XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Real-Time Rate of Penetration Prediction Analysis using LSTM Networks Under a Continuous Learning Scenario

Antonio P. A. Ferro[1], Erasmo A. B. Silva[1], Francisco de A. V. Binas Jr.[1], Lucas P. de Gouveia[1], Aline da S. R. Barboza[1], Aliel F. Riente[2]

[1]*Laboratory of Scientific Computing and Visualization, Federal University of Alagoas*
Avenida Lourival Melo Mota, S/N, Tabuleiro dos Martins, 57072-970, Maceió-AL, Brazil
antonio.ferro@lccv.ufal.br, erasmo.bezerra@lccv.ufal.br, francisco.junior@lccv.ufal.br, lucasgouveia@lccv.ufal.br, aline@lccv.ufal.br
[2]*Petróleo Brasileiro, CENPES*
Avenida Horácio de Macedo, 950, Cidade Universitária, 21941-915, Rio de Janeiro - RJ, Brazil
aliel@petrobras.com.br

Abstract. Recent technological advancements in data transmission and acquisition systems for drilling operations, along with the growing global energy demand, have led the Oil & Gas industry to pursue data-driven solutions in real-time to improve performance and reduce well construction costs. In this context, the Rate of Penetration (ROP) serves as a key drilling performance metric, reflecting the effective speed of the drill string as the bit penetrates the rock formation. Predictive models for ROP can be employed in real time to suggest appropriate operational parameters that optimize ROP, potentially reducing drilling time and overall operating costs. This work investigates the application of Long Short-Term Memory (LSTM) networks for ROP prediction within a continual learning scenario based on drilling data. LSTM networks are well-suited for this task due to their ability to model dependencies in sequential data. The results obtained using the LSTM model are compared with benchmark results from Multilayer Perceptron (MLP) networks. In two simulated scenarios, the models are incrementally trained on new data received during drilling to predict ROP in subsequent intervals. Given the transient nature of drilling, abrupt changes in data distribution are expected and may significantly impact predictive performance. Therefore, the model's performance in the test intervals is analyzed in relation to statistical characterizations of the newly received data. The dataset used comprises public data from three wells in the Volve field in the North Sea. It includes both operational parameters and lithological information, the latter being particularly important for analyzing the results obtained in this study. Model performance is evaluated using the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The results demonstrate the potential of the adopted strategies using the LSTM model. It is expected that the comparative analyses presented in this study will contribute to key aspects of developing data-driven solutions for real-time drilling applications.

Keywords: Drilling, ROP Prediction, Real-Time, Deep Learning, LSTM

1 Introduction

Drilling oil wells is a complex and costly process, particularly in offshore deep reservoirs, requiring careful risk analysis and mitigation. The advent of real-time monitoring has enhanced control and the early detection of operational issues. According to Xue [1], data processing now plays a crucial role, providing both technical and technological benefits. In this context, digital twin systems have emerged as promising solutions for drilling monitoring, often employing AI for predictive purposes [2], and are continuously updated to improve the virtual representation of the physical system.

The Rate of Penetration (ROP) — the speed at which the drill bit advances — is influenced by controllable parameters such as Weight on Bit (WOB), Revolutions per Minute (RPM), and drilling fluid Flow Rate (FR), as well as by lithology characteristics. Accurate ROP models play an important role in real-time monitoring, enabling recommendations that maximize drilling speed while maintaining well integrity and reducing costs. Several studies have applied Machine Learning (ML) to this task, including Soares and Gray [3], Barbosa et al. [4], Tunkiel et al. [5], Elmgerbi et al. [6], Cheng et al. [7], and Ferro [8].

This study applies Long Short-Term Memory (LSTM) networks for ROP prediction using public data from

three wells in the Volve oil field [9]. The dataset, enriched with drilling reports, real-time measurements, and lithological descriptions was processed and aligned by depth (described in Ferro [8]). A continuous learning approach simulating real-time data streams is adopted. The approach involves splitting the three wells into intervals and retraining with data batches to make predictions for the next intervals. Two strategies for using these data in training the LSTM model are compared. Given that LSTM networks are designed to capture sequential dependencies, unlike traditional models such as the Multilayer Perceptron (MLP), results from both models are also analyzed. The prediction error curves for each well are analyzed, and the impact of different input features is evaluated.

2 On Real-Time prediction of ROP using Deep Learning

As noted by Barbosa et al. [4], ROP modeling initiatives have a long history, traditionally relying on analytical formulations and multivariate linear regression. More recent studies [3, 5, 7] have shown that classical ML algorithms and deep learning networks outperform these traditional approaches. Commonly used features include WOB, RPM, Depth, Fluid Flow Rate (FR), and Mud Weight [4], with performance further enhanced when rock property measurements such as *Gamma Ray* and *Delta-T* are incorporated, when available in real time [3, 5].

The main application of ROP prediction is drilling optimization, often coupling predictive models with optimization algorithms. For example, Soares and Gray [3] employ Multilayer Perceptron (MLP) and Random Forest (RF) models based on WOB, RPM, and FR, while Tunkiel et al. [5] use XGBoost, AdaBoost, and other ML methods. A particularly relevant strategy for real-time scenarios is continuous learning: after drilling an initial well interval, measured data are used to train a model that predicts ROP in subsequent intervals [3, 5], as illustrated in Fig. 1. As drilling advances, new data may be incorporated into the model.

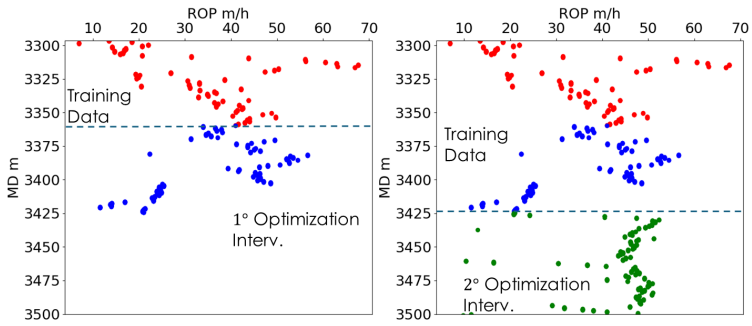


Figure 1. Illustrating the data stream for ROP prediction in a continuous learning scenario and its application

Because the drilling process is complex and involves multiple factors, the relationship between the main drilling variables and the resulting ROP is nonlinear, not yet fully understood, and subject to change during operations. As noted by Cheng et al. [7], achieving accurate ROP predictions remains a considerable challenge due to the non-trivial interdependencies among the relevant drilling variables. Recent research has increasingly focused on deep learning models for this task, with growing interest in architectures such as LSTM, given the inherently sequential nature of the drilling process, as well as in approaches incorporating attention mechanisms [7]. An important consideration in data streams (continuous, high-volume sequential data) is that the underlying data distribution may shift over time. Detecting these shifts involves identifying changes in the statistical properties of recent data, which can have a significant impact on model performance.

2.1 Brief overview of relevant techniques

Long-Short Term Memory (LSTM):. LSTM is a neural network architecture specifically designed to handle sequential data and time series, representing an advancement over traditional recurrent neural networks (RNNs). LSTM networks are capable of selectively retaining or discarding information through gating mechanisms, thereby effectively capturing long-term dependencies within data. The architecture of an LSTM cell comprises an internal cell state, which preserves long-term information (C_t), and a hidden state (h_t), responsible for modeling short-term dependencies in the sequences. The three gating mechanisms are: the forget gate, which determines which information should be removed from the cell state; the input gate, which regulates the incorporation of new information into the memory; and the output gate, which controls the information propagated to the next time step. These gates employ activation functions such as the sigmoid and hyperbolic tangent (tanh) to selectively filter and update the cell states. This iterative process continues until all LSTM cells have processed all input sequences [10].

Population Stability Index (PSI): According to Yurdakul and Naranjo [11], the PSI is a statistical measure of how much a variable has changed its distribution over time. This metric is used, for example, to detect the so-called Data Drift phenomenon in scenarios involving continuous data streams and the deployment of machine learning models in production. The expression is given by the eq. (1).

$$PSI = \sum_{i=1}^{B} (\hat{p}_i - \hat{q}_i) \cdot (\ln \hat{p}_i - \ln \hat{q}_i) \tag{1}$$

PSI measures the magnitude of the change in a distribution relative to the original distribution. $\hat{p}_i$ corresponds to the proportion of the current population, while $\hat{q}_i$ corresponds to the proportion of the expected or reference population (proportion of values within a prescribed bin). The logarithm of the ratio between the proportions indicates the magnitude of the difference.

3 Methodology

3.1 The Dataset

The public dataset released by Equinor [9] contains data from wells in the Volve oil field. Among the available operational data are geophysical logs, including measurements of drilled rock formation properties, such as compressional Delta-T (DT), which records the transit time of mechanical waves and is relevant for characterizing parameters such as Unconfined Compressive Strength (UCS). The data were processed in Ferro [8], and the wells analyzed in this study are detailed in Tab. 1.

Table 1. Relevant information about the three analyzed wells

Wells	Points	Initial MD	Final MD	Bit Type
F05	5459	2931.17	3765.91	PDC
F14	5457	2780.28	3465.85	PDC
F04	4474	2750.06	3483.86	PDC

Table 2 presents the set of features used in the investigated models. The different features compared in this study help identify which variables truly contribute to improving model performance.

Table 2. Different sets of features

Features	Measured Parameters
Set 1	WOB, RPM, FR
Set 2	WOB, RPM, FR, DT
Set 3	WOB, Torque, FR, DT
Set 4	WOB, RPM, Torque, FR, DT

3.2 Model Design, Training Process and Performance Evaluation

Each well was divided into intervals so that each data segment contained 500 points. This was considered suitable for training and corresponded, on average, to sections of 79 m, 65 m, and 83 m for F05, F14, and F04, respectively. As previously mentioned, the two continuous learning approaches used are illustrated in Fig. 2: one in which the training data are accumulated, and another in which the training data come from the most recent interval.

The metrics used in this study are the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE), as defined in eq. (2). Both share the same unit, and when errors are uniform, their values tend to be similar.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - h_i| \,, \qquad RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - h_i)^2} \tag{2}$$

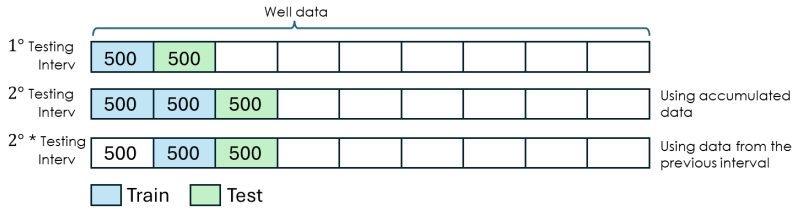


Figure 2. Detail of the two ROP prediction strategy through continuous learning

In this work, PyTorch was used to implement the LSTM models. Data were split into a 10% validation set, preserving the temporal sequence; the validation set corresponds to the final portion of the training data. An attention mechanism was incorporated into the LSTM models, enabling the network to focus on specific parts of the input during prediction, which is a common approach for such architectures [7, 10]. The attention mechanism applied corresponds to a simple additive attention. No hyperparameter tuning was performed; instead, the LSTM topology was defined based on empirical analysis. Continuous retraining with computationally complex models, such as LSTM, may hinder real-time ROP prediction. Regularization techniques were included to mitigate overfitting, such as early stopping and L2 regularization (weight decay equal to 0.01). Table 3 presents important model specifications.

Table 3. Model topology and important specifications

Parameter	Value
Hidden Size	64
Number of Recurrent Layers	1
Sequence Length	5
Optimizer	Adam
Loss Function	Mean Square Error (MSE)
Learning Rate	0.0001
Epochs	Early Stopping (Patience 15)
Batch Size	60

Given the ability of LSTM networks to model sequential dependencies, comparing them with models such as MLPs is relevant, as it allows assessing the influence of the sequential nature of the phenomenon on the data. In this comparison, results from Ferro [8] were considered, where MLPs with hyperparameter tuning (using GridSearchCV) were employed for each training.

4 Results and Discussion

Figure 3 shows the mean error metrics computed for each test interval of every well. LSTM 1 represents predictions for the next interval using models trained with accumulated data, while LSTM 2 represents ROP predictions using only data from the immediately preceding interval. Figure 3 also presents the MLP results trained with accumulated data, as reported in Ferro [8]. As previously mentioned, MLP training involved hyperparameter tuning; thus, no single topology was used for all predictions. The results for LSTM 1 and LSTM 2 were obtained using the architecture and specifications described in Tab. 3.

Overall, LSTM 1 shows comparable performance to the MLP model, although the refined MLP achieves slightly lower errors. When examining the average error metrics per well, we observe close agreement between models even without LSTM hyperparameter tuning. Notably, predictions for well F14 demonstrate lower RMSE values compared to MLP. Well F05 emerges as the most challenging case across all scenarios (showing the largest discrepancies between metrics), suggesting intervals with greater prediction difficulty that yield higher errors. The MLP exhibits performance improvement with additional features, where feature Set 4 proves most representative on average. Conversely, for LSTM models, Set 3 yields better results with lower average errors. This result may indicate model sensitivity to increased dimensionality or the presence of possible redundancy between Torque and RPM. The incorporation of attention mechanisms in LSTM models appears to enable more effective feature utilization compared to MLP architectures, as evidenced by reduced prediction errors. This improvement particularly

highlights the positive contribution of Torque and DT parameters to model performance.

It is observed that LSTM 2 generally exhibits smaller errors compared to LSTM 1, indicating that training the model with accumulated data is disadvantageous in the investigated LSTM cases. It is possible to identify that, overall, the prediction errors with the LSTM 2 model are on average smaller than those with the refined MLP. For instance, the average prediction for wells F05 and F14 was better when using LSTM 2. This result suggests that LSTM 2 is able to generalize better with limited information. Furthermore, the comparison between the results of MLP and LSTM 1 (which on average show similar results) indicates that long sequential dependency is not dominant. The fact that LSTM 1 and MLP are close in performance suggests an advantage in using the LSTM model, as the results were comparable without requiring refinement, and the sequential modeling of LSTMs favored, for example, the prediction of well F05. In this sense, it is also noted that the use of the LSTM 2 approach was even more relevant for well F05.

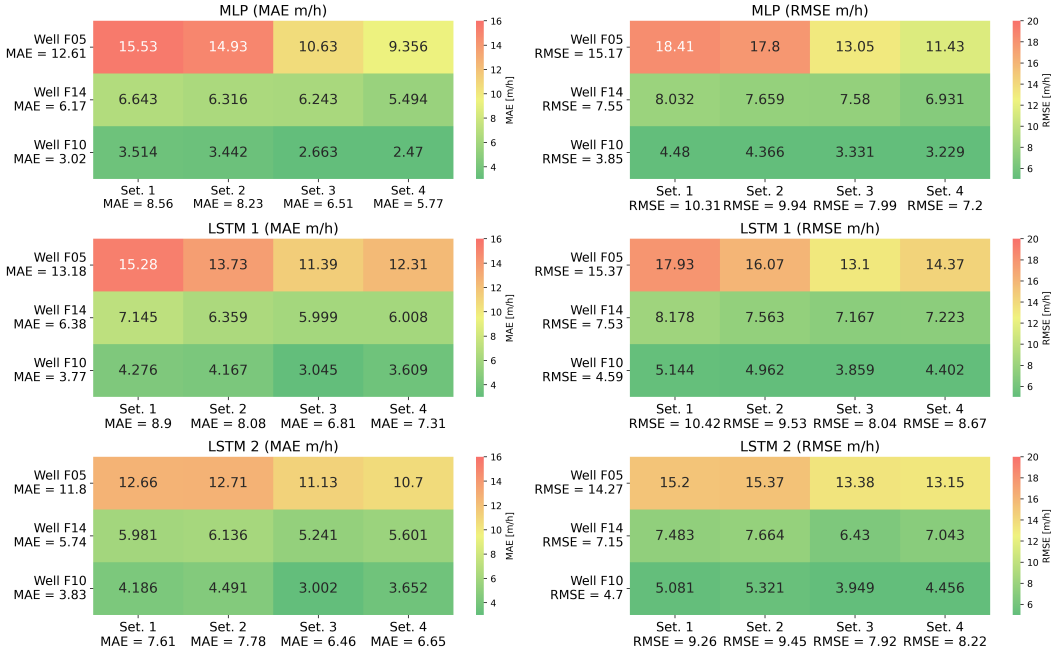


Figure 3. Heatmap of the average error metrics for each well across three models and considering different feature sets

Figures 4, 5, and 6 show the prediction error profiles along the depth of the analyzed wells. The well lithology is also presented, along with the PSI indices, which indicate the statistical distance metrics between the distribution of two datasets — in this case, corresponding to the statistical distance between two close intervals.

It is observed that well F05 exhibits the greatest lithological variation, followed by well F14, which shows less variation and contains thicker layers of the same lithology, and finally well F10, which is predominantly composed of a single lithology. Regarding the PSI indices, eq. (1) was applied with the number of bins determined by Sturges' empirical criterion (10 bins). In the case of well F05, there is a change in statistical distance for most features, with Flow Rate showing the smallest variation in this well. Notably, in intervals 2, 4, and 6, there is a sudden increase in more than one feature (WOB, Torque, RPM, and DT), followed by more stable subsequent intervals. Changes in the distribution of features affect predictions, and in this case, short sequential dependencies would be more appropriate. Significant variations can be seen in the predictive capacity of the models along depth, which may be related to the aforementioned distribution shifts. Among the compared sets, LSTM-based profiles produce lower maximum error in general; however, for the MLP profile using Set 4, there is greater stability in the error profile, which supports the use of efficient hyperparameter tuning—for instance, in LSTM models using Set 3, which is more suitable for the LSTM case.

For wells F14 and F10, the prediction curves produced by LSTM 2 are close to the actual results. In the case of well F14, as shown in Fig. 5, an increase in the distance of more than one variable can be observed between intervals 4 and 6, including the lithological variable (DT), although to a lesser extent than in well F05. For well F10, in Fig. (6), a closer alignment between the error curves is evident, with the largest changes in statistical distribution occurring in the features included in Set 1, which explains the small difference in performance among all feature sets for F10.

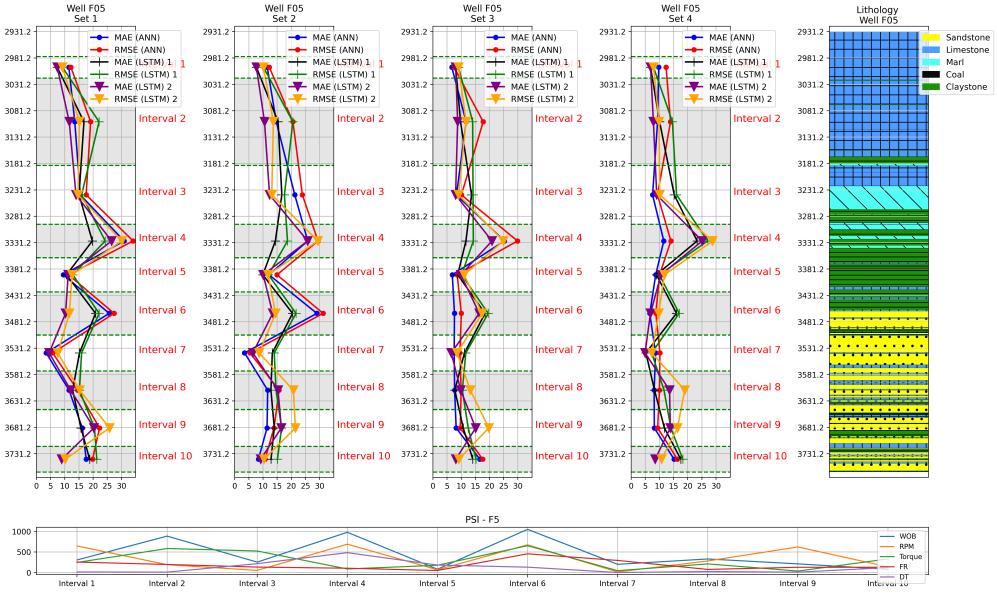


Figure 4. Error metrics and PSI index across depth for the analyzed models with different features in well F05

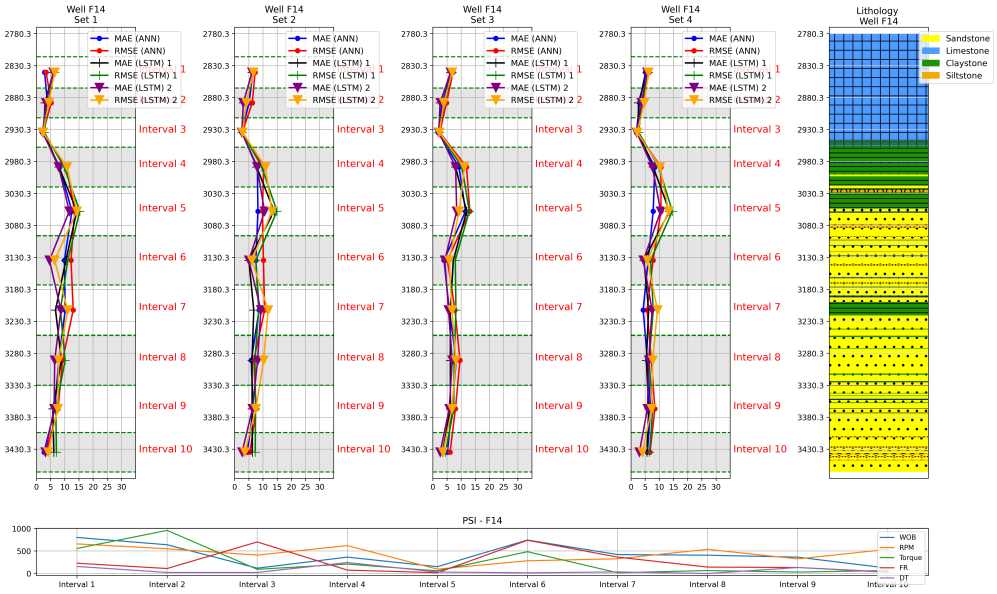


Figure 5. Error metrics and PSI index across depth for the analyzed models with different features in well F14

5 Conclusions

The results show that the choice of input variables has a more significant impact on model performance than extending the temporal history. Replacing RPM with Torque while keeping DT benefited both static models (MLP) and sequential models (LSTM), whereas including both RPM and Torque simultaneously (Set 4) yielded additional gains only for the fine-tuned MLP, but not for LSTMs without hyperparameter tuning, likely due to redundancy and increased complexity. Regarding temporal modeling, the LSTM trained solely on the last interval (without data accumulation) maintained good performance compared to other approaches, suggesting that long-term temporal dependency is limited for this problem and that short windows may suffice for real-time operation. The MLP, benefiting from hyperparameter tuning during training, achieved results comparable to or even superior to those of the LSTMs in some scenarios, indicating that much of the ROP behavior can be modeled from current conditions alone, considering the granulometry of the analyzed data. Conversely, in more complex wells, there is potential for performance gains in real-time LSTM models when hyperparameters are efficiently tuned, provided their operational use is viable.

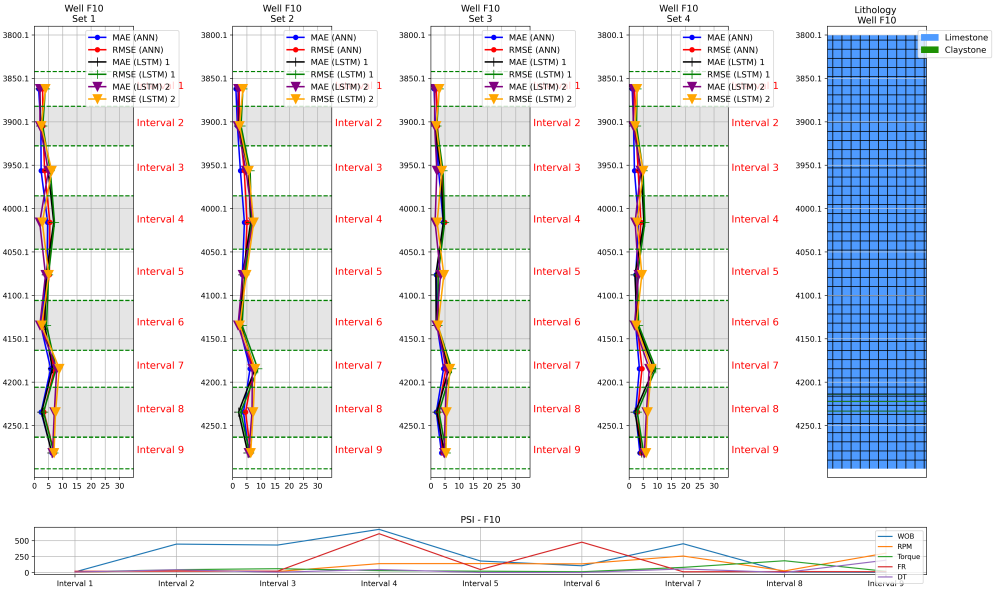


Figure 6. Error metrics and PSI index across depth for the analyzed models with different features in well F10

Acknowledgements. The authors would like to thank Petrobras for providing technical and financial support.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Q. Xue. *Data Analytics for Drilling Engineering: Theory, Algorithms, Experiments, Software*. Springer International Publishing, 2020.

[2] R. V. d. Almeida, M. V. G. Jacinto, G. C. De Medeiros, L. C. De Montalvão, and G. S. Bassani. Bayesian deep learning and evolutionary algorithms applied to drilling optimization: an approach combining long short-term memory and genetic algorithms. *Rio Oil and Gas Expo and Conference*, vol. 20, n. 2020, pp. 415–416, 2020.

[3] C. Soares and K. Gray. Real-time predictive capabilities of analytical and machine learning rate of penetration (rop) models. *Journal of Petroleum Science and Engineering*, vol. 172, pp. 934–959, 2019.

[4] L. F. F. M. Barbosa, A. Nascimento, M. H. Mathias, and de J. A. Carvalho. Machine learning methods applied to drilling rate of penetration prediction and optimization - a review. *Journal of Petroleum Science and Engineering*, vol. 183, 2019.

[5] A. T. Tunkiel, D. Sui, and T. Wiktorski. Reference dataset for rate of penetration benchmarking. *Journal of Petroleum Science and Engineering*, vol. 196, 2021.

[6] A. M. Elmgerbi, C. P. Ettinger, P. M. Tekum, G. Thonhauser, and A. Nascimento. Application of machine learning techniques for real time rate of penetration optimization. In *SPE/IADC Middle East Drilling Technology Conference and Exhibition*, 20MEDT. SPE, 2021.

[7] Z. Cheng, F. Zhang, L. Zhang, S. Yang, J. Wu, T. Li, and Y. Liu. A sequential feature-based rate of penetration representation prediction method by attention long short-term memory network. *SPE Journal*, vol. 29, n. 02, pp. 681–699, 2023.

[8] A. P. A. Ferro. *Predictive models for ROP as support for real-time optimization of operational drilling parameters in oil well drilling. (In Portuguese).* Dissertation (master in structural engineering), Federal University of Alagoas, Maceió, 2024.

[9] Equinor. Disclosing all volve data. Disponível em: `https://www.equinor.com/news/archive/14jun2018-disclosing-volve-data`. Acesso em: 09 de Jul. 2022, 2018.

[10] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press. `http://www.deeplearningbook.org`, 2016.

[11] B. Yurdakul and J. Naranjo. Statistical properties of the population stability index. *The Journal of Risk Model Validation*, 2020.