Machine learning-based well log conditioning enhances facies classification in Fandango Field

Felipe Melo*, Fred Jenson and Chiranjith Ranganathan, GeoSoftware

This paper was prepared for presentation during the 18th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 16-19 October 2023.

Abstract

Facies classification is a process used to describe the different rock types and sedimentary structures. This process can be performed using responses of well logs due to geology. We selected nineteen wells from Fandango field to demonstrate the benefit of machine learning methods as a pre-step for facies classification. First, we flag anomalous productive zones as pay zones, and then we identify outliers with an unsupervised machine learning technique and flag them with null values. The null values are replaced by synthetic data generated from a multiple linear regression with a subset of wells, thus generating the final corrected log curve data. Then, we use hierarchical agglomerative clustering to perform facies analysis on a subset of six wells. Six classes were generated and found to be geologically sound when compared with the well logs, and each class was then assigned to specific facies. A random forest model for supervised facies classification was trained on five of the previous labeled wells, one was set aside as a test well. Finally, random forest was employed to classify all nineteen wells based on the facies labels. The results of supervised facies classification on the raw and corrected data show the improvement of the latter with machine learning-based well log conditioning. A north-to-south stratigraphic cross-section of eight conditioned wells shows the consistency of the classification.

Introduction

The term facies is used to describe either a specific volume of sediment or the inferred depositional environment of that sediment (Anderton, 1985). Facies classification can be performed with the description of cored sediments (Anderton, 1985) or by analyzing well log responses (Wolf and Pelissier-Combescure, 1982). The facies log ultimately can be used to drive deterministic (Pendrel and Schouten, 2020) and geostatistical (Filippova et al., 2013) inversions.

Machine learning has been applied to facies classification for decades (Baldwin et al., 1990; Wong et al., 1995; Dubois et al., 2007). Hall (2016) and Bormann et al. (2020) promoted machine learning competitions that made facies classification a popular topic. The achievements of these competitions show not only advanced machine learning techniques, but also the need of pre-processing techniques and data augmentation prior to facies classification.

Well log editing may be necessary for three potential reasons: sticking tools, poor borehole conditions, and noisy data (Cannon 2016). Akkurt et al. (2018) presented a machine learning workflow to perform quality control, outlier detection and well log correction. Recently, Rivera et al. (2022) used machine learning to show the impacts of outlier detection and log editing in determining acoustic impedance and porosity logs.

In this work, we follow the workflow proposed by Ranganathan et al. (2021) to detect and correct anomalies in sonic and density well log data from Fandango field. Then, we perform unsupervised facies classification with hierarchical agglomerative clustering (Pedregosa et al., 2011) using a subset of six available wells. Six classes are generated and labelled according to geology. Thus, we generate a supervised facies classification model with random forest (Pedregosa et al., 2011) using five wells, one of the wells used for the unsupervised facies classification is left as test well and is not used in the model. We apply the model to perform facies classification to all the nineteen wells and the results are found to be geologically sound when compared with the well logs. Finally, we compare the results of the supervising facies classification in the test well using the raw and the corrected logs. Results show an eight percent increase in quantitative metrics (confusion matrix, F1, precision, recall, and accuracy) for corrected well logs compared to raw logs. A north-to-south stratigraphic cross-section of eight conditioned wells flattened in a sand of interest shows the consistency of the classification.

Dataset

Fandango field, located in South Texas (red star in Figure 1), was the first deep Wilcox discovery in the continental US and the first onshore field that acquired 3D Seismic data. The productive sands in Fandango field are encountered at depths of 14000 to 18000 feet. These deep reservoirs in the Wilcox Group occur within predominantly upward-coarsening shallow-marine parasequences comprised of sandstones, siltstones, and shales, that were deposited during the Eocene (Ambrose, et al., 2016; Dutton et al., 2016).

The dataset consists of nineteen wells sampled every 0.5 feet and the working interval ranges from 13000 to 18000 feet. The well logs used in this work are density (RHOB), gamma ray (GR), porosity (NPHI), resistivity (ILD), sonic (DT), volume of clay (VCLAY), and water saturation (SW). Figure 2 shows a type well, the stratigraphic column with tops and interest sands, and the well logs GR (filled), ILD (brown), NPHI (red), RHOB (blue) and DT (green). The

sands of interest are shown in the Upper Wilcox formation, of which R Sand and T5 highlights.

Figure 1 *– The red star denotes the location of Fandango Field in South Texas.*

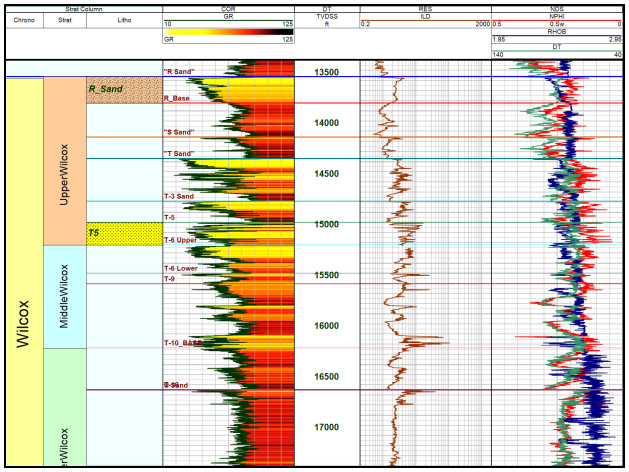

Figure 2 *– Type well showing the stratigraphic column with tops and interest sands, and the well logs GR (filled), ILD (brown), NPHI (red), RHOB (blue) and DT (green).*

Method

This work is divided into three parts: well log editing including anomaly detection and well log correction, unsupervised facies classification in a subset of the wells and supervised facies classification in all available data.

The set of well logs were qualitatively analyzed, and the first data conditioning step was the two-point gamma ray normalization (Shier, 2004). From now on, the output logs of this step will be referred to as GR_NRM.

We created a pay flag, to identify hydrocarbon zones, where SW is lower than 0.5, NPHI is higher than 0.12 and VLCAY is lower than 0.6. The pay flag will prevent the edition of these anomalous zones in the log correction process. The outliers are automatically detected (Ranganathan et al., 2021) in sonic and density logs based on their impurity. The data points where the values are assigned as outliers will be corrected.

High quality synthetics are generated with multiple linear regressions. The input logs for the sonic synthetic are GR (normalized), ILD (in log scale), NPHI, RHOB, and true vertical depth. The input logs for the synthetic density are the same as before, except density is now the target curve and has been replaced by the sonic log.

The values assigned as outliers are replaced by the synthetic logs, except when a pay flag is present in the same position. In this case, the values are kept. The sonic and density logs are then corrected and referred to as DT_merge and RHOB_merge.

Six wells were used in the model for unsupervised facies classification: Muzza 2, Muzza 3, Rancho Blanco State 5, Hinojosa 13, De Garza Paula Es and Leyendecker 2. The input curves are DT_merge, GR_NRM, ILD, NPHI, and RHOB_merge. We used dendrograms to determine the optimal number of clusters in these well logs. Then, the hierarchical agglomerative clustering method (Pedregosa et al., 2011) was used to perform the unsupervised classification and facies were assigned to the six wells.

The next step is to create a supervised facies classification model. In this case, the target is the facies curve generated in the previous step. All the wells used for unsupervised facies classification are used in this step except by Leyendecker 2. This well is set as a test well and is used to check the performance of the model. We used the random forest classifier (Pedregosa et al., 2011) with twenty estimators in the five wells and the same five well logs to perform the supervised facies classification. Then, the model was applied to all the nineteen wells.

The results of the supervised facies classification are quantitatively and qualitatively analyzed in the test well. The supervised facies classification was performed in the test well twice, with the raw logs and with the corrected logs. The performance of the model was evaluated on the test well with the confusion matrix, F1, precision, recall and accuracy.

Results

Figure 3a shows the histograms of the raw gamma ray logs and Figure 3b shows the histograms of the normalized gamma ray logs (GR_NRM). The normalization process allows comparison between multiple wells on the same scale.

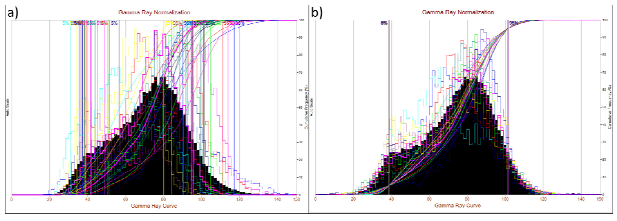

Figure 3 *– Histograms of the gamma ray logs: a) raw gamma ray logs, and b) two-point normalized gamma ray logs.*

Figure 4 shows the cross-plots of sonic and density logs with outlier detection and the synthetic generated for the

Rancho Blanco State 5 well. Figure 4a shows the detected outliers in red and inliers in blue. Figure 4b shows the pay flag in green and the inliers in blue. Figure 4c shows the corrected log, after outlier replacement by synthetic data generated with the multiple linear regression. Some outliers shown in red in Figure 4a are also flagged as pay and shown in green in Figure 4b. These points are not replaced by the generated synthetic data.

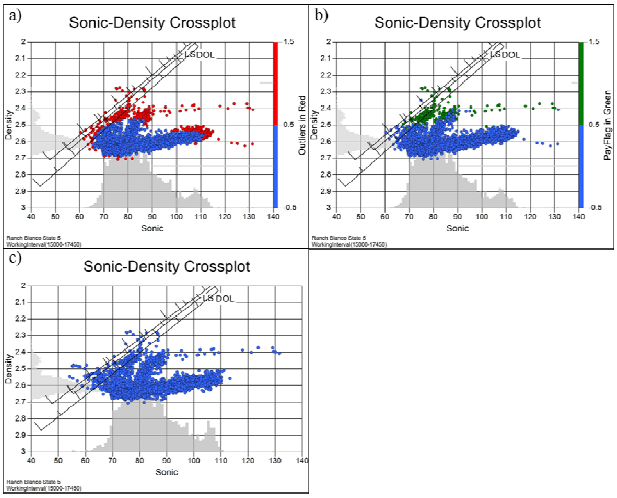

Figure 4 *– Sonic-Density cross-plots from well Rancho Blanco State 5. a) Raw data with outliers in red, b) raw data with pay flag in green, and c) corrected data.*

Figure 5 shows the well Muzza 8. The first track shows the stratigraphic column, and the second track shows the colored GR_NRM log, where the lowest values highlight the sands in yellow. The third track shows the depth curve, the outliers detected in red, and the pay flag in green. Notice that some points flagged as outliers and pay are in the same position in the T-3 Sand. The points in these positions are not edited, this is another view of the logs shown in Figures 4a and 4b. Additionally, virtually the entire sand range is flagged as pay and the well logs will not be edited in this area. The third track shows the ILD and the fourth track shows input logs NPHI, RHOB and DT. The fifth track shows the input log NPHI and the corrected logs RHOB_merge and DT_merge. Note that these logs have been corrected where outliers are flagged on the third track, around 14200 and 14400 ft, but not where the pay flag is present, in the T3- sand zone.

Figure 6 shows the dendrogram computed with the six wells and five well logs in each. Six clusters are identified in this dataset and defined by the black line.

Figure 7 shows the well Rancho Blanco State 5 with the results of the unsupervised facies classification after running the hierarchical agglomerative clustering. The first track shows the stratigraphic column, the second track shows the colored GR_NRM log, where lowest values highlight the different sands in yellow. The third track shows the depth curve and tops, the fourth track shows ILD, and the fifth track shows NPHI, RHOB_merge and DT_merge. The sixth track shows the unsupervised facies classification results, the classified facies are geologically sound compared to the well logs. They were correlated with the lithologies: light shale, grey shale, silt, fine grained sandstone, medium grained sandstone, and coarse sand. The last track of Figure 7 shows the shaly sand analysis. In the last track, the volume of clay is filled with dashed gray as the lithology for light shale and doted yellow for sandstone, the blue filled curve is the bulk volume water and the red filled curve is the sandstone porosity. In this curve, the highest porosities agree with the unsupervised facies in yellow.

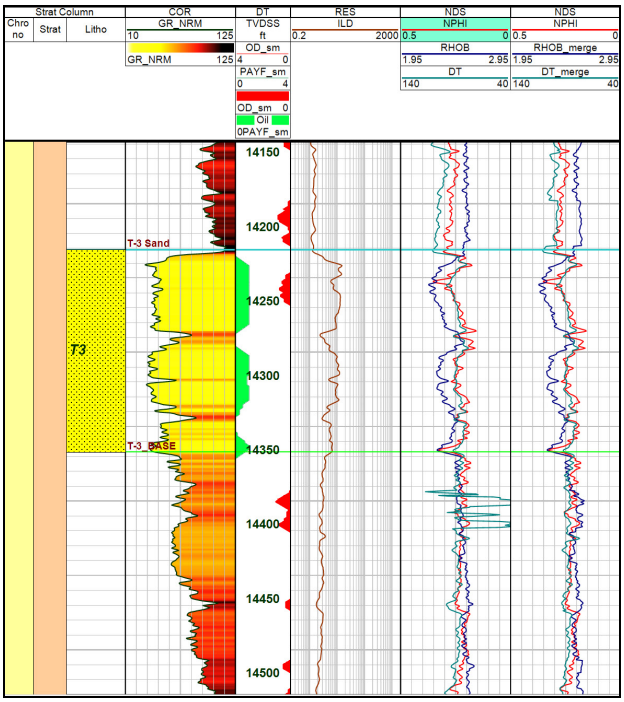

Figure 5 *– Well view of Muzza 8. Stratigraphic column, GR_NRM (filled), outlier (red) and pay (green) flags, ILD, NPHI, raw RHOB (dark blue) and DT (green), and corrected RHOB_merge (in dark blue) and DT_merge (in green).*

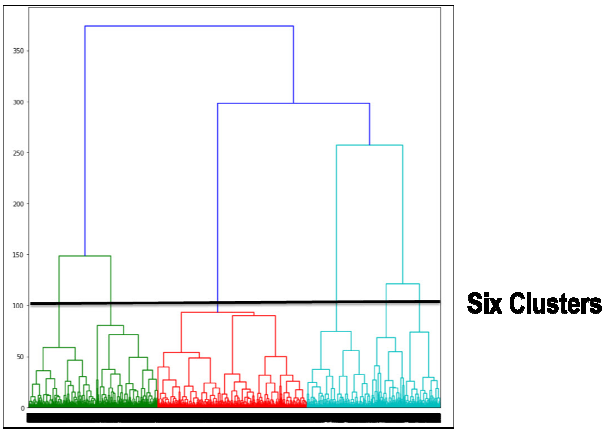

Figure 6 *– Dendrogram. Six clusters are identified and defined by the black line.*

Figure 8 shows the confusion matrix after performing the supervised classification in the Leyendecker 2 well, the test well. In this figure, the rows represent the actual class and the columns represent the predicted class. Figure 8a

and Figure 8b show the confusion matrix with classification results using raw logs and corrected logs as input, respectively. In general, facies are misclassified in the same classes. However, facies classification has better results in corrected well logs, with more samples correctly classified in all classes.

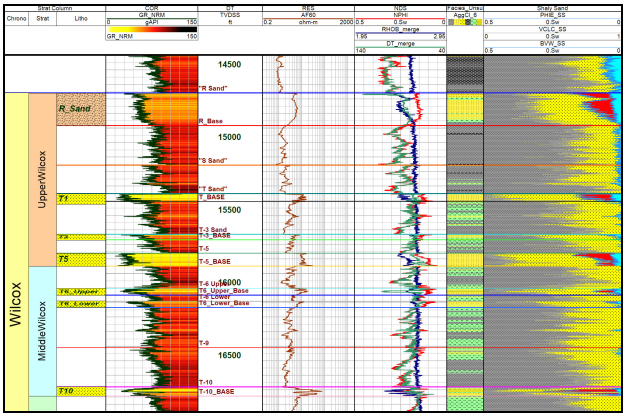

Figure 7 *– Well Rancho Blanco 5. Stratigraphic column, GR_NRM (filled), depth curve and tops, ILD, NPHI, corrected RHOB_merge and DT_merge. The two last tracks show the unsupervised facies classification and the shaly sand analysis.*

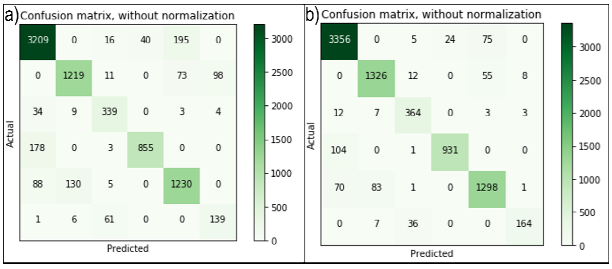

Figure 8 *– Confusion matrix in the Leyendecker 2 well. a) raw data as input, and b) corrected data as input.*

Table 1 shows the metrics of supervised facies classification in the Leyendecker 2 well: F1, precision, recall and accuracy. The first row shows the classification metrics using the raw data as input and the second row shows the results using the corrected data as input. Classification of corrected data produced higher quantitative results, the overall increase in the metrics is about eight percent.

Table 1 *– Metrics of supervised facies classification in the Leyendecker 2 well, using raw and corrected well logs.*

Data	F1	Precision	Recall	Accuracy
Raw	0.881	0.883	0.879	0.879
Corrected	0.936	0.937	0.937	0.936

Figure 9 shows results of unsupervised and supervised facies classification in the Leyendecker 2 well, the test well. The first track shows the stratigraphic column, the second track shows the colored GR_NRM log and tops, and the third track shows the depth curve. The three last tracks show the unsupervised facies classification, supervised facies classification in corrected data and supervised facies classification in raw data. As previously shown, quantitatively, the facies classification in corrected well logs has better results than in raw well logs. There is virtually no place where the classification on raw data is correct and classification on corrected data is misclassified. Namely, classification with raw data misclassifies where classification with corrected data misclassifies and in more samples.

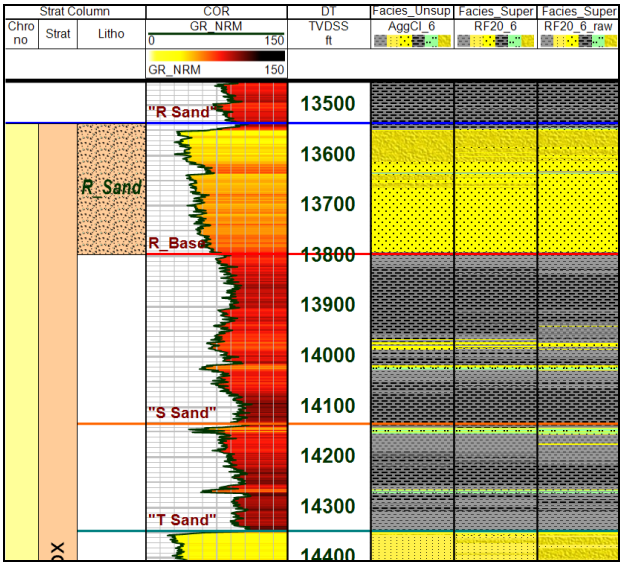

Figure 9 *– Leyendecker 2, the test well. The first tracks show the stratigraphic column, GR_NRM and tops, and depth. The three last tracks show the unsupervised facies classification, supervised facies classification in corrected data and supervised facies classification in raw data.*

Figure 10 shows multiple well cross-plots of the supervised facies classification in wells Muzza 2, Muzza 3, Rancho Blanco State 5, Hinojosa 13, De Garza Paula Es and Leyendecker 2. The wells have different symbols, and the classified classes have different colors, shown in the legends. Figures 10a-10d show the cross-plot between DT and GR_NRM, ILD and GR_NRM, RHOB_merge and DT_merge, and RHOB_merge and NPHI, respectively. The facies classification is consistent in all wells and well logs.

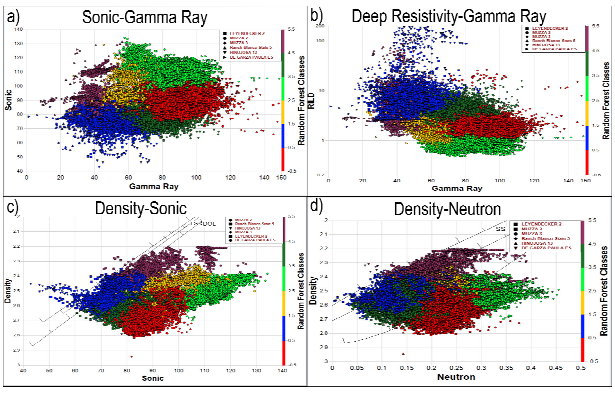

Figure 10 *– Multiple well cross-plots of the supervised facies classification, the wells are shown in the figure. a) DT x GR_NRM, b) ILD x GR_NRM, c) RHOB_merge x DT_merge, and d) RHOB_merge x NPHI.*

Figure 11 shows a stratigraphic cross section north to south with eight wells flattened on R Sand. The wells shown are Zachry 1A, Zachry 4, Zachry 1, Hinojosa 13, Leyendecker 1-R, Leyendecker 2, De Garza Paula Es and Longoria 1, respectively. The tracks show the stratigraphic column, GR_NRM, depth, the supervised classified facies, ILD, and NPHI, RHOB_merge and DT_merge. This figure shows good agreement with the tops throughout the section and is geologically sound.

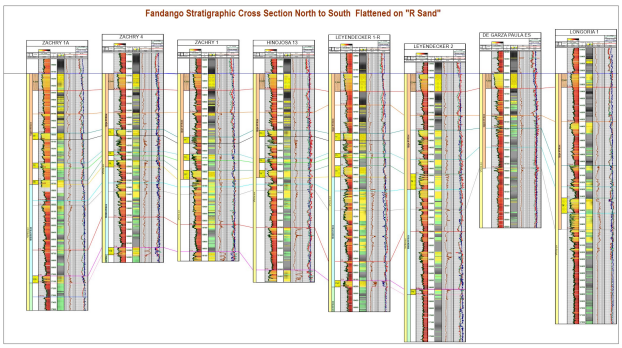

Figure 11 *– Cross section north to south flattened on R Sand showing eight wells. The tracks show the stratigraphic column, GR_NRM, depth, supervised facies classification, ILD, and NPHI, RHOB_merge and DT_merge.*

Conclusions

We compared supervised facies classification in raw and corrected well logs of Fandango field. GR normalization was applied to all the wells as the first conditioning step. The well logs DT and RHOB were corrected using data generated from multiple linear regressions in positions where the outlier flag was present, except where the pay flag was present. A subset of wells was used to perform unsupervised facies classification on six classes defined by the dendrogram. Then, a supervised classification model was trained with facies labeled in all six wells except the test well. This model was applied to all the wells. The results in the test data show an increase of eight percent in all metrics of the supervised facies classification using the corrected well log data compared to the raw data. The application of the supervised facies classification model to the whole dataset is geologically sound.

Acknowledgments

The authors thank GeoSoftware and the PowerLog development team for supporting their work.

References

Akkurt, R.; Conroy, T.; Psaila, D.; Paxton, A.; Low, J.; Spaans, P. 2018. Accelerating and Enhancing Petrophysical Analysis with Machine Learning: A Case Study of An Automated System for Well Log Outlier Detection and Reconstruction. *In:* Society of Petrophysicists and Well Log Analysts (SPWLA) Annual Logging Symposium, London, United Kingdom.

Anderton, R. 1985. Clastic facies models and facies analysis. *Geological Society, London, Special Publications*, 18(1), 31-47.

Ambrose, W. A.; Dutton, S. P. Loucks, R. G. 2016. Depositional systems, facies variability, and reservoir quality in shallow-marine reservoirs in the Eocene Upper Wilcox Group in Fandango Field, Zapata County, Texas. *Gulf Coast Association of Geological Societies Journal*, 5, 73-94.

Baldwin, J.L.; Bateman, R.M.; Wheatley, C.L. 1990. Application of a neural network to the problem of mineral identification from well logs. *The Log Analyst*, 31, 279–293.

Bormann, P.; Aursand, P.; Dilib, F.; Dischington, P.; Manral, S. (2020). FORCE Machine Learning Competition.

Cannon, S., 2016. *Petrophysics: A Practical Guide*. John Wiley & Sons, Ltd.

Dubois, M. K.; Bohling, G. C.; Chakrabarti, S. 2007. Comparison of four approaches to a rock facies classification problem. *Computers & Geosciences*, 33 (5), 599–617.

Dutton, S. P.; Ambrose, W. A.; Loucks, R. G. 2016. Diagenetic controls on reservoir quality in deep upper Wilcox sandstones of the Rio Grande delta system, south Texas. *Gulf Coast Association of Geological Societies Journal*, 5, 95-110.

Filippova K.; Ponomarenko P.; Gazaryan Z.; Molokanov S.; Vingalov V.; Shkondratov V. 2013. Geostatistical Inversion as a Tool for the Accurate Updates of the Hydrodynamic Models – Case Study. *In:* EAGE annual conference.

Hall, B., 2016. Facies classification using machine learning. *The Leading Edge*, 35, pp. 906–909.

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.

Pendrel, J.; Schouten, H. J. 2020. Facies – the drivers for modern inversions. *The Leading Edge*, 39, 102-109.

Ranganathan, C.; Johnston, J.; Jenson, F. 2021. U.S. Patent Application No. 17/265,589.

Rivera, D.; Melo, F.; Ranganathan, C.; Jenson, F.; Pérez, V. 2022. The impact of outlier detection and well log editing in determining P-impedance and Porosity. *In:* 11º Congreso de Exploración y Desarrollo de Hidrocarburos Evaluación de Formaciones.

Shier, D. E. 2004. Well log normalization: methods and guidelines. *Petrophysics*, 45(03), 268-280.

Wolf, M. ; Pelissier-Combescure, J. 1982. Faciolog - Automatic Electrofacies Determination. *In:* SPWLA 23rd Annual Logging Symposium.

Wong, P.M.; Jian, F.X.; Taggart, I.J. 1995. A critical comparison of neural networks and discriminant analysis in lithofacies, porosity and permeability predictions. *Journal of Petroleum Geology*, 18(2), 191-206.

Filename: SBGf_2023_Fandango
Directory: C:\Users\fmelo\OneDrive - GEOSOFTWARE LP\Research\2023_SBGf\PL_Fandango
Template: C:\Users\user\AppData\Local\Temp\template_2011.dot
Title:
Subject:
Author: user
Keywords:
Comments:
Creation Date: 1/27/2023 4:55:00 PM
Change Number: 35
Last Saved On: 5/10/2023 5:22:00 PM
Last Saved By: Felipe Melo
Total Editing Time: 1798 Minutes
Last Printed On: 5/10/2023 5:37:00 PM
As of Last Complete Printing
Number of Pages: 5
Number of Words: 3089 (approx.)
Number of Characters: 17608 (approx.)