OTC-32167-MS

A New Hybrid Data-Driven and Model-Based Methodology for Improved Short-Term Production Forecasting

Vitor Hugo de Sousa Ferreira, Manuel Castro, Renato Moura, Rafael de Oliveira Werneck, Marcelo Ferreira Zampieri, Maiara Moreira Gonçalves, Oscar Linares, and Soroor Salavati, Unicamp; Leopoldo Andre Dutra Lusquino Filho, Unesp/Unicamp; Pedro Ribeiro Mendes Júnior, Alexandre Mello Ferreira, Alessandra Davolio, Denis Jose Schiozer, and Anderson Rocha, Unicamp

This paper was prepared for presentation at the Offshore Technology Conference held in Houston, TX, USA, 1 – 4 May, 2023.

Abstract

Model-based (MB) solutions are widely used in reservoir management and production forecasting throughout the life-cycle of oil fields. However, such approaches are not often used for short-term (up to six months) forecasting due to the immediate-term productivity missmatch and the large number of models required to honor uncertainties. Recently developed data-driven (DD) techniques have shown promising performance in immediate term forecasting (from days to months) while losing performance as the timeframe increases. This work, proposes and investigates a hybrid methodology (HM) that combines MB and DD techniques focusing on improving the short-term production forecast. A common practice in reservoir management to understand the impact of uncertainties, is to build an ensemble of simulation model scenarios to assess the impact of these uncertainties on production forecasts. The proposed HM relies on the DD-assisted selection of a subset of models from the set of assimilated (posterior) models. Specifically, the pool of MB models is ranked based on their similarities to the DD production forecasts in the immediate term (e.g., one month), followed by the selection of the top models. The selected MB models are then used in the short-term forecasting task. In a case study for an offshore pre-salt reservoir benchmark, the proposed HM is compared to two baselines: one purely DD and another fully MB. The case study considered two forecasting conditions: human intervention-free with restrictions (HIF-R), with no intervention in the controls except to follow physical restrictions, and with human interventions (WHI), following optimization rules. Our results showed that the HM significantly outperformed the MB baseline, regardless of forecasting condition (HIF-R and WHI) or variables (pressure and oil/water/gas rates) for all evaluation metrics (time series similarity and rank-based) and top-selected models tested. The hybrid approach also helped improve the well productivity uncertainty that emerged from the data assimilation. Such results indicate that the performance of MB short-term forecasts can be enhanced when assisted by DD techniques, such as in our proposed HM. Comparing these two approaches, the best forecasts were split between the HM and the DD baseline. In the partially idealized HIF-R conditions, the DD baseline was best when the forecast trend was steady. However, the HM was superior for the more complex production behaviors. In the more realistic

WHI conditions, the HM outperformed the DD baseline in almost every aspect tested given the inability of the chosen DD technique to leverage known interventions. This work is the first effort to improve MB short-term production forecasts, using production data, with a machine learning technique through a proposed HM. The proposed DD-assisted selection of models proved successful in a benchmark case study, which means it is promising for application in other fields and for further development.

Introduction

Reservoir management can be simply defined as optimizing the economic recovery of oil and gas from a reservoir bounded by capital and operational expenses (Fanchi 2002; Thakur 1996). This multi-disciplinary activity relies on decision making based on data and models and considers financial, technological, and human resources (Schiozer et al. 2019; Thakur 1996). Production forecasting is an essential part of reservoir management as reliable forecasts of future reservoir performance can be used to make informed decisions in the present, to maximize economic return or minimize risks in the future. Many methods and techniques have been developed to forecast oil production, which can be divided into two categories: Model-Based (MB) and Data-Driven (DD) (Yu et al. 2021).

MB methods rely on the mathematical modeling of the physics involved in oil production. The industry standard numerical reservoir simulation is an example of such MB methods. The numerical reservoir simulators have been developed since the 1950s (Aziz and Settari 1979) and have been very successful in reservoir management and production forecasting throughout the life cycle of oil fields (Mirzaei-Paiaman et al. 2021). Such simulation models can be very complex and include many different equations to model physical phenomena such as high-fidelity fine geological grid meshes, fluid compositional models, fracture modeling, rock-fluid iteration representation, phase-behavior modeling, integration with surface facilities, geomechanical models, among others. In general, the more complex the model, the more resources are needed to collect information, build and compute the simulations (Ligero et al. 2003; Maschio et al. 2022). In light of this, lower-fidelity numerical reservoir simulation models have been developed as a way to speed up this time-consuming specialist work (Schiozer et al. 2019).

Recently, many DD approaches have been proposed to tackle the production forecast task (Cao et al. 2016; Makinde 2017; Werneck et al. 2022; Xiong et al. 2020; Yu et al. 2021). Such methods rely on statistical and machine learning techniques to leverage collected production data into meaningful forecasts. These methods present advantages over the standard model-based methods due to their lower complexity, being less reliant on expert analyses, requiring less information from hard-to-get sources, and producing good short-term forecasts (Werneck et al. 2022). However, long-term oil production forecasts are very challenging and often result in a loss of granularity in the prediction, or they require similar long-term training data (Tadjer et al. 2021; Werneck et al. 2022). Additionally, due to the difficulty in generating new scenarios unforeseen by the data, reservoir management based on purely data-driven methods is still an open problem, even though such techniques can support decision making during reservoir management in the short-term.

Another class of predictive models available is the physics-based DD approaches, such as proxy models (Silva et al. 2020) and Capacitance-Resistance Model (CRM) (Kansao et al. 2017; Nguyen et al. 2011). These models have been proposed as a middle-ground between the aforementioned MB and non-physics–based DD approaches.

Given the inherent characteristics of MB and DD models, combining such approaches into hybrid methods (HM) has been proposed across various fields of study (Borisova et al. 2021; Cai et al. 2022; Krasnopolsky and Fox-Rabinovitz 2006; Liao and Kottig 2016; Netto et al. 2019). However, the investigation of HMs to solve reservoir forecasting and management is still fledgling (Nikitin et al. 2022; Temirchev et al. 2019). Thus, the potential of such hybrid methods and models for reservoir management and forecasting should be investigated further to mitigate the drawbacks associated with dedicated MB or DD solutions individually.

A common practice in reservoir management under uncertainties is to build many numerical reservoir simulation models to consider the range of parameter uncertainties in the forecasts. Such models undergo a process of data assimilation using the production history to reduce uncertainty and obtain a set of statistically well-represented simulation models to be used in forecasting tasks (Schiozer et al. 2019). The models obtained through this data assimilation process have been tested in numerous conditions and proved valuable in life-cycle reservoir management. However, not all the models obtained through the traditional data assimilation process are suitable for predicting short-term forecasts (e.g., 6 months or less) because well productivities are not necessarily matched and given the change of boundary conditions from data assimilation mode (rates) to forecast mode (pressure) (Almeida et al. 2018). Additionally, the life-cycle reservoir management requires the generation of many models (often 100 or more) as the different uncertainties need to be statistically well-represented. However, these uncertainties may not promote notable differences in short-term forecasts. Therefore, one could select only a set of the most appropriate models generated through the data assimilation process to be used in short-term forecasts, resulting in improved forecasts and reduced computational time.

Our objective is to investigate and propose a HM suitable for short-term (e.g., 6 months ahead) production forecasts, combining numerical reservoir simulation (MB) and recurrent neural network (DD) techniques. Specifically, such hybrid methodology is accomplished by the data-driven–assisted selection of a subset of assimilated simulation models.

Hybrid Methodology

The general definition of the proposed hybrid method is shown in Figure 1.

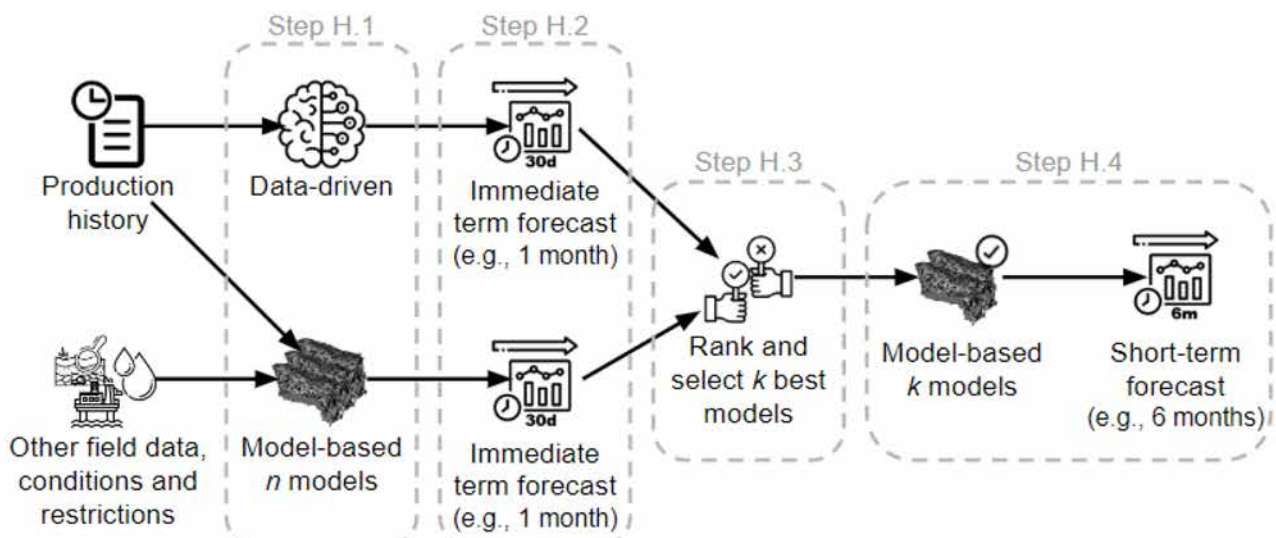

Figure 1—General flowchart of the proposed hybrid method.

The method considers the following steps:

H.1. Build the predictive models

a. Build the MB models

   i. A set of *n* reservoir simulation models are built using petrophysical, geological, and geophysical data, while considering the restrictions of the facilities and the operating conditions

   ii. Such *n* models undergo the process of data assimilation using the oil/gas/water production history.

b. Build the DD model

i. The same production history data used in the data assimilation process is employed to train and validate a DD model

H.2 Perform an immediate term production forecast of variables *V* with period t_{it}, and granularity Δt_{it}, using all the MB models and the DD technique

H.3. Rank and select the *k* best MB models according to metric *M*

a. Compare each of the *n* MB immediate term forecasts with the ones obtained by the DD technique using metric *M*
b. Rank the *n* MB models from best to worst according to metric *M*
c. Select the top *k* MB models

H.4. The *k* selected MB models can then be used in short-term production forecasting and reservoir management with period t_{st}, and granularity Δt_{st}

Application

We applied the HM to a reservoir benchmark and compared the results with two baselines; one fully MB and another purely DD. In this section, we describe the benchmark, define the specifics of the HM for this application, explain the methodologies for each baseline, describe the result evaluation procedure, and delineate the performed experiments.

Dataset

This work used the UNISIM-IV-24 benchmark dataset (Botechia et al. 2022), which is a synthetic case based on a carbonate reservoir with typical features of Brazilian pre-salt reservoirs. This benchmark was built using public data (ANP – Brazilian National Agency of Petroleum, Natural Gas and Biofuels) (Correia et al. 2020). The geological model contains open fracture networks and karst features (both high permeability features), while the fluid model considers 5 pseudo-components with high CO_2 content (Correia et al. 2020). The UNISIM-IV-24 benchmark has a defined production strategy considering 6 producers (P11, P12, P13, P14, P15, and P16), 6 Water Alternating Gas (WAG) injectors (I11, I12, I13, I14, I15, and I17) and a dedicated gas injector (I16), with 3 years and 3 months of production history (starts on April 27th, 2021 and ends on August 2nd, 2024).

We use three resources from UNISIM-IV-24:

- the reference model
- the simulation models
- the production history

The reference model represents the true earth model, it has a fine grid and is used to extract reservoir data to create the simulation models, mimicking what happens in a real field under production. In our case, the reference model is used solely as the final evaluation of the short-term forecasts and no design decision is based on this model. The simulation models have coarse grids built considering data extracted from the reference model. They contemplate multiple uncertainties and have several different realizations. These models are used in the MB approach.

The production history aims to be realistic and emulates the field operations. It therefore considers partial and total production shut-ins and is based on the reference model with noise included. This production history can be used to train, validate, and be assimilated in the MB and DD techniques.

Specific Choices for the Hybrid Method

The application of the HM for this work requires the definition of several constants and techniques to be applied in the general HM defined in Figure 1.

The immediate term and short-term forecasting periods were 30 and 184 days, respectively (t_{it} = *30 days*; t_{st} = *184 days*), while the granularity was 1 day for both forecasts ($\Delta t_{it} = \Delta t_{st} = 1$ *day*). The production forecast variables (*V*) considered were production Bottom Hole Pressure (BHP), Oil Production Rate (QO), Gas Production Rate (QG), and Water Production Rate (QW) for each well. The metric used to select the best MB models (*M*) was the Normalized Quadratic Deviation with Sign (NQDS) (see section **Result Evaluation Metrics** for a description of the metric).

We used a compositional commercial reservoir simulator software to build the MB models. We performed the data assimilation with the Ensemble Smoother with the Multiple Data Assimilation (ESMDA) method, considering the local objective functions of NQDS for BHP, QO, QG, and QW for each well. A total of 100 assimilated models were considered ($n = 100$). We used a simulation time step of 1 day to correspond to $\Delta t_{it} = \Delta t_{st} = 1$ *day*.

A Gated Recurrent Unit (GRU) Recurrent Neural Network (RNN) proposed by Werneck et al. (2022), named *GRU2_10*, was used as the DD technique. The neural network (NN) considered input and output sizes of 85 and 30 days, respectively. For each well, the DD technique used the BHP, QO, QG, and QW information as input to forecast each *V* for that well. Therefore, one NN was trained for each well and each variable *V*, all with the same architecture. The DD technique also employed anomaly detection through z-score and data augmentation of 6 hours (i.e., 4 points for each day) in its pipeline (Werneck et al. 2022). Post-processing was performed in the DD forecasts to ensure that they honored the well and platform restrictions. For the wells, we verified if the liquid rates were lower than the maximum limit (8,000 m^3). If not, the oil and water rates were reduced proportionally to their forecasted values following Equation (1).

$$V_{pp} = V_{raw}\min(1,\ V_{lim}/V_{raw});\ V \in \{BHP, QO, QG, QW\} \quad (1)$$

where V_{pp} is the post-processed forecast for variable *V*, V_{raw} is its raw forecasted value, and V_{lim} is its maximum production limit.

After the well restrictions were applied, we checked if the fields' daily production of liquid, oil, water, and gas were within the platform processing capabilities (28,617 m^3, 28,617 m^3, 23,848 m^3, and 12,000,000 m^3, respectively). Otherwise, the non-conforming variables were reduced for each well proportionally to their forecasted rates following Equation (1).

We considered different numbers for the selection of the *k* best models, which are detailed in the **Experiments** section.

Baselines for Result Evaluation

The results obtained by the hybrid methodology were compared to two baselines; one fully MB and another purely DD.

Model-Based Baseline. The MB baseline followed a similar flowchart for selecting the best models as the hybrid one, however, instead of a DD-assisted selection of the models, where the DD production forecast is used, the baseline selected the models according to the best metrics in the history period. The set of n simulation models is the same for the MB baseline and the hybrid method and, like the hybrid method, the MB baseline used the NQDS as the rank and selection metric.

The MB baseline selection process followed the flowchart of Figure 2 and steps below.

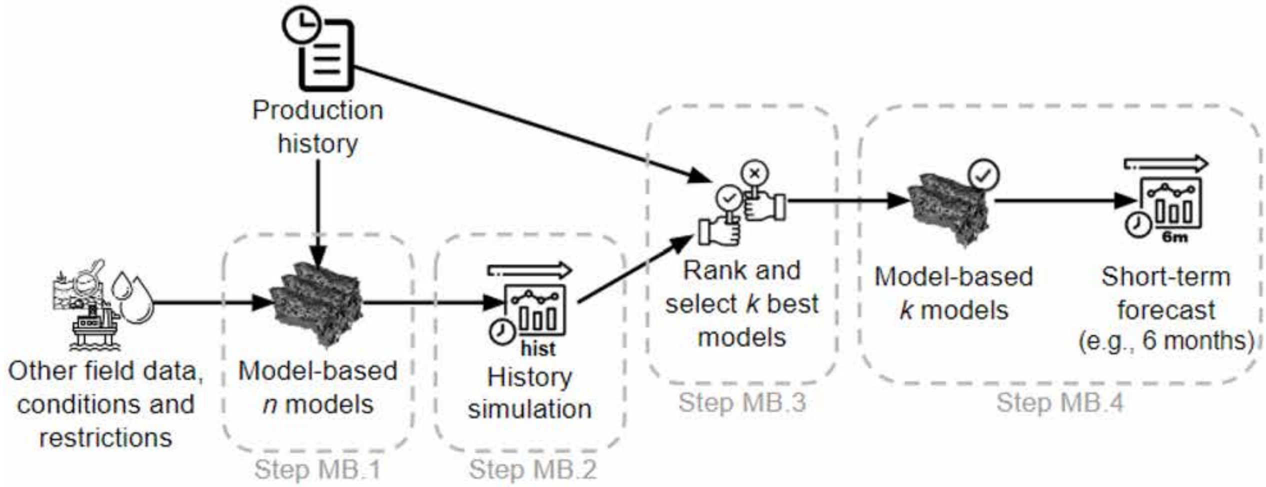

Figure 2—Flowchart of the fully MB model selection method (baseline). The main difference between the MB baseline and the hybrid method (Figure 1) is the third step.

MB.1. Build the predictive MB models

a. A set of *n* reservoir simulation models are built using petrophysical, geological, and geophysical data, while considering the restrictions of the facilities and the operating conditions
b. Such *n* models undergo the process of data assimilation using the production history

MB.2. Run the production simulation of variables *V* in the history period and granularity Δt_{it}, using all the MB models

MB.3. Rank and select the *k* best MB models according to metric *M*

a. Compare each of the *n* MB simulation results with the history using metric *M*
b. Rank the *n* MB models from best to worst according to metric *M*
c. Select the top *k* MB models

MB.4. The *k* selected MB models can then be used in short-term production forecasting and reservoir management with period t_{st}, and granularity Δt_{st}

Data-Driven Baseline. The DD baseline used similar GRU-RNN architectures (Werneck et al. 2022) to the one used in the hybrid method. For the variables BHP, QO, and QG, we used a network named stateful BigGRU2_10, while for the QW, we used the stateful_BigGRU_Seq01_AR. The need for a different network for QW arises from the typical upward trend present in QW curves after breakthrough, which is very dissimilar to the other variables. The network stateful BigGRU2_10 is characterized by two stateful GRU layers with 1024 neurons each, followed by 10 dense layers with reducing number of neurons. In this configuration, the input variables were BHP, QO, QG, and QW, for each target output (target ∊ {BHP, QO, QG}). This DD technique also employed anomaly detection through z-score and data augmentation of 6 hours (4 points for each day) in its pipeline (Werneck et al. 2022).

The network *stateful_BigGRU_Seq01_AR* contains two stateful GRU layers with 1024 neurons each, followed by a time distributed layer. In this configuration, we performed an autoregression, i.e., the input variable was QW, and the target was also QW. This DD technique also employed data augmentation of 6 hours (4 points for each day) in its pipeline (Werneck et al. 2022), but no anomaly detection.

For both networks, the input and output sizes were 184 days and the same post-processing described in section **Specific Choices for the Hybrid Method** (Equation 1) was applied.

Result Evaluation Metrics

The short-term forecasts were compared with the reference model (used as ground truth), and the results were evaluated through time series forecast error and rank similarity. We used the NQDS and the Normalized Quadratic Deviation (NQD) as the metrics to measure the time series forecast error, as defined by Equations (2) and (3), respectively.

$$NQDS = (QD/AQD)(LD/|LD|) \tag{2}$$

$$NQD = (QD/AQD) \tag{3}$$

where QD, AQD, and LD are defined by Equations (4), (5), and (6), respectively.

$$QD = \sum_{i=1}^{n_i} (F_i - A_i)^2 \tag{4}$$

$$AQD = \sum_{i=1}^{n_i} (Tol\ A_i + C)^2 \tag{5}$$

$$LD = \sum_{i=1}^{n_i} (F_i - A_i) \tag{6}$$

in which F is the forecasted value, A is the actual value, n_i is the number of value pairs, Tol is an acceptable tolerance for the variable, and C is a constant employed to avoid division by zero (Mesquita et al. 2015).

The *Tol* and *C* constants should be carefully chosen considering the specificity of each variable. In our evaluations, we chose the values for each variable as defined in Table 1.

Table 1—Values of *Tol* and *C* used in the forecast evaluations.

Variable	*Tol*	*C*
BHP	0.03	0.01
QO	0.1	0.01
QG	0.1	0.01
QW	0.1	40

Both the HM and the MB baseline select simulation models based on a rank, as explained in the earlier sections. Additionally, the simulation models can be evaluated against the reference model to generate a ground-truth rank of the best models for the short-term forecast. The ranks generated by each methodology can then be compared with the ground-truth rank to evaluate how well each methodology selects the available models from the initial set of assimilated models. Therefore, evaluating the results using these rank similarity metrics is complementary to the time series forecast error.

We used the Rank-Biased Overlap (RBO) (Webber et al. 2010) metric to evaluate the rank similarity, as defined by Equation (7).

$$r_{RBO} = (1-p)\sum_{d=1}^{k} (p^{d-1}|I_d|/d); \quad r_{RBO} \in [0, 1] \tag{7}$$

where k is the number of items in the ranked lists, p is a tunable parameter proportional to the contribution of the top d ranks to the similarity measure, and I_d is the intersection of the two ranked lists up to depth d. For all our r_{RBO} calculations, we used $p = 0.9$.

Experiments

Our experiments considered two different forecasting conditions, one Human Interference-Free with Restrictions (HIF-R) and another With Human Interferences (WHI). We also tested the selection of different top k simulation models, with k # $\{1, 5, 10, 20\}$ for each condition.

When dealing with production forecasts in a real-case scenario, the production is subject to a series of restrictions and condition changes. These restrictions come from physical capabilities (e.g., the maximum liquid rate for a well) or operational boundaries (e.g., minimum BHP allowed in a producer), while condition changes often come in the form of interventions. Such interventions can be either scheduled (e.g., periodic closure of wells for testing, de-scaling, etc.) or unscheduled (e.g., replacement of downhole equipment due to failure) procedures. Furthermore, the scheduled interventions can be performed according to well-defined periodic rules (e.g., WAG cycle length) or condition-based (e.g., well closure due to excessive water production). Some of those different classes of restrictions and interventions can be hard to predict, leading to inaccurate forecasting in real-case scenarios. However, tackling a real case production forecast with an incipient technique may be challenging and can lead to wrong conclusions about the potential of such a technique. Therefore, the use of a series of idealized forecasting scenarios can aid the evaluation and coverage of different production forecast schemes.

With the aforementioned motivation in mind, we define the following scenarios of production forecasting with increasing levels of realistic representations:

Human Interference-Free with No Restrictions (HIF-NR). The wells bind production to one condition (e.g., minimum BHP) and no other production system restrictions can be in place. There are no interventions of any kind.

Human Interference-Free with Restrictions (HIF-R). The production system has to operate bounded by all its production restrictions (e.g., well and platform limits) and only well-defined interventions are permitted (e.g., fixed WAG cycles).

With Human Interferences (WHI). All restrictions and interventions (scheduled and unscheduled) are to be considered.

We ran experiments of production forecasting in HIF-R and WHI conditions. For evaluation purposes, we consider that all interventions in WHI are scheduled (i.e., known in advance) so these can be informed in the forecasting approaches. Additionally, we consider only partial and total well closures as the intervention in the WHI. These closures follow the same distribution as the ones present in the production history period.

As previously mentioned, we considered 4 different values for top models to select (*k # {1, 5, 10, 20}*). Therefore, we have a total of 8 experiments.

Results and Discussions

The production forecast results are discussed in relation to the idealized HIF-R condition and the realistic WHI context. Each section is divided into discussions about the best selected model (top 1) and multiple selected models (top 5, 10, and 20). The results and discussions pertain only to short-term production forecasting.

Human Interference-Free with Restrictions

The HIF-R conditions keep production restrictions and well-defined periodic interventions but do not consider any discrete intervention. Therefore, the analyses under this idealized condition aim to evaluate the potential of the techniques in forecasting the general production behavior but not measure their ability to generalize responses to interventions.

Best Selected Model (Top 1). The objective of this evaluation is to measure how the best simulation models selected by the hybrid method and MB baseline compare with the DD baseline's performance in terms of production forecast. The 6-month forecast average NQD against the reference model is shown in Figure 3.

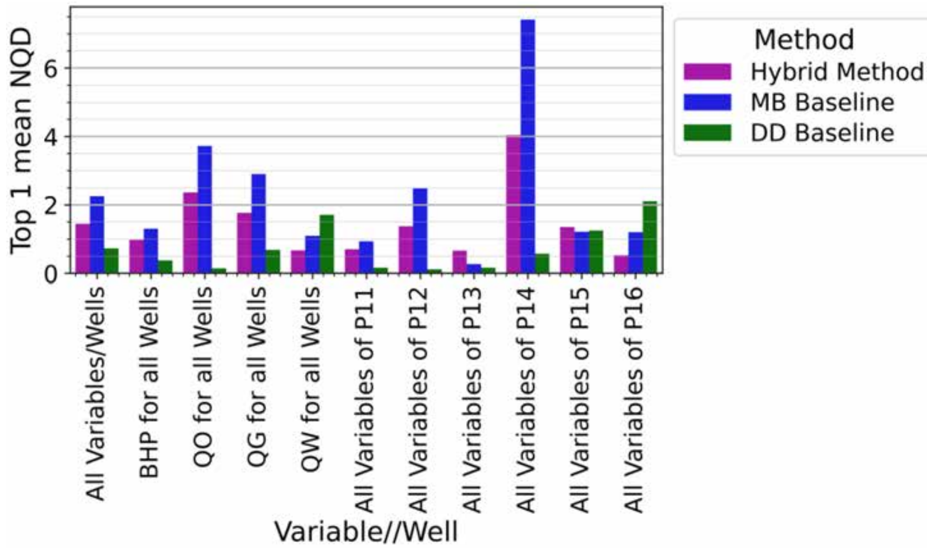

Figure 3—Average NQD versus the reference model obtained by the hybrid method and MB/DD baselines for the top 1 selection in the task of 6-month production forecast under HIF-R conditions. The closer to zero, the better the result. We present the global average results (all variables/wells) as well as the results grouped by each variable (BHP, QO, QG, QW) and each production well (P11, P12, P13, P14, P15, P16).

Overall, for this task, the DD baseline outperformed both the HM and the MB baseline. However, the forecast of variable QW and well P16 were challenging for the DD baseline, being outperformed by the hybrid and fully MB approaches.

This benchmark represents a high-production field in its early production life and, due to platform limitations, its production is restricted. This characteristic lead to many wells having long periods of constant production and, when an event occurs (e.g., water breakthrough), it is hard to predict its behavior due to the lack of historical data. Specifically, the variable QW and well P16 are the time series with the most variation close to the end of the production history period, so they are more challenging to predict than variables/wells that hold almost constant levels. Such a situation is exemplified in Figure 4, with the variables in well P11 being almost constant (albeit noisy) over time, while P16 has some variation. The P16 variations are associated with the support it receives from nearby injectors, therefore well P16 produces large volumes of water.

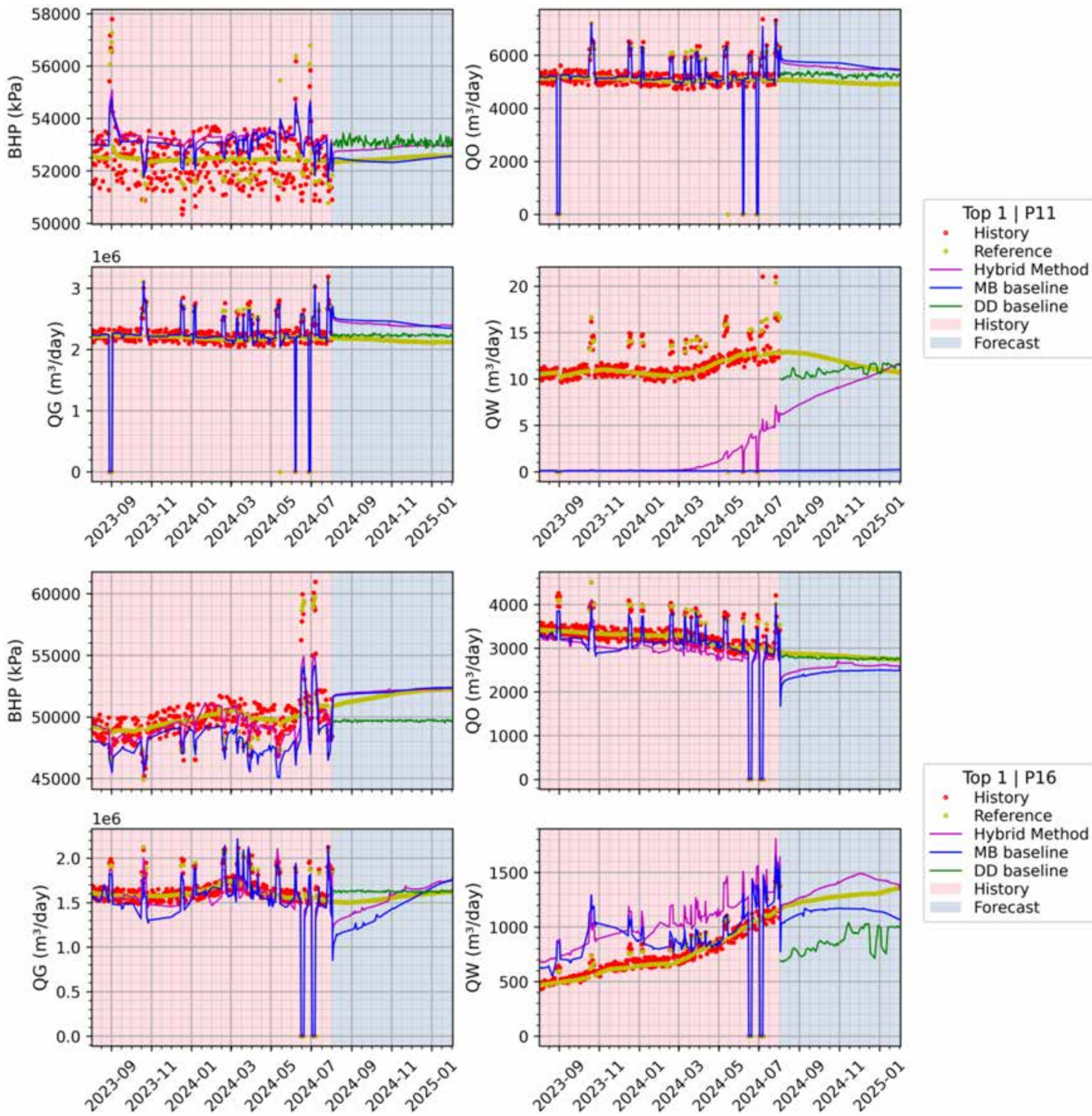

Figure 4—Time series of all the variables for wells P11 and P16 in HIF-R conditions, containing the last 12 months of production history (red background) and 6 months forecast (blue background): history (red), reference model (yellow), best hybrid model (purple), best MB baseline (blue), and DD baseline (green).

In Figure 4, one can assert the effectiveness of the purely DD method in the production forecast of variables when they remain stable in the forecast period. It is worth highlighting that the DD baseline was trained on a noisy production history containing several interventions and it was still able to capture the overall trend of production variables. However, the DD baseline was not particularly effective in forecasting the production in wells with considerable water production (i.e., P16), as the variables had long-time oscillations along with the aforementioned noise and interventions. In such cases, the HM and fully MB were more effective.

Careful observation of Figure 4 reveals that no technique was able to achieve a perfect forecast. We will discuss the issues in the forecasts and address the future improvements in a dedicated section at the end of the paper.

Another relevant aspect to be highlighted in Figure 3 is that the model selected by the HM outperformed the one selected by the fully MB approach (MB baseline) in almost all cases. Such results are promising for the hybrid approach and a more in-depth analysis of these two approaches is presented in the multiple selected models' discussion.

Multiple Selected Models (Top 5, 10, and 20). The objective of this evaluation is to measure how good both model selection methods (HM and MB baseline) are in the task of ranking the simulation models available in the set of assimilated models. The rank similarity scores for the top 5, 10, and 20 are presented in Figure 5, where top k refer to the best k MB models selected in step 3 of each methodology (see Figure 1 and Figure 2).

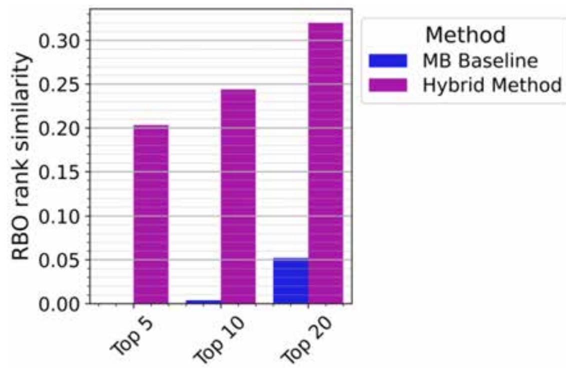

Figure 5—Rank similarity measure through RBO against the ground-truth rank (reference model), for a 6-month production forecast under HIF-R conditions. The higher the value, the better the result. Hybrid method (purple) and MB baseline (blue) selection methods considering top 5, 10, and 20 models.

Figure 5 shows that the HM consistently ranked the models better than the MB baseline, regardless of the top *k* selections. This indicates that the DD-assisted selection of simulation models (hybrid) is superior to their selection based on the history (MB baseline) for the short-term forecasting task. Further analyses can be drawn from the 6-month time series forecasting error, as measured by the NQDS against the reference model shown in Figure 6.

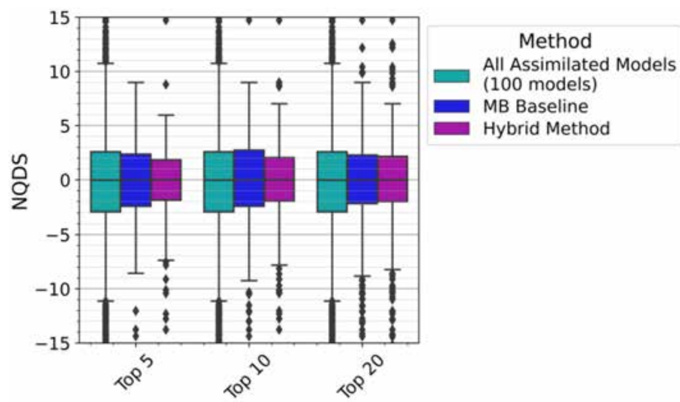

Figure 6—Average NQDS versus the reference model, considering all the variables (BHP, QO, QG, QW), for all the 100 assimilated simulation models (cyan), top models selected by the MB baseline method (blue), and top models selected by the HM (purple) in the task of a 6-month production forecast under idealized HIF-R conditions. Selections considering the top 5, 10, and 20 models. The closer to zero, the better the result.

The first characteristic to be aware of in Figure 6 is that, for all cases, the median of the boxplots is very close to zero. This means that the original and selected models are not biased towards over or underestimation in the forecasts. Such characteristics are a good indicator that the assimilation process achieved a well-represented set of models, even for short-term forecasts. Additionally, both selection methods kept this desirable characteristic for the top 5, 10, and 20 selected models.

Another characteristic is that the Interquartile Range (IQR) in the hybrid method is consistently lower than the one in the MB baseline for the top 5, 10, and 20. Therefore, we can say that the analysis of the NQDS confirms the observation made based on the rank similarities. That is, the data-driven–assisted selection of simulation models (hybrid) is superior to the history-based selection (MB baseline) for the short-term forecasting task under HIF-R conditions.

With Known Human Interferences

Production forecasting in these conditions aims to evaluate how the techniques deal with the real-life field operations, with partial and total well closures. Note that these well closures in the forecast were distributed following the same rules as those present in the production history.

It is worth highlighting that the DD technique used in this work cannot receive those interventions as input, nor is it able to predict them. Therefore, these conditions are expected to hinder the DD baseline forecasts.

Best Selected Model (Top 1). The 6-month forecast average NQD against the reference model for the hybrid model and MB/DD baselines are shown in Figure 7.

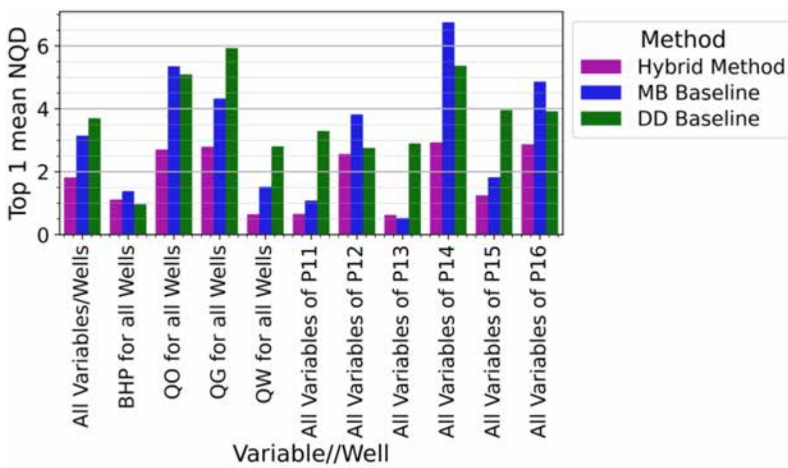

Figure 7—Average NQD versus the reference model obtained by the hybrid method and MB/DD baselines for the top 1 selection in the task of a 6-month production forecast under WHI conditions. The closer to zero, the better the result. We present the global average results (all variables/wells) as well as the results grouped by each variable (BHP, QO, QG, QW) and each production well (P11, P12, P13, P14, P15, P16).

Unlike what is observed for HIF-R conditions (Figure 3), in WHI conditions (Figure 7), the HM is by far the best, beating both baselines in overall behavior (see the results labeled as "All Variables/Wells") and in most other groups. Additionally, the DD baseline had the worst forecast in most cases given its inability to capitalize on it knowing the interventions. Even so, the DD baseline was able to outperform the MB baseline in some situations (BHP, QO, P12, P14, and P16) and it slightly outperformed the hybrid for BHP.

Overall, all the techniques had worsened metrics from HIF-R to WHI, which is expected, as this condition is more challenging to predict. However, by far, the DD baseline had the highest performance loss among the tested methods due to its inability to use known intervention information, as observed in Figure 8.

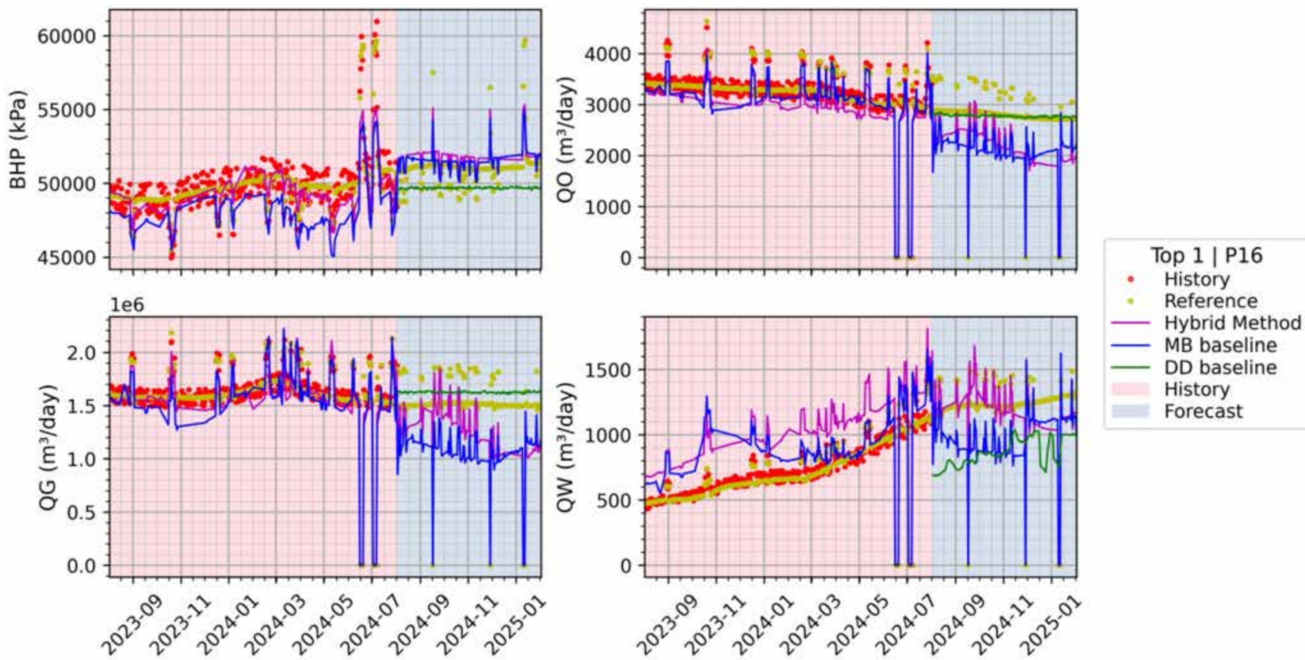

Figure 8—Time series of all the variables for well P16 in WHI conditions, containing the last 12 months of production history (red background) and a 6 months forecast (blue background): history (red), reference model (yellow), best hybrid model (purple), best MB baseline (blue), and DD baseline (green).

Multiple Selected Models (Top 5, 10, and 20). The rank similarity scores for the top 5, 10, and 20 under WHI conditions are presented in Figure 9.

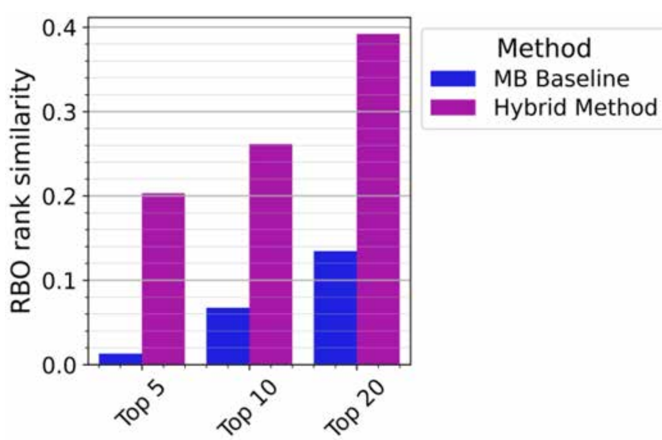

Figure 9—Rank similarity measure through RBO against the ground-truth rank (reference model), for a 6-month production forecast under WHI conditions. The higher the value, the better the result. Hybrid method (purple) and MB baseline (blue) selection methods considering the top 5, 10, and 20 models.

Following the same behavior under HIF-R conditions (Figure 5), the HM outperformed the MB baseline for ranking the top 5, 10, and 20 models under WHI conditions. The superiority of the HM over the MB baseline is also observed in the 6-month time series forecasting error, as measured by the NQDS against the reference model shown in Figure 10.

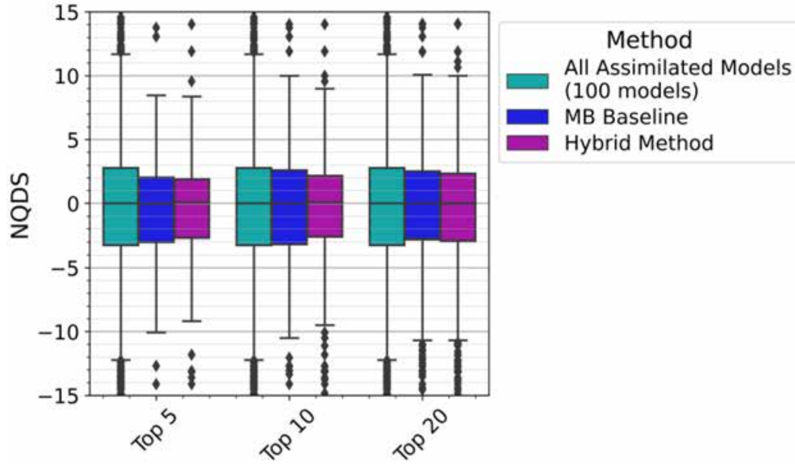

Figure 10—Average NQDS versus the reference model, considering all the variables (BHP, QO, QG, QW), for all the 100 assimilated simulation models (cyan), top models selected by the MB baseline method (blue), and top models selected by the HM (purple) in the task of a 6-month production forecast under realistic WHI conditions. Selections considering the top 5, 10, and 20 models. The closer to zero, the better the result.

The analysis of Figure 10 is similar to that under HIF-R conditions (Figure 6), with both MB baseline and HMs improving the IQR over all the 100 assimilated models, and the HM is superior to the MB baseline. However, a noticeable difference between HIF-R and WHI cases is that the performance gain of the HM over the MB baseline, while still existent, is less evident in the WHI.

Such results further solidify the evidence that the HM can select good models for short-term production forecasting. Moreover, the data-driven-aided selection of models (HM) is superior to the selection based on production history performance (MB baseline).

Issues and Future Improvements

This successful case study revealed that the proposed HM offers benefits over isolated MB and DD approaches. However, we can identify potential improvements in the fully MB and DD approaches, which will likely ensue better results for those techniques. Therefore, we discuss these potential improvements to tackle the current limitations of the methods and we outline some of our next steps in this research line. One important aspect to highlight of the proposed HM is that it is built on top of the fully MB and DD techniques, so any improvements in either of those techniques will benefit the HM.

Discontinuity in the History-Forecast Transition of the MB Results. The results based on models (hybrid and MB baseline) have a discontinuity in the transition from production history to forecast in most production curves. Such behavior can be seen in the time series of Figure 4 and is a typical characteristic of such models due to the change in boundary conditions in this transition and the inability to honor the correct well productivities. This characteristic severely hinders the ability of the model-based solutions to perform short-term forecasts. However, the longer the forecast period, the better such approaches tend to perform. The QO of P11 is a good example of such characteristic, as the discontinuity in the production history-forecast transition changes to an almost asymptotic behavior towards the reference model as time approaches the 6 months of forecast (see blue and purple curves of Figure 4).

There are more complex data assimilation techniques proposed in the recent literature to improve the production history-forecast transition of reservoir simulation models (Almeida et al. 2018; Formentin et al. 2019). Those techniques rely on the coupling of additional objective functions in the data assimilation process and have achieved good results in the past-future transition performance.

Performance of the DD Methods for Enlarged Forecast Horizons. Unlike the MB solutions, the DD forecasts tended to worsen the larger the output layer was. Such behavior is expected, as forecasting becomes harder with the more points one has to predict and the farther it gets from a known value (Chantry et al. 2021).

We can exemplify this challenge by comparing the DD technique used in the hybrid method (GRU2_10), containing an output layer of 30 days, with a version of the same GRU2_10 but retrained for an enlarged output layer of 184 days, as shown in Figure 11.

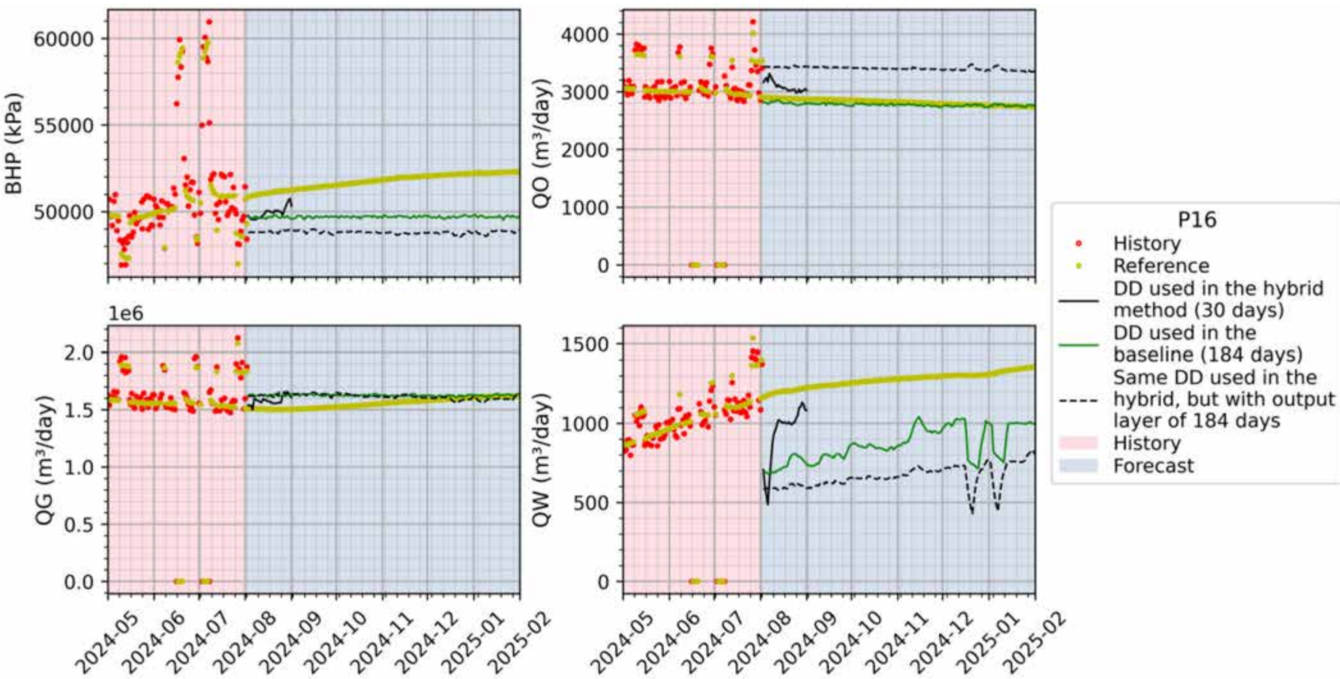

Figure 11—Time series of all the variables for well P16 in HIF-R conditions, containing the last 3 months of production history (red background) and 6 months of forecast (blue background): production history (red), reference model (yellow), DD used in the hybrid method (output layer of 30 days, solid black line), DD baseline (output layer of 184 days, green), and the same DD technique used in the hybrid method but retrained with an output layer increased to 184 days (dashed black line).

Figure 11 also contains a plot of the DD baseline. By comparing the DD baseline (solid green line) with the enlarged output *GRU2_10* (dashed black line), we can see that the baseline achieves a better 184 days forecast performance. This comparison demonstrates that the DD techniques should be tailored to the application. That is, we chose different techniques for the DD part of the HM and the DD baseline to maximize the task-specific performance.

It is also important to highlight that such techniques were validated in the final period of the production history to choose the ones for each task, i.e., no choice was made using the reference model.

Improvements for the DD Solution. Along with the previously mentioned improvements to the MB performance, we can identify the current challenges for the DD techniques explored in this paper as being twofold.

Firstly, the DD technique employed here cannot deal with information about future interventions, e.g., well closures, changes in restrictions, or production limitations. Such information should help its performance in realistic forecasting scenarios (e.g., WHI), if it is so desired. This kind of information could be exploited by employing an architecture which receives the planned interventions as inputs or by developing a workflow to predict any intervention.

Secondly, another critical aspect of the DD forecasts presented in this paper is that the first points predicted after the production history period often do not hold similar values to the history if the history data is trending upwards or downwards. It is intuitive for the first days of forecast to be close to the known history, within the expected noise and excluding interventions, therefore, a large discrepancy between history and the first forecasted points is undesirable. Therefore, this aspect of the DD forecasts should be improved in future implementations, which can be achieved by changing the training/validation scheme or exploring different NN architectures.

Conclusions

We proposed a new hybrid methodology (HM) for improved short-term production forecasting, combining standard numerical reservoir simulation (MB) and recurrent neural networks (DD). The HM is based on the DD-assisted selection of a subset of simulation models from a pool of assimilated models. The method was tested in a Brazilian pre-salt benchmark reservoir model case (UNISIM-IV-24) under two different forecasting conditions and compared with MB and DD baselines.

For all the tested cases, the HM outperformed the MB baseline. This indicates that the DD-assisted selection of models leads to superior model picking than its selection based on the production history performance (MB baseline). The purely DD forecast outperformed the hybrid method in the partially idealized HIF-R condition, while the hybrid outperformed this DD baseline in the more realistic WHI forecast. In addition, the poor DD performance in the WHI case was due to its inability to leverage known interventions in the forecast period. Some of the poor performance of the solutions based on models (MB and hybrid) can be attributed to the models not having the productivity indices adjusted, resulting in a discontinuity in the history-forecast transition. It is also worth noting that the HM is built on top of the fully MB and DD techniques and, thus, should benefit from any future improvements in the forecasting capabilities of the DD or MB solutions.

Since the proposed HM (i.e., DD-assisted selection of models) proved successful in the benchmark case study, we deem it promising for application in other fields and for further development.

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 Pre-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. We also would like to thank Frances Abbots for the valuable discussions during the development of this work.

References

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

Aziz, K., Settari, A. 1979. *Petroleum reservoir simulation*. Springer.

Borisova, J., Aladina, A., Nikitin, N. O. 2021. Hybrid modelling of environmental processes using composite models. *Procedia Computer Science*. **193**: 256–265. https://doi.org/10.1016/j.procs.2021.10.026

Botechia, V., Correia, M., Rios, V., Davolio, A., Santos, S., Hohendorff, J.Filho, Schiozer, D. 2022. UNISIM-IV Benchmark Proposal for Light Oil Carbonate Reservoir with High CO_2 Content. *Brazilian Journal of Petroleum and Gas*, **16** (4): 149–160. https://doi.org/10.5419/bjpg2022-0012

Cai, Q., Luo, X., Wang, P., Gao, C., Zhao, P. 2022. Hybrid model-driven and data-driven control method based on machine learning algorithm in energy hub and application. *Applied Energy*. **305**: 2022. https://doi.org/10.1016/j.apenergy.2021.117913

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

Chantry, M., Christensen, H., Dueben, P., Palmer, T. 2021. Opportunities and challenges for machine learning in weather and climate modelling: hard, medium and soft AI. *Philosophical Transactions of the Royal Society A*, 379. https://doi.org/10.1098/rsta.2020.0083

Correia, M., Botechia, V., Pires, L., Rios, V., Santos, S., Rios, V., Hohendorff, J., Chaves, M., Schiozer, D. J. 2020. UNISIM-III: Benchmark case proposal based on a fractured karst reservoir. Presented at the European Conference on the Mathematics of Oil Recovery, September. https://doi.org/10.3997/2214-4609.202035018

Fanchi, J. R. 2002. *Shared Earth Modeling*. Chapter 1: Introduction to shared earth modeling. Gulf Professional Publishing. https://doi.org/10.1016/B978-0-7506-7522-2.X5000-9

Formentin, H. N., Almeida, F. R., Avansi, G. D., Maschio, C., Schiozer, D. J., Caiado, C., Vernon, I., Goldstein, M. 2019. Gaining more understanding about reservoir behavior through assimilation of breakthrough time and productivity deviation in the history matching process. *Journal of Petroleum Science and Engineering*, **173**: 1080–1096. https://doi.org/10.1016/j.petrol.2018.10.045

Kansao, R., Yrigoyen, A., Haris, Z., Saputelli, L. 2017. Waterflood performance diagnosis and optimization using data-driven predictive analytical techniques from capacitance resistance models crm. Presented at the SPE Europec featured at 79th EAGE Conference and Exhibition, Paris, France, June. https://doi.org/10.2118/185813-MS

Krasnopolsky, V. M., Fox-Rabinovitz, M. S. 2006. Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. *Neural Networks*, **19**: 122–134. https://doi.org/10.1016/j.neunet.2006.01.002

Liao, L., Kottig, F. 2016. A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. *Applied Soft Computing*, **44**: 191–199. https://doi.org/10.1016/j.asoc.2016.03.013

Ligero, E. L., Maschio, C., Schiozer, D. J. 2003. Quantifying the impact of grid size, upscaling, and streamline simulation in the risk analysis applied to petroleum field development. Presented at the SPE Reservoir Simulation Symposium, Houston, Texas, February. https://doi.org/10.2118/79677-MS

Makinde, I. 2017. Statistical, data-driven approach to forecasting production from liquid-rich shale reservoirs. *Open Access Library Journal*, 4. https://doi.org/10.4236/oalib.1104053

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

Mesquita, F. B., Davolio, A., Schiozer, D. J. 2015. A systematic approach to uncertainty reduction with a probabilistic and multi-objective history matching. Presented at the EUROPEC 2015, Madrid, Spain, June. https://doi.org/10.2118/174359-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

Netto, M., Krishnan, V., Mili, L., Susuki, Y., Zhang, Y. 2019. A hybrid framework combining model-based and data-driven methods for hierarchical decentralized robust dynamic state estimation. Presented at the IEEE Power & Energy Society General Meeting (PESGM). https://doi.org/10.1109/PESGM40551.2019.8973772

Nguyen, A. P., Kim, J. S., Lake, L. W., Edgar, T. F., Haynes, B. 2011. Integrated capacitance resistive model for reservoir characterization in primary and secondary recovery. Presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA, October. https://doi.org/10.2118/147344-MS

Nikitin, N. O., Revin, I., Hvatov, A., Vychuzhanin, P., Kalyuzhnaya, A. V. 2022. Hybrid and automated machine learning approaches for oil fields development: The case study of volve field, north sea. *Computers & Geosciences*, 161. https://doi.org/10.1016/j.cageo.2022.105061

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

Silva, L. M. D., Avansi, G. D., Schiozer, D. J. 2020. Development of proxy models for petroleum reservoir simulation: a systematic literature review and state-of-the-art. *International Journal of Advanced Engineering Research and Science*, 7. https://dx.doi.org/10.22161/ijaers.710.5

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

Temirchev, P., Gubanova, A., Kostoev, R., Gryzlov, A., Voloskov, D., Koroteev, D., Simonov, M., Akhmetov, A., Margarit, A., Ershov, A. 2019. Reduced order reservoir simulation with neural-network based hybrid model. Presented at the SPE Russian Petroleum Technology Conference, Moscow, Russia, October. https://doi.org/10.2118/196864-MS

Thakur, G. 1996. What is reservoir management? *Journal of Petroleum Technology*, 48. https://doi.org/10.2118/26289-JPT

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

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

Xiong, H., Kim, C., Fu, J. 2020. A data-driven approach to forecasting production with applications to multiple shale plays. Presented at the SPE Improved Oil Recovery Conference, August. https://doi.org/10.2118/200365-MS

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