SPE-220095-MS

Enhancing Short-Term Production Forecast in Oil Fields: Integrating Data-Driven and Model-Based Approaches to Reduce Uncertainty

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

Reservoir simulation models are usually applied to optimize oil field production across its life cycle but face challenges in short-term forecasting. Data-driven techniques (DD) show promise for short-term predictions but lack reliability over extended periods. This study introduces a Hybrid Methodology (HyM) combining the optimal features of model-based (MB) with DD approaches to select the best simulation models to make short-term decisions, effectively reducing uncertainty in short-term production forecasts for a real field. Using DD techniques, such as Transformers, we forecast 150 days of cumulative water, oil, and gas production for each well of a real field based only on historical production. We also run production forecasts using 200 history-matched simulation models for this field. In our HyM methodology, we propose to select the best simulation models according to their fit to the DD forecasts, assuming the DD forecasts to be more representative of the short-term behavior of the real field. We compare our HyM to a conventional MB model selection based on history-matching quality, referred to as MB. The proposed approach was applied to a heavy oil field from the Campos basin (Brazil), a deep-water turbiditic reservoir. We separated five months of real history data as the ground truth to rank the selected models through the forecast error and rank similarity. With this evaluation, we measured how accurate the forecast of the chosen simulation models was and the efficacy of each approach (HyM and MB) in ranking the simulation models. Our results showed that the subset of models selected using HyM successfully outperformed those chosen by the MB approach since they are better fitted for short-term forecasts for oil, water, and gas rates. The proposed HyM method yields a 75% error reduction, considering the rank similarity, with respect to the MB approach. Consequently, we conclude that better results may be achieved by using a DD forecast as a reference to reduce the uncertainty of oil, water, and gas rates (HyM approach) instead of using the actual history (MB approach). The new methodology showed promising results by selecting models with attenuated errors for oil/water rates during the history-forecast transition, complementing the data assimilation procedure. This study presents an innovative methodology merging the strengths of machine learning (ML) and simulation models to enhance the reliability of short-term production forecasts. The proposed approach is versatile, not

tied to any specific ML algorithm, and effectively minimizes uncertainties, particularly in complex fields with numerous wells.

Introduction

Reservoir management aims to maximize oil and gas recovery and minimize costs associated with capital investments and operations. It is an interdisciplinary process that uses data and models to make decisions, integrating geology, petroleum engineering, reservoir engineering, geophysics, and economics. Accurate production forecasting is crucial for optimizing reservoir performance, guiding economic planning, and mitigating operational risks in reservoir management. Two broad categories of forecasting oil production include Model-Based (MB) and Data-Driven (DD) approaches.

MB approaches utilize simulation models to capture the physics of oil production, serving various purposes such as predicting reservoir production performance under diverse production strategies, estimating ultimate oil recovery, and designing optimal well completions. As a result, they have been very successful in reservoir management and production forecasting through the life cycle of oil fields (Mirzaei-Paiaman et al., 2021). The simulation models are complex and involve multiple equations to model physical phenomena, including fine geological grid meshes, fluid compositional models, fracture modeling, rock-fluid iteration representation, phase-behavior modeling, integration with surface facilities, geomechanical models, and more. As models become more complex, they demand increased data and computational resources (Maschio et al., 2022). However, hundreds to thousands of simulation runs may be necessary to run field management studies, requiring simplified models designed to enhance computational efficiency while still adhering to the reservoir's fundamental physical processes (Schiozer et al., 2019). These simplified models, commonly referred to as lower-fidelity models, maintain computational efficiency while capturing the essential reservoir behaviors.

Numerous works have proposed data-driven approaches for production forecasting, such as deep learning models (Cao et al., 2016; Werneck et al., 2022; Yu et al., 2021). New forecasting models are arising, such as the novel transformer architecture in natural language and time series applications to better handle long-term temporal dependencies using attention layers. Al-Ali and Horne (2023) used a Temporal Fusion Transformer (TFT) probabilistic model for multivariate oil rate forecasting in oil and gas. All these approaches use machine learning algorithms, statistical models, and data analytics techniques to analyze patterns and trends within the data, enabling production forecasting and optimizing production strategies. These methods have advantages over standard model-based ones due to lower complexity, less reliance on expert analysis, less information needed from hard-to-get sources, and good short-term forecasts (Werneck et al., 2022). Forecasting long-term oil production is challenging and often lacks granularity or requires extensive training data (Tadjer et al., 2021; Werneck et al., 2022).

Approaches that combine DD and MB methods are called hybrid methods (HyMs). However, it's worth noting that HyMs encompass different types of techniques. Some of these include (1) the use of simulation models to train DD methods and then replace them as proxy models and (2) the use of ML short-term forecasts to reduce uncertainty in model forecasts. In the present work, we will be considering only the second type. Xiao et al. (2023) presented a HyM to production forecast and uncertainty quantification. They used deep-learning-based methods with trained samples simulated from high-fidelity models. Borisova et al. (2021) used a HyM to combine capacitance-resistance models (CRM) and data-driven (Random Forest) to forecast oil production for a very short-term period of 10 days.

Ferreira et al. (2023) have proposed a HyM to select the best simulation models through DD forecasts. The simulation models are ranked based on their similarities to the DD daily production forecast. When comparing two simulation models, the best model is the one with the lowest maximum absolute error metric among all variables for all wells. The method was applied in a reservoir benchmark and considered two forecasting conditions, one with an idealized scenario without human interventions and the other with human interventions. The methodology was tested on a synthetic reservoir, and the production curves showed a stable trend even with the simulated interventions. This result suggests that, for this case, predicting the daily DD forecast will not be difficult. However, the simulation models still show a mismatch in production curves between history and forecast, even after data assimilation, resulting in uncalibrated models. It is crucial to carefully analyze the selection of the best models, as all the models selected will be based on the lowest maximum error in a challenging well scenario. Even using a mean error selection method among all curves may not solve the potentially biased rank created. Therefore, more research is needed to rank the simulation models for real scenarios where the well's production curves will always be challenging.

Further exploration of HyMs' potential and applicability in reservoir management and forecasting is essential, given their capacity to address and mitigate the limitations inherent in traditional MB or DD approaches. However, its application demands that both the simulation models and data-driven solutions be correctly calibrated to produce reliable results.

Data assimilation is a technique for incorporating observed data into simulation models, typically from well-production histories. This process helps reduce uncertainties in the model's predictions by updating the initial conditions and parameters to match observed behavior (Schiozer et al., 2019). By assimilating real data into the simulation models, they become valuable tools for life-cycle reservoir management. However, the simulation models obtained through the traditional data assimilation process present an issue characterized by a discontinuity in the transition from production history to the forecast in production curves (Almeida et al., 2018). This behavior is typical in such models, arising from the change in boundary conditions at the transition and the inability to represent well productivities accurately. Consequently, this common aspect negatively impacts model-based solutions' short-term forecast performance. Maschio and Schiozer (2023) proposed a two-stage data assimilation procedure to overcome this challenge. The first stage focuses on matching well rates and pressure, while the second stage aims to calibrate well productivity.

The data assimilation procedure developed by Maschio and Schiozer (2023) uses the productivity index (PI) calibration to improve the liquid rate curve in the history-forecast transition period. However, not all the models obtained through this procedure are suitable for short-term forecasts (e.g., 6 months or less) since the procedure only calibrates liquid rates. Thus, the phase flow rates (Qo, Qw, and Qg) are not necessarily matched. Hence, selecting a subset of the most appropriate models derived from the data assimilation process for short-term forecasting is necessary. This results in improved forecasts and reduced computational requirements.

This paper introduces a novel approach to aggregating individual ranks to retain information while ranking the best simulation models for forecasting. The proposed method extends upon and enhances the methodology presented by Ferreira et al. (2023) by implementing additional improvements, such as utilizing Maschio and Schiozer's (2023) calibrated simulation models during the history-forecast transition and accounting for human interventions via cumulative production curves. Furthermore, we have tested the proposed methodology in a real field.

Our main objective is to propose a HyM that combines the optimal features of numerical reservoir simulation (MB) with transformer models (DD) to make short-term decisions, effectively reducing uncertainty in short-term production forecasts for a real field. The HyM approach involves selecting the most suitable simulation models through a data-driven-assisted selection process. In this work, the employed DD technique is the TFT model.

Methodology

This section outlines our HyM, which is designed to improve the accuracy of short-term production forecasting within simulation models. The method is specifically designed for this purpose and employs a combination of numerical reservoir simulation and TFT models. The pivotal aspect of this HyM is the data-driven-assisted selection, wherein a subset of assimilated models is carefully chosen to optimize the forecasting process. Furthermore, we introduce a novel rank aggregation method called "SumRank Aggregation," which consolidates multiple ranks by adding the positions of individual elements to construct a composite rank.

Hybrid Approach

Figure 1 presents a summarized scheme of the proposed hybrid methodology.

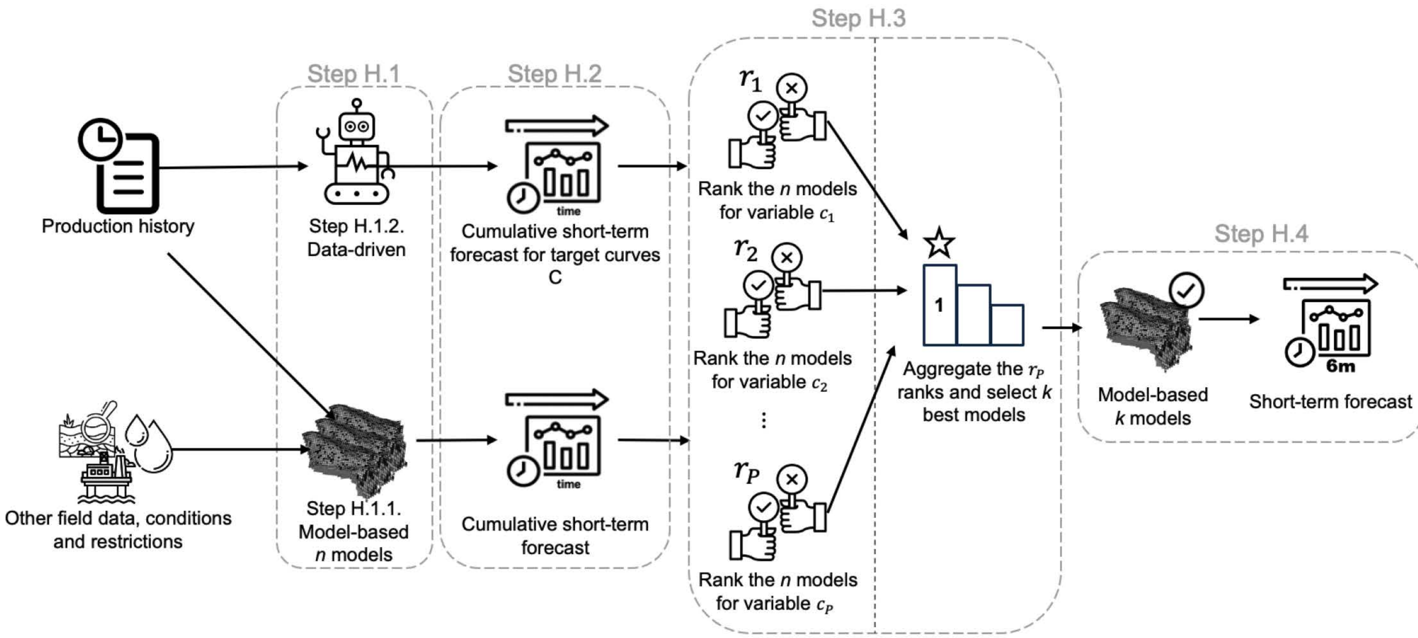

Figure 1—General flowchart of the proposed hybrid method.

The following steps guide the method:

H.1. Build the predictive models:

H.1.1. Built the MB models:

H.1.1.1. A set of *n* reservoir simulation models is constructed incorporating petrophysical, geological, and geophysical data, considering facility constraints and operational conditions.

H.1.1.2. Perform the data assimilation process on *n* models in two stages (as done by Maschio et al., 2023).

H.1.2. Build the DD model:

H.1.2.1. The DD model is trained with the same production history data used in the data assimilation process.

H.2. Perform the short-term production forecast of target curves C with period t_{st} and granularity Δ_{st} using all the simulation models and the DD technique. C is the set$(c_1, c_2, \dots, c_P)$ of production curves across the wells in the field (such as oil, water, and gas rates), where P corresponds to the scalar product between the number of wells and production curves.

H.2.1. Calculate the cumulative daily short-term forecasts obtained in step H.2 for all the simulation models and the DD technique.

H.3. Apply the SumRank Aggregation, considering the predicted values to be all MB cumulative short-term forecasts and the reference data to be the DD cumulative short-term forecasts obtained in step H.1.2. Then, select the top *k* best simulation models (see section 2.2. SumRank Aggregation for a description of the rank method).

H.4. The *k*-top selected simulation models can be used in short-term forecasting and reservoir management with period t_{st} and granularityΔ_{st}.

SumRank Aggregation

We develop a rank aggregation method to rank the simulation models. This methodology aims to include all relevant information from production curves for all wells (C) into the ranking process while avoiding a biased rank that could result from generating a single rank from an aggregated metric. We define individual rank as a single rank for one well and one production curve (e.g., P1-QW).

The SumRank Aggregation approach is summarized in Figure 2. It is composed of the following steps:

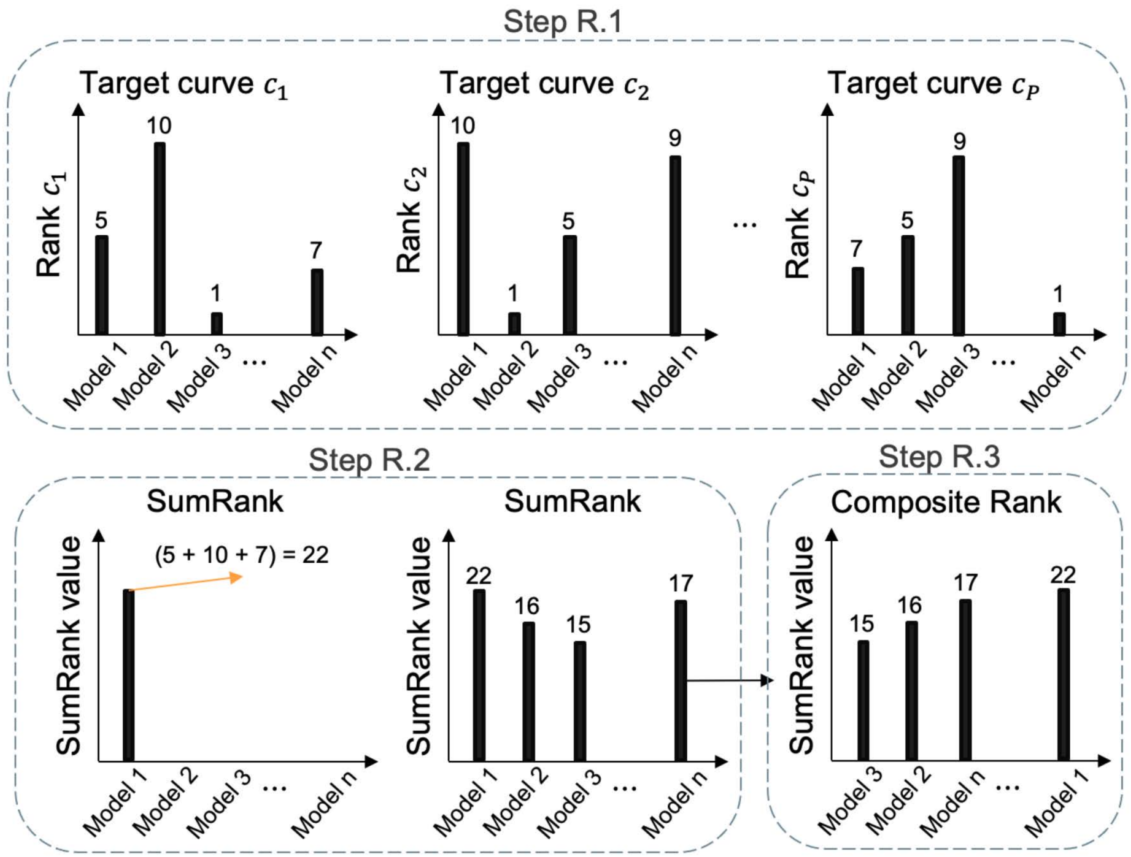

Figure 2—General flowchart of the proposed SumRank Aggregation. C is the set of production curves (Qo, Qw, and Qg) across the wells in the field. In R.1, the horizontal axis presents every simulation model, and the vertical axis represents its rank position for target curve c. R.2 presents a simple case where p = 3 (case where we have three target curves), such that the SumRank of a given model is the sum of each individual rank presented in Step R.1 for that model. In R.3, the horizontal axis presents the composite rank (r) of the simulation models, and in the vertical axis, the SumRank value for each model.

R.1. For each target curve in C:

R.1.1. Compare the predicted values for each n simulation model with reference data according to metric M.

R.1.2. Rank the models from the best to worst according to metric M, creating an individual rank.

R.2. Sum the positions across all individual ranks for each model. The number of individual ranks equals the total number of target curves in C.

R.3. Create the composite rank by sorting the models from the lowest to the highest according to the SumRank value (as presented in Figure 2 R.3).

Application

The proposed methodology was applied to a real dataset from a field from the Campos basin (Brazil) named S-Field. The results were compared with two MB baselines. In this section, we delineate the S-Field, specify the setup of the HyM for this application, elucidate the methodologies for each baseline, outline the procedure for evaluating results, and detail the experiments conducted.

Dataset

The S-Field is a deep-water heterogeneous turbiditic reservoir containing heavy oil. A porosity realization is illustrated in Figure 3. The reservoir comprises 8 horizontal production wells, and its pressure is supported by 4 horizontal injection wells and two aquifers. The production data span 2359 days and are divided into three distinct periods. The first part encompasses the production history, comprising the initial 2206 days used in data assimilation of the simulation models. The second part corresponds to the validation period from day 2207 to 2236. The last part is the forecasting period, from day 2237 to 2359. We have two sources for S-Field: the production history and 200 simulation models. The simulation models contemplate multiple uncertainties and have several different realizations, and they are used in the MB approach.

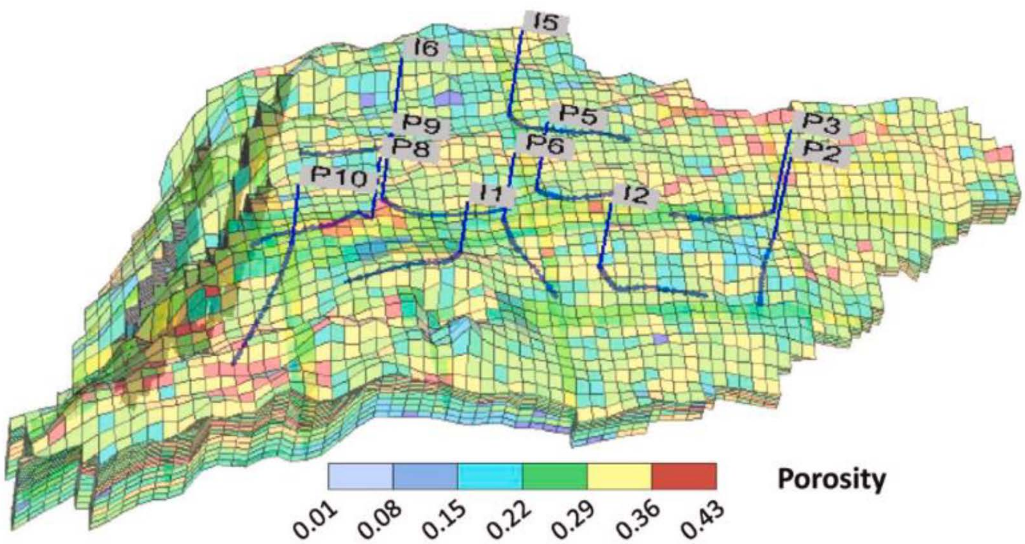

Figure 3—S-Field 3D porosity showing well locations from Maschio and Schiozer (2023).

Setup for the Hybrid Method

This section presents several setups and techniques to be applied in the general HyM workflow. The short-term forecasting period will vary based on the baseline, which will be compared (see section **Model-Based History Baseline** and **Model-Based Validation Baseline**). Still, the granularity is fixed and set to 1 day (Δst=1 day). The production forecast variables (C) analyzed were the Oil Production Rate (QO), Gas Production Rate (QG), and Water Production Rate (QW) for each well. The metric M used to rank the MB models was Normalized Quadratic Deviation with Sign (NQDS) (M=NQDS).

The fluid flow simulation for the MB models was performed in the commercial black-oil reservoir simulator IMEX from the CMG version 2017. We set the observed liquid rates as target constraints for injector and producer wells during the production history. We performed the first stage of data assimilation using the Ensemble Smoother with Multiple Data Assimilation (ESMDA) method (Emerick and Reynolds, 2013). The ESMDA was conducted considering 8 data assimilation steps with 200 models (n = 200). The data to be assimilated were QO and QW for producers and BHP for producers and injectors. In the second stage of data assimilation, we applied the PI calibration approach developed by Maschio and Schiozer (2023) to all simulation models in the transition from production history to validation period, which considerably improved the short-term forecast of the simulation models.

We used a TFT network with attention to perform the DD forecast (Lim et al., 2021; Vaswani et al., 2017). As known future inputs of TFT, we used the distribution of partial and total well closure calculated in the production history. The network's input and output sizes will vary based on the compared baseline, which will be treated in each section. For each well, the DD technique used the QO, QG, and QW information as input to forecast each curve *c* for that well. Therefore, one network was trained for each well and each curve *c*, all with the same architecture. In its pipeline, the DD technique also employed data augmentation of 6 hours (i.e., 4 points for each day). To ensure DD forecasts honor the well and platform restrictions, we applied a post-processing approach proposed by Ferreira et al. (2023). The wells' liquid rate should be lower than the maximum limit (2400 m^3); if this condition is not present, the oil and water rates are reduced proportionally. After the well restrictions were applied, the field's daily production of liquid, oil, water, and gas was within the platform processing capabilities (14400 m^3, 14400 m^3, 9920 m^3, and 12000000 m^3, respectively).

Baselines for Result Evaluation

The top *k* models selected through the HyM were compared to those selected by two model-based baselines: one that used historical data (MBHIST) and the other considering the validation period for an immediate-term production forecast (MBVAL). Figure 4 illustrates how a subset of simulation models (blue lines) is selected for the HyM (Figure 4. A), for MBHIST (Figure 4. B), and for MBVAL (Figure 4.C).

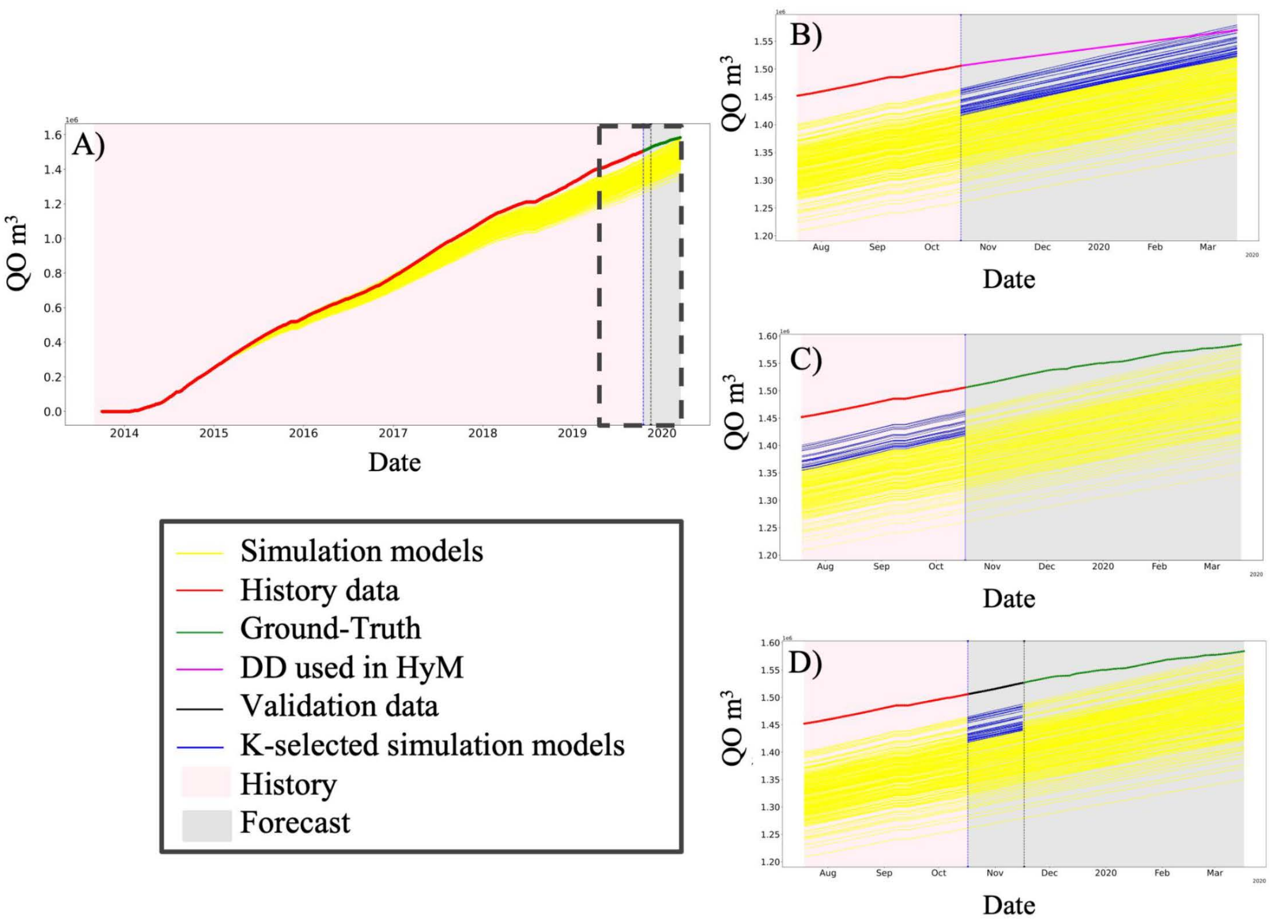

Figure 4—Comparison between the HyM, MBHIST, and MBVAL methodologies. A) All the cumulative production data is divided into history (red curve) and forecast (green curve, the ground-truth) periods, time series of the cumulative simulation models (yellow lines); B) HyM selecting the k simulation models (blue curves) using DD forecast (magenta curve); C) MBHIST selecting the k simulation models (blue curves) using the cumulative history data (red curve) and D) MBVAL selecting the k simulation models using the cumulative validation data (black curve).

To ensure comparability of results, the same data availability had to be guaranteed for both HM and MB approaches. To assure this, the training period for the TFT ranges until the end of the production history when compared to MBHIST and until the end of the validation period when compared to MBVAL.

In the HyM, the output of the DD technique also varied depending on the size of the data used in the MB baselines, corresponding to 152 days when compared to MBHIST and 121 days when compared to MBVAL. The input used was the same size as the output for both cases. The same set of *n* simulation models was used in both MB baselines and the HyM. Additionally, the NQDS was used as the rank metric M in both MB baselines and the HyM, and they followed a similar methodology flowchart.

Model-Based History Baseline. The MB history baseline selects the best models using the history period, following the process shown in Figure 5 and the steps below.

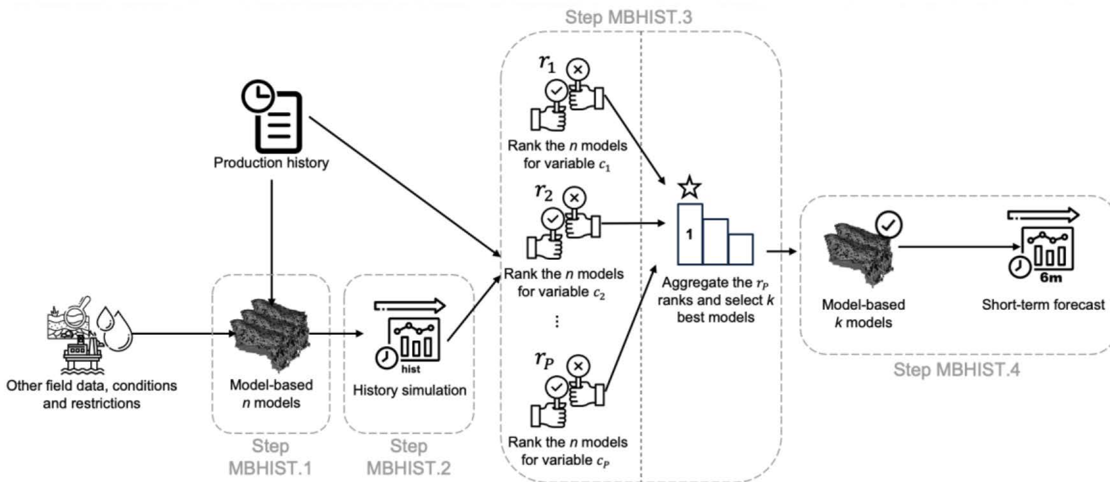

Figure 5—Flowchart of the full MBHIST history model selection method (baseline). The main difference between the MBHIST baseline and the hybrid method (Figure 1) is the second step, in which we did not consider data-driven forecasting.

MBHIST.1 Build the predictive models (see section **Hybrid Methodology** in item H.1. for a detailed description of the step MBHIST.1.).

MBHIST.2. Run the production simulation for all target curves C with period t_{hist} and granularity Δ_{hist} using all the simulation models.

MBHIST.2.1. Calculate the cumulative history for each target curve in C for all the simulation models.

MBHIST.3. Apply the SumRank Aggregation for all MB cumulative simulation results, assuming the predicted values as the simulation model results and the reference data as the real field cumulative data during production history. Then, select the top *k* best simulation models (see section **SumRank Aggregation** for a description of the rank method).

MBHIST.4. The *k*-selected simulation models can be used in short-term forecasting and reservoir management with period t_{st} and granularityΔ_{st}.

Model-Based Validation Baseline. The MBVAL followed a similar flowchart for selecting the best models as the MB approach. Still, with a few differences: instead of using the production history to choose the models, this approach selected the models according to an immediate-term production forecast performed during a validation period, as presented in Figure 6. The validation period (t_{val}) is 30 days long.

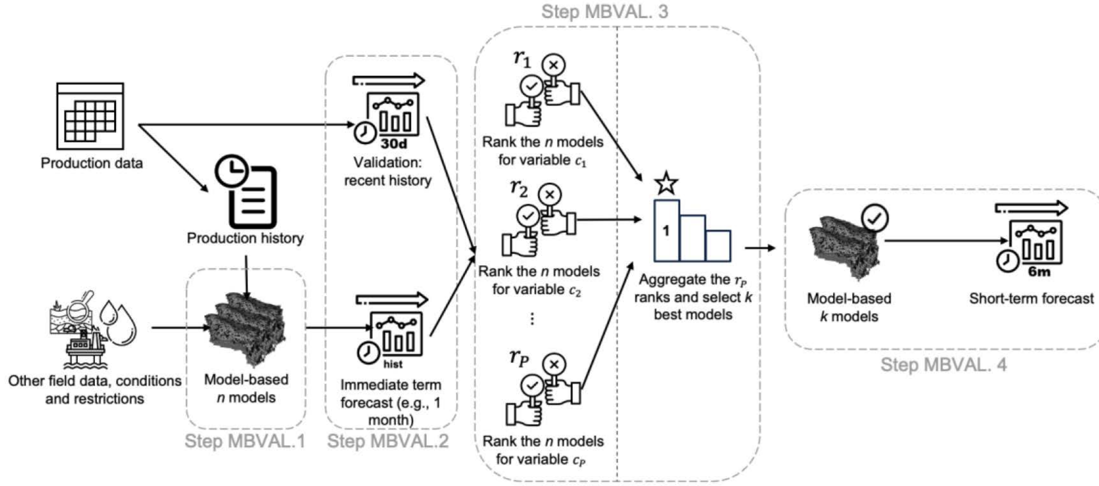

Figure 6—Flowchart of MBVAL model selection method (baseline). The main difference between the MBVAL and the MBHIST (Figure 5) is the second step (MBVAL. 2), in which we perform an immediate-term production forecast using the production history.

MBVAL.1. Build the predictive models (see the section **Hybrid Methodology** in item H.1. for a detailed description of this step). This step is performed in the same manner as for MBHIST.

MBVAL.2. Perform an immediate-term production forecast for all target curves in C with period t_{it} and granularity Δ_{it} using all the simulation models.

MBVAL.2.1. Calculate the cumulative immediate-term forecast for all target curves in C for all the simulation models.

MBVAL.3. Apply the SumRank Aggregation for all MB cumulative immediate-term forecasts, assuming the predicted values as the immediate-term forecast results and the reference data as the real field cumulative data during the validation period. Then, select the top k best simulation models (see section **SumRank Aggregation** for a description of the rank method).

MBVAL.4. The k-selected simulation models can be used in short-term forecasting and reservoir management with period t_{st} and granularityΔ_{st}.

Result Evaluation Metrics

The short-term forecasts were compared with the reference data (used as ground truth), and the results were evaluated through the rank similarity method proposed (Indicator GT).

NQDS. The NQDS represents the quality matching to the measurement errors, and it is helpful to determine whether the distribution of models is centered on the observed data. For all evaluations in the present work, in step R.1.1. of the SumRank Aggregation methodology, we used the NQDS as the metric M to measure the time series forecast error. The $NQDS_i$ for curve c_i is defined as in Equation (1):

$$NQDS_i = \frac{\sum_{j=1}^{\frac{t_{si}}{\Delta_{si}}}\left(prediction_{i,j} - observed_{i,j}\right)^2}{\sum_{j=1}^{\frac{t_{si}}{\Delta_{si}}}\left(Tol_{i,j} * observed_{i,j} + \gamma\right)^2} \frac{\sum_{j=1}^{\frac{t_{si}}{\Delta_{si}}}\left(prediction_{i,j} - observed_{i,j}\right)}{\left|\sum_{j=1}^{\frac{t_{si}}{\Delta_{si}}}\left(prediction_{i,j} - observed_{i,j}\right)\right|} \tag{1}$$

where the *prediction* is the forecasted value, either by the simulation models or DD method; the *observed* is the reference value, the ratio t_s / Δ_s is the total number of timesteps; *Tol* is an acceptable tolerance for the variable and γ is a constant employed to avoid division by zero. The indexes i and j represent the target curve and the timestep, respectively. In our evaluations, we chose the values of γ and *Tol* to be 0.01 and 0.1, respectively.

GT Rank. As explained earlier, the HyM and the MB baselines select simulation models based on ranks. In the HyM approach, we rank the models according to their similarity to the data-driven forecast. In contrast, we rank these models in the MB approach according to the production history (MBHIST) or validation period (MBVAL). Both these alternatives use references that aim to proxy the behavior of the future production date.

To evaluate the quality of the ranks, we also compare the simulation to the reference (future) data during the short-term forecast period, generating a Ground Truth (GT) rank of the best models. First, SumRank aggregation is applied, using the real data during the forecast period as the reference data and the MB short-term cumulative forecast as the predicted values; after that, the top *k* best models are selected.

Indicator GT. There are two principal metrics for comparing the ranked list: RBO and Kendall Tau (Webber, 2010; Fagin et al., 2003). While Kendall Tau focuses on pairwise comparison of ranks, RBO emphasizes the overlap between top-ranked elements, considering their positions within the lists. However, these metrics do not consider the magnitude of differences between ranks for each position. Therefore, we suggest an indicator that considers this characteristic and shows how reliable the rank is for certain positions and how much it was hindered for other positions.

We created an indicator to compare the composite ranks, called Indicator GT. Figure 7 shows an example of how the Indicator GT is calculated. For each of the *n* simulation models, we calculate the SumRank of the GT Rank (shown in the table "Composite Rank - GT) in Figure 7. Then, for each position of the rank generated by either the HyM or MB, we attribute the GT SumRank of the model that occupies that position. For example, the highlighted model in red (model 5) has a GT SumRank of 11. We attribute this value (Indicator GT) to the position the model occupies in the rank we expect to evaluate, so both the HyM and MB ranks in Figure 7 have an Indicator GT of 11 for model 5.

Composite Rank - HyM			Composite Rank - MB			Composite Rank – GT		
Rank	Models	Indicator GT	Rank	Models	Indicator GT	Rank	Models	Indicator
1	Model 1	7	1	Model 1	7	1	Model 1	7
2	Model 2	8	2	Model 2	8	2	Model 2	8
3	Model 3	9	3	Model 9	10	3	Model 3	9
4	Model 5	11	4	Model 5	11	4	Model 4	10
5	Model 4	10	5	Model 3	9	5	Model 5	11

Figure 7—An example of how the Indicator GT is formed for HyM and MB baselines.

The comparisons we make are position-wise between the Indicator GT of a given rank and the SumRank of the GT rank in that same position. With the indicator GT, we compared the similarity of the ranks to the GT Rank using several metrics (Symmetric Mean Absolute Percentage Error (SMAPE), Dynamic Time Warping (DTW), Chi-square Distance (Chi distance), and statistical tests). These metrics compare the quality of the ranks generated by the HyM to those generated by the MBHIST and MBVAL.

Results

The HyM was applied to a real dataset from a Brazilian oil field (S-Field) and was validated against two different model selection approaches, one consisting of all the history data (MBHIST) and the other one using the most recent data (MBVAL). We used NQDS to measure the time series forecast error to create the individual ranks and then combined these ranks into composite ranks using the SumRank Aggregation methodology. First, we applied the hybrid method to the *k* simulation models of the S-Field. Then, we compare the HyM, MB baselines (MBHIST and MBVAL) composite ranks against the GT Rank in terms of indicator GT versus rank order, considering k simulation models. Our experiments tested the selection of different top K simulation models, with $k = \{30 \text{ and } 200\}$. In this section, we show the results in a group manner, considering all wells in C for ranking and individually considering the curves for each well to better understand how each well behaves.

Multiple Selected Models (k = 30)

This evaluation aims to measure the performance of the best simulation models selected by the HyM and MBHIST baseline regarding production forecast compared with reference data. The 5-month forecast average NQD (absolute value of NQDS) against the reference data is shown in Figure 8. Overall, the HyM outperformed the MBHIST baseline. Among wells, P6 exhibits the most significant difference in NQD, while water rate has the greatest difference for variables. Figure 9 displays the time series for all variables for P6. The DD method used in HyM in the production forecast presents a very good fit to the reference data (green curve) for the cumulative QO, QG, and QW.

Compared with the MBHIST approach, the forecasts of models selected by the hybrid approach are more similar to reference data for target curves QO and QW but perform worse for QG, even with a good fit of the data-driven forecast. This worse result happens due to the composite nature of the rank, which considers all target curves C. On average, compared to the MB approach, the models chosen by HyM will perform better for most but not all target curves.

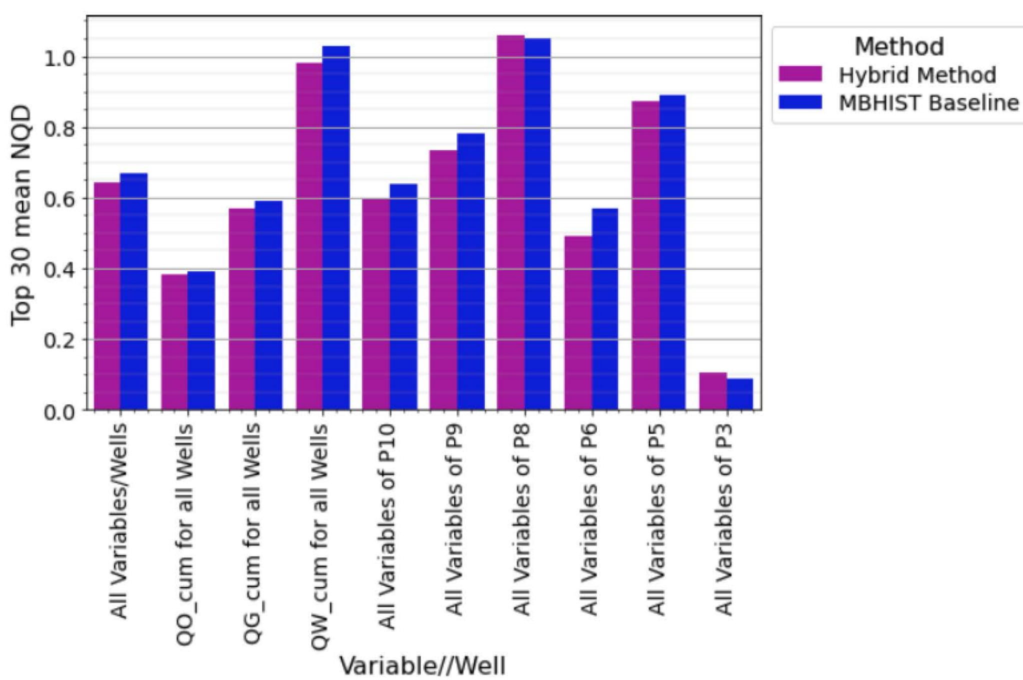

Figure 8—Average NQD versus the reference model obtained by the hybrid method and MBHIST baseline for the top 30 selection in the 5-month production forecast task. The closer to zero, the better the result. We present the global average results (all variables/wells) as well as the results grouped by each cumulative variable (QO, QG, QW) and each production well (P10, P9, P8, P6, P5, P3).

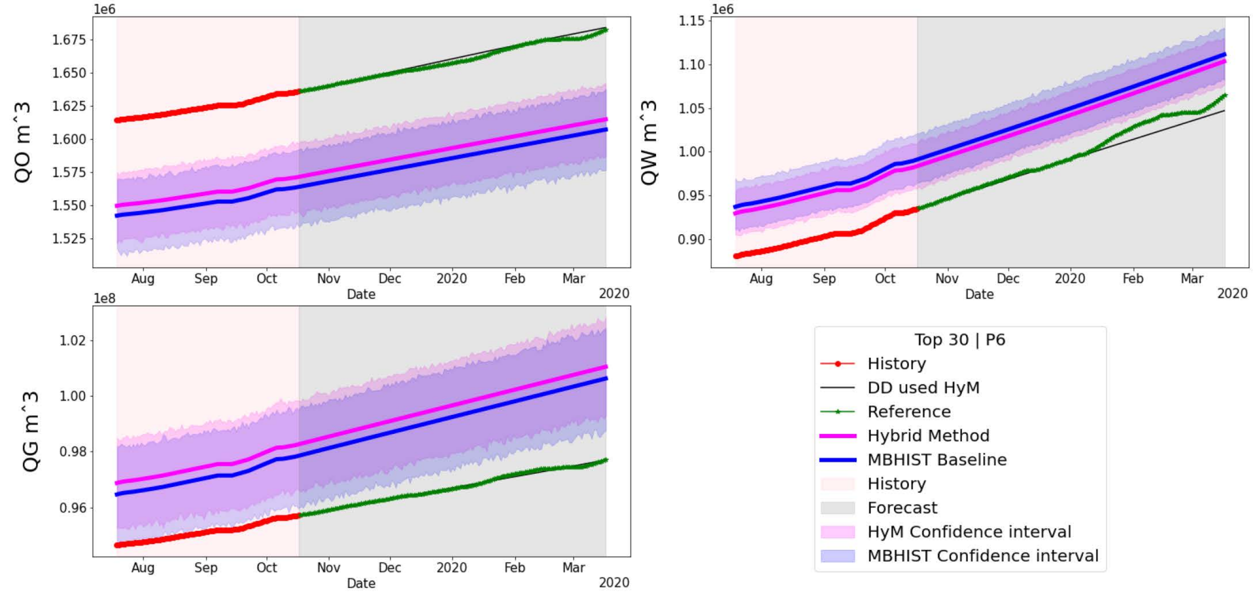

Figure 9—Time series of all the variables for wells P6, containing the last three months of production history (red background) and five months' forecast (gray background): history (red), reference model (green), best hybrid models (purple), best MBHIST baseline (blue), and DD used in HyM (black).

All Models (k=200)

In this section, we investigate HyM's rank quality and compare it with MB baselines using the indicator GT (see section Result Evaluation Metrics).

The composite rank comparison against the GT rank is shown in Figure 10 for HyM and MBHIST and Figure 11 for HyM and MBVAL. Overall, the models selected by the rank generated using the HyM successfully outperformed those selected using either the MBHIST or MBVAL ranks for all metrics (SMAPE, DTW, and Chi distance), see Table 1. We can see in these metrics that the difference between HyM and MBVAL (see Figure 11) is reduced compared to the difference between HyM and MBHIST (see Figure 10). This reduction leads us to identify that selecting models using the MBVAL approach was significantly better than using MBHIST, indicating that more recent data has a greater weight in reducing uncertainty.

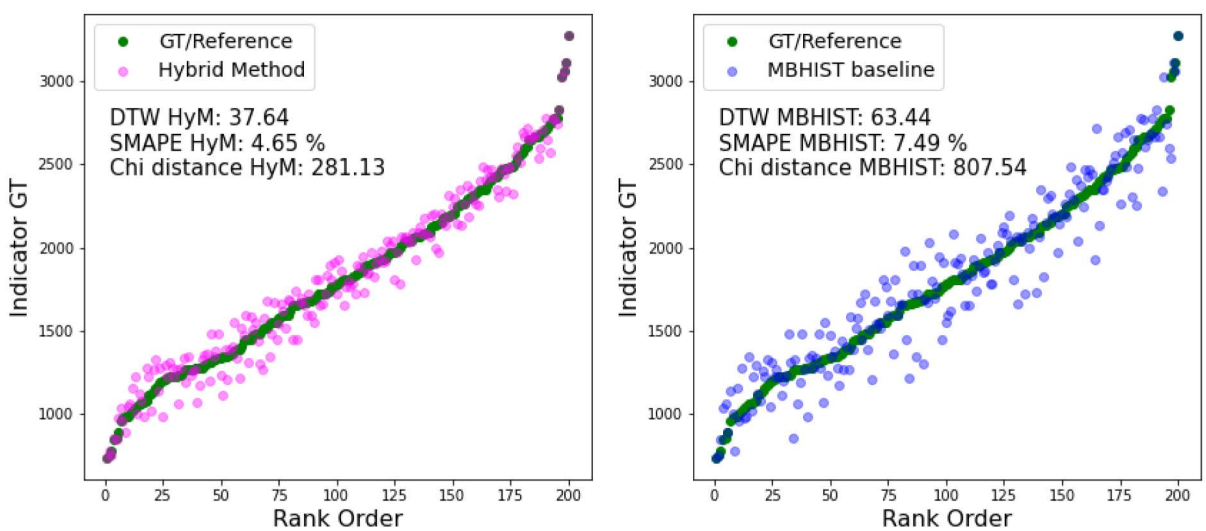

Figure 10—Comparison between GT and HyM ranks on the left and GT and MBHIST ranks on the right. We use the indicator GT versus the rank order to compare the approaches.

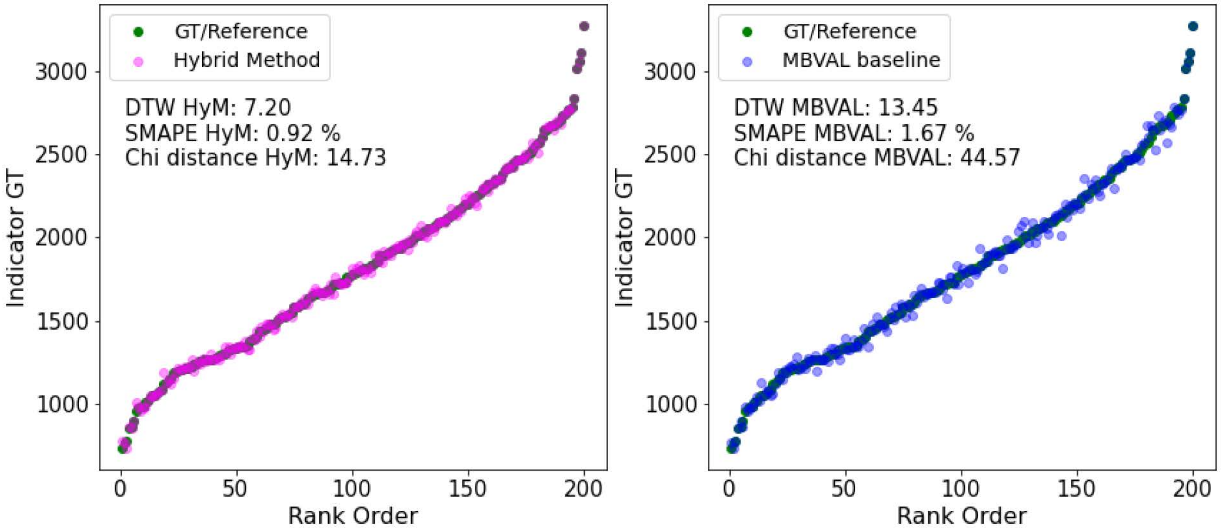

Figure 11—Comparison between GT and HyM ranks on the left and GT and MBVAL ranks on the right. We compare the approaches using the indicator GT versus the rank order.

Table 1—Metrics of comparison between GT, HyM, and MB baselines ranks.

Metrics	Hybrid	MBHIST	Hybrid	MBVAL
DTW	**37.64**	63.44	**7.20**	13.45
SMAPE	**4.65**	7.49	**0.92**	1.67
Chi distance	**281.13**	807.54	**14.73**	44.57

One interesting point is the difference between the HyM results (comparing Figure 10 and Figure 11, left graph). The DD-assisted selection used in HyM compared with MBVAL seems more effective in selecting the simulation models than the DD-assisted selection used in HyM compared with MBHIST. This effectiveness is because, compared with the MBVAL approach, an additional month (validation month) was available as a training set for the DD forecast. This extra month had a significant impact due to two wells that had just gone through a closure period. One example is well P9, which can be seen in Figure 12. The cumulative data constant (red line) indicates the well is closed, and just after (vertically dotted blue line) is the DD forecast used in HyM (blue line) to compare with MBHIST. The DD forecast used in HyM (black line) to compare with MBVAL start (vertically dotted black line) one month after the previously mentioned closure.

In Figure 12, we present the transition period between history and forecast. Only the last months of history (training data) are shown for better clarity. We see that the DD forecast (blue curve) used in the HyM used for the comparison with MBHIST is performed just after the well closure (constant values for cumulative curves). This closure causes the forecast to have underrated daily production (represented by the lower inclination in the forecast curve compared to observed future data). Since the closure data is recent, it strongly impacts the DD forecasted trend. The DD model can recalibrate the trend according to the recent production history when an additional month is provided for training. For both cases, however, the HyM approach performed better than MB, but this demonstrates how the quality of the training data is a crucial aspect of robust model selection.

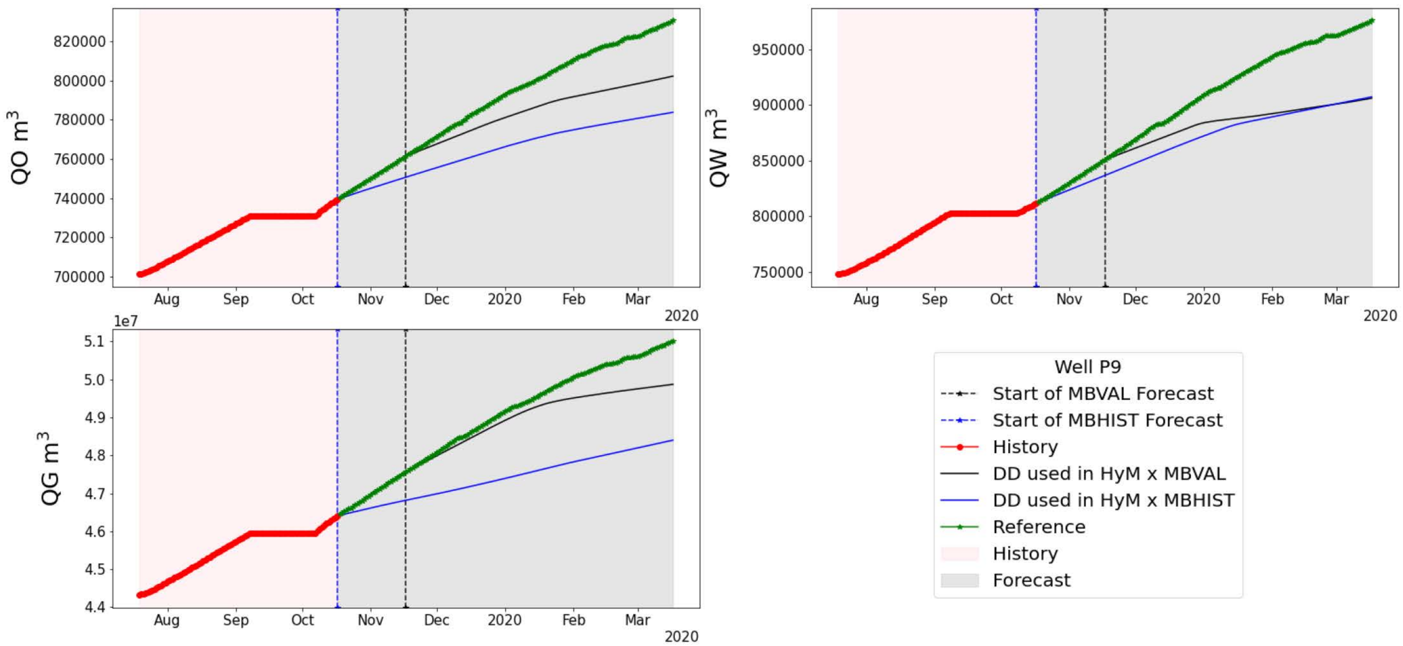

Figure 12—Time series of all the variables for wells P9, containing the last three months of production history (red background) and five months' forecast (gray background): history (red), reference model (green), DD used in HyM × MBHIST (black) and DD used in HyM × MBVAL (blue)

Well analysis

This section presents the rank similarity analysis between HyM and MBVAL made per well. For each well, all target curves and simulation models are considered. All the composite rank comparisons against the GT rank for all wells are shown in Figure 13. For wells P5, P6, P8, P9, and P10, the ranks generated by HyM overperformed the ranks generated by the MBVAL approach for all metrics (see Table 2; highlighted numbers indicate the approach performs better). The exception was the well P3, in which MBVAL performed slightly better in all metrics.

Table 2—Metrics of comparison between GT and HyM ranks and MBVAL baseline ranks for wells P3, P5, P6, P8, P9, P10.e

Metrics	P3		P5		P6		P8		P9		P10	
Approach	HyM	MB	HyM	MB	HyM	MB	HyM	MB	HyM	MB	HyM	MB
DTW	3.9	**1.5**	**0**	3.7	**1**	2	**0**	1.2	**3.8**	12	**1.3**	4.2
SMAPE	3.6	**1.4**	**0**	4.1	**1**	1.8	**0.1**	1.4	**5.8**	11.8	**3.14**	5.5
Chi distance	33	**4.5**	**0**	27	**3.8**	9	**0.1**	4.1	**72**	252	**26**	40

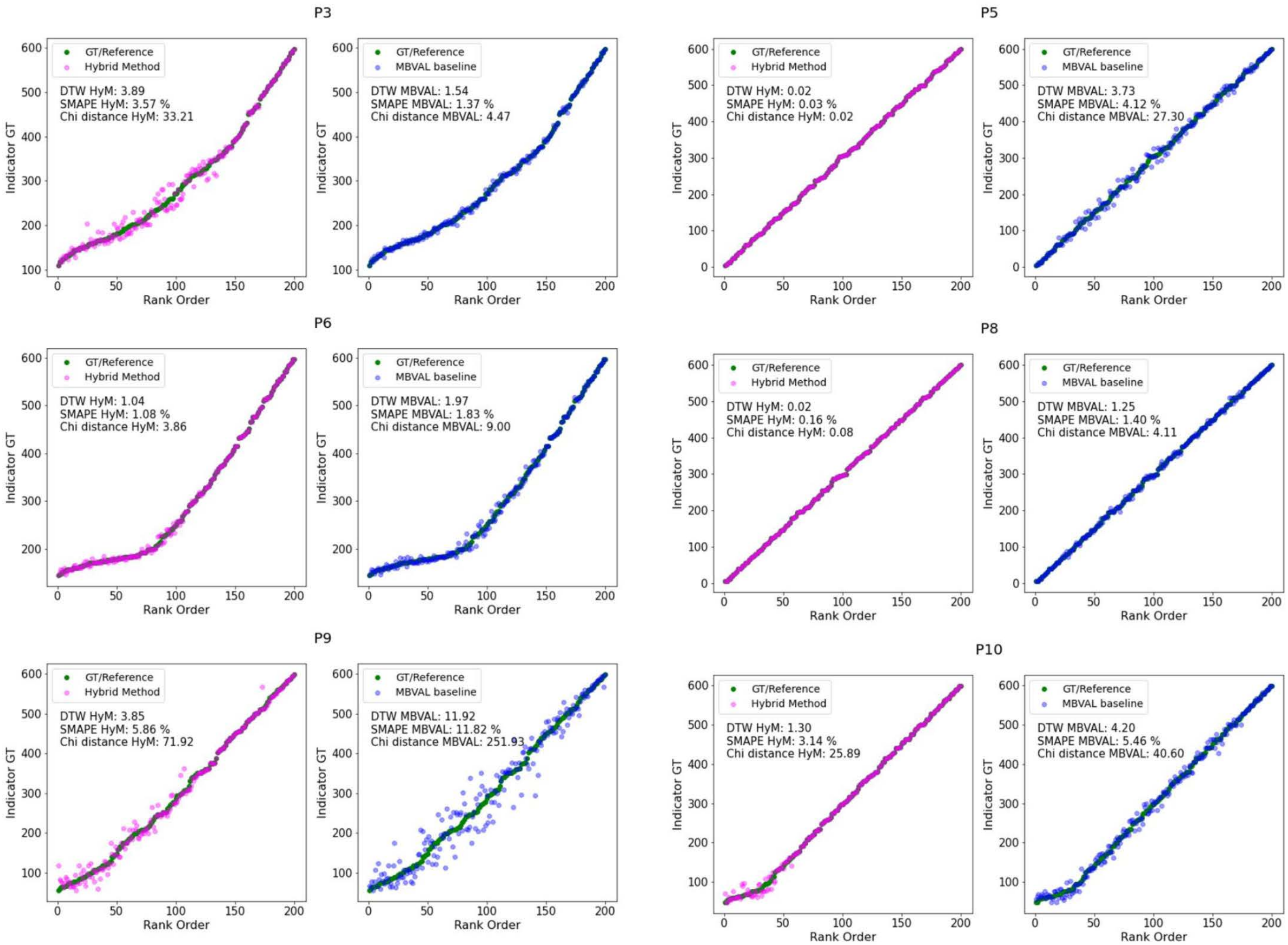

Figure 13—Comparison between GT and HyM ranks on the left and GT and MBVAL ranks on the right for wells P3, P5, P6, P8, P9 and P10. We compare the approaches using the GT indicator with the rank order for each well.

Wells P9 and P10 were challenging because, as discussed in the previous section, they had an interruption time (e.g., well closure) window right before the forecast period, making it difficult for DD to forecast. In Figure 14, we also show the rank similarity between HyM and MBHIST. As noted, the well closure causes a large discrepancy between the HyM (left) and MBHIST (right) to the ground truth, also adding dispersion to the results of the aggregated rank shown in Figure 10. But even in this extreme case, the rank aggregate produced by HyM follows the ground truth reasonably well (SMAPE<5%, see left graph in Figure 10); this demonstrates that SumRank is a robust indicator to aggregate the ranks.

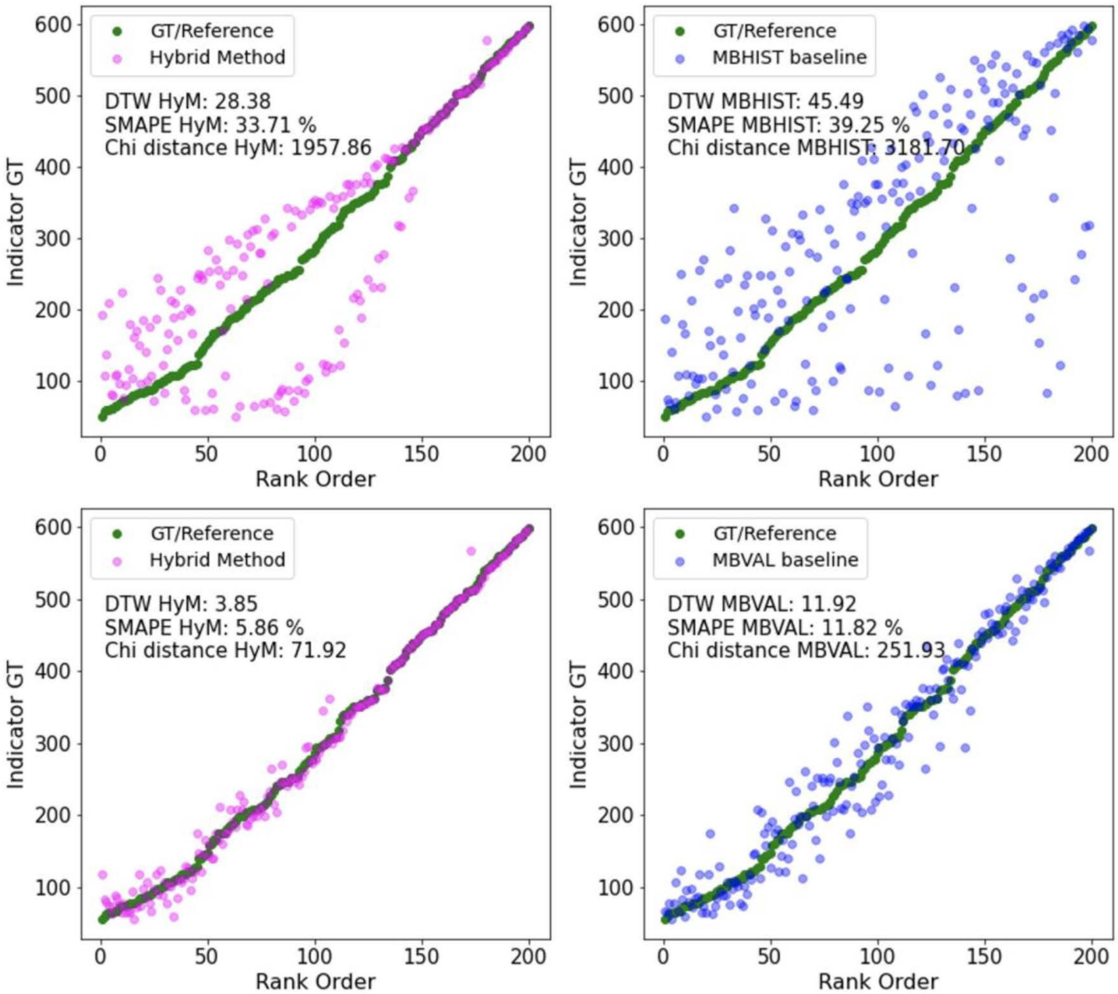

Figure 14—Comparison between GT and HyM ranks on the left and GT and MB baselines ranks on the right. We compare the approaches using the GT indicator with the rank order for well P9.

Conclusions

We have developed an innovative approach combining numerical reservoir simulation (MB) with transformer models (DD) to mitigate uncertainty in short-term production forecasts for actual field operations. Our hybrid methodology (HyM) leverages the data-driven (DD) approach to select a subset of simulation models from a pool of assimilated models.

We tested the HyM against two MB baselines, referred to as MBHIST and MBVAL. For both cases, using a DD forecast as a reference for selecting simulation models for short-term forecasting has shown to be more effective than using real historical data. To compare the results in these two cases, we analyzed production data for each well, revealing that well intervention strongly affects the DD forecasted trend in selecting simulation models. The proposed SumRank aggregation is a robust alternative for ranking the quality of the models, even in these extreme cases with well closures in the time window right before the forecast period. It produced an aggregate rank that reasonably follows the ground truth.

Tests have shown that our approach outperforms the MB baselines for model selection. The new methodology showed promising results by selecting models with attenuated errors for oil/water rates during the history-forecast transition, complementing the data assimilation procedure. The approach is versatile, not tied to any specific ML algorithm, and effectively minimizes uncertainties, particularly in complex fields with numerous wells.

Further development is needed to explore methodologies for dealing with human intervention, which will help in daily forecast tasks, as the proposed HyM has proven successful in real case studies using cumulative data.

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." The authors thank Schlumberger and CMG for software licenses and the support of Energi Simulation.

References

Bryan Lim, Sercan Ö. Arık, Nicolas Loeff, Tomas Pfister, Temporal Fusion Transformers for interpretable multi-horizon time series forecasting, *International Journal of Forecasting*, Volume **37**, Issue 4, 2021, Pages 1748–1764, ISSN 0169-2070, https://doi.org/10.1016/j.ijforecast.2021.03.012.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). *Curran Associates Inc.*, Red Hook, NY, USA, 6000–6010.

Alexandre A. Emerick, Albert C. Reynolds, Ensemble smoother with multiple data assimilation, *Computers & Geosciences*, Volume **55**, 2013, Pages 3–15, ISSN 0098-3004, https://doi.org/10.1016/j.cageo.2012.03.011.

Cao, Q., Banerjee, R., Gupta, S., Li, J., Zhou, W., Jeyachandra, B. 2016. Data driven production forecasting using machine learning. Presented at the SPE Argentina Exploration and Production of Unconventional Resources Symposium, Buenos Aires, Argentina, June. https://doi.org/10.2118/180984-MS

Mirzaei-Paiaman, A., Santos, S. M., Schiozer, D. J. 2021. A review on closed-loop field development and management. *Journal of Petroleum Science and Engineering*, **201**. https://doi.org/10.1016/j.petrol.2021.108457

Schiozer, D. J., Santos, A. A. S., Santos, S. M. G., von Hohendorff, J. C. Filho 2019. Model-based decision analysis applied to petroleum field development and management. *Oil & Gas Science and Technology – Revue d'IFP Energies nouvelles*, **74**. https://doi.org/10.2516/ogst/2019019

Maschio, C., Avansi, G. D., Silva, F. B. M., Schiozer, D. J. 2022. Data assimilation for uncertainty reduction using different fidelity numerical models. *Journal of Petroleum Science and Engineering*, **209**. https://doi.org/10.1016/j.petrol.2021.109851

Cong Xiao, Shicheng Zhang, Xingfang Ma, Tong Zhou, Tengfei Hou, Fushan Chen, Robust production forecast and uncertainty quantification for waterflooding reservoir using hybrid recurrent auto-encoder and long short-term memory neural network, *Geoenergy Science and Engineering*, Volume **230**, 2023, 212252, ISSN 2949-8910, https://doi.org/10.1016/j.geoen.2023.212252.

Julia Borisova, Anastasia Aladina, Nikolay O. Nikitin, Hybrid Modelling of Environmental Processes using Composite Models, *Procedia Computer Science,Volume* **193**, 2021, Pages 256–265, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2021.10.026

Almeida, F. L. R., Formentin, H. N., Maschio, C., Davolio, A., Schiozer, D. J. 2018. Influence of additional objective functions on uncertainty reduction and history matching. Presented at the SPE Europec featured at 80th EAGE Conference and Exhibition, Copenhagen, Denmark, June. https://doi.org/10.2118/190804-MS

Maschio C. and D. J. Schiozer. A new procedure for well productivity and injectivity calibration to improve short-term production forecast. *Geonergy Science and Engineering*, **229**, 7 2023.

A.-A. H. Al-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.

Yu, Y., Liu, S., Liu, Y., Bao, Y., Zhang, L., Dong, Y. 2021. Data-driven proxy model for forecasting of cumulative oil production during the steam-assisted gravity drainage process. *ACS Omega*, **6**. https://doi.org/10.1021/acsomega.1c00617

Webber, W., Moffat, A., Zobel, J. 2010. A similarity measure for indefinite rankings. *ACM Transactions on Information Systems*, **28**. http://doi.acm.org/10.1145/1852102.1852106

Fagin, Ronald, Ravi Kumar, and Dakshinamurthi Sivakumar. "Comparing top k lists." *SIAM Journal on discrete mathematics* **17.1** (2003): 134–160.

Tadjer, A., Hong, A., Bratvold, R. B. 2021. Machine learning based decline curve analysis for short-term oil production forecast. *Energy Exploration & Exploitation*, **39**. https://doi.org/10.1177/01445987211011784

Werneck, R. O., Prates, R., Moura, R., Goncalves, M. M., Castro, M., Soriano-Vargas, A., Mendes-Junior, P. R., Hossain, M. M., Zampieri, M. F., Ferreira, A. M., Davolio, A., Schiozer, D. J., Rocha, A. 2022. Data-driven deep-learning forecasting for oil production and pressure. *Journal of Petroleum Science and Engineering*, **210**. https://doi.org/10.1016/j.petrol.2021.109937

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. Lusquino 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.