SPE-220059-MS

Watch the Reservoir! Improving Short-Term Production Forecast Through Transformers

R. Werneck, Universidade Estadual de Campinas; L. A. Lusquino Filho, São Paulo State University; A. Lustosa, A. Loomba, and M. M. Gonçalves, Universidade Estadual de Campinas; A. Esmin, Federal University of Lavras; S. Salavati, E. Morais, P. Ribeiro Mendes Junior, M. Zampieri, M. Amaral, O. C. Linares, M. Castro, R. Moura, D. J. Schiozer, A. Mello Ferreira, A. Davolio, and A. Rocha, Universidade Estadual de Campinas

This paper was prepared for presentation at the SPE Europe Energy Conference and Exhibition held in Turin, Italy, 26 - 28 June 2024.

Abstract

Data-driven methodologies have been used in reservoir management and production forecasting, particularly demonstrating remarkable efficacy in short-term oil production forecasts. However, there is space to improve its prediction, especially in tackling the complexities of challenging reservoirs, such as the heterogeneous carbonate reservoirs from Brazilian Pre-salt fields. Methods for oil production forecasting in the petroleum literature generally consider linear correlations or recurrent neural networks (RNNs). In this paper, we propose a new strategy to improve short-term forecasting for oil production through attention mechanisms that boost state-of-the-art methods. Traditional data-driven techniques generally do not consider static data or planned activities. However, we address this critical gap by leveraging the Temporal Fusion Transformer (TFT) to integrate such information into our short-term forecasting. Transformers, the architectural inspiration behind ChatGPT, employ attention mechanisms to establish relationships between different time series data points, assigning weights to these connections. We jointly explore oil, gas, and water production, pressure, and the ratios between them. This method includes static data (e.g., geographical coordinates) as well as known side reservoir information. Such side information can be, for instance, another predicted future production or planned well shut-ins. We also investigate which side information improves the obtained forecasting. This paper presents two main findings. First, it shows how using certain side information can improve the overall predictive capability of a model. For example, using predicted gas production as side information can significantly improve the oil production forecast. This is logical and in line with expectations, as there is an intimate connection between oil and gas production. In the second application of TFT, we considered well closures as the side information. We used an anomaly detection tool to identify well closures in the history period and converted it to usable side information for the TFT model. The distribution of these well closures is used as a guide to predict our target oil production. As we considered the distribution of well closures as side information, we framed our results in terms of cumulative oil production rather than daily forecast rates. The results of this work show that the cumulative production gets very close to the ground-truth data, better than linear and proposed baselines. In summary,

the second key result shows and underscores the significance of incorporating side information within our TFT approach.

Introduction

Hydrocarbon reservoir management is challenging because it integrates different areas of knowledge, such as geoscience and production engineering. A paramount part of this management is the production's accurate forecasting. Traditional tools are reservoir simulation models, known for their precision in long-term predictions but human and computational time-consuming for complex reservoirs. Moreover, these models have drawbacks for short-term prediction (Ertekin and Sun, 2019). These particularities are problematic for some Brazilian pre-salt fields since they are formed by highly heterogeneous carbonate rocks and they operate with a complex production strategy such as Water Alternating Gas (WAG) injection. Accurate short-term forecasts are necessary for fast decision-making for reservoir management. On the other hand, a promising alternative for the forecast is to focus on the short-term, in which analytics data-driven and machine-learning approaches offer an agile solution to help the management of the reservoir (Liu et al., 2020; Sun et al., 2018; Tadjer et al., 2021).

The use of data-driven approaches for production forecasting has been in the literature for a while. Differently from model-based approaches that consider general information from the reservoir, such as seismic, well logs, and more, data-driven approaches only consider field response observational data obtained from sensors and well produced (e.g., oil, water, and gas rates, bottom-hole pressure) and machine-learning techniques for performing short-term forecasts (Davtyan et al., 2020; Kubota & Reinert, 2019; Martínez & Rocha, 2023; Zhong et al., 2020).

Accurate forecasting is an essential part of a reservoir's management to help the responsible engineer propose the next steps for the development of the field (Liu et al., 2020). However, predicting accurately the well production is not an easy task, as we have a dynamic field in operation, with injection of water and gas, interference of other wells, and maintenance of reservoir pressure, together with the chaotic nature of the monitoring data and the unpredictability of closures (Werneck et al., 2022).

This work focuses on data-driven approaches to perform short-term forecasting of oil production. We propose a new approach with Transformers that uses predicted side information to improve the overall forecasting. We also investigate a new idea for the Temporal Fusion Transformer (TFT) (Lim et al., 2021) method to use closure information to help approximate our forecasting to the actual oil production.

This paper is organized as follows. The **Related Work** section presents the theory and related approaches in the literature. In the **Proposed Methodology** section, we present our approach to forecasting oil production. This is followed by the section **Experimental Protocol**, which details our experiments and datasets. Our results are presented in the section **Experiments and Results**. Finally, the section **Conclusion and Future Work** presents our conclusions and future work.

Related Work

This section presents related works from the literature on oil forecasting, especially on the advances of attention mechanisms with Transformers. Additionally, it introduces the Temporal Fusion Transformer (TFT) (Lim et al., 2021), which serves as the method employed in our experiments.

Oil Forecasting

Data-driven approaches and algorithms have been present in the literature of oil forecasting for a while. Kubota and Reinert (2019) applied linear regression techniques with recurrent neural networks for forecasting. The authors only considered time series from injection history, production history, and the number of producers. They showed that a reliable production forecast could be made with data-driven models without geological or numerical simulators.

Liu et al. (2020) proposed to use an ensemble empirical mode decomposition (EEMD) followed by Long-Term Short-Term learning. They split two daily datasets to use one as the training set, which was decomposed using an EEMD to choose the basis functions, applied a Genetic Algorithm to select the hyperparameters of their methods, and tested on the other dataset as the test set. There needs to be detailed information on how they performed the forecasting, whether daily or using a longer time frame.

Pan et al. (2019) proposed to apply Recurrent Neural Networks (RNNs) to long-term forecasting. They considered two scenarios that use specific types of Long Short-Term Memory (LSTM) networks trained considering hidden periods of the time series. In the first scenario, a Denoising Long Short-Term Memory (DeLSTM) only considers the oil production rate, followed by a Decline Curve Analysis (DCA) to perform the forecasting. The second scenario uses Savitzky-Golay Cascaded Long Short-Term Memory (SG-CLSTM) to smooth, fill hidden periods, and forecast. In their case, they used both oil production rate and pressure as input data.

Werneck et al. (2022) proposed the combination of stacked Gated Recurrent Units (GRUs) with Dense layers for short-term forecasting. They also proposed a new forecasting setup to consider the most challenging data in the prediction metrics. They showed that more than off-the-shelf methods are needed for oil forecasting, and a designed network for the problem is crucial.

Martínez and Rocha (2023) proposed The Golem, a general RNN-based model for oil and gas forecasting composed of two RNN layers and a set of dense layers in a trapezoidal shape. They provided an in-depth study of the network's hyperparameters for two datasets: one benchmark and one actual field. The authors found the best hyperparameters for all wells and four different targets (oil, gas, and water production, as well as bottom-hole pressure).

Our proposed approaches differ from the previous works mentioned above in several aspects. First, they incorporate self-attention mechanisms to identify correlations between different points in time series data, and these correlations help to improve the predictions, furthermore, the attention focuses the model on relevant information. Second, unlike the aforementioned methods, our approaches incorporate interventions in the data as informative factors for forecasting purposes.

Transformers

Recurrent models focus on the sequential aspect of the input data; however, as the sequences grow in length, memory constraints limit their performance. To address this, attention mechanisms were proposed as part of the sequence modeling to model dependencies without considering the distance between the input and output sequences.

Vaswani et al. (2017) proposed a new model architecture that discarded recurrent layers and relied solely on attention mechanisms to capture dependencies between input and output sequences. They proposed an encoder-decoder model with sequential layers of multi-head self-attention mechanisms and point-wise feed-forward networks with residual connections. The authors showed that this architecture was trained faster in Natural Language Processing problems.

Shih et al. (2019) developed a new method called Temporal Pattern Attention (TPA), an attention mechanism for multivariate time series, which combined a set of filters to extract time-invariant temporal patterns, achieving state-of-the-art performance in several real-world tasks.

Temporal Fusion Transformers (TFT)

With the success of attention mechanisms in Natural Language Processing, the literature on forecasting devised different approaches to integrate them in this domain. One approach was proposed by Lim et al. (2021) called the Temporal Fusion Transformer (TFT). The TFT architecture leverages attention mechanisms to achieve high-performance multi-horizon forecasting while providing interpretable insights into the temporal dynamics of the time series. The authors demonstrated a substantial improvement over existing benchmarks by incorporating recurrent layers for local processing and self-attention mechanisms for capturing long-term dependencies.

Al-Ali and Horne (2023) applied the TFT method to obtain probabilistic forecasting of well oil production. They considered historical bottom-hole pressure, wellhead pressure and temperature, and choke size opening as inputs to the model. Compared with conventional BlockRNN-based models, the TFT model outperforms them in the Volve Field dataset. Our proposed approaches also consider the TFT method for forecasting; however, they considered the wellhead pressure as the method's side information and only predicted the next time step. Our approaches use a production feature and a closure information as side information and, to calculate the metrics, the last day of the forecasting window of 30 days, which is a more difficult time step to predict (Werneck et al., 2022).

Proposed Methodology

This section presents our proposed methodology for working with TFT in oil production forecasting. We have two main findings in this paper: first, we explore which side information to use with TFT, showing the improvement of the forecasting in this approach; and then, we present another side information, which considers the well closures in the testing set to represent the valleys of non-production in the data. Figure 1 shows our proposed pipeline considering both findings.

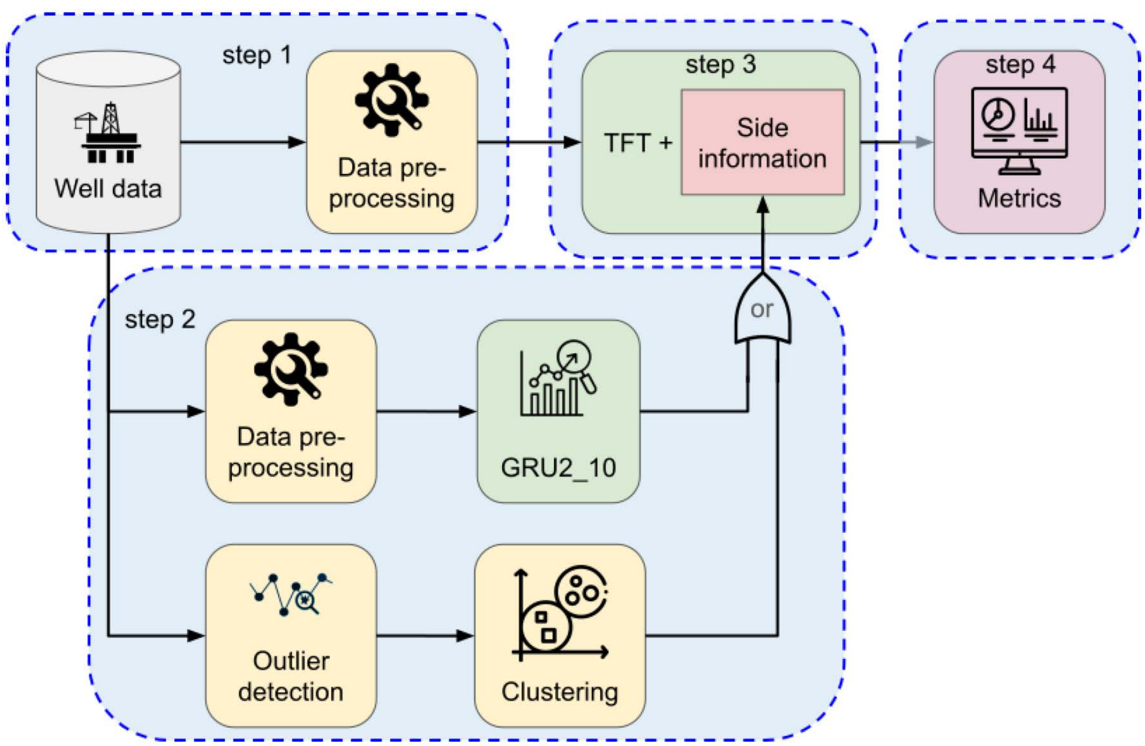

Figure 1—Proposed methodology for our experiments with TFT with two options. First, we obtain the well data and perform preprocessing on it (step 1). Then, in step 2, we have two options: forecasting with GRU2_10 defines one possible side information for our TFT method or another experiment that considers closures from the training data as the side information. In step 3, we have our Transformer method with the selected side information. Finally, step 4 presents the metrics of our final results.

Choosing Side Information Data

Temporal Fusion Transformer is a forecasting method that considers different information. It considers historical and static data as well as known information in the historical and future inputs. However, we have yet to have a piece of future information available to use with this method in oil production. Thus, we propose using another forecasting method to predict one of our features and then use this prediction as the known future input for the TFT.

First, we must decide which feature to choose as our known future input. For this, we perform experiments to see which of our other features would perform better in oil production forecasting. For this, we use the ground-truth information to identify the feature that exhibits the best predictive capability. Then, we forecast this selected feature for the duration of the testing set. We propose using GRU2_10 (Werneck et al., 2022) as our first forecasting model to obtain the future known input of the TFT.

Well Closure Side Information

Even when using predicted future information, it is impossible to forecast the well closures during the testing time, as these well closures could be scheduled. Scheduled well closures and partial closures could happen for different reasons, such as periodic closures for testing, fixed WAG cycles, excessive water production, descaling, among others. Unscheduled interventions could happen due to the limits of the platform and the well or replacement of equipment due to failure, among others (Ferreira et al., 2023).

We can include scheduled closures in the future information of TFT, but not the unscheduled ones. To address this problem and acknowledge both closures, we propose considering possible closures in the testing set, using the same distribution of closures in the training set. To accomplish this, we used the outlier detection approach proposed by Soriano et al. (2021). We consider a moving time window using z-score (Yadav et al., 2018), considering only the negative cut-off, as we are only considering the closures of the production well in this work. Later, we bundle the outliers into a few clusters representing different well closures in oil production. With this information in hand, we incorporate this cluster information as a new feature in our dataset, assigning values to non-outlier points accordingly. Then, we get the distribution of these values in the training set and apply it to the testing set to be our known future information. This way, we have an approximation of closures in the testing set, similar to the training set.

However, as we consider the distribution of closures and do not have information on the exact closure in the testing set, we could not consider the daily production forecasting, as our daily predictions would not fit in the ground-truth daily production curve. To resolve this, we propose considering the cumulative oil production of this forecasting. In this case, as we consider the closures in the testing data, the cumulative curve will be closer to the ground-truth cumulative oil production, avoiding overestimation of this curve because of an unrealistic non-stop production.

Experimental Protocol

In this section, we present details on the dataset used in our experiments, the parameters for considering the outliers/closures, and how we performed the experiments using the TFT, considering the parameters of the method and the selected features to use on it.

Datasets

We chose two datasets to perform the experiments using our approaches. The first is a proprietary dataset from a pre-salt oil reservoir. This dataset comprises production data from a Brazilian pre-salt reservoir. It provides information on fluid production (oil, gas, and water rates), pressure (bottom-hole), and the ratio between them (water cut, gas-oil ratio [GOR], and gas-liquid ratio [GLR]). The reservoir has 16 producers and 16 injector wells, divided into nine dedicated water and seven WAG injectors. The oldest producer well has five years of historical daily data.

The second dataset is the UNISIM-IV-24 benchmark (Botechia et al., 2023). It is a synthetic benchmark based on a carbonate reservoir with typical Brazilian pre-salt reservoir features, built on public data (ANP - Brazilian National Agency of Petroleum, Natural Gas and Biofuels) (Correia et al., 2020). This benchmark has six Producers (P11, P12, P13, P14, P15, and P16), six WAG injectors (I11, I12, I13, I14, I15, and I17), and a dedicated gas injector (I16), with a total of three years and three months of production history.

Besides the fluid production, their ratios, and pressure from the datasets, we also consider some information for our TFT method, such as geographical coordinates and region of the reservoir (as static data). The following experiments will select the known future information data from our other features.

Outlier/Closure Detection

For the outlier/closure detection step to obtain the well closure in the training set, we considered the daily oil production with a sliding window of 15 days. As we only consider the well closures, we only deal with the negative cut-off of one standard deviation. With the defined outlier data, we cluster them with the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (Campello et al., 2013) method to bundle these outlier points into a few clusters, representing different partial and complete closures of the production well.

Oil Forecasting

We performed oil forecasting following this protocol: we selected the *N-th day* setup proposed by Werneck et al. (2022), considering a forecasting window of 30 days with 85 days of input and a testing window of 365 days (one year). For data preprocessing, we apply an anomaly outlier detection using a z-score to remove any outlier above one standard deviation. We also perform experiments comparing results with and without data augmentation, in which we included interpolated values between two-time quanta to generate points of 3-hours spaced.

All our experiments use Huber loss as our loss function and are performed ten times, and their mean is considered our final answer to avoid cherry-picking any execution. We evaluate our results considering two main metrics: Symmetric Mean Absolute Percentage Error (SMAPE), which measures the percentage error of the predicted values:

$$SMAPE(X, \quad h) = \frac{100}{m}\sum_{i=1}^{m} \frac{|h(x^i) - y^i|}{\frac{(|y^i| + |h(x^i)|)}{2}} \tag{1}$$

where $h(x^i)$ are the predicted values, y^i the ground truth, and m is the total number of observations; and Dynamic Time Warping (DTW), employed as a similarity metric between two sequences. We selected DTW because SMAPE and other error-based metrics do not consider trends in the data, i.e., when the predicted time series follows the trend of the ground truth when there is a phase shift between them.

To compute the DTW, we consider two sequences, A and B, of length n and m, respectively. First, we construct a D matrix of size $n \times m$, where $D(i,j)$ represents the "cumulative distance" to align a_i withb_j. We populate this matrix as follows:

1. Initialize the matrix D with $D(0,0)=0, D(i,0)=\infty$ and $D(0,j)=\infty$, for $i,j>0$;
2. Compute, iteratively, the remaining cells, as $D(i,j)=dist(a_i,b_j)+min(D(i-1,j),D(i,j-1),D(i-1,j-1)$, where $dist(a_i,b_j)$ is the distance between a_i andb_j :
3. DTW metric would be the minimum cumulative distance after alignment: $D(n,m)$.

This way, DTW aligns both sequences with a minimum accumulated cost, representing the optimal warp between them, showing their similarities, and accounting for phase shifts and local variations.

Experiments and Results

In this section, we detail our experiments and discuss the obtained results. Our first set of experiments (in the section **Side Information Data**) refers to different side information on the forecasting of oil production using the TFT method. The second experiment (in the section **Quantile Forecasting**) provides insight into the quantile forecast performed by the TFT method and how we could use it to improve our forecasting.

Finally, our last experiment (in the section **Well Closure**) provides our proposed approach for oil production forecasting considering well closure information.

Side Information Data

Our first experiment was necessary to discover which side feature would help to improve our oil production forecasting when selected as the known side information. To achieve that, we selected the ground truth data of each other feature in our datasets to be used as the known side information on the proprietary dataset. Table 1 compares the SMAPE metric for six features and five production wells in the private dataset.

Table 1—Comparison of different features as side information for the TFT approach for oil production forecasting using the SMAPE metric in the private dataset.

Features	P1	P2	P3	P4	P5
Gas production	**18.76**	42.10	**5.80**	13.01	**59.35**
Water production	29.74	39.21	11.10	17.23	69.49
Bottom-Hole Pressure (BHP)	29.84	41.38	7.83	**8.63**	64.99
Water cut	31.59	**37.28**	11.65	9.89	67.16
Gas Oil Rate (GOR)	30.99	45.24	9.26	12.74	65.19
Gas Liquid Rate (GLR)	30.86	37.58	10.85	17.25	63.63

In Table 1, we can see that *Gas production* was the feature that best helped the oil production forecasting, improving the results in three of the five production wells, with more than ten percentage points below the second place for well P1. This is expected and logical, as there is an intimate connection between gas and oil production.

Thus, we can perform our forecasting with the TFT approach. As we cannot use the ground truth information of gas production as the side information of TFT, we need to calculate this side information first. For that, we proposed to use the GRU2_10 (Werneck et al., 2022) to forecast gas production and then use this forecasted data as the side information input of the TFT method to forecast our target: oil production. For comparison, we used the GRU2_10 forecasting oil production as our baseline, considering all the same features, as it outperformed off-the-shelf techniques in the same problem (Werneck et al., 2022). Tables 2 and 3 compare the metrics in the testing set considering the methods TFT and GRU2_10 for the private and UNISIM-IV datasets, respectively.

Table 2—GRU2_10 and TFT results for the private dataset.

Augmentation	Model	Metric	P1	P2	P3	P4	P5
No augmentation	GRU2_10	SMAPE	34.29	46.01	14.55	**40.32**	**42.20**
		DTW	541.57	321.91	188.13	**303.99**	**431.18**
	TFT	SMAPE	**32.49**	**40.33**	**14.53**	46.24	42.67
		DTW	**424.96**	**276.46**	**187.92**	658.16	439.00
Augmentation 3h	GRU2_10	SMAPE	32.82	42.70	15.32	40.44	45.03
		DTW	433.83	**184.55**	**178.87**	**295.66**	**343.25**
	TFT	SMAPE	**32.49**	**40.01**	**14.88**	**40.16**	**43.58**
		DTW	**424.96**	233.52	184.50	325.89	364.97

Table 3—GRU2_10 and TFT results for the UNISIM-IV dataset.

Augmentation	Model	Metric	P11	P12	P13	P14	P15	P16
No augmentation	GRU2_10	SMAPE	**9.18**	**11.47**	14.14	10.01	12.92	15.50
		DTW	**193.98**	**204.51**	231.93	131.88	**226.78**	215.58
	TFT	SMAPE	11.54	12.01	**13.36**	**9.56**	**12.75**	**14.26**
		DTW	248.61	214.54	**221.75**	**128.66**	226.81	**189.11**
Augmentation 3h	GRU2_10	SMAPE	10.44	**12.05**	14.64	10.49	13.57	13.48
		DTW	**167.24**	197.98	213.70	**124.65**	**221.44**	**152.79**
	TFT	SMAPE	**10.11**	12.21	**13.54**	**9.77**	**12.85**	**13.38**
		DTW	195.93	**197.10**	**212.51**	128.72	222.63	170.70

As we can observe in previous tables (2 and 3), TFT improved the SMAPE metric in 17 of 22 experiments, while the DTW metric demonstrated improvement in 9 of the same 22 experiments. The improvements in SMAPE are not shown in DTW mainly because most of the production data is almost linear, with slight variation, as shown in Figure 2. At the same time, the SMAPE metric improves with the variation provided by the side information included in the TFT method.

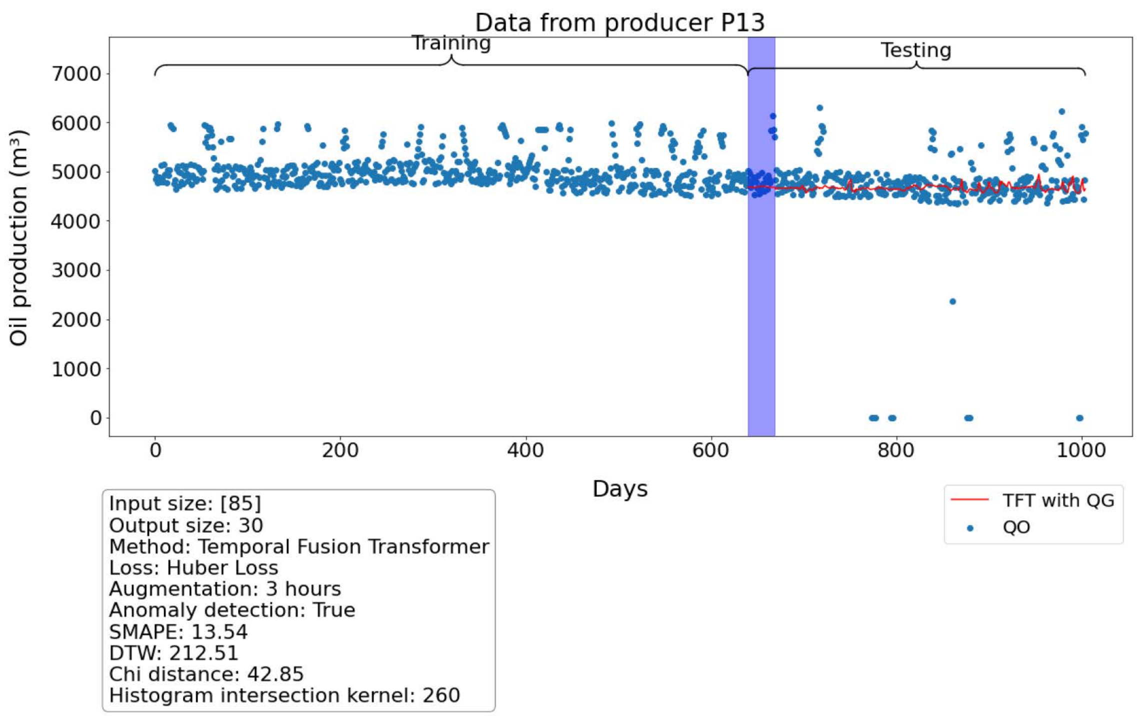

Figure 2—TFT with QG prediction (red line) for well P13 from the UNISIM-IV dataset. UNISIM-IV wells are mostly linear, with a noise in the daily data provided by the simulator. QO (blue dots) is the ground truth data, after we apply a pre-processing only in the training data to clean outliers.

Quantile Forecasting

The TFT method provides three outputs with different results from three quantile losses (0.1, 0.5, and 0.9). The quantile loss equation used in this method is determined as follows:

$$pred_{diff} = y_{gt} - y_{pred} \tag{2}$$

$$Q_{Loss} = q \times max(pred_{diff}, 0) + (1-q) \times max(-pred_{diff}, 0) \tag{3}$$

where y_{gt} is the ground truth value, y_{pred} is the predicted value, and q is the quantile in consideration. Figure 3 shows the weight of the loss for six different quantile options (0.1, 0.2, 0.3, 0.4, 0.5, and 0.9).

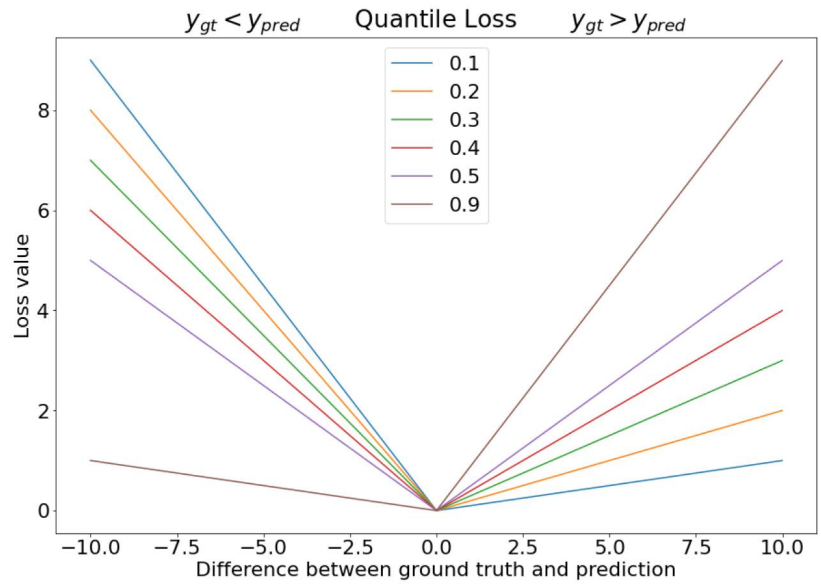

Figure 3—Quantile loss of TFT method. On the left side, when the ground truth value is smaller than the predicted value, the loss penalizes smaller quantile options. On the right side, we have the opposite. When the ground truth value is greater than the predicted value, the loss penalizes greater quantile options.

This quantile loss weights differently if the predicted values are above or below the ground truth value. For quantile 0.1, when the predicted value is greater than the ground truth, the loss is bigger than when the opposite occurs. For quantile 0.9, we have greater losses when the ground truth is greater than the predicted values. Finally, they are the same for quantile 0.5. The results presented in the tables (1, 2, and 3) from Section **Side information** data are considered to be in the quantile of 0.5.

Well Closure

In previous experiments, closure within the training data was not considered and was mostly removed during the preprocessing step. However, the scenario differs for the testing data, where the unexpected occurrence of production closures presents a difficult challenge for prediction and could potentially affect the efficacy of our proposed approaches.

To address this problem of closures in the testing set, we propose a new approach to deal with them, considering possible closures in the testing data originating from the closures in the training set.

To obtain the closures in the training data, we applied the closure detection method presented in the section **Outlier/closure detection** together with the cluster method HDBSCAN (Campelo et al., 2013). Thus, the possible closures are clustered into a few sets of similar data points. Thus, these aggregated data represent the same information, e.g., complete well closure, half closure (halving the production), or just a small closure. The closures identified by the outlier detection are present in Figure 4, and their clusters are shown in Figure 5.

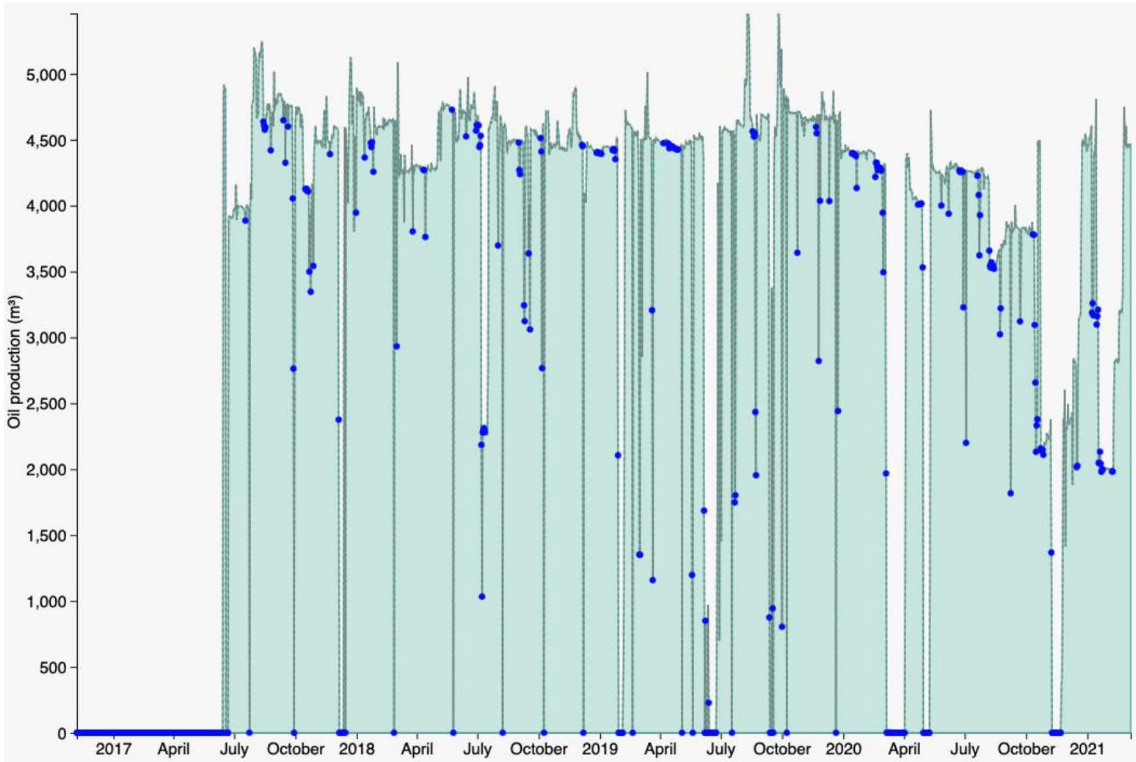

Figure 4—Partial and total closures found in well P1 from the private dataset.

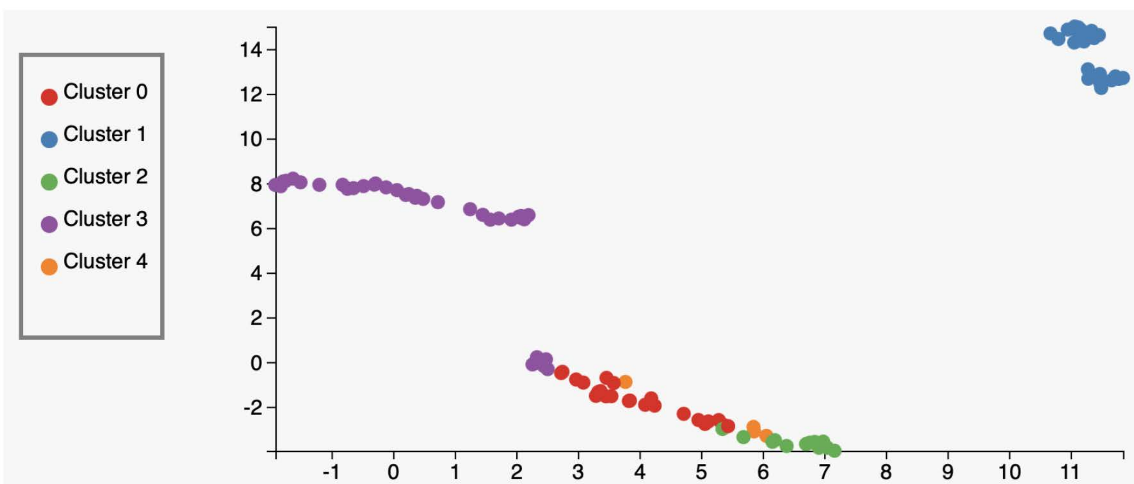

Figure 5—Projection in two dimensions using UMAP of the closures in well P1 with its clusters.

These points and their defined clusters are complemented with the rest of the training data points allocated to a non-closure cluster. Then, we have a new feature for the training data, representing different closures (or not) of the oil production data. From this new feature, we obtain the distribution frequency of the closures in the training set, which will be applied to the testing set.

It is difficult to consider unknown closures in the testing data because we do not have this information. We propose randomly allocating closures and non-closures to the testing set, following the same data distribution in the training set. This way, we include closure information in the testing data and maintain the frequency of closures in the training data.

Finally, as we randomly included this new data information in the testing set, we could not guarantee the correct place of the closure, so we could not use the daily forecasting results. In this way, we also propose to present its results as cumulative oil production, in which we will see the evolution of oil production in our results.

With that in hand, we can perform a new experiment with the TFT method, in this case, considering the closure as the side information. Table 4 and Table 5 show the SMAPE metric of our proposed method with three baselines: two Hyperbolic DCA (Arps, 1945), one with a data preprocessing to remove outliers (treated), and another without any preprocessing, maintaining the closures in the data (closures); and GRU2_10 (Werneck et al., 2022), as proposed in the reference paper. All these baselines are predicted daily and then summed up to obtain their cumulative.

Table 4—SMAPE comparison between TFT with closures with three baselines (DCA treated, DCA with closures, and GRU2_10) for the private dataset.

Models	**P1**	**P2**	**P3**	**P4**	**P5**
DCA treated	1.20	5.39	0.25	2.78	0.55
DCA with closures	1.31	6.32	0.71	2.01	1.46
GRU2_10	1.52	2.96	**0.22**	1.35	**0.40**
TFT with closures	**0.80**	**0.25**	2.53	**0.32**	1.44

Table 5—SMAPE comparison between TFT with closures with three baselines (DCA treated, DCA with closures, and GRU2_10) for the UNISIM-IV dataset.

Models	**P11**	**P12**	**P13**	**P14**	**P15**	**P16**
DCA treated	**0.52**	3.19	1.57	0.59	**0.38**	1.52
DCA with closures	2.00	1.13	**0.63**	1.24	0.95	0.41
GRU2_10	0.58	0.77	1.28	0.72	0.49	0.80
TFT with closures	0.88	**0.49**	0.67	**0.53**	1.07	**0.25**

We can see from the above tables (4 and 5) that TFT with closures achieved the best results in half (50%) of the wells of both the private and UNISIM-IV datasets. Figure 6 shows a challenging example of the daily production of the P2 well from the private dataset, in which the TFT with closures performed better than the baselines when considering the cumulative prediction of the same well (Figure 7). In this case, the DCA could not follow the declining trend; however, the TFT with closures, even without information on the closures in the testing set, had a good approximation of the cumulative curve. In the other cases, the wells of the private dataset were well-behaved in the testing data, with little to no closure. As the UNISIM-IV dataset originated from a simulated model, it includes noise when providing daily data, making the prediction difficult, for which the baseline GRU2_10 is close to the best achieved SMAPE.

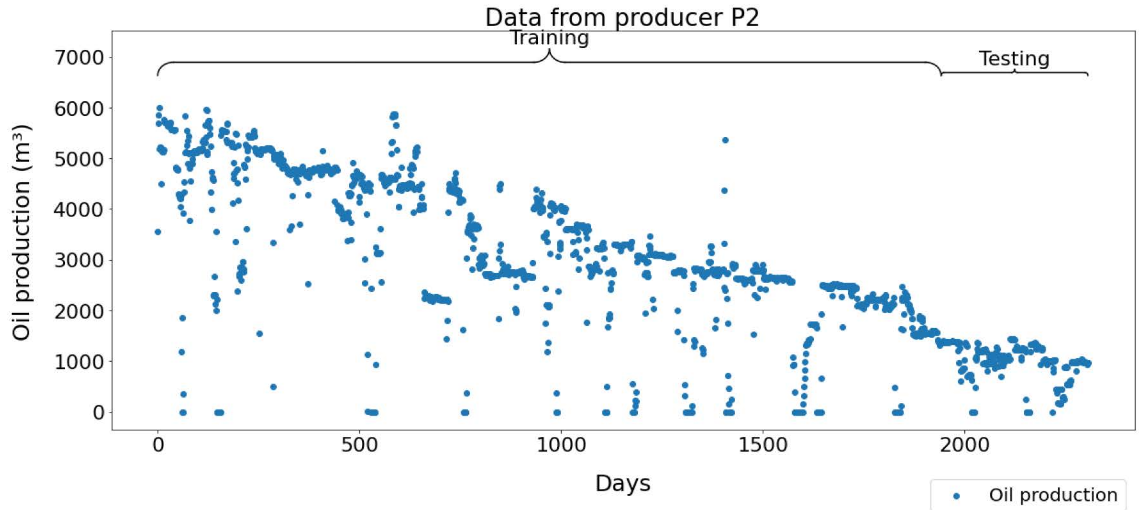

Figure 6—Private dataset well P2 with declined daily production.

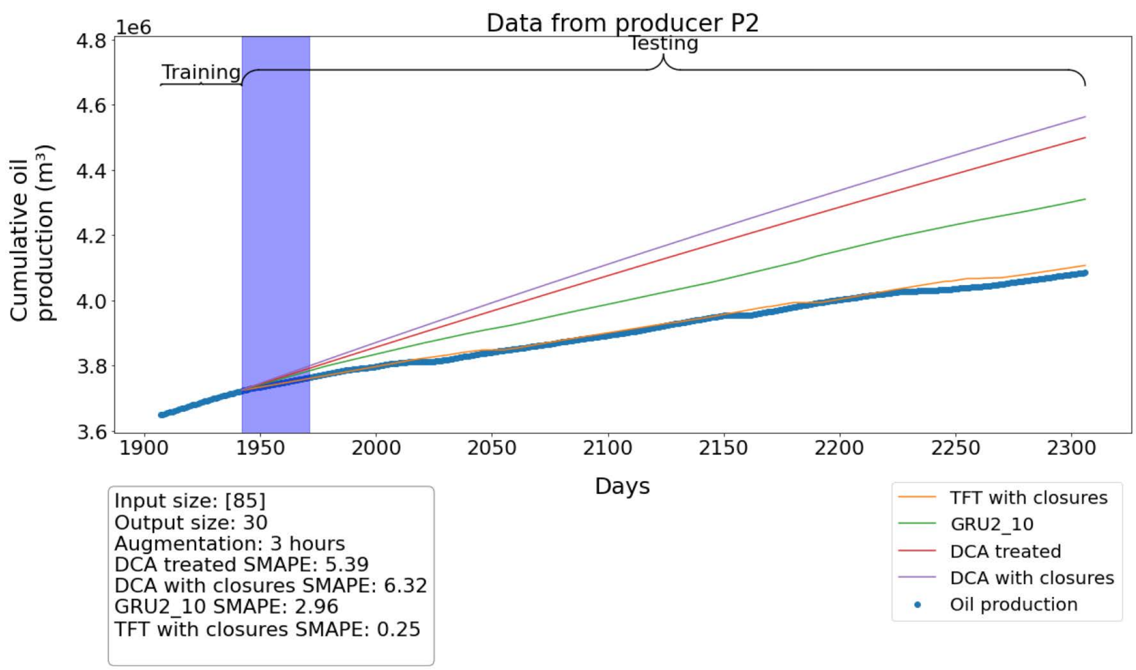

Figure 7—Comparison between all methods in the oil production cumulative curve for the private dataset well P2.

Conclusion and Future Work

This work presents data-driven short-term forecasting approaches in which we apply state-of-the-art natural language methods when considering attention mechanisms to relate different points in the prediction model. First, we explored which side information to include in our approach, discovering that including predicted gas production as a side information is paramount to improving oil forecasting. This was expected, as gas production is intimately connected to oil production in general. Still, the knowledge of a specialist about the target well could provide even better side information and improve the results with TFT.

Second, we also presented a new method for considering historical closure information in the cumulative oil production obtained from our daily prediction model. This method is useful, as we did not ignore part of the data points (closures), as traditional approaches do, incorporating their information in our method. Thus, we deal with the complete real data from the reservoir, as it was obtained in production. The proposed approach was successfully applied in two datasets, especially with challenging testing data, showing that predicting the cumulative curve is complex and simple methods are not enough for its prediction.

Additionally, this research holds substantial promise for enhancing strategic decision-making within the oil and gas industry. By providing more accurate forecasts, companies can better manage their resources, optimize production schedules, and effectively plan their maintenance operations to avoid unplanned downtimes. Moreover, our proposed adaptabilities to include different forms of side information further enriches the model's predictive capabilities.

Future Work

Our proposed approach for short-term oil production forecasting can be improved by considering, for example, another loss function for the training step, which relates to our metrics, like SMAPE and DTW, instead of using a more general loss, such as Huber loss. Another foreseen future work is to consider information from other wells during the training step of the target well. In this case, we will have more data to better understand the evolution of well production and its behaviors, as deep-learning approaches expect and perform better with more data for it to learn from.

Acknowledgments

This research was performed as part of the ongoing Project registered under ANP number 21373-6 as "Desenvolvimento de Técnicas de Aprendizado de Máquina para Análise de Dados Complexos de Produção de um Campo do Pré-Sal" (UNICAMP/Shell Brazil/ANP) – "Machine-Learning Development for Analysis of Complex Production Data in a Pre-Salt Carbonate Field" – funded by Shell Brazil Technology under the ANP R&D Levy as "Compromisso de Investimentos com Pesquisa e Desenvolvimento."

References

A.-Al A. H.-Ali and R. Horne. Probabilistic Well Production Forecasting in Volve Field Using Temporal Fusion Transformer Deep Learning Models. volume Day 3 Wed, March 15, 2023 of SPE Gas, page D031S046R003, 03 2023.

Arps J.. Analysis of Decline Curves. *Transactions of the AIME*, **160**(01):228–247, 12 1945.

V. Botechia, M. Correia, V. de Souza Rios, A. Davolio, S. Santos, J. von Hohendorff Filho, and D. Schiozer. Unisim-iv: Benchmark proposal for light oil carbonate reservoir with high CO2 content. *Brazilian Journal of Petroleum and Gas*, **16**:149–160, 01 2023.

R. J. G. B. Campello, D. Moulavi, and J. Sander. Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, and G. Xu, editors, Advances in Knowledge Discovery and Data Mining, pages 160–172, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

M. Correia, V. Botechia, L. Pires, V. Rios, S. Santos, V. Rios, J. Hohendorff, M. Chaves, and D. Schiozer. Unisim-iii: Benchmark case proposal based on a fractured karst reservoir. **2020**(1):1–14, 2020.

A. Davtyan, A. Rodin, I. Muchnik, and A. Romashkin. Oil production forecast models based on sliding window regression. *Journal of Petroleum Science and Engineering*, **195**:107916, 2020.

R. de Oliveira Werneck, R. Prates, R. Moura, M. M. Gonçalves, M. Castro, A. Soriano-Vargas, P. Ribeiro Mendes Júnior, M. M. Hossain, M. F. Zampieri, A. Ferreira, A. Davólio, D. Schiozer, and A. Rocha. Data-driven deep-learning forecasting for oil production and pressure. *Journal of Petroleum Science and Engineering*, **210**:109937, 2022.

Ertekin T. and Q. Sun. Artificial intelligence applications in reservoir engineering: A status check. *Energies*, **12**(15):2897–2919, Jul 2019.

V. H. d. S. Ferreira, M. Castro, R. Moura, R. d. O. Werneck, M. F. Zampieri, M. M. Gonçalves, O. Linares, S. Salavati, L. A. D. Filho, P. Ribeiro Mendes Júnior, A. Mello Ferreira, A. Davolio, D. J. Schiozer, and A. Rocha. A New Hybrid Data-Driven and Model-Based Methodology for Improved Short-Term Production Forecasting. volume Day 4 Thu, May 04, 2023 of OTC Offshore Technology Conference, page D041S049R007, 05 2023.

Kubota L. and D. Reinert. Machine Learning Forecasts Oil Rate in Mature Onshore Field Jointly Driven by Water and Steam Injection. In SPE Annual Technical Conference and Exhibition, volume Day 2 Tue, October 01, 2019, pages 1–18, 09 2019.

B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, **37**(4):1748–1764, 2021.

W. Liu, W. D. Liu, and J. Gu. Forecasting oil production using ensemble empirical model decomposition based long short-term memory neural network. *Journal of Petroleum Science and Engineering*, **189**:107013, 2020.

Mart V. ínez and A. Rocha. The Golem: A general data-driven model for oil & gas forecasting based on recurrent neural networks. *IEEE Access*, **11**:41105–41132, 2023.

Y. Pan, R. Bi, P. Zhou, L. Deng, and J. Lee. An effective physics-based deep learning model for enhancing production surveillance and analysis in unconventional reservoirs. In SPE/AAPG/SEG Unconventional Resources Technology Conference. OnePetro,07 2019.

S.-Y. Shih, F.-SunK., and H.-y. Lee. Temporal pattern attention for multivariate time series forecasting. *Machine Learning*, **108**(8):1421–1441, Sep 2019.

A. Soriano-Vargas, R. Werneck, R. Moura, P. Mendes Júnior, R. Prates, M. Castro, M. Gonçalves, M. Hossain, M. Zampieri, A. Ferreira, A. Davólio, B. Hamann, D. J. Schiozer, and A. Rocha. A visual analytics approach to anomaly detection in hydrocarbon reservoir time series data. *Journal of Petroleum Science and Engineering*, **206**:108988, 2021.

J. Sun, X. Ma, and M. Kazi. Comparison of decline curve analysis DCA with recursive neural networks RNN for production forecast of multiple wells. In SPE Western Regional Meeting, volume Day 4 Wed, April 25, 2018, 04 2018.

A. Tadjer, A. Hong, and R. B. Bratvold. Machine learning based decline curve analysis for short-term oil production forecast. *Energy Exploration & Exploitation*, **0**(0):01445987211011784, 2021.

A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, u. Kaiser, and I. Polosukhin. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.

S. S. Yadav, V. Vijayakumar, and J. Athanesious. Detection of anomalies in traffic scene surveillance. In 2018 Tenth International Conference on Advanced Computing (ICoAC), pages 286–291, 2018.

Z. Zhong, A. Y. Sun, Y. Wang, and B. Ren. Predicting field production rates for waterflooding using a machine learning-based proxy model. *Journal of Petroleum Science and Engineering*, **194**:107574, 2020.