando atributos sísmicos y electrofacies. 1° Jornadas Geológicas de la Cuenca del Golfo San Jorge, Libro de resúmenes, p. 89-90. Comodoro Rivadavia, Argentina.

Solomatine, D. P. and Ostfeld, A., 2008. Data-driven modelling: some past experiences and new approaches. Journal of Hydroinformatics, 10: 3-22.

Szerbiak, R. B., McMechan, G. A., Corbeanu, R. and Forster, C., 2001. 3-D characterization of a clastic reservoir analog: From 3-D GPR data to a 3-D fluid permeability model. Geophysics, vol. 66 (4): 1026-1037.

P. Bestagnini, V. L. (2017). A Machine Learning Approach to Facies Classification Using Well Logs. SEG International Exposition and 87th Annual Meeting, 2137-2142.