SPE Brazil Flow Assurance Technology Congress
Rio de Janeiro, Brazil, November 06 – 08 2024

Impact of Fluid Properties on Flow Assurance Events in Offshore Oil Production: A Data-Driven Study

Afrânio Melo[1*], Rogério Leite[2], Márcia Khalil [1]
[1] Centro de Pesquisas Leopoldo Américo Miguez de Mello – CENPES, Petrobras, Brasil.
[2] Exploração e Produção, Petrobras, Brasil.

*afranio.melo@petrobras.com.br

Abstract

Assumptions regarding the criticality of flow assurance events are typically based on operational experience and laboratory studies. However, with the availability of large amounts of online and historical field data and advancements in data science methodologies, there are opportunities to utilize this data to establish more reliable criticality criteria. This study aims to analyze fluid property data from various offshore operational units in order to gain new insights into the occurrence and factors influencing flow assurance events. Statistical techniques, such as non-parametric tests for comparing distributions, are employed to assess different scenarios. Additionally, machine learning models based on GradientBoosting are trained, resulting in predictive models that improves risk assessment compared to traditional methodologies. The results demonstrate the value of integrating data from multiple production units to enhance our understanding of the risk associated with flow assurance events.

Keywords
Flow assurance; exploratory data analysis; machine learning

Introduction

One of the emerging trends in the industry is the growing use of digital technologies to create value, as demonstrated by concepts such as Industry 4.0 and digital transformation [1]. A key application of this trend is the use of data-driven techniques, including machine learning and statistical models, to monitor processes, evaluate risks, and minimize losses [2]. In the energy sector, flow assurance management faces critical challenges that can significantly benefit from advancements in digital transformation [3].

A substantial body of research focuses on applying data-driven techniques to address flow assurance challenges. Lal et al. [4] compiled a comprehensive list of works that apply machine learning methodologies to tackle various issues, such as gas hydrates [5,6], wax [7], asphaltenes [8] and scaling [9]. While most of these studies were conducted using experimental laboratory data, some incorporated pilot-scale and actual field measurements [5]. Notably, Vargas et al. [10] developed a public dataset containing real and synthetic data featuring anomalies in oil wells, including flow assurance challenges such as hydrate formation and severe slugging.

Despite the growing interest in this field, to the best of our knowledge, the literature lacks integrated studies that comprehensively analyze multiple flow assurance issues within a broad context. The present work aims to compile and analyze fluid property information from various operational units and over an extended period, encompassing multiple flow assurance events, with the goal of enhancing criticality criteria for different events. The remainder of this article is organized as follows: Section 2 describes the methodology employed, Section 3 presents the results, and Section 4 concludes the article.

Methodology

Fluid data from three operational units of Petrobras (referred to as A, B, and C) were collected from various systems and combined into a consolidated dataset. The dataset includes Basic Sediment and Water (BSW), Gas-to-Liquid Ratio (RGL), various water ionic concentrations, and metadata associated with each well, such as platform and field. The wells were divided into two categories: those that did not experience flow assurance losses (labeled as 0) and those that did experience flow assurance losses (labeled as 1). Three types of flow assurance issues were analyzed: hydrates, wax, and scale.

For the statistical analysis, two strategies were used. Firstly, measures of central tendency and dispersion were calculated. Specifically, the median and mean absolute deviations (MAD) were chosen as they are robust statistical measures that are less affected by outliers compared to the mean and standard deviations, respectively. Secondly, a distribution analysis was conducted by plotting histograms to visually compare the distributions of the two groups (0 and 1). Additionally, a statistical test was performed to determine if there was a significant difference between the two distributions. The Mann-Whitney U rank test was chosen for this purpose, which tests the null hypothesis that the two distributions being compared are equal [11].

For machine learning modeling, the selected approach was a histogram-based Gradient Boosting Classification Tree from the scikit-learn Python package [12]. This estimator has built-in support for handling missing values, which is advantageous for this problem. The objective of the model is to predict the criticality of events, which can be represented as probabilities. However, boosted tree-based models often suffer from distortions in predicted probability distributions. To address this issue, the classifier needed to be calibrated, which involves mapping the output of the model to a calibrated probability ranging from 0 to 1 [13].

Results and Discussion

The results of the statistical analysis are presented in Figs 1 to 6. Figure 1 highlights the significant differences in statistical summaries when comparing calcium concentration in water for hydrate formation. This finding is unexpected since the commonly assumed variable contributing to hydrate formation criticality is BSW (Basic Sediment and Water). Moreover, while sodium chloride is typically analyzed for its influence on hydrate formation [14], the data suggests that other ions may be even more significant in this context. Additionally, the behavior across operational units varies, with operational unit A showing a more pronounced difference in calcium concentration medians between the groups 0 and 1. Figure 2 displays histograms for comparing distributions, with the p-values resulting from the Mann-Whitney U rank test shown at the top of each plot. A lower p-value means a stronger indication that the null hypothesis of equal distributions is false. Once again, calcium concentration appears to be the most effective discriminator, as it consistently exhibits low p-values across all three operational units.

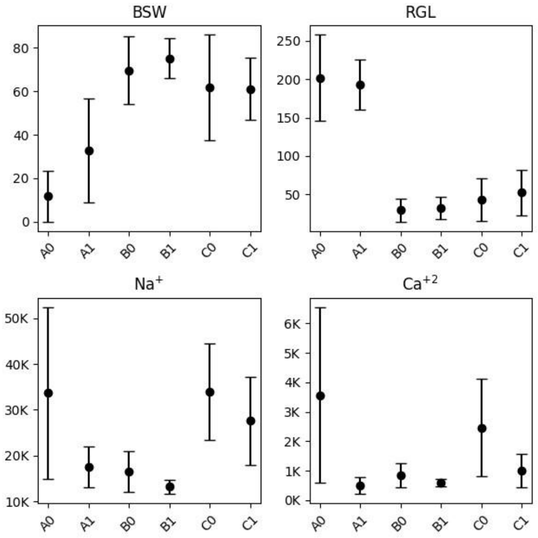

Figure 1. Medians and Mean Absolute Deviations (MAD) for four variables across three operational units (A, B, and C) under two conditions (0: no loss, and 1: loss due to hydrates).

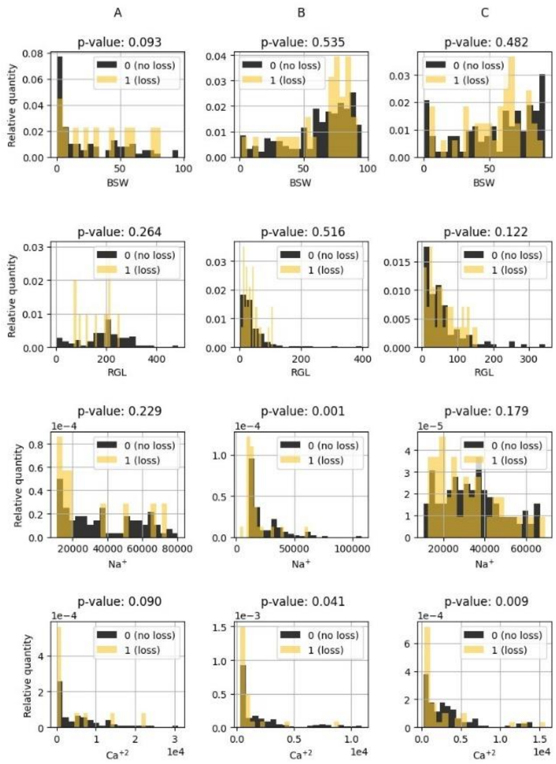

Figure 2. Histograms for four variables across three operational units (A, B, and C) under two conditions (0: no loss, and 1: loss due to hydrates).

Regarding wax formation, there is a reversal in the behavior of operational unit B compared to units A and C. The results of the Mann-Whitney U rank test indicate that RGL (Gas-to-Liquid Ratio) is the most effective discriminator between scenarios with and without loss in general. For scale formation, due to the limited number of recorded events, histograms are plotted only for units B and C. As expected, sodium concentration emerges as the most prominent discriminator for this type of event.

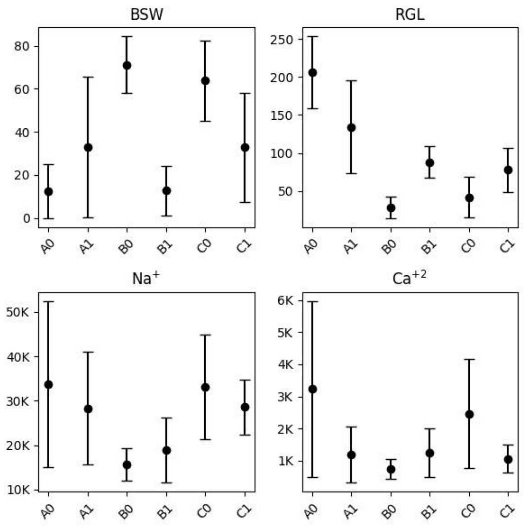

Figure 3. Medians and Mean Absolute Deviations (MAD) for four variables across three operational units (A, B, and C) under two conditions (0: no loss, and 1: loss due to wax).

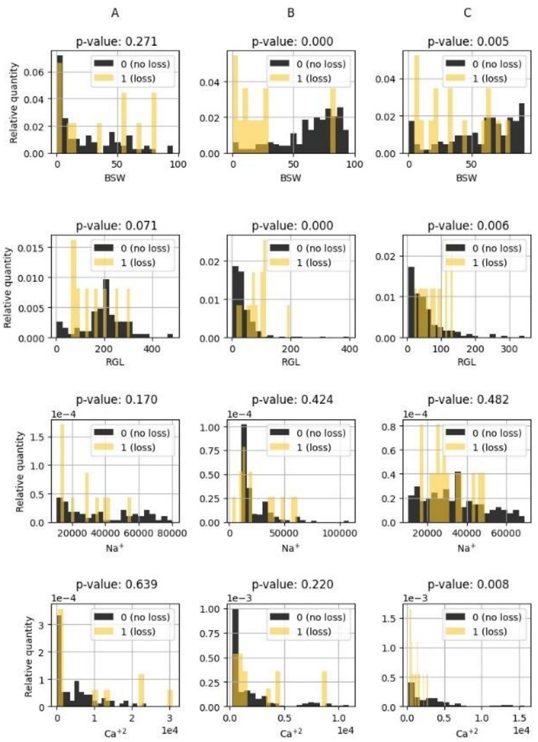

Figure 4. Histograms for four variables across three operational units (A, B, and C) under two conditions (0: no loss, and 1: loss due to wax).

It is important to exercise caution when interpreting results solely based on data-driven analyses, as correlation alone does not indicate causation. While these findings can offer valuable evidence, they should be viewed as a preliminary stage for further investigation. This may involve conducting additional experimental and theoretical studies to explore specific findings in more depth, such as the significant impact of calcium concentration on hydrate formation. Additionally, extending this analysis to include data over varying time periods could provide even more robust evidence regarding the impact of different variables on flow assurance events.

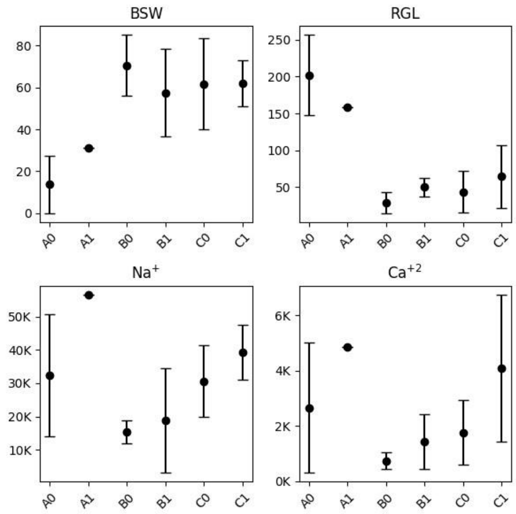

Figure 5. Medians and Mean Absolute Deviations (MAD) for four variables across three operational units (A, B, and C) under two conditions (0: no loss, and 1: loss due to scale).

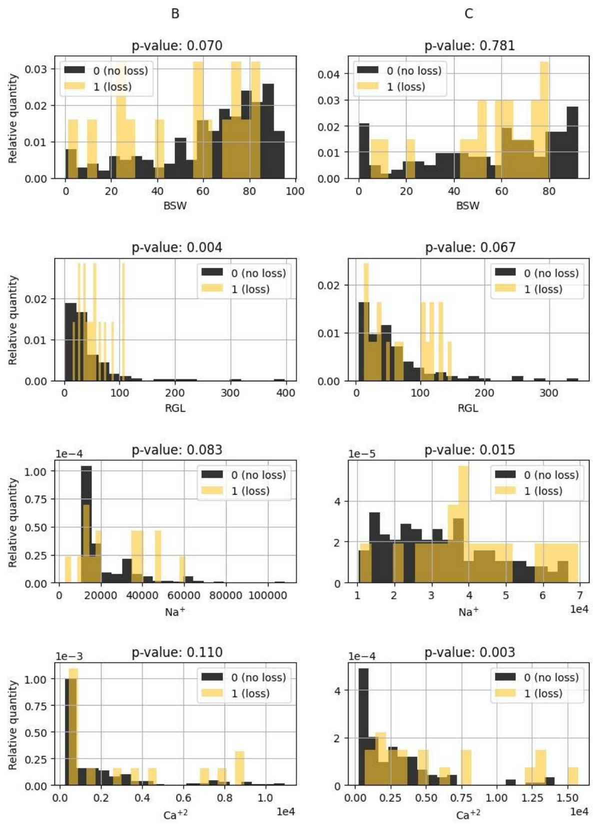

Figure 6. Histograms for four variables across three operational units (A, B, and C) under two conditions (0: no loss, and 1: loss due to scale).

The machine learning modeling results are shown in Figure 7, which represents the criticality values calculated by the machine learning algorithm for the test datasets. The data points representing wells with flow assurance events have a higher percentage of values above the threshold (arbitrarily set at 0.15 in this case). This observation is further supported by Tab. 1, which presents the precision lift for each event. The precision lift measures how much the model's accuracy in identifying positives exceeds the overall proportion of positives in the dataset. In simpler terms, precision lift tells how much more accurate the model is in identifying true positive events compared to just guessing based on the overall rate of these events in the dataset. In all three cases, there is an improvement in precision compared to the global baseline. This indicates that the models effectively assign higher risks to wells with flow assurance events.

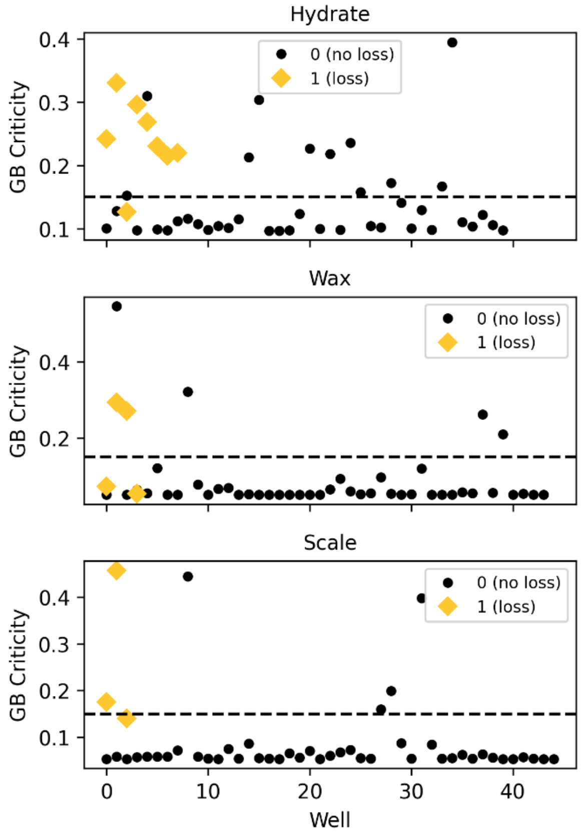

Figure 7. Gradient Boosting criticalities for test samples.

Table 1. Results from machine learning classification.

Event	Precision lift
Hydrate	2.33
Wax	2.63
Scale	6.86

Conclusions

The present study demonstrates the potential of data-driven analysis in enhancing the understanding of flow assurance phenomena in offshore operations. The integration of data from multiple production units offered insights into the influencing factors of flow assurance events. By conducting statistical analysis on fluid property data from different operational units and training machine learning models, new criticality criteria were established, resulting in improved risk assessment. This research highlights the significance of incorporating online and historical field data to inform decision-making and enhance flow assurance strategies in the oil and gas industry.

Responsibility Notice

The authors are the only responsible for the paper content.

References

[1] Udugama, I. A.; Bayer, C.; Baroutian, S.; Gernaey, K. V.; Yu, W.; Young, B. R. *Educ. Chem. Eng*. 39, 94–107, 2022.
[2] Melo, A.; Câmara, M. M.; Pinto, J. C. *Processes* 12(2), 251, 2024.
[3] Kumar, A. Energy Fuels 37(12), 8142-8159.
[4] Lal, B.; Bavoh, C.; Sayani, K. Machine Learning and Flow Assurance in Oil and Gas Production. Springer, 2023.
[5] Qin, H.; Srivastava, V.; Wang, H.; Zerpa, L. E.; Koh, C. A. *OnePetro*, 2019.
[6] Xu, H.; Jiao, Z.; Zhang, Z.; Huffman, M.; Wang, Q. *Comput. Chem. Eng*. 151, 107358, 2021.
[7] Amar, A.; El-Gohary, W.; El-Sherif, M.; El-Adl, G. *Petroleum* 8, 75–88, 2022.
[8] Gesho, M.; Chaisoontornyotin, W.; Elkhatib, O.; Goual, L. *Ultramicroscopy* 217, 113074, 2020.
[9] Al-Hajri, N. M.; Al-Ghamdi, A.; Tariq, Z.; Mahmoud, M. *SPE Prod. Oper.* 35, 987–1009, 2020.
[10] Vargas, R. E. V.; Munaro, C. J.; Ciarelli, P. M.; Medeiros, A. G.; Amaral, B. G.; Barrionuevo, D. C.; Araújo, J. C. D.; Ribeiro, J. L.; Magalhães, L. P. *J. Pet. Sci. Eng.* 181, 106223, 2019.
[11] Mann, H. B.; Whitney, D. R. *Ann. Math. Statist.* 18, 50–60, 1947.
[12] 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, É. *J. Mach. Learn. Res.* 12, 2825–2830, 2011.
[13] Niculescu-Mizil, A.; Caruana, R. *Proc. 22nd Int. Conf. Mach. Learn.* 625–632, 2005.
[14] Teixeira, A.; Valim, L. S.; Sum, A. K.; Barreto, A. G.; Tavares, F. W. *Energy Fuels* 38, 3609–3621, 2024.