THE IMPACT OF OUTLIER DETECTION AND WELL LOG EDITING IN DETERMINING P-IMPEDANCE AND POROSITY

Diana Rivera[1], Felipe Melo[1], Chiranjith Ranganathan[1], Fred Jenson[1], Verónica Pérez[1]

1: GeoSoftware. diana.rivera@geosoftware.com, felipe.melo@geosoftware.com, chiranjith.ranganathan@geosoftware.com, fred.jenson@geosoftware.com, veronica.perez@geosoftware.com

Keywords: Outlier detection, Log editing, Synthetic generation, Machine learning

RESUMEN

Presentamos un flujo de trabajo de edición de registros que utiliza la detección de valores anómalos para corregir problemas con registros de pozos. Este flujo de trabajo incluye análisis de datos, detección de valores anómalos para identificar datos de baja o mala calidad, generación de curvas sintéticas y reconstrucción de las curvas corregidas. El conjunto de datos contiene ocho pozos del Campo Amberjack, ubicado costa fuera en el Golfo de México. Los principales intervalos anómalos se identifican en los registros de densidad y sónico de onda compresional P en varios pozos. El objetivo de este trabajo es mejorar la calidad de los registros de pozos. La detección de valores anómalos se realiza utilizando algoritmos de aprendizaje automático no supervisados que genera una curva indicadora de valores anómalos donde se detectan desviaciones de la norma. Para corregir los registros, generamos curvas sintéticas para reemplazar muestras identificadas como valores anómalos. Para la generación de curvas sintéticas, el conjunto de datos de entrenamiento son los registros de seis pozos del modelo, seleccionando curvas que excluyen datos anómalos. Elegimos construir el modelo entrenando un regresor multi-lineal con los datos de los registros seleccionados. El modelo se utiliza para generar densidad y registros sónicos sintéticos en los pozos que requieren corrección. La reconstrucción de las curvas corregidas se realiza sustituyendo únicamente los valores catalogados como no confiables por las muestras correspondientes del registro sintético. Finalmente, analizamos los efectos del uso de registros medidos o sin modificar en los cálculos de Impedancia P y porosidad, para luego comparar los resultados de los cálculos con las curvas corregidas. Los resultados muestran cómo los cálculos con valores anómalos introducen errores en la interpretación y el impacto de este flujo de trabajo para producir resultados más precisos.

INTRODUCTION

Quality control of data and its validation is a first step in a reservoir characterization project. This step is necessary to establish that the loaded information represents what they are supposed to represent and that the data is numerically correct, precise and of usable quality for the purposes of the project, Cannon (2016). Well log processing workflows are usually a time-consuming repetitive task that can be slow due to the several iterations that may be needed to get a good result during data preparation. There might be three likely reasons for log editing: sticking tools, poor borehole conditions and noisy data, Cannon (2016). As a part of the workflow, if data is unreliable or missing, we might also need to generate a log in a section of a well using another log as input.

Machine learning applications in the Oil and Gas industry has become more popular in recent years (Dramsch 2020; Yu and Ma 2021). Traditional well log processing and interpretation workflows have been modified to include machine learning (ML) based methods to automate key workflows, such as zonation assignment, outlier detection, synthetic generation, log prediction, facies classification and formation property interpretation (Hall 2016; Akkurt *et al.* 2018; Wu *et al.* 2018; Ippolito *et al.* 2021; Jenson *et al.* 2021; Ranganathan *et al.* 2021).

The objective of this work is to address typical challenges of the well log processing workflow by implementing ML techniques to improve the efficiency and productivity of the traditional workflow in key steps such as outlier detection and synthetic data generation. For these steps, we follow the approach described by Ranganathan *et al.* (2021) for automatically detecting and correcting anomalies in log data, this method includes depth merging the synthetic curve and raw curve to replace erroneous values along the curves. We show that ML applications can make the petrophysicist's job of reviewing and editing a large amount of data more efficient and effective. This workflow produces quality assessments based on exploratory analysis of input data, flagging the wells and logs that need to be carefully reviewed by the expert. Our focus in this paper is the correction of bulk density (density) and compressional slowness (sonic) logs since they are used in P-impedance and porosity calculations. We use several metrics for quantitative analysis and visual inspection in the logs and crossplots for qualitative analysis. The results of the P-impedance and porosity curves computed with the raw data and the corrected data are compared and show dramatic difference.

METHODOLOGY

The dataset contains eight wells from the Amberjack Field, located offshore in the Gulf of Mexico (Fig. 1). The reservoir type is a clastic reservoir within a shelf-edge delta system. The available well logs are Gamma Ray, Density, Sonic, Resistivity and Neutron Porosity. Our methodology consists of outlier detection, synthetic generation, and replacement of outlier's values by high quality synthetics. We then compare logs computed with raw data against logs computed with corrected data. In the following, we will describe the steps to perform our workflow.

The first step of this workflow is the exploratory data analysis to identify wells with outliers. Outliers are defined as the data points that are inconsistent with the rest of the data, Akkurt *et al.* (2018). In Step 1, we perform a conformity check to verify that the basic logs exist in most wells as suggested by Ranganathan *et al.* (2021). Then, to identify wells with outliers, we use boxplot

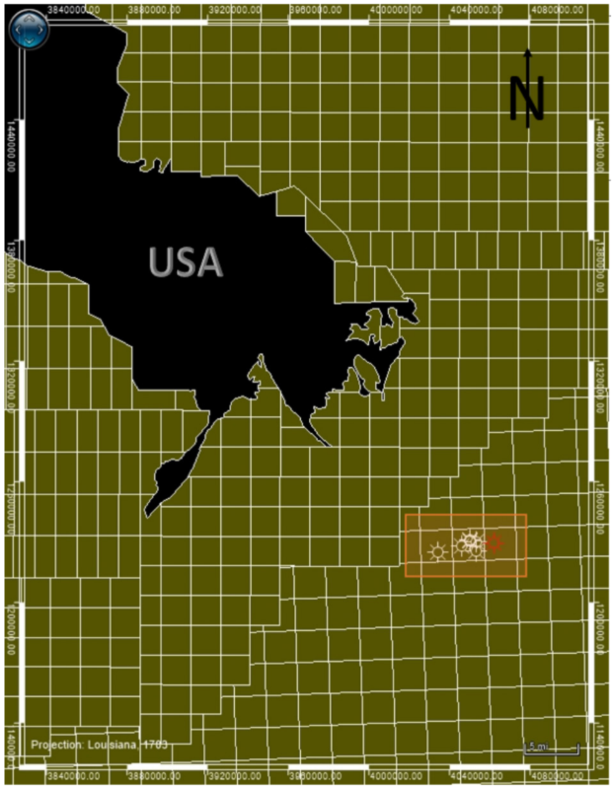

Figure 1. Location of the wells from the Amberjack Field in Gulf of Mexico, the area of study in represented by the orange area

whisker diagrams for visual and quick inspection of the density and sonic logs for each well. This analysis allows us to understand spread, variability, and skewness of the data. Values that differ significantly from the rest of the dataset might appear in the boxplot as separate and individual points beyond the whisker's lines. Wells identified to have questionable data are revised carefully in log plots, evaluating each curve, and confirming that outliers are present.

Step 2 is the application of an unsupervised ML algorithm, Ranganathan *et al.* (2021) to identify outliers in the wells and logs identified in Step 1. We analyze those wells and curves separately with the neutron log for the outlier detection. We identify the depths at which anomalies and flag them in an anomaly log that has binary values for outliers and inliers, that will be used for the synthetic generation. Subsequently, it is necessary to create a pay flag in a first pass interpretation. We use the logs Effective Porosity, Clay Volume and Water Saturation to generate the pay flag. It is important not to perform correction on pay zones, as the response of well curves due to the presence of hydrocarbons might look like outliers. Thus, the pay flag will preserve the original values in pay zones, even if it is identified as an outlier.

Then in step 3, we run the Jaccard similarity analysis to identify wells that would be good candidates for the synthetic generation, wells with high similarity scores will be the training group. In this case, all well logs are used for the similarity analysis.

In step 4, the well group selected from the previous step is used for the training of a multilinear regressor using the available logs with data from the inlier depths only. With this procedure, we ensured that all erroneous data are excluded from the model. We evaluate the model based on the coefficient of determination, coefficient of multiple correlation and the standard error estimate.

For the synthetic curve generation in step 5, we use the model from step 4 to predict the density and sonic curves in the wells identified with outliers in steps 1 and 2, across all depths.

The well log correction is made in step 6 performing a depth merging using the raw curve and the synthetic curve. The original values flagged as outliers are replaced with a high-quality synthetic. Thus, preserving the original values from pay zones using the pay flag.

Finally, we compute the P-impedance and porosity curves with the raw data and with the corrected curves. Then, we analyze the impact of outlier detection and synthetic generation in the process.

RESULTS

Figure 2a shows the boxplot whisker diagram from density log and Figure 2b shows the diagram from sonic log. Questionable data are identified in the density and sonic log curve for several wells.

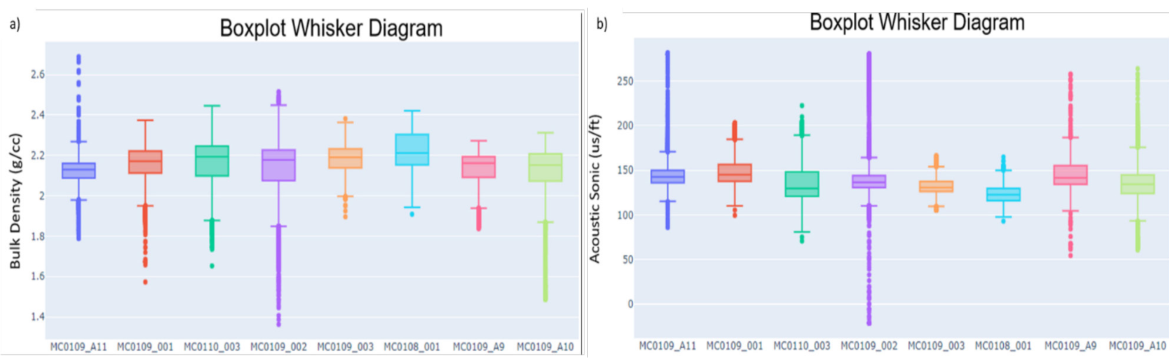

Figure 2. Boxplot whisker diagram for all wells in the project using: a) the bulk density and b) the acoustic sonic. The corresponding well names are indicated below each boxplot.

We perform outlier detection to flag the depths where outliers are present, and the quality control of the outlier identification was made using Neutron-Density and Neutron-Sonic crossplots. In addition, we checked the accuracy of the intervals detected as outliers in log plot for all wells with questionable data. However, we selected two of those wells with great number of anomalies present for further discussion of the results. Figure 3a and 3b show the Neutron-Density crossplots for wells MC109_002 and MC109_A10, respectively, and Figure 3c and 3d show the Neutron-Sonic crossplots from wells MC109_002 and MC109_A10, respectively. We identify these wells as the most problematic wells, and we analyze the correction of their curves at the end of this workflow. Notice, in Figure 3 how the ML algorithm successfully captured the outliers in the density and sonic curve using the neutron porosity log. We can use this ML algorithm in similar wells simultaneously for outlier detection, but in problematic wells that have been previously identified, the outlier detection should be performed separately as a different percentage of impurities might be needed for a proper detection of erroneous data.

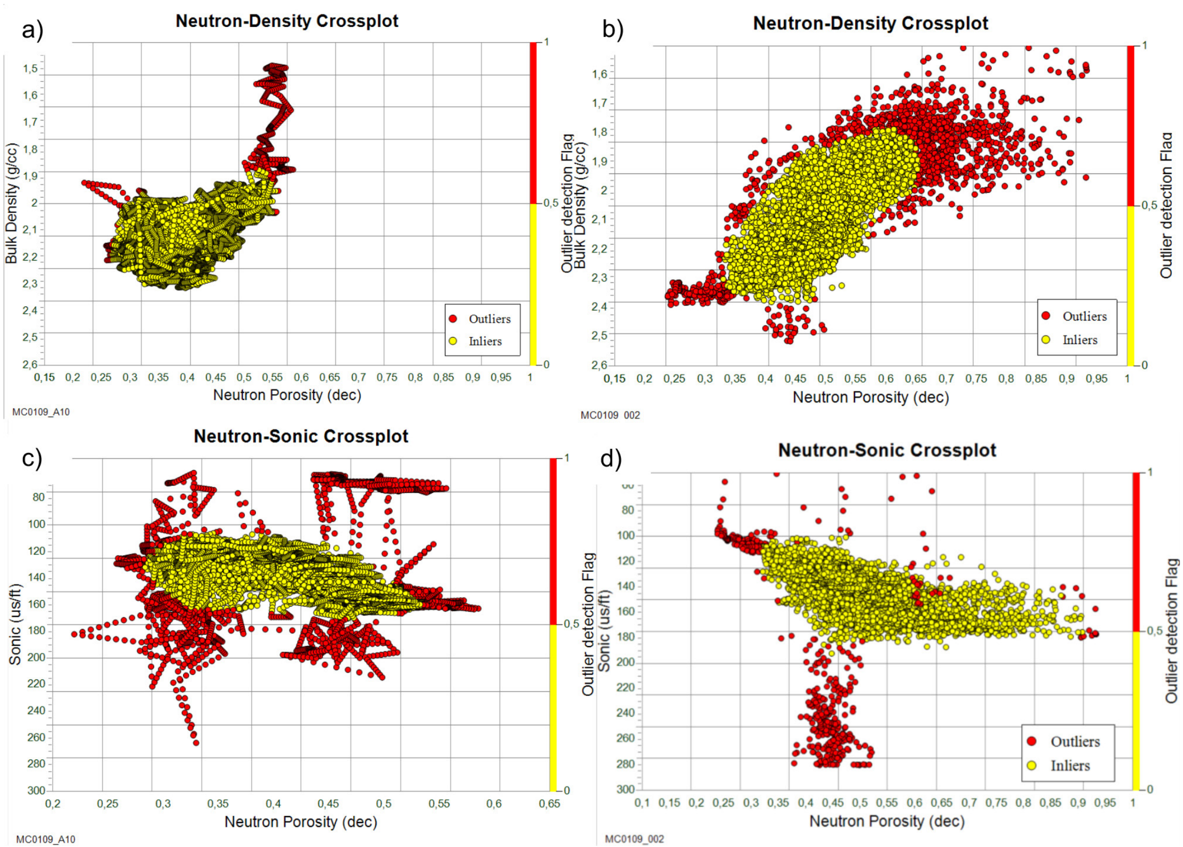

Figure 3. Neutron-Density crossplots from wells a) MC109_A10 and b) MC109_002. Neutron-Sonic crossplots from wells c) MC109_A10 and d) MC109_002. The anomaly flag log is red for outliers and yellow for inliers.

Figure 4a shows the Jaccard similarity scores and Figure 4b shows Overlap similarity scores. Well MC109_001 has lower Jaccard similarity scores for half of the wells in the dataset, leading to the decision of leaving this well outside of the training group for a better result. Next, we check the Overlap similarity to understand better in which cases one well might be a subset of another well, this might guide to the decision of including the subset well into the training group.

The Overlap score is high for most wells, however this result doesn't give us an idea of which well might be a subset of the other, Akkurt *et al.* (2018). We checked the relationship between pairs of wells and several wells prove to be a subset for well MC110_003, leading to the decision of leaving also that well outside of the training group for a better result. These two wells will be used to check the corrections as test wells.

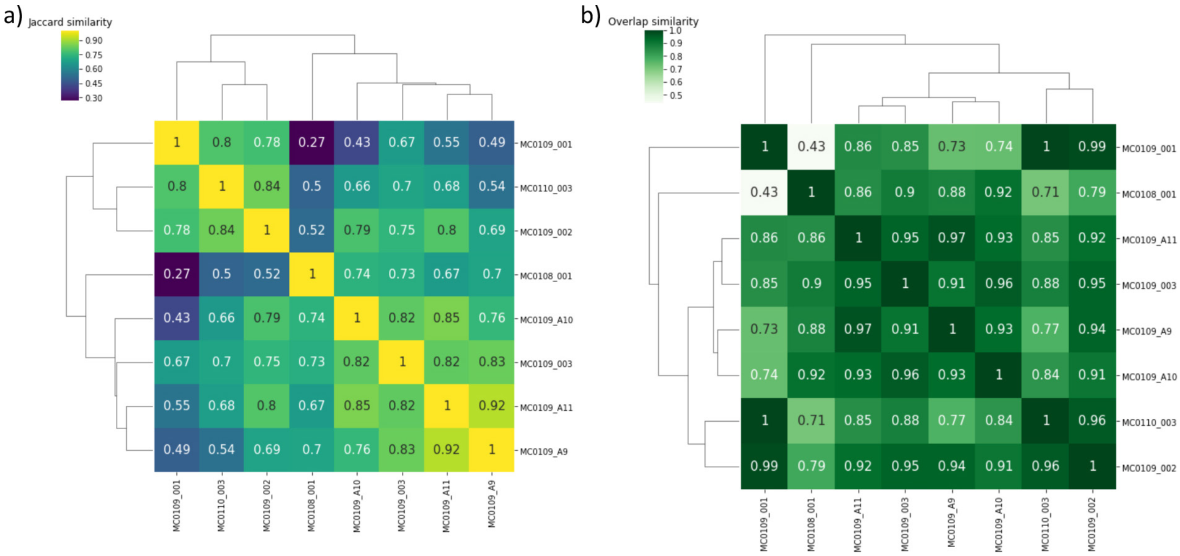

Figure 4. Similarity between wells. a) Jaccard Similarity scores and b) Overlap Similarity scores.

To correct the problematic curves, we replaced the outliers with the synthetic curves generated using six wells as model wells. The synthetic curves were generated with a multilinear regression algorithm using the available logs plus a depth curve we created based on the True Vertical Depth. We used the depth curve because shales are very important due to their thickness and lateral extension. Since the compaction has a great effect in offshore data and in unconsolidated rocks and it is valuable to model the effect of the compaction in the shale density, this process can be quite challenging given poorer quality logs due to rugosity and washouts, Akkurt *et al.* (2018). For the density model the values of the coefficient of determination, coefficient of multiple correlation and the standard error estimate were 0.75, 0.86 and 0.04, respectively; and for the sonic model the values were 0.73, 0.86 and 6.49, respectively. We correct the density log first to use this corrected log as another input in the synthetic sonic generation.

Figure 5a and 5b show the raw logs in black overlayed by the corrected logs in blue dashed lines for the test wells MC110_003 and MC109_001, respectively. The outlier flag is in red, and the pay flag is in green. In regions where the pay flag and the outlier flag are at the same depth the outlier is kept, for example in Figure 4b between 5500 and 6000 ft for the sonic log. In other regions where the pay flag is not presented and the values are defined as outliers, the intervals are successfully replaced with a high-quality synthetic, resulting in better well log curves in both test wells.

In the beginning of the workflow, the wells MC109_002 and MC109_A10 were identified as the most problematic wells. Now, we are going to evaluate the corrections in these wells. Figure 6a shows the well MC109_002 and Figure 6b shows the well MC109_A10, with the raw (black) and corrected (blue) sonic and density curves. For each well, the figures also show the logs Gamma Ray, Caliper, Pay Flag (green), Outlier Flag (red) and the computed P-impedance and porosity, with raw data (black) and with the corrected data (blue).

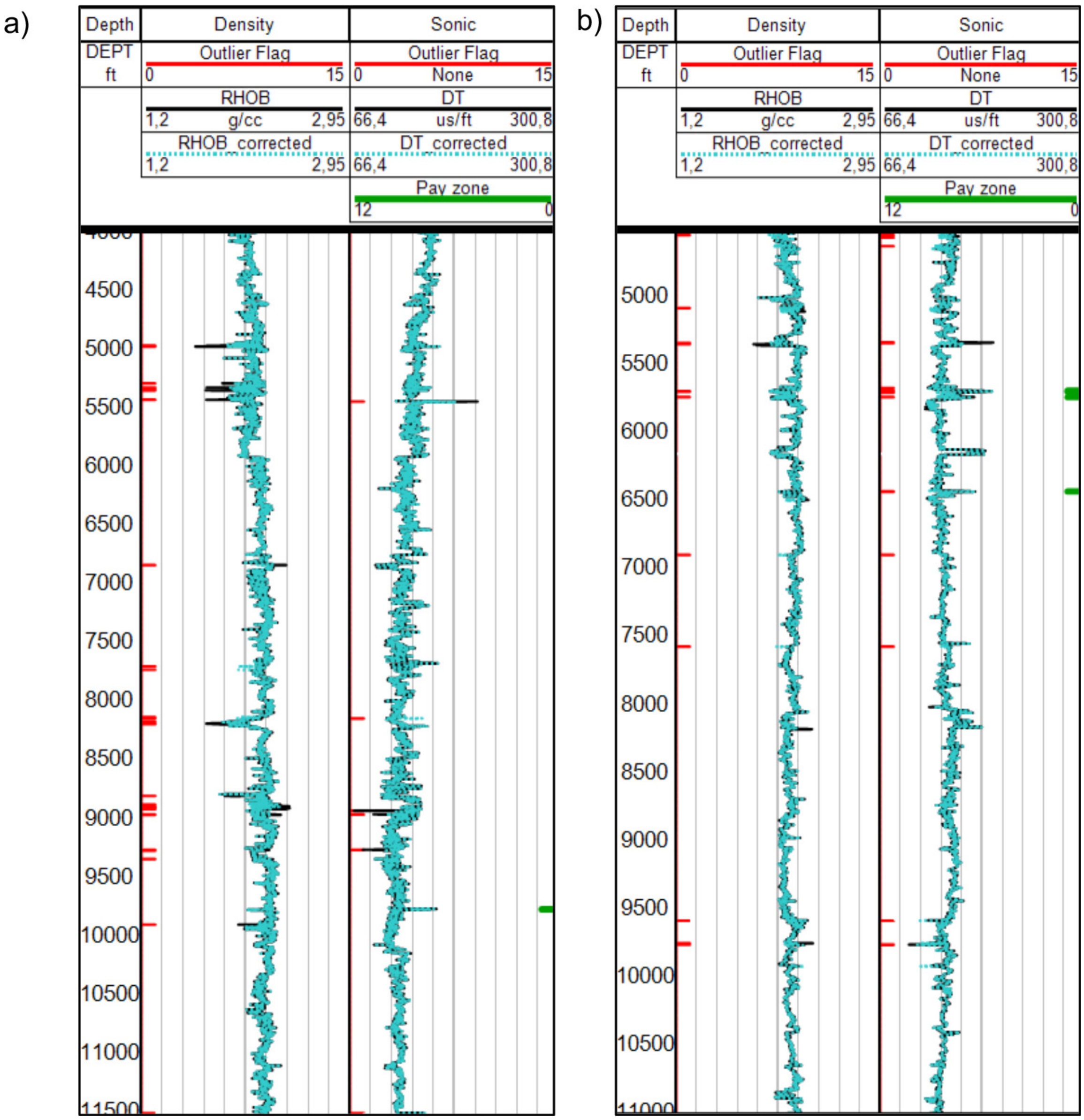

Figure 5. Comparison of raw (black) and corrected curves (blue) for the density in the first track and sonic log in the second track for wells a) MC110_003 and b) MC109_001. The outlier flag is in red, and the pay flag is in green.

The caliper in well MC109_002 shows that from 2600 to 4600 ft there are challenging borehole conditions encountered, resulting in a noisy raw density log. The outlier detection captured properly this washout section together with some minor intervals along the density curve and for the sonic log, fewer anomaly values are detected. Comparing the corrected curves (blue) against the raw curves (black), we see above 4600 ft lower density values in the washout section in the corrected curve, this came from the synthetic curve generated using the multilinear regressor. The sonic and density corrected curves show an outstanding behavior, all the spikes present in the log curve were corrected successfully by replacing those intervals with the high-quality synthetic (Fig. 6a). We can also see how the uncorrected density and sonic logs will derive in a bad P-impedance and porosity computations if outliers are not handled. Below 11000 ft we can see high compressional slowness and without any correction, this might translate into a low P impedance interval.

Well MC109_A10 shows erroneous intervals for the sonic and density logs deriving into erroneous P-impedance and porosity values above 5000 ft. In addition, around 8000 ft the same erroneous intervals are observed leading to wrong P-impedance computations. The pay flag has an important role to avoid corrections in interest zones also flagged as outliers. In this case, we didn't have a caliper log to have a better understanding of borehole conditions. You can see the ML algorithms are doing a good job detecting the outliers and the high-quality synthetic curve generated to correct those erroneous intervals.

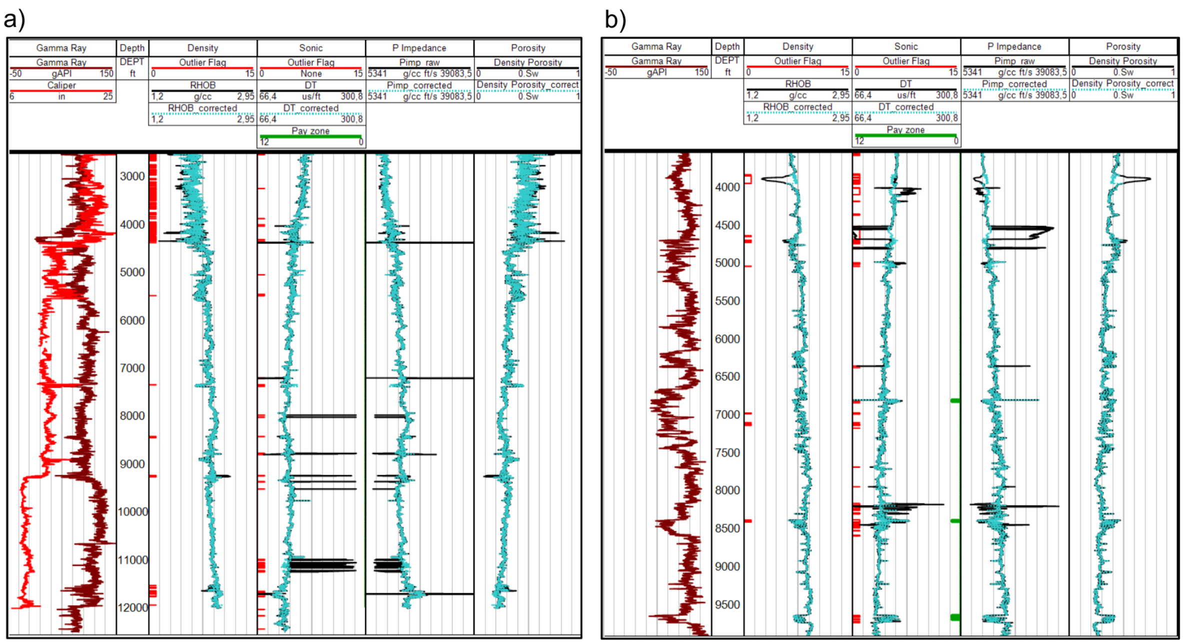

Figure 6. Comparison of raw (black) and corrected (blue) curves for the density in second track, sonic log in third track, P impedance in fourth track and porosity in fifth track; for wells a) MC109_002 and b) MC109_A10. The outlier flag is in red, and the pay flag is in green.

CONCLUSIONS

We presented a workflow using machine learning to identify, edit and create high-quality well logs. The corrected logs show excellent results where outliers are properly and quickly handled using ML algorithms. The erroneous computation of these logs can lead to incorrect results in reservoir characterization. The automated ML workflow includes data analysis, outlier detection, synthetic generation and editing of well logs. The main steps of the workflow are: exploratory analysis to identify wells or logs with outliers and highlight those log curves that need correction;

creation of a pay flag in a first pass interpretation to prevent editions in pay zone; using a ML algorithm in the previously identified curves to capture outliers; similarity analysis to compare wells and chose inputs for the synthetic generation and ML algorithm for the prediction of the synthetics and logs correction by replacing intervals with outliers by high-quality synthetics. We also demonstrate the effect of log editing for crucial parameters used in reservoir characterization such as P-impedance and porosity. This automated log editing workflow improves time efficiency, consistency, and objectivity for evaluating large and complex datasets, resulting in more time available for the petrophysicist to concentrate on interpretation and not spending a large amount of time on log processing.

ACKNOWLEDGEMENT

The authors wish to thank Talos Energy for permission to show these data and GeoSoftware for supporting the work.

REFERENCES

Akkurt, R., Conroy, T., Psaila, D., Paxton, A., Low, J. and 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, Society of Petrophysicists and Well Log Analysts (SPWLA) Annual Logging Symposium, London, United Kingdom.

Cannon, S., 2016. Quality Control of Raw Data. Petrophysics: A Practical Guide. John Wiley & Sons, Ltd., 204pp.

Dramsch, J.S., 2020. 70 years of machine learning in geoscience in review, Advances in Geophysics, 61, pp. 1-55.

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

Ippolito, M., Ferguson, J. and Jenson, F., 2021. Improving facies prediction by combining supervised and unsupervised learning methods, Journal of Petroleum Science and Engineering, 200, 108300.

Jenson, F., Ranganathan, C., Xiuping, S. and Holden, T., 2021. Machine learning for predicting stochastic fluid and mineral volumes in complex unconventional reservoirs, World Oil, April, pp. 45-47.

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

Wu, P., Jain, V., Kulkarni, M. and Abubakar, A., 2018. Machine learning–based method for automated well log processing and interpretation, Society of Exploration Geophysicist SEG International Exposition and 88th Annual Meeting, Anaheim, California.

Yu, S. and Ma, J., 2021. Deep learning for geophysics: Current and future trends, Reviews of Geophysics, Vol 59, Issue 3, e2021RG000742.