

Sensitivity of Reservoir Property Predictions to Loss Function Geometry: A Robustness and Bias Analysis

Bahram Lavi

State University of Campinas

João Roberto Bertini Junior

bertini@unicamp.br

State University of Campinas

Luis Oliveira Pires

State University of Campinas

Denis José Schiozer

State University of Campinas

Article

Keywords: Production Forecasting, Asymmetric Loss Functions, Offshore Reservoirs, Data-Driven

Posted Date: August 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10503152/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Sensitivity of Reservoir Property Predictions to Loss Function Geometry: A Robustness and Bias Analysis

Bahram Lavi^{1*}, João Roberto Bertini Junior^{2*},
Luis Oliveira Pires¹ and Denis José Schiozer^{1,3}

¹Centro de Estudos de Energia e Petróleo, Universidade Estadual de Campinas, Campinas, 13083-970, São Paulo, Brazil.

²Faculdade de Tecnologia, Universidade Estadual de Campinas, Limeira, 13484-332, São Paulo, Brazil.

³Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, Campinas, 13083-860, São Paulo, Brazil.

*Corresponding author(s). E-mail(s): blavi@unicamp.br;
bertini@unicamp.br;

Contributing authors: pires@cepetro.unicamp.br; denis@unicamp.br;

Abstract

Accurate production forecasting in offshore reservoirs is essential for short-term reservoir management and production planning. However, forecasting remains challenging because production responses are affected by complex transient behavior, interwell interference, and operational interventions. In particular, shutting in a producing well alters the reservoir pressure field and redistributes fluid flow, potentially increasing production rates in neighboring active wells. Under these conditions, conventional data-driven forecasting models may exhibit reduced accuracy and substantially underestimate production rates. Despite the importance of these transient responses, previous studies have largely focused on model architecture and input selection, while the influence of the training loss function and the unequal consequences of underprediction and overprediction have received limited attention. To address this gap, this study systematically compares thirteen symmetric and asymmetric loss functions for short-term production forecasting across four recurrent and convolutional deep learning architectures. The evaluation uses a complex offshore reservoir benchmark that incorporates various combinations of production and injection data. For oil production rates, asymmetric loss functions reduce forecasting errors by

approximately 50% compared with conventional losses. For the more challenging water-rate forecasting problem, the best asymmetric formulation reduces error by approximately 5% relative to the corresponding baseline. Asymmetric losses also reduce the occurrence and magnitude of underprediction during positive production transients while maintaining stable forecasts under rapidly changing operating conditions. These findings demonstrate that accounting for directional forecast errors through the training objective can improve short-term production forecasting under interwell interactions and operational interventions.

Keywords: Production Forecasting, Asymmetric Loss Functions, Offshore Reservoirs, Data-Driven

1. Introduction

Robust data-driven (DD) models for short-term production forecasting are essential for supporting rapid operational optimization and strategic reservoir management in the petroleum industry. However, the predictive reliability of these models is often compromised by the inherent variability of multiphase production data (Pires et al. 2026). In offshore production environments, time-series signals are characterized by complex nonlinear behavior and rapid transients arising from the combined effects of reservoir dynamics and surface operational constraints (Aranha et al. 2024).

Offshore production systems operate under dynamically varying well configurations. While relatively stable operating conditions may produce consistent production trends when control settings remain unchanged, the system is periodically perturbed by operational interventions, particularly well shut-ins. When a producing well is shut in, its production rate immediately drops to zero, while the resulting redistribution of reservoir pressure and fluid flow may induce production increases in neighboring active wells through interwell pressure interference and fluid diversion (Sturm et al. 2004; Peng et al. 2022). Accurately forecasting these intervention-induced responses is particularly challenging because models must represent both abrupt production changes and the underlying temporal evolution of production. The forecasting challenge is further compounded by the distinct dynamic behavior of the produced fluid phases. During waterflooding, reservoir heterogeneity, mobility changes, and preferential water breakthrough through high-permeability regions can substantially alter oil–water flow behavior and well productivity (Fu et al. 2024). These evolving multiphase-flow conditions can produce production responses with distinct transient characteristics, imposing different forecasting requirements for oil- and water-production rates.

Recent studies have addressed the effects of complex production dynamics through different modeling strategies. Kim et al. (2025b) employed instance-based variance scaling and ensemble screening to mitigate unphysical extrapolation during abrupt operational changes, reduce prediction uncertainty, and improve model reliability beyond the range represented in the training data. Zhang et al. (2025a) developed a DD approach for production forecasting during well interventions by directly incorporating

real-time operational parameters, such as flow-control valve settings, into the forecasting architecture. As multiphase geoenergy production systems become increasingly complex, however, DD models require further advances beyond conventional modeling configurations to adequately represent interconnected temporal and spatial scales and nonlinear physical heterogeneities (Kim et al. 2025a). Wang et al. (2025) addressed the prediction of wellhead choke oil flow rates under critical flow conditions by evaluating multiple DD methods. Although several surface operational parameters were considered during model training, the framework employed a leverage-based technique to identify potential outliers and assess model sensitivity to unusual observations. AlRasas et al. (2022), in turn, improved oil-production forecasting by addressing limitations associated with parameter tuning in conventional artificial neural networks through an integrated adaptive fuzzy system. Other studies have investigated the effects of operational interventions on petroleum production systems (Chahar et al. 2022; Yang et al. 2024; Hu et al. 2025; Lu et al. 2025; Pires et al. 2026).

Within this body of research, existing studies have primarily improved production forecasting through network architecture, input-variable selection, uncertainty quantification, parameter optimization, and outlier handling, while comparatively little attention has been given to the role of the training objective in representing intervention-induced production transients. This gap is particularly relevant for volatile and non-stationary production signals because widely used symmetric loss functions, such as Mean Squared Error (MSE), may present important limitations under these conditions. MSE is highly sensitive to large errors and abrupt observations and, due to its symmetric formulation, penalizes underprediction and overprediction equally despite their potentially different consequences during transient production responses (Olsen et al. 2022; Ghazvini et al. 2024; Huang and Ren 2024; Zhang et al. 2025b; Loutfi et al. 2022; Wang et al. 2022). Consequently, the suitability of asymmetric loss formulations for improving transient forecasting and reducing underprediction during positive production responses remains insufficiently investigated in short-term reservoir production forecasting.

Motivated by the findings of Jadon et al. (2024), who evaluated 14 regression loss functions for general time-series forecasting and showed that no single formulation is universally optimal, this study investigates loss-function selection for short-term petroleum production forecasting. Given the distinct transient characteristics of multiphase production signals, thirteen symmetric and asymmetric loss formulations are systematically evaluated for forecasting oil (Q_o) and water (Q_w) production rates across four widely used DD architectures: Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Convolutional Neural Network (CNN), and Temporal Convolutional Network (TCN) (Hochreiter and Schmidhuber 1997; Schuster and Paliwal 1997; Lawrence et al. 1997; Lea et al. 2017). The objective is to determine how the choice of training loss affects forecasting accuracy and transient-response representation under fluid-flow transients and operational variability, with particular attention to the distinct response characteristics of oil and water production and to underprediction during intervention-induced positive transients.

Extending the forecasting framework established by Lavi et al. (2025), the present study advances the analysis along three dimensions: dataset complexity, diversity of

DD architectures, and systematic evaluation of training loss functions. The experimental framework considers multiple dataset configurations combining producer and injector data to represent complex interwell interactions. The evaluated loss functions comprise symmetric formulations (MSE, Mean Absolute Error (MAE), Log-Cosh, Fair, and Cauchy) and alternative formulations designed to address robustness, scale dependence, or directional error penalties (Weighted Mean Squared Error (WMSE), Weighted Squared Error (WSE), Asymmetric Linear Loss (AsL), Huber-like Asymmetric Loss (HSL), Poisson, Mean Absolute Percentage Error (MAPE), Mean Squared Percentage Error (MSPE), and Quantile Loss (QL)). Figure 1 presents the end-to-end experimental framework, including the four DD architectures, loss-function evaluation, and performance assessment used to identify suitable model-loss configurations for short-term reservoir production forecasting.

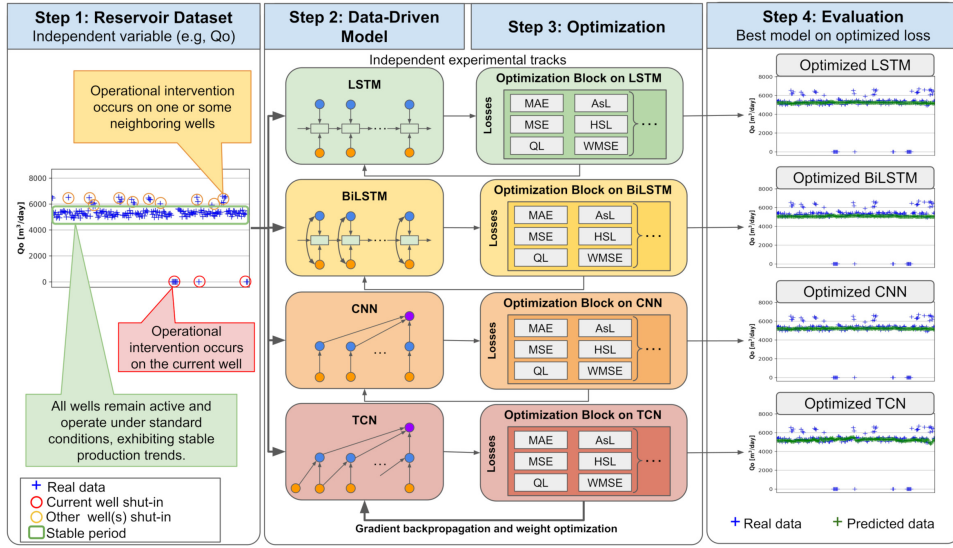


Fig. 1. Schematic workflow of the loss-function-driven reservoir production forecasting framework.

The main contributions of this study can be seen in fourfold:

- Providing a rigorous comparative analysis of symmetric and asymmetric loss functions, demonstrating that selecting an appropriate loss function is crucial for accounting for the distinct transient behavior and operational risk profiles associated with specific production rates in short-term forecasting tasks.
- Quantifying the ability of DD models to track the main production trends while identifying the specific challenges associated with capturing positive production transients induced by neighboring-well shut-ins.

- Demonstrating how asymmetric loss functions can reduce high-magnitude residuals and improve model performance under the nonlinear production transients and operational variability characteristic of offshore production environments.
- Demonstrating, through quantitative and statistical analyses, that asymmetric loss functions improve forecasting performance and reduce underprediction for oil and water production rates.

The remainder of this paper is structured as follows. Section 2 introduces the case study, providing a comprehensive description of the offshore reservoir environment and the operational characteristics of the production well ensemble. Section 3 outlines the methodological framework, focusing on the mathematical formulations of the regression loss functions and their capacity to handle operational uncertainty. Section 4 delineates the DD regression architectures implemented for the time-series forecasting task. Section 5 details the experimental setup and preprocessing protocols established in this work. Section 6 presents the results and discussion, incorporating empirical analyses of error metrics, statistical ranking frameworks, and regression assessments to evaluate model stability. Finally, Section 7 concludes the paper by highlighting key findings and outlining future research directions for robust production forecasting in complex reservoir environments.

2. Case Study: UNISIM-IV-2026

This paper focuses on the UNISIM-IV-2026 dataset (Botechia et al. 2023) as a case study to evaluate DD modeling approaches on short-term production rate forecasts. UNISIM-IV is an open-source benchmark case based on a light-oil carbonate Brazilian pre-salt field with a high oil-gas ratio and CO₂ content. In the production strategy, the database comprised six vertical producers (P11-P16) and seven vertical WAG-CO₂ injectors. The dataset encompasses 2891 days of authentic production history. Although production history data are available up to day 3123. The history data collection period began in September 2018. Figure 2 demonstrates a three-dimensional structural grid and well configuration of the UNISIM-IV-2026 benchmark reservoir model. The dataset for each producer well comprises daily historical records of the oil, water, and gas production rates (Q_o , Q_w , and Q_g , respectively). Analogously, to account for reservoir energy maintenance and inter-well connectivity, the dataset incorporates daily injection profiles from the neighboring injector wells, specifically capturing the water injection rate for water injectors and the gas injection rate for gas injectors.

The field production data exhibit three distinct behavioral patterns that influence the development of forecasting models. The first behavior is the occurrence of zero-value production rates, which are caused by a well being temporarily shut in. This phenomenon can result from various reasons, including planned maintenance, equipment failure, or adherence to safety protocols. A second, more typical behavior is observed during periods when all six wells are operating normally. In these cases, the time series displays a stable, predictable main signal trend, as the wells consistently contribute to the field’s production. The third and most complex behavior is

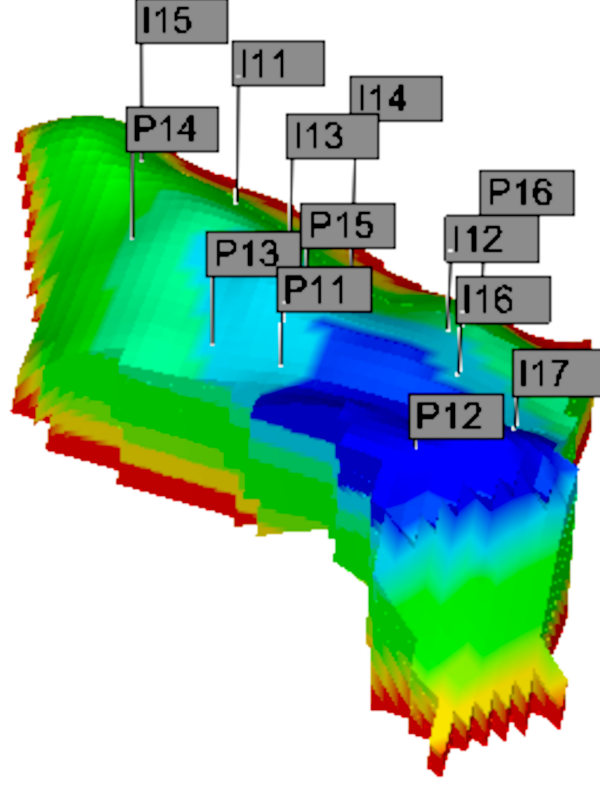


Fig. 2. Structural 3D visualization of the UNISIM-IV-2026 benchmark (Botechia et al. 2023).

the appearance of positive outliers, in which a well's production rate rises significantly above the main signal trend. The shut-in of another well often causes this.

This study focuses on developing a robust DD model to forecast the Q_o and Q_w . Figure 3 shows an overview of the Q_o and Q_w production history across six producer wells, highlighting the distinct behaviors of oil and water rates over time. The data spanning days 1 to 1200 were deemed unsuitable for model training, as they limit the effective extraction of discriminant information. This initial period exhibits significant production fluctuations, primarily due to the opening of multiple wells, which introduced inter-well variability and uncertainty. For instance, operational events such as well shut-ins generated both zero-production values and upward-trending patterns in neighboring wells. Well shut-in events can induce pressure redistribution in the reservoir, producing transient increases in the production rates of neighboring wells. If the production facility operates near its capacity constraints, these operational events may further amplify the effects on the production behavior of other wells. Consequently, the presence of zero-production values and positive outliers can introduce higher-error

distributions in fluid rate forecasting, potentially biasing the learning phase of DD models. Therefore, the analysis focuses on data from day 1200 onward to develop robust DD models. The training phase utilizes data from days 1201 to 2892, a period characterized by operational stability—consistent activity across all wells—and high discriminant value for capturing core temporal dynamics.

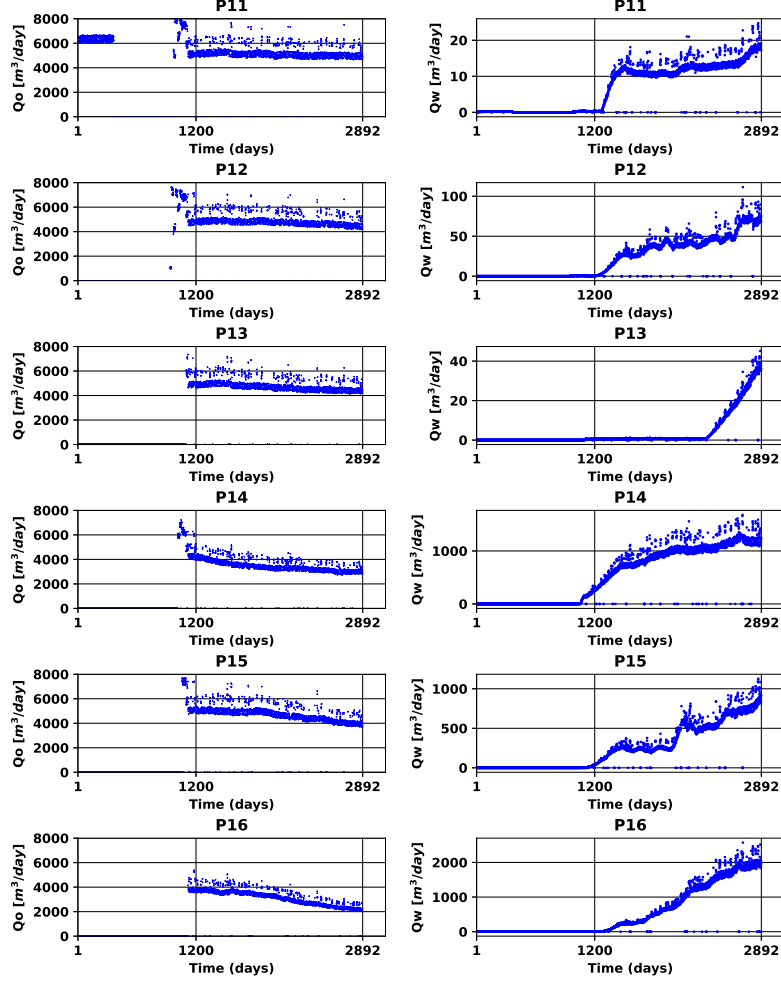


Fig. 3. History production time-series data for Q_o (left-hand side column) and Q_w (right-hand side column) across six producer wells from UNISIM-IV dataset.

Q_o patterns divide the wells into two distinct groups. Wells P11 through P14 maintain stable, predictable outputs with narrow distributions centered at a median of 5,100 m^3/day . In contrast, P15 and P16 exhibit significantly wider spreads, indicating greater operational fluctuations, with P16 showing the greatest variability within the

group. Qw behavior reflects varying degrees of well maturity. Wells P11, P12, and P13 maintain low rates (near zero), concentrated distributions typical of initial production phases. Conversely, P14, P15, and P16 exhibit substantially higher and more variable water rates; P14 shows a distribution that peaks at 1,500 m³/day, while P16 displays the broadest spread and the highest water cut in the group. Notably, all wells exhibit positive outliers, driven by production shifts resulting from the shut-in of neighboring wells.

3. Regression Loss Functions

The performance of a regression model in petroleum engineering applications is heavily influenced by the choice of loss function during optimization. While MSE remains the industry standard due to its computational efficiency and alignment with Gaussian noise assumptions, it often fails to provide reliable results when faced with the nonlinearities and high outlier density characteristic of field-acquired data.

The loss functions assessed in this study are categorized into two primary groups based on their mathematical geometry. First, this work examines symmetric losses, which treat over- and underestimation equally, including MSE, MAE, Fair Loss, Cauchy, and Log-Cosh. Second, an analysis of asymmetric losses is conducted to introduce directional bias and scale sensitivity, accounting for the petroleum field. This category encompasses WSE, WMSE, AsL, HSL, Quantile, Poisson, MAPE, and MSPE. This section details the formulation of these loss functions to help the DD model improve predictive performance under varying oilfield data conditions.

As summarized in Table 1, the loss functions selected for this study represent a spectrum of behaviors tailored to the complexities of petroleum datasets. The following describes the calculation of each loss function between the predicted value ($\hat{\mathbf{y}} \in \mathbb{R}^n$) and the actual value ($\mathbf{y} \in \mathbb{R}^n$) for n samples, tailored for specific predictive challenges in short-term forecasting tasks. For clarity of notation consider the error as $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$.

Table 1. Summary and characteristics of regression loss functions.

Category	Loss Function	Focus	Engineering Highlights
Symmetric	MSE	Baseline	Used for high-precision data and sensitive to large outliers.
	MAE	Baseline	Robust for noisy, general-field data.
	Log-Cosh	Robustness	Stable convergence for noisy and continuous streams.
	Fair	Robustness	Tunable transition point via parameter c for limiting outliers.
	Cauchy	Robustness	Maximum outlier rejection for extreme telemetry errors.
Asymmetric	WMSE/WSE	Risk	Directional weighting; prioritizes specific error directions.
	AsL	Risk	Asymmetric linear error for risk-weighted optimization.
	HSL	Risk	To prioritize either under- or over-estimation.
	QL	Risk	Fundamental for P10/P50/P90 uncertainty.
	Poisson	Risk	Ideal for discrete-event frequencies.
	MAPE	Risk	Focuses on relative well performance.
	MSPE	Risk	Focuses on Squared relative well performance.

3.1. MSE

MSE is one of the widely used loss functions in regression analysis. It is a convex, differentiable function that provides a unique minimum and facilitates gradient-based optimization. It is defined as the average of the squared differences between the predicted values and the actual values in all data points and is expressed as follows.

$$\text{MSE} = \frac{1}{n} \|\mathbf{e}\|_2^2 \quad (1)$$

3.2. MAE

MAE is particularly suitable for petroleum analysis because it provides a linear, transparent measure of prediction accuracy. By calculating the average of the absolute differences between predicted and actual values, MAE assigns a constant penalty to every unit of error regardless of its magnitude. The MAE is calculated as

$$\text{MAE} = \frac{1}{n} \|\mathbf{e}\|_1 \quad (2)$$

3.3. Log-Cosh

The Log-Cosh loss is particularly advantageous for petroleum data modeling, as it provides a smooth, twice-differentiable alternative to the MAE, combining the stability of squared error for small residuals with the robustness of absolute error for large outliers. This hybrid characteristic ensures that the model remains sensitive to flow fluctuations near the target value while preventing extreme sensor noise and spiky telemetry data from disproportionately biasing the optimization process. The Log-Cosh loss is twice differentiable with respect to the data sample, facilitating more stable gradient descent and smoother convergence in complex deep learning architectures. The Log-Cosh loss is calculated as:

$$\text{Log-Cosh} = \frac{1}{n} \mathbf{1}^\top \log(\cosh(\mathbf{e})) \quad (3)$$

where $\cosh(\cdot)$ and $\log(\cdot)$ are applied element-wise, and $\mathbf{1}$ is a vector of ones.

3.4. Fair

The Fair loss is an advanced, robust loss designed to mitigate the impact of outliers while maintaining numerical stability. It is particularly effective for petroleum datasets where positive outliers can lead to large, non-Gaussian residuals. Unlike the MSE, which grows quadratically, the Fair Loss transitions from a quadratic behavior for small errors to a linear behavior for large errors, governed by a tuning constant c . The Fair Loss is symmetric and twice-differentiable, providing a smoother optimization path than the MAE. It is calculated as:

$$\text{Fair} = \frac{1}{n} \mathbf{1}^\top \left[c^2 \left(\frac{|\mathbf{e}|}{c} - \log \left(1 + \frac{|\mathbf{e}|}{c} \right) \right) \right] \quad (4)$$

3.5. Cauchy

The Cauchy loss is a robust loss derived from the Cauchy distribution, specifically engineered to handle datasets with extreme outlier contamination. In petroleum engineering applications, such as high-frequency monitoring, data is often plagued by positive spikes that do not mainly follow a stable distribution. While the MSE would be overwhelmed by these anomalies, the Cauchy Loss utilizes a logarithmic penalty that significantly suppresses the influence of large residuals. By incorporating a scale parameter γ , the function allows precise calibration near the target while ensuring that errors beyond a threshold have diminishing impact on the model's weight updates. The Cauchy Loss is calculated as:

$$\text{Cauchy} = \frac{1}{n} \mathbf{1}^\top \left[\frac{\gamma^2}{2} \log \left(1 + \left(\frac{\mathbf{e}}{\gamma} \right)^2 \right) \right] \quad (5)$$

3.6. WSE

WSE is a quadratic loss function that applies asymmetric weighting to the squared errors. Unlike standard MSE, which penalizes all errors equally, this loss function multiplies the squared error by a specific weight based on whether the prediction is an over- or under-estimation. The primary goal is to assign a higher penalty to one type of error. The loss is defined as:

$$\text{WSE} = \mathbf{e}^\top W \mathbf{e} \quad (6)$$

The weight parameter $W = \text{diag}(w_1, \dots, w_n)$ is set to values $w_i = \{0.1, 0.3, 0.5\}$ for $i = 1, \dots, n$, to examine the model's behavior under varying degrees of asymmetric penalty.

3.7. WMSE

WMSE is a simple extension of MSE where a constant weight (W) is applied to each squared error term. This function is typically used when certain data points are considered more important than others, and a higher weight is assigned to the error associated with them. It can be used to emphasize the importance of specific regions of the data or to address imbalances within a dataset. The loss is computed as:

$$\text{WMSE} = \frac{1}{n} \mathbf{e}^\top W \mathbf{e} \quad (7)$$

the weights in W are incorporated analogously to the WSE formulation.

3.8. AsL

AsL is a piecewise loss function that differentially penalizes over- and under-predictions. It is a variant of Mean Absolute Error (MAE) in which the error term is scaled by different factors depending on its sign. Given a weighting factor $\alpha \in [0, 1]$,

and $\mathbf{e}_+ = \max(\mathbf{e}, 0)$ the loss is defined as:

$$\text{AsL} = \frac{1}{n} \mathbf{1}^\top [\alpha \mathbf{e}_+ + (1 - \alpha) (-\mathbf{e})_+] \quad (8)$$

The parameter α is selected from $\alpha = \{0.1, 0.3, 0.5\}$.

3.9. HSL

HSL is a robust, hybrid loss function that combines the benefits of squared and linear penalties. For small errors, it behaves quadratically like MSE and provides a smooth gradient. For large errors, it switches to a linear penalty, making it less sensitive to outliers. It defined as

$$\text{HSL} = \frac{1}{n} \mathbf{1}^\top \left[\frac{1}{2} \mathbf{e}^2 \odot \mathbf{I}_{|\mathbf{e}| \leq \delta} + \left(\delta w \mathbf{e} - \frac{1}{2} \delta^2 \right) \odot \mathbf{I}_{\mathbf{e} > \delta} + \left(\delta (w - 1) \mathbf{e} - \frac{1}{2} \delta^2 \right) \odot \mathbf{I}_{\mathbf{e} < -\delta} \right] \quad (9)$$

The parameters for this function include both the weight w and a threshold δ , $\odot$ is the Hadamard (elementwise) product and $\mathbf{I}$ is the indicator vector. w is evaluated through the set $w = \{0.1, 0.3, 0.5\}$ to control the asymmetry, and $\delta = 1.0$ is set to different positive values to adjust the transition point from quadratic to linear loss.

3.10. QL

QL is a specialized loss function for quantile regression. Instead of predicting the conditional mean, it allows a model to predict the conditional τ -th quantile, where $\tau \in [0, 1]$. The loss function is asymmetric, penalizing errors differently based on whether the prediction is an overestimation or underestimation relative to the specified quantile. It is expressed as:

$$\text{QL} = \frac{1}{n} \mathbf{1}^\top [\tau \mathbf{e}_+ + (1 - \tau) (-\mathbf{e})_+] \quad (10)$$

The value of τ controls the level of asymmetry, and it serves to control the asymmetry and is selected from $\tau = \{0.1, 0.3, 0.5\}$.

3.11. Poisson

The Poisson loss is a statistical function derived from the Poisson distribution, primarily used to model non-negative, skewed data in which the variance is expected to scale with the mean. In petroleum engineering, this characteristic is frequently observed in production flow rates and event-based data, where higher production magnitudes are associated with greater absolute fluctuations. The Poisson Loss optimizes the likelihood of the observed data under the assumption that the target variable represents a rate. This makes it inherently asymmetric, as it imposes a stricter penalty on under-predictions and prevents the model from predicting negative values, which are physically impossible in reservoir contexts. The Poisson Loss is calculated as:

$$\text{Poisson} = \frac{1}{n} \mathbf{1}^\top [\hat{\mathbf{y}} - \mathbf{y} \odot \log(\hat{\mathbf{y}})], \quad \hat{\mathbf{y}} > 0 \quad (11)$$

3.12. MAPE

MAPE is a relative loss function that measures prediction accuracy as a percentage of the actual values. In petroleum engineering, datasets often involve variables spanning several orders of magnitude—such as flow rates from high-productivity horizontal wells versus marginal stripper wells. While absolute losses like MSE or MAE are dominated by the high-volume wells, MAPE normalizes the error, ensuring that a 5% deviation is penalized equally regardless of the well’s absolute scale. This relative sensitivity makes it an essential tool for multi-well analysis where maintaining consistent fractional accuracy is more important than minimizing total volumetric error. However, due to its denominator-dependent structure, MAPE is inherently asymmetric, penalizing under-predictions more heavily than over-predictions of the same absolute magnitude. For n sample data and $\odot$ the elementwise division, the MAPE is calculated as:

$$\text{MAPE}_{\%} = \frac{100}{n} \mathbf{1}^\top |\mathbf{e} \oslash \mathbf{y}| \quad (12)$$

3.13. MSPE

The MSPE is the average of the squared percentage differences between the predicted and observed values. Unlike MSE, which measures absolute squared error, MSPE evaluates the relative error. This makes it robust when dealing with heterogeneous data, where different wells have vastly different flow rates. For n sample data, the MSPE is calculated as:

$$\text{MSPE}_{\%} = \frac{100}{n} \mathbf{1}^\top [(\mathbf{e} \oslash \mathbf{y})^2] \quad (13)$$

4. DD Regression Methods

This section provides an overview of the DD architectures employed for time-series tasks, focusing on their structural designs and core parameters. These models are specifically engineered to capture the complex temporal and spatial dependencies inherent in production data. To optimize performance, the hyperparameters for each architecture are tuned via an exhaustive grid search. The specific hyperparameters and their corresponding search spaces are detailed in Table 2. A detailed description of each DD model is provided below.

4.1. LSTM

LSTM networks ([Hochreiter and Schmidhuber 1997](#)) represent a specialized class of recurrent neural networks engineered for robust sequence modeling and long-term trend forecasting. Standard recurrent architectures often suffer from vanishing or exploding gradients when processing extended sequences, which limits their capacity to

Table 2. Hyperparameter settings employed in the DD models.

Model	Hyperparameter	Value
LSTM	number of units	{8, 16, 64, 128}
BiLSTM	LSTM number of units	{8, 16, 64, 128}
CNN	Conv1: filter size	{3, 5, 7, 9}
	Conv1: kernel size	{2, 4, 6}
	Conv2: filter size	{3, 5, 7}
	Conv2: kernel size	{2, 4, 6}
	Drop out rate	{0.25, 0.50}
TCN	Dense: number of units	{3, 5, 7}
	Filter size	{5, 10, 15, 20}
	Kernel size	{3, 5, 7}
	Dilation	{1, 2, 4, 8}
Common DD	Batch size	{128, 256}
	Epoch	{50, 100, 150}
Hyperparameters	Optimizer	ADAM

retain long-range dependencies. LSTM networks mitigate this limitation by introducing a specialized cell state that acts as a continuous pathway for information, regulated by three distinct gating mechanisms: a forget gate that discards irrelevant historical data, an input gate that selects new valuable information to store, and an output gate that determines the final hidden state passed to the next time step. By seamlessly integrating these long- and short-term temporal pathways, the architecture effectively regulates gradient flow across deep unrolled operations, making it highly proficient at capturing the non-linear dynamics, multi-timescale variations, and prolonged lag phases inherent in complex production time-series data.

4.2. BiLSTM

BiLSTM networks (Schuster and Paliwal 1997) extend the capabilities of standard sequential models by deploying two separate, independent hidden layers to process input sequences simultaneously in both forward and backward temporal directions. This dual-directional architecture enables the network to capture comprehensive context from both historical (past) and look-ahead (future) states within a given sequence window, providing a symmetrical understanding of temporal dynamics that unidirectional frameworks cannot achieve. The forward and backward hidden states are typically concatenated at each time step to construct a comprehensive feature representation, with the model’s overall capacity and expressive power directly governed by the total number of allocated BiLSTM units. This bidirectional processing is particularly advantageous for deciphering complex boundary conditions, underlying trends, and transient pressure-rate dependencies across complete production history profiles.

4.3. CNN

CNNs (Lawrence et al. 1997) leverage specialized weight-sharing convolutional operations to exceptionally identify local hierarchical patterns and extract translation-invariant features from sequential datasets. When adapted for time-series analysis, one-dimensional convolutional layers slide across the temporal axis to automatically capture localized feature representations without requiring manual feature engineering. This extraction process is fundamentally governed by two critical hyperparameters: the number of filters, which dictates the depth and variety of the generated feature maps, and the kernel size, which establishes the temporal receptive field or window size for localized feature detection. Following these convolutional layers, pooling operations are applied sequentially to downsample the spatiotemporal dimensions, thereby reducing computational complexity and enforcing spatial robustness by retaining only the most salient features across the production profile.

4.4. TCN

TCN (Lea et al. 2017) is a pair of architectural principles that define the model’s structure, employing a fully convolutional network with all hidden layers. To effectively capture long-range dependencies inherent in complex time series, the TCN leverages dilated convolutions within its causal framework. Dilation introduces a fixed, non-contiguous skip between filter taps, thereby exponentially increasing the network’s receptive field with depth without increasing the number of layers or the number of kernel parameters. The architecture is constructed using Residual Blocks, each typically comprising a sequence of two dilated causal convolutional layers, followed by a non-linearity, weight normalization, and dropout. The inclusion of the residual connection, which adds the block’s input to its output, is critical for mitigating exploding gradients and stabilizing the training of deep temporal models. Where necessary, a 1×1 convolution is applied to the skip connection to ensure dimensional consistency with the block’s output. The empirical performance of the TCN is highly dependent on the selection of its hyperparameters. The optimal configuration, determined through a systematic grid search, included filter and kernel sizes and increasing dilation rates $\{d_1, d_2, \dots, d_L\}$ across the L stacked layers. In this study, two CNNs are considered as stacked layers.

5. Experimental Design and Setup

All DD models were developed using Python 3.7, with primary dependencies on TensorFlow 2.13. For the more computationally intensive deep learning models, training was performed on a server equipped with an NVIDIA A100 GPU and 80 GB of video memory. The GPU drivers and CUDA toolkit versions used for this experiment were NVIDIA-SMI 535.104.12 and CUDA Version 12.2, respectively.

5.1. Dataset Construction Scenario

To develop robust DD models for short-term production forecasting in the pre-salt petroleum field, this study constructs three distinct datasets. These datasets are

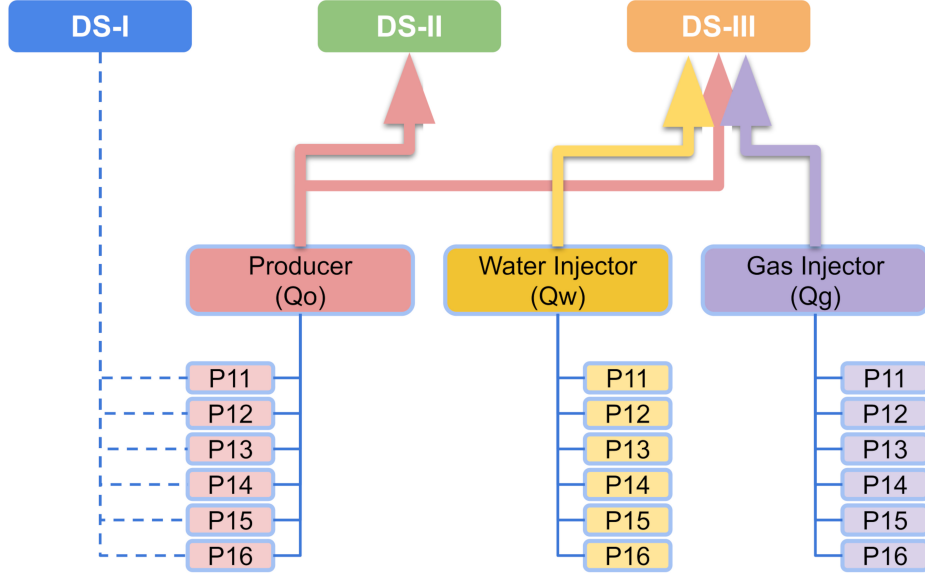


Fig. 4. The diagram illustrates different approaches to constructing datasets (DS-I, DS-II, and DS-III) from six wells (P11–P16). The dashed lines for DS-I indicate that each well’s data is treated individually as a single-attribute predictor, rather than as a combined multi-well system.

designed to capture the increasing complexity of data and will be used to predict the production rate of individual producer wells (P11–P16). Figure 4 illustrates the three scenarios used to construct the training datasets for wells P11–P16.

5.1.1. DS-I: Univariate Prediction

DS-I utilizes a univariate forecasting strategy, leveraging historical production rates from a target well to predict its future performance. For each well producer (P11–P16), the input feature is the production rate from the immediately preceding time step, while the targets are the rates for the same well at specified future time steps.

5.1.2. DS-II: Field-Wide Data-Stream

DS-II expands the feature space to incorporate the production history of all active producer wells in the reservoir. For each target well, the input vector includes the synchronized production rates of all producer wells (P11–P16). This configuration helps the model comprehend inter-well dependency and learning patterns across the field.

5.1.3. DS-III: Multimodal Production and Injection

DS-III provides a comprehensive reservoir dataset by integrating production rates with field-wide injection data. The input feature vector for each target well combines the history rates of all producers with the water and gas injection rates from all injector wells. By capturing the pressure support and displacement dynamics between injectors and producers, DS-III enables the model to characterize complex fluid-flow interaction effects. This dataset aggregates history production rates from all available producer wells with the water-and-gas injectors' history well data (water injector - water rate, and gas injector - gas rate) in the system.

5.2. Data Lag

The analysis of multivariate time-series data is a cornerstone of effective reservoir management and accurate production forecasting. Within this context, selecting an appropriate lookback window (a.k.a. lag size) is a critical optimization step; this hyperparameter dictates the depth of temporal context provided to the machine learning architectures. Determining the optimal lag size is essential for capturing complex subsurface reservoir dynamics. While a larger window can capture prolonged lag phases, a more compact lag offers distinct advantages, including a sharper focus on short-term, highly relevant temporal dependencies. Ultimately, to identify the most robust configuration for each architecture, the optimal lag size is determined empirically via an exhaustive grid search evaluated over the discrete set $\{2, 4, 6, 8\}$.

5.3. Model Prediction Horizon

Each DD architecture is configured to evaluate a specific forecasting horizon. The primary objective of this study is to develop robust DD frameworks that accurately forecast two distinct fluid production streams: Q_o and Q_w . To evaluate the predictive stability and temporal adaptability of the proposed models under different operational conditions, the architectures are tested across two distinct prediction window sizes: 7-day and 30-day ahead, representing short-term operational monitoring horizons. Furthermore, to prevent artificial bias from non-representative production anomalies, such as zero-value entries typically caused by sudden operational events, isolated zero-value sequences in the target variables are imputed using the localized mean of the active production periods immediately preceding and following the disruption, thereby maintaining the structural and temporal continuity of the dataset.

5.4. Dataset Normalization and Split

To optimize algorithmic convergence and ensure stable training dynamics, feature normalization is applied as a critical preprocessing step. In this study, the dataset attributes are scaled using Min-Max normalization, mapping all feature values to the range $[0, 1]$. To preserve the underlying temporal dependencies of the time series and prevent data leakage, the data is partitioned chronologically into three mutually exclusive subsets: 80% of the historical sequence is allocated to the training set; the

subsequent 10%, immediately following the training block, is designated as the validation set for hyperparameter tuning; and the final 10% is reserved strictly as a holdout test set for evaluating the definitive forecasting performance.

6. Results and Discussions

This section provides a detailed evaluation of the proposed predictive framework across diverse reservoir scenarios and fluid phases. The assessment comprises three primary analytical layers designed to rigorously validate the data-driven models and their corresponding loss functions. Initially, Section 6.1 describes the physical interpretation associated with production data in the case study. Section 6.2 presents an empirical analysis that establishes a quantitative performance baseline using scale-aware and relative metrics. The results are evaluated using two primary metrics to quantify forecasting accuracy on the test set and report two metrics: the MAE to provide a direct measurement of the average magnitude of the residuals in the same units as the production rate, and the MAPE to measure the relative error metric to express the forecasting deviation as a percentage of the actual production. This is followed by Section 6.3, which presents the statistical tests used to determine whether the performance differences between symmetric and asymmetric loss functions are statistically significant. Finally, Section 6.4 presents a regression analysis of the best-performing model, providing a granular examination of the time-series forecasts to assess how the model navigates signal trends and high-magnitude transients.

6.1. Physical Interpretation of Production Data

Before evaluating the predictive performance of different loss functions, it is important to characterize the production signals. As illustrated in Figure 3, the Q_o and Q_w production time series exhibit distinct behaviors that reflect the reservoir’s underlying multiphase flow dynamics. Oil production signals generally follow relatively smooth depletion-driven trends, with moderate variability around a stable main production level. Occasional positive spikes are observed, associated with operational events such as temporary shut-ins of neighboring wells and the resulting redistribution of reservoir pressure. These events introduce isolated positive outliers while preserving the overall production trend. Water production signals, however, display different characteristics. Moreover, several wells exhibit abrupt, sustained increases in water rate, consistent with breakthrough events and mobility-driven channeling through high-permeability pathways. These phenomena produce pronounced nonlinear variations in production data and may lead to regime-shifting behavior.

These differences suggest that the statistical structure of the residuals associated with oil and water production may differ significantly. Oil production signals exhibit stable trends, occasionally perturbed by positive outliers, indicating that loss functions that are robust to outliers and highly stable are necessary to maintain a reliable baseline without being disproportionately skewed by isolated transient points. In contrast, water production signals exhibit abrupt transitions associated with breakthrough dynamics, suggesting that loss formulations that incorporate data risk profiles are required to better navigate high variance and minimize the impact of large residuals

Table 3. Comparative Qo forecast performance metrics showing the mean and standard deviation ($\pm$) of MAE and MAPE error on the test set (final forecast set) across three distinct petroleum datasets. The best-performing loss is highlighted in bold.

Loss	DS-I		DS-II		DS-III	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
AsL(0.1)	1086.46 $\pm$ 202.19	0.31 $\pm$ 0.14	1095.67 $\pm$ 272.90	0.32 $\pm$ 0.17	1115.78 $\pm$ 255.56	0.32 $\pm$ 0.17
AsL(0.3)	616.90 $\pm$ 230.68	0.19 $\pm$ 0.12	605.99 $\pm$ 293.61	0.19 $\pm$ 0.15	616.70 $\pm$ 312.64	0.19 $\pm$ 0.16
AsL(0.5)	465.24 $\pm$ 195.14	0.14 $\pm$ 0.10	435.82 $\pm$ 223.21	0.14 $\pm$ 0.11	454.95 $\pm$ 237.64	0.14 $\pm$ 0.12
Cauchy(0.1)	448.04 $\pm$ 195.91	0.14 $\pm$ 0.10	424.06 $\pm$ 223.30	0.13 $\pm$ 0.11	440.30 $\pm$ 242.23	0.14 $\pm$ 0.12
Cauchy(0.3)	475.88 $\pm$ 177.85	0.14 $\pm$ 0.10	446.37 $\pm$ 203.58	0.14 $\pm$ 0.11	475.20 $\pm$ 226.89	0.15 $\pm$ 0.12
Cauchy(0.5)	496.69 $\pm$ 175.25	0.15 $\pm$ 0.10	467.58 $\pm$ 198.20	0.14 $\pm$ 0.11	480.16 $\pm$ 208.77	0.14 $\pm$ 0.11
Fair(0.1)	483.44 $\pm$ 186.84	0.15 $\pm$ 0.10	449.53 $\pm$ 214.18	0.14 $\pm$ 0.11	470.50 $\pm$ 226.17	0.14 $\pm$ 0.12
Fair(0.3)	493.16 $\pm$ 179.42	0.15 $\pm$ 0.10	465.26 $\pm$ 204.09	0.14 $\pm$ 0.11	484.81 $\pm$ 226.70	0.15 $\pm$ 0.12
Fair(0.5)	501.87 $\pm$ 176.71	0.15 $\pm$ 0.10	470.85 $\pm$ 204.46	0.14 $\pm$ 0.11	489.50 $\pm$ 213.87	0.15 $\pm$ 0.12
HSL(0.1)	914.06 $\pm$ 201.23	0.27 $\pm$ 0.13	903.26 $\pm$ 260.01	0.27 $\pm$ 0.16	914.58 $\pm$ 264.30	0.27 $\pm$ 0.16
HSL(0.3)	652.82 $\pm$ 190.45	0.19 $\pm$ 0.11	628.54 $\pm$ 230.13	0.19 $\pm$ 0.13	640.73 $\pm$ 238.12	0.19 $\pm$ 0.14
HSL(0.5)	525.93 $\pm$ 175.29	0.16 $\pm$ 0.10	495.71 $\pm$ 206.07	0.15 $\pm$ 0.11	506.66 $\pm$ 200.34	0.15 $\pm$ 0.11
Log-Cosh	521.08 $\pm$ 171.23	0.16 $\pm$ 0.10	491.55 $\pm$ 202.73	0.15 $\pm$ 0.11	496.91 $\pm$ 202.55	0.15 $\pm$ 0.11
MAE	460.08 $\pm$ 191.28	0.14 $\pm$ 0.10	433.98 $\pm$ 226.69	0.13 $\pm$ 0.12	450.26 $\pm$ 240.97	0.14 $\pm$ 0.12
MAPE	344.00 $\pm$ 155.47	0.10 $\pm$ 0.07	332.82 $\pm$ 145.90	0.10 $\pm$ 0.07	348.16 $\pm$ 176.86	0.10 $\pm$ 0.08
MSE	525.51 $\pm$ 177.82	0.16 $\pm$ 0.10	497.15 $\pm$ 206.03	0.15 $\pm$ 0.11	504.69 $\pm$ 200.80	0.15 $\pm$ 0.11
MSPE	311.52 $\pm$ 125.10	0.09 $\pm$ 0.06	308.38 $\pm$ 121.16	0.09 $\pm$ 0.06	328.48 $\pm$ 144.69	0.09 $\pm$ 0.07
Poisson	638.86 $\pm$ 555.48	0.19 $\pm$ 0.16	685.13 $\pm$ 593.83	0.20 $\pm$ 0.17	644.69 $\pm$ 587.82	0.19 $\pm$ 0.21
QL(0.1)	253.32 $\pm$ 58.13	0.07 $\pm$ 0.03	253.20 $\pm$ 58.96	0.07 $\pm$ 0.03	252.66 $\pm$ 55.31	0.06 $\pm$ 0.03
QL(0.3)	349.24 $\pm$ 128.76	0.10 $\pm$ 0.07	328.33 $\pm$ 129.92	0.10 $\pm$ 0.07	332.27 $\pm$ 137.49	0.10 $\pm$ 0.08
QL(0.5)	462.92 $\pm$ 191.00	0.14 $\pm$ 0.10	433.72 $\pm$ 221.42	0.13 $\pm$ 0.11	448.82 $\pm$ 239.93	0.14 $\pm$ 0.12
WMSE(0.1)	522.47 $\pm$ 169.77	0.16 $\pm$ 0.10	499.17 $\pm$ 202.18	0.15 $\pm$ 0.11	509.16 $\pm$ 208.30	0.15 $\pm$ 0.11
WMSE(0.3)	528.26 $\pm$ 177.12	0.16 $\pm$ 0.10	495.88 $\pm$ 196.23	0.15 $\pm$ 0.11	504.14 $\pm$ 208.66	0.15 $\pm$ 0.11
WMSE(0.5)	528.72 $\pm$ 180.04	0.16 $\pm$ 0.10	491.97 $\pm$ 199.66	0.15 $\pm$ 0.11	509.42 $\pm$ 206.64	0.15 $\pm$ 0.11
WSE(0.1)	910.58 $\pm$ 197.93	0.26 $\pm$ 0.13	911.76 $\pm$ 269.87	0.27 $\pm$ 0.16	914.19 $\pm$ 262.87	0.27 $\pm$ 0.16
WSE(0.3)	650.64 $\pm$ 189.84	0.19 $\pm$ 0.11	623.84 $\pm$ 238.96	0.19 $\pm$ 0.13	643.00 $\pm$ 237.11	0.19 $\pm$ 0.13
WSE(0.5)	524.80 $\pm$ 170.11	0.16 $\pm$ 0.10	498.76 $\pm$ 200.71	0.15 $\pm$ 0.11	504.95 $\pm$ 203.03	0.15 $\pm$ 0.11

during sharp signal transitions. Therefore, the physical and statistical characteristics observed already provide qualitative insight into why different loss functions may be better suited to modeling oil and water production signals.

6.2. Empirical Analyses

The evaluation of the forecasting framework across the six producer wells (P11–P16) highlights significant performance variances across the eleven loss functions tested. Table 3 presents the comparative performance of the Qo forecast, reporting the mean and standard deviation for both MAE and MAPE on the test set, averaged across all wells, prediction windows, hyperparameters, and DD models evaluated on three distinct dataset constructions (DS-I, DS-II, and DS-III).

For the Qo forecast, standard baselines show that while MSE provides a common starting point, the average MAE of 509.12 ± 11.99 differs from the average MAE of 448.11 ± 10.76 across those three datasets on the test set. While the MAE loss delivers a more robust baseline test average, the linear penalty of MAE is better suited to the transient-heavy nature of offshore production data, as it prevents over-influenced residuals from squaring during abrupt shut-in events. Most remarkably, the QL at $\tau = 0.1$, QL(0.1), yielded the lowest overall error with an average test MAE of 253.06 ± 0.29 across those datasets. Furthermore, when evaluating the models using

MAPE, the difference in predictive reliability becomes even more pronounced. The loss achieved an average test MAPE of 0.067 ± 0.03 , indicating that the relative error in the average was consistently lower than for the other losses. These results suggest that by focusing on the lower decile of the data distribution, the framework can establish a highly stable production rate that remains resilient to the volatile upward data points (positive outliers) and interference effects observed in the field benchmark.

Other robust asymmetric losses, tested with tuning parameters $\{0.1, 0.3, 0.5\}$, showed varying degrees of success in mitigating offshore operational noise. The Cauchy(0.1) loss achieved a low test MAE of 437.47 ± 9.99 , and the Fair(0.1) loss achieved a low test MAE of 467.82 ± 13.97 across three datasets. These results position the robust family as a superior alternative to the MSE baseline, although it remains less effective than MAE in this specific application. The superior performance of the Cauchy function, in particular, suggests that its aggressive logarithmic penalty is well-suited for isolating the sharp, nonlinear transients and well interference surges common in reservoir signals. As the symmetric tuning parameters for Cauchy, Fair, and HSL increased from 0.1 to 0.5, the performance metrics shifted, generally indicating that a tighter threshold (0.1) is necessary to effectively filter high-magnitude outliers inherent in the production signals. This shift indicates that a more restrictive threshold (0.1) is essential for effectively filtering the high-magnitude outliers inherent in production data. Although these robustness-focused losses achieve significant error reduction, they lack the specialized risk focus of the Asymmetric group and exhibit stable footprints in the overall comparative analysis.

To ensure a rigorous and unbiased comparison across the producer ensemble, it is necessary to account for the substantial variability in water production across individual wells. In this study, wells P11–P13 exhibit lower water rates, resulting in naturally lower absolute error magnitudes, whereas wells P14–P16 exhibit substantially higher production volumes, which inherently inflate the standard MAE. Because MAE is a scale-dependent metric, a simple arithmetic average across all wells would disproportionately reflect the errors of the high-rate producers, potentially obscuring the model’s actual predictive performance on the smaller wells. To address this bias and provide a fair evaluation of the loss functions, the Normalized MAE (NMAE) is used for the water rate analysis. By normalizing the MAE for each well to its corresponding mean production value, it transforms the error into a scale-invariant ratio, enabling an equitable assessment of model accuracy across the entire field.

Table 4 reports the results of the water production rate, Qw. The experimental results demonstrate a shift in the optimal loss function compared to oil rate forecasting. The DD models using the AsL loss function, with its tuning parameter set to 0.1, achieved the lowest error across all three datasets, with an average NMAE of 0.62 ± 0.86 and MAPE of 0.19 ± 0.15 on the test set, indicating that AsL(0.1) yielded the lowest test MAE and MAPE. This asymmetric approach is particularly effective for Qw, as it provides the necessary sensitivity to account for the unique statistical distributions of water breakthroughs and transients in offshore reservoirs. The standard MAE and MSE models exhibit significant variance during testing, particularly in the more complex DS-II and DS-III datasets. This suggests that symmetric loss functions

Table 4. Comparative Qw forecast performance metrics showing the mean and standard deviation ($\pm$) of NMAE and MAPE error on the test set (final forecast set) across three distinct petroleum datasets. The best-performing loss is highlighted in bold.

Loss	DS-I		DS-II		DS-III	
	NMAE	MAPE	NMAE	MAPE	NMAE	MAPE
AsL(0.1)	0.49 $\pm$ 0.74	0.15 $\pm$ 0.11	0.64 $\pm$ 0.88	0.20 $\pm$ 0.16	0.73 $\pm$ 0.97	0.24 $\pm$ 0.18
AsL(0.3)	0.54 $\pm$ 0.79	0.17 $\pm$ 0.12	0.70 $\pm$ 0.98	0.21 $\pm$ 0.15	0.77 $\pm$ 1.11	0.23 $\pm$ 0.18
AsL(0.5)	0.58 $\pm$ 0.85	0.19 $\pm$ 0.13	0.81 $\pm$ 1.20	0.23 $\pm$ 0.19	0.83 $\pm$ 1.23	0.24 $\pm$ 0.20
Cauchy(0.1)	0.60 $\pm$ 0.84	0.20 $\pm$ 0.13	0.81 $\pm$ 1.17	0.23 $\pm$ 0.19	0.90 $\pm$ 1.39	0.25 $\pm$ 0.22
Cauchy(0.3)	0.56 $\pm$ 0.79	0.18 $\pm$ 0.12	0.73 $\pm$ 1.02	0.22 $\pm$ 0.16	0.78 $\pm$ 1.12	0.23 $\pm$ 0.18
Cauchy(0.5)	0.57 $\pm$ 0.81	0.19 $\pm$ 0.13	0.74 $\pm$ 1.02	0.22 $\pm$ 0.16	0.74 $\pm$ 1.07	0.22 $\pm$ 0.17
Fair(0.1)	0.57 $\pm$ 0.80	0.19 $\pm$ 0.13	0.75 $\pm$ 1.08	0.22 $\pm$ 0.17	0.75 $\pm$ 1.05	0.23 $\pm$ 0.17
Fair(0.3)	0.58 $\pm$ 0.80	0.19 $\pm$ 0.13	0.76 $\pm$ 1.06	0.22 $\pm$ 0.17	0.79 $\pm$ 1.14	0.23 $\pm$ 0.18
Fair(0.5)	0.59 $\pm$ 0.82	0.19 $\pm$ 0.13	0.73 $\pm$ 1.01	0.22 $\pm$ 0.16	0.77 $\pm$ 1.11	0.23 $\pm$ 0.18
HSL(0.1)	0.55 $\pm$ 0.88	0.16 $\pm$ 0.13	0.73 $\pm$ 1.01	0.22 $\pm$ 0.16	0.80 $\pm$ 1.09	0.24 $\pm$ 0.18
HSL(0.3)	0.56 $\pm$ 0.79	0.18 $\pm$ 0.13	0.74 $\pm$ 1.05	0.22 $\pm$ 0.16	0.76 $\pm$ 1.07	0.23 $\pm$ 0.17
HSL(0.5)	0.61 $\pm$ 0.85	0.20 $\pm$ 0.13	0.75 $\pm$ 1.04	0.22 $\pm$ 0.16	0.79 $\pm$ 1.17	0.23 $\pm$ 0.18
Log-Cosh	0.58 $\pm$ 0.80	0.19 $\pm$ 0.13	0.74 $\pm$ 1.07	0.21 $\pm$ 0.17	0.77 $\pm$ 1.12	0.23 $\pm$ 0.17
MAE	0.57 $\pm$ 0.79	0.19 $\pm$ 0.12	0.81 $\pm$ 1.17	0.24 $\pm$ 0.19	0.83 $\pm$ 1.24	0.24 $\pm$ 0.20
MAPE	1.42 $\pm$ 2.12	0.44 $\pm$ 0.42	1.24 $\pm$ 1.74	0.36 $\pm$ 0.30	1.22 $\pm$ 1.79	0.34 $\pm$ 0.29
MSE	0.60 $\pm$ 0.84	0.20 $\pm$ 0.13	0.76 $\pm$ 1.10	0.22 $\pm$ 0.17	0.78 $\pm$ 1.18	0.22 $\pm$ 0.18
MSPE	1.56 $\pm$ 2.19	0.52 $\pm$ 0.43	1.30 $\pm$ 1.67	0.39 $\pm$ 0.31	1.33 $\pm$ 1.79	0.40 $\pm$ 0.36
Poisson	1.00 $\pm$ 2.04	0.35 $\pm$ 0.51	1.07 $\pm$ 1.74	0.30 $\pm$ 0.32	1.52 $\pm$ 2.51	0.45 $\pm$ 0.60
QL(0.1)	0.78 $\pm$ 0.91	0.28 $\pm$ 0.14	1.17 $\pm$ 1.55	0.35 $\pm$ 0.25	1.25 $\pm$ 1.72	0.37 $\pm$ 0.28
QL(0.3)	0.64 $\pm$ 0.83	0.22 $\pm$ 0.13	0.90 $\pm$ 1.26	0.27 $\pm$ 0.20	1.00 $\pm$ 1.49	0.28 $\pm$ 0.24
QL(0.5)	0.59 $\pm$ 0.83	0.19 $\pm$ 0.13	0.77 $\pm$ 1.12	0.22 $\pm$ 0.18	0.81 $\pm$ 1.23	0.23 $\pm$ 0.20
WMSE(0.1)	0.58 $\pm$ 0.81	0.19 $\pm$ 0.13	0.73 $\pm$ 1.01	0.21 $\pm$ 0.16	0.81 $\pm$ 1.15	0.24 $\pm$ 0.19
WMSE(0.3)	0.59 $\pm$ 0.81	0.19 $\pm$ 0.13	0.78 $\pm$ 1.09	0.22 $\pm$ 0.17	0.77 $\pm$ 1.09	0.23 $\pm$ 0.17
WMSE(0.5)	0.59 $\pm$ 0.83	0.19 $\pm$ 0.13	0.76 $\pm$ 1.07	0.22 $\pm$ 0.17	0.74 $\pm$ 1.07	0.22 $\pm$ 0.17
WSE(0.1)	0.55 $\pm$ 0.83	0.16 $\pm$ 0.12	0.72 $\pm$ 1.03	0.21 $\pm$ 0.17	0.77 $\pm$ 1.05	0.24 $\pm$ 0.18
WSE(0.3)	0.58 $\pm$ 0.86	0.18 $\pm$ 0.13	0.74 $\pm$ 1.07	0.21 $\pm$ 0.17	0.74 $\pm$ 1.08	0.22 $\pm$ 0.17
WSE(0.5)	0.59 $\pm$ 0.81	0.19 $\pm$ 0.13	0.75 $\pm$ 1.08	0.22 $\pm$ 0.17	0.77 $\pm$ 1.13	0.23 $\pm$ 0.19

struggle to handle the nonlinear transients and breakthrough volatility characteristic of offshore water production.

Additionally, the empirical evaluation of the various loss functions across the three dataset scenarios (DS-I, DS-II, and DS-III) is synthesized in the dual radar charts presented in Figure 5 for the Qo forecast. The radar visualization shows a clear performance footprint for each loss function, with the left-hand-side plot showing the validation (Val.) set and the right-hand-side plot showing the test set.

In both the validation and test sets, the QL(0.1) is the superior estimator, consistently occupying the innermost position on the radar plots, with a test mean of 253.32. The tight shaded band around the QL(0.1) line—reflecting a low standard deviation of 58.13—visually confirms its remarkable stability across the entire experimental setup. The MSE serves as a baseline, exhibiting a pronounced outward extension from the center and visually indicating a higher error magnitude than most specialized losses. In contrast, the MAE yields a more constrained and stable profile, indicating that its linear penalty is more naturally robust to outliers than the quadratic penalty of the MSE.

Other robust symmetric losses (Cauchy, Fair, Log-Cosh) fall into the middle tier but are more affected by outliers than standard baselines. It lacks the specialized directional focus required to fully mitigate the non-normal error distributions of this

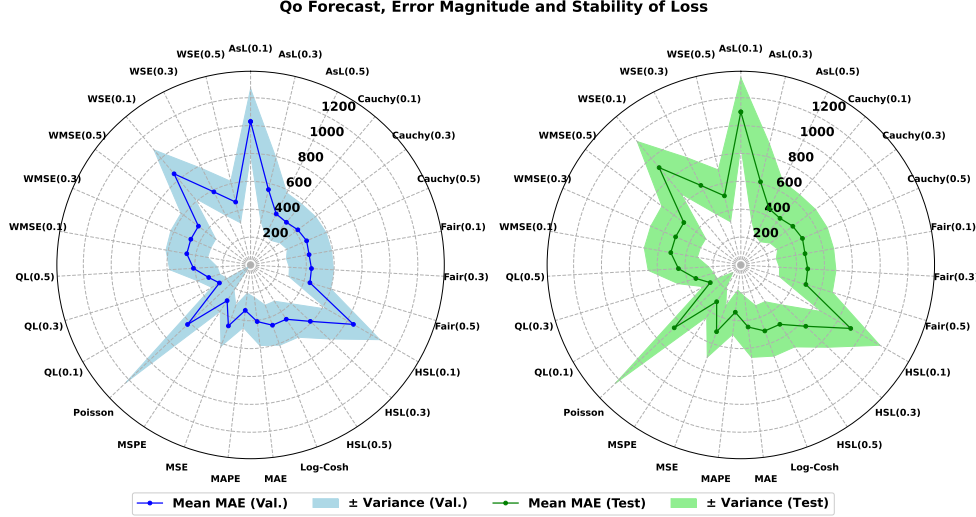


Fig. 5. Radar charts on Qo forecast illustrating the mean MAE and stability (variance) across three datasets for validation (left-hand-side plot) and test (right-hand-side plot) sets across various loss functions.

field. However, this tier is significantly more stable than asymmetric variants using quadratic penalties (WSE, AsL) and unstable losses such as Poisson, which exhibit erratic, high-variance footprints and large shaded variance regions. Overall, the most effective generalization is observed in the QL, and the MSPE remains tightly circular around the plot’s center. Specifically, QL(0.1) shows a robust footprint and a narrow shaded region of variance in both the validation and test sets. The results confirm that asymmetric and scale-invariant losses are not only more accurate but also significantly more stable than standard symmetric estimators when handling the volatile transients and heterogeneous scales inherent in offshore reservoir forecasting.

Figure 6 shows the radar charts for water rate, Qw, forecast. The plot illustrates the inherent complexity of the water rate variable, characterized by greater variance and sensitivity than oil rate predictions. While oil rate forecasts show highly compressed footprints for asymmetric functions, the Qw stability profiles exhibit more pronounced variance bands, reflecting the volatility of water breakthrough and multiphase flow transients. Traditional symmetric metrics such as MAE and MSE occupy a much larger share of the radar, indicating significant performance degradation and greater instability under the non-linear surges typical in water production.

6.3. Statistical Tests

To further analyze and gain insights into model performance, a statistical test is conducted to evaluate the algorithms under comparison. Following the approach outlined by Demšar (2006) for comparing multiple algorithms across datasets, each algorithm is ranked within each dataset using a Friedman test based on its MAE performance. The best-performing algorithm is assigned rank 1, the second-best rank 2, and so on.

Qw Forecast, Error Magnitude and Stability of Loss

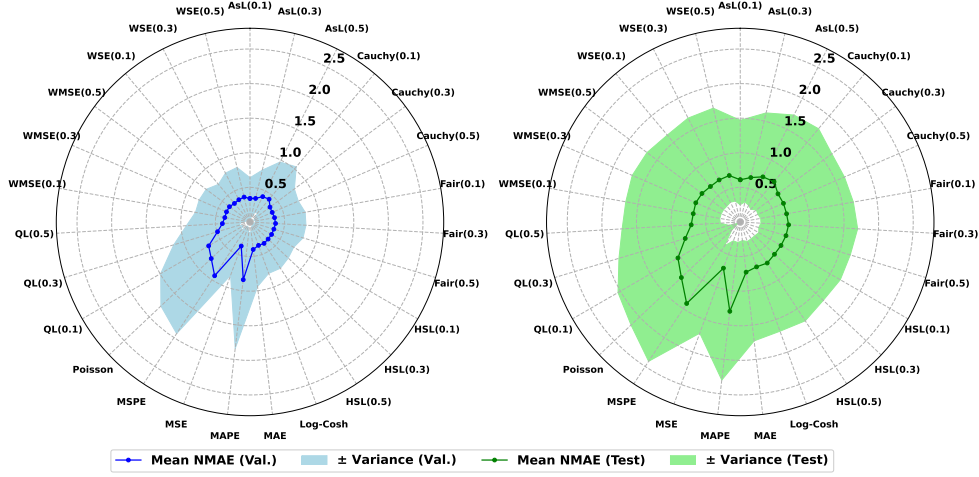


Fig. 6. Radar charts on Qw forecast illustrating the mean NMAE and stability (variance) across three datasets for validation (left-hand-side plot) and test (right-hand-side plot) sets across various loss functions.

In cases of tied performance, the algorithms receive the average rank. In this test, a Critical Difference (CD) is calculated for a given significance level. If the performance difference between any two algorithms exceeds the CD, the lower-ranked algorithm is considered to perform significantly better than the other. Otherwise, the statistical conclusion is that the experimental data are insufficient to draw any conclusions about the algorithms being compared.

The statistical test analysis of the evaluated loss functions is rigorously established through the CD diagram. Figure 7 summarizes the average rankings across three distinct datasets on the Qo forecast. The results identify QL(0.1) as the most effective loss function, achieving a premier rank of 1.00, followed closely by MSPE at 2.00. These top-tier performers belong to the asymmetric loss family, demonstrating that objectives designed to handle the non-linear nature of petroleum production data and ensure scale invariance are significantly more reliable than standard baselines. The superior rankings of QL(0.3) and MAPE further reinforce the conclusion that accounting for the specific risk profiles of offshore wells enhances forecasting accuracy.

A closer examination of the CD diagram reveals that while there are clear leaders, certain models share a horizontal red bar, indicating that the differences in their average rankings are not statistically significant. For instance, the top-performing group, including QL(0.1), MSPE, QL(0.3), and MAPE, exhibits overlapping performance, suggesting that they provide similar predictive ability. In contrast, the Symmetric robustness group—led by Cauchy(0.1) with a rank of 5.00—is statistically separated from the leading asymmetric group. On the other hand, the CD diagram shows significant performance degradation for certain parameterized models, such as HSL(0.1) and AsL(0.1), which rank last with scores of 25.67 and 27.00, respectively. These low

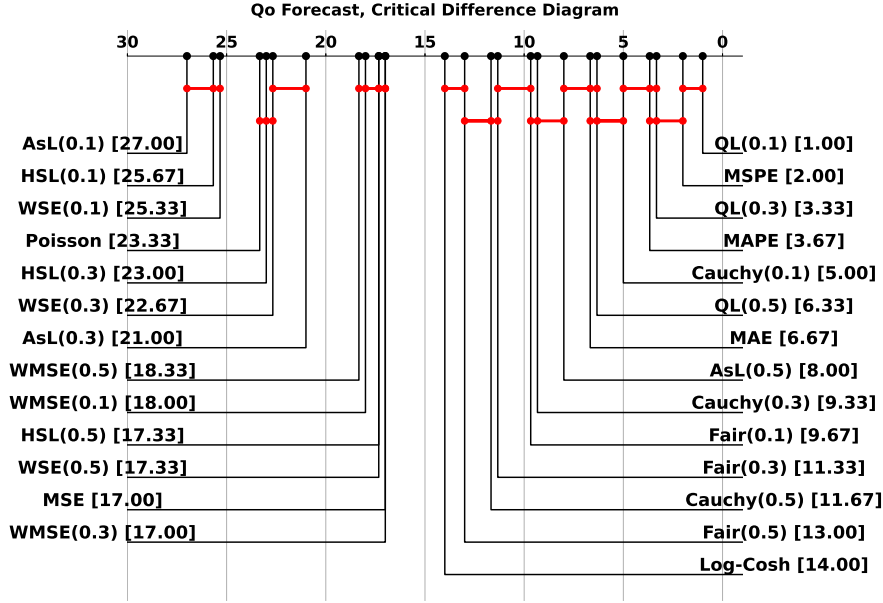


Fig. 7. CD Diagram for statistical ranking of loss functions on Qo forecast; average ranking across three datasets. The top-performing losses are identified from right to left.

rankings correspond directly to the outliers and high-variance footprints observed in the radar plots, where these specific loss functions failed to maintain stability across the validation and test sets. Ultimately, the statistical evidence from the three datasets underscores that the QL and MSPE frameworks offer the most robust and accurate solution for characterizing complex offshore production data.

On the Qw forecast, the statistical evaluation of loss functions illustrated in Figure 8 confirms the heightened complexity of this variable compared to oil rate forecasting. While the Qo analysis showed a relatively clear separation between top-tier models and baselines, the Qw diagram reveals significantly more overlap among algorithms, as evidenced by the extensive, intersecting red horizontal bars. This statistical analysis underscores the challenges of predicting water-phase transients, in which multiple loss functions often fail to yield statistically significant differences despite differing average ranks.

6.4. Regression Analysis

The regression analysis underscores the superior performance of QL(0.1), which emerged as the most robust loss function across the experimental trials. This is particularly evident in the DS-III test set, where QL(0.1) achieved a test MAE of 252.66 ± 55.31 compared to the other datasets. By applying QL(0.1) as the primary loss function, Figure 9 shows the regression performance of the top-performing DD model on the Qo forecast, which achieved the minimum training error for each of the

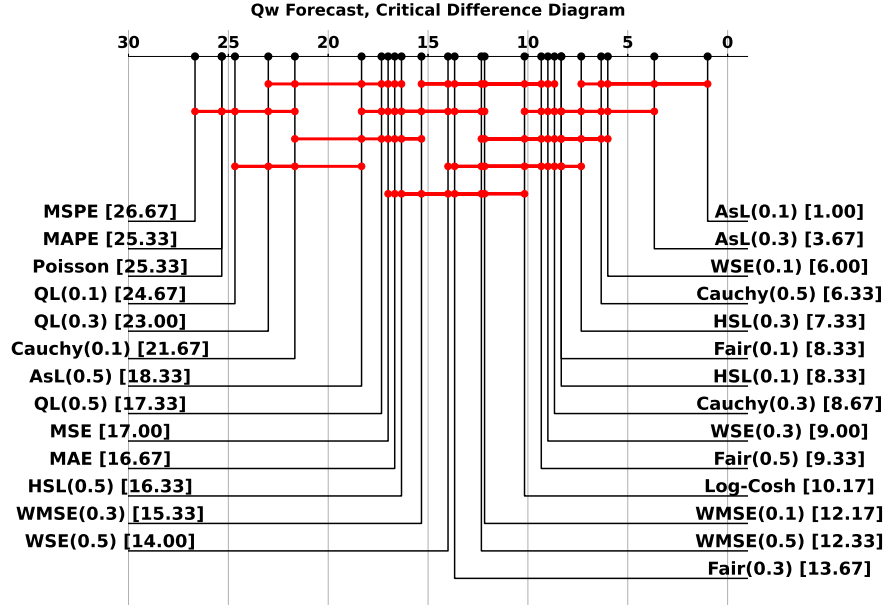


Fig. 8. Critical Difference (CD) diagram for statistical ranking of loss functions on Qw forecast; average ranking across three datasets. The top-performing losses are identified from right to left.

six producer wells (P11-P16) for the 30-Day forecast. It also compares the performance of the DD model with a similar configuration trained with baseline losses (MSE and MAE). The plots present the DD model's performance from the Validation (Val.) to the final forecast (Test) set.

The DD model trained on QL(0.1) (red line) consistently tracks the actual test data (blue dots) with the highest precision. By focusing on the 0.1 quantile, this asymmetric loss effectively filters out the operational noise and fluctuations. The similar model trained on both MAE (green) and MSE (orange) exhibits a consistent overall overestimation bias. This behavior confirms that symmetric baseline functions are highly sensitive to the uneven nature of petroleum production data, often failing to capture the true underlying signal when faced with the non-normal residuals typical of offshore reservoirs.

The regression analysis illustrated for Qo further elucidates the performance of the selected loss functions during transient operational events, specifically the well shut-ins at P14 and P15, which lasted three and six days, respectively. These shut-ins induced localized pressure redistribution, resulting in anomalously high production rates in neighboring wells P11, P12, P13, and P16. While these spikes represent physical deviations from steady-state production, they serve as outliers relative to the primary regression trend. Notably, the QL(0.1) loss function demonstrates superior tracking of the underlying signal across all wells, effectively filtering these transient perturbations

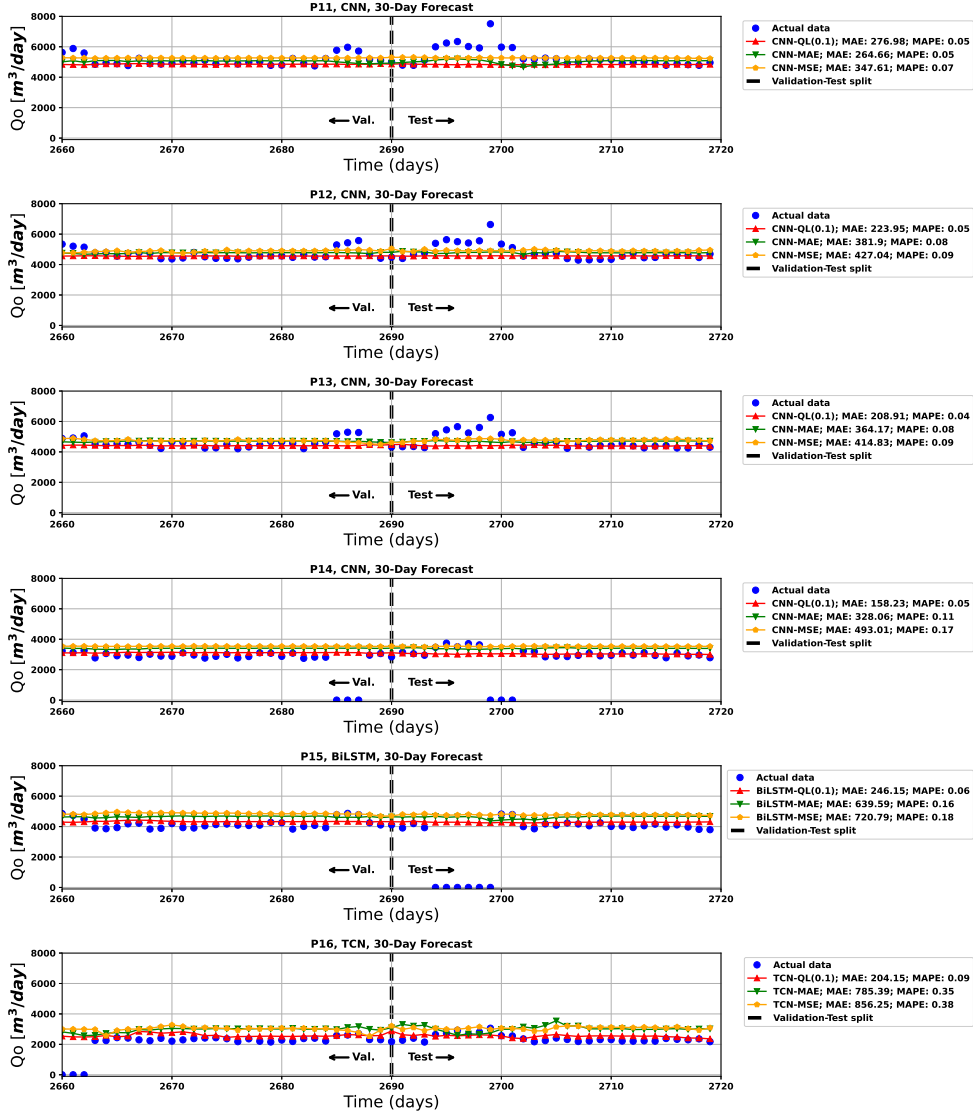


Fig. 9. Regression analysis for Q_o forecasting using the DS-III configuration across all wells over a 30-day forecast horizon. The plots show the transition from the validation phase to the final prediction (Test) set.

to maintain the main production trend. In contrast, the baseline symmetric loss functions—MAE and MSE—exhibit pronounced sensitivity to these fluctuations, leading to a visible overestimation of the primary signal trend as they attempt to minimize residuals associated with the shut-in-induced outliers.

On the Qw analysis, Figure 10 presents the regression performance across the producer wells using the AsL(0.1) loss function. The visualization offers a critical perspective on the operational strengths and inherent limitations of current DD models, particularly regarding their ability to characterize positive outliers. While the statistical analysis suggests that performance variations among the evaluated loss functions are not significantly divergent for water rates, the regression plots reveal a persistent underestimation of peak production events, mainly with baseline loss functions, whereas AsL(0.1) captures the main signal trend in most wells.

The Qw analysis during the identical shut-in periods appeared for Qo at P14 and P15 reveals similar transient behavior, particularly the surge in production at P11, P12, P13, and P16. While the AsL(0.1) loss function effectively captures the primary signal trend, it exhibits a unique sensitivity to positive deviations observed in wells such as P13. Unlike symmetric baselines that merely average these fluctuations, consistently underestimating them in most wells, the asymmetric approach captures higher-magnitude values while preserving the underlying production trajectory.

Overall, this systematic bias analysis indicates that while DD models are proficient at identifying and tracking the primary signal trends, high-magnitude transients remain a significant challenge. Consequently, selecting an appropriate loss function is a critical design decision; the optimization objective should align precisely with the physical nature of the target signal to enhance the model’s robustness and ensure reliable convergence during training of DD architectures. Therefore, integrating an appropriate loss function enables the forecasting DD model to maintain high fidelity across individual well transients, effectively isolating the actual production signal from operational noise. The results suggest that capturing these volatile spikes—often triggered by external operational factors like neighborhood well shut-ins—requires more sophisticated methodological approaches beyond standard regression.

7. Conclusions

This study systematically evaluated and compared the influence of different training loss functions on the performance of Data-Driven (DD) architectures for offshore reservoir production forecasting. The main contributions of this work are reflected in the following aspects:

- By analyzing oil and water production forecasts across three distinct data scenarios, the investigation demonstrated that loss-function selection is a critical modeling decision for representing the distinct transient characteristics of multiphase production rates.
- The results reveal that traditional symmetric baselines, such as Mean Squared Error (MSE) and Mean Absolute Error (MAE), may provide limited performance under abrupt reservoir production transients and operational interventions, particularly when underprediction and overprediction have different consequences.
- Among the pool of evaluated loss functions, asymmetric losses exhibited superior performance for oil and water rates forecasting, reducing high-magnitude forecast errors and improving the representation of transient production responses. Their benefits were particularly pronounced for oil-rate forecasting, with error reductions

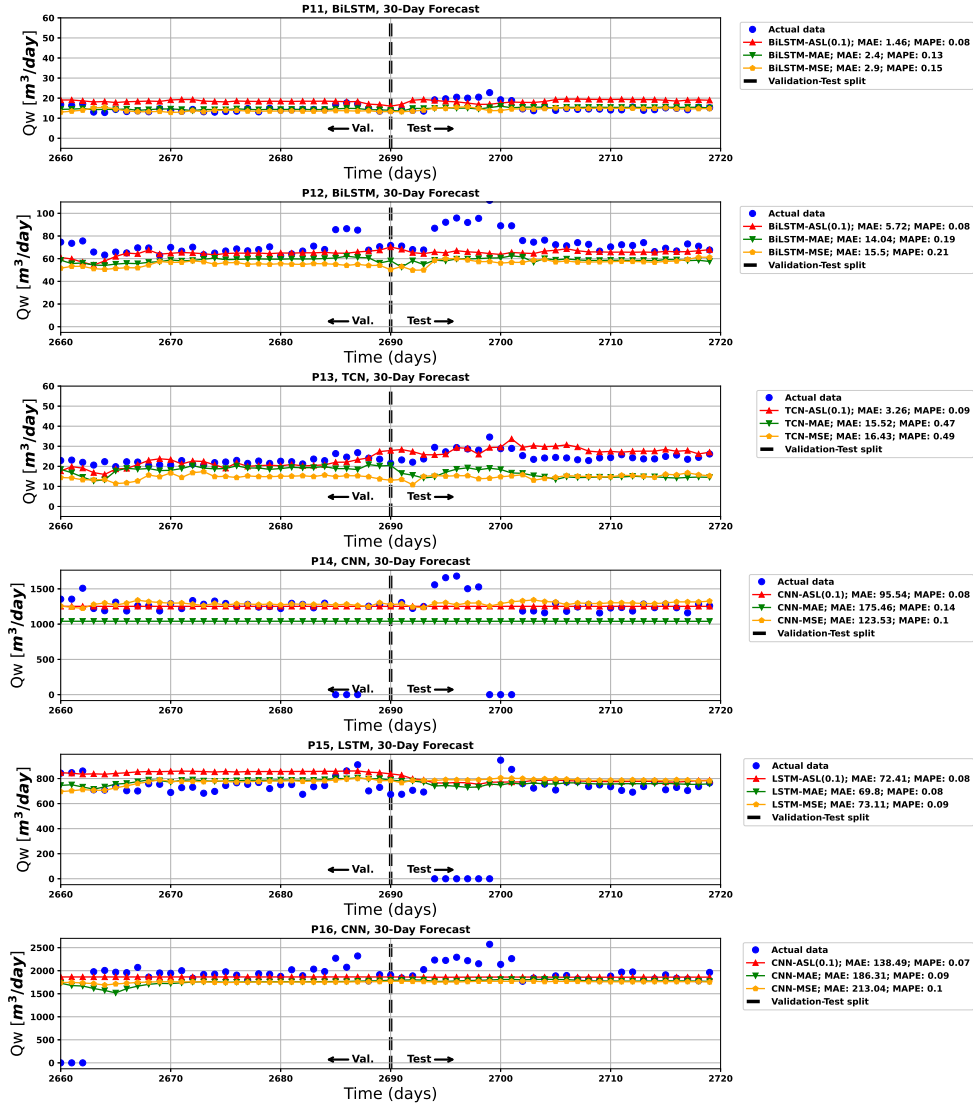


Fig. 10. Regression analysis for Q_w forecasting using the DS-III configuration across all wells over a 30-day forecast horizon. The plots show the transition from the validation phase to the final prediction (Test) set.

of approximately 50%, compared with approximately 5% for water-rate forecasting relative to conventional loss-function baselines.

- Aligning the training loss with the specific characteristics of each production variable provides a promising direction for improving forecasting components that may support future digital twins and optimization workflows for complex multi-well offshore assets.

Reliably forecasting abrupt nonlinear production responses induced by operational interventions remains challenging in complex reservoir systems and motivates further integration of reservoir engineering knowledge into DD forecasting approaches. Future research should investigate the combination of specialized loss functions with domain-specific expert systems to incorporate operational knowledge and engineering rules into forecasting frameworks. Another direction is the integration of asymmetric training objectives with existing physics-informed and physics-constrained neural networks, incorporating reservoir constraints and material-balance relationships to improve the representation of production transients and interwell interactions.

Nomenclatures

Latin Letters

c	Tuning constant
d	Dilation parameter
e	Prediction error
$\mathbf{I}$	Indicator vector
L	Number of stack layers
n	Number of data samples
Q_g	Gas production rate [m ³ /d]
Q_o	Oil production rate [m ³ /d]
Q_w	Water production rate [m ³ /d]
w	Weighting parameter
W	Diagonal weight matrix
$\mathbf{y}$	Actual production value [m ³ /d]
$\hat{\mathbf{y}}$	Predicted value [m ³ /d]

Greek Letters

α	Weighting factor
γ	Tuning constant
δ	Threshold value
τ	Controller of the level of asymmetry

Mathematical Operators

$\odot$	Hadamard elementwise product
$\oslash$	Elementwise division

List of Acronyms

AsL	Asymmetric Linear
BiLSTM	Bidirectional Long Short-Term Memory
CD	Critical Difference
CUDA	Compute Unified Device Architecture
CNN	Convolutional Neural Network
DD	Data-Driven
GB	Gigabyte
GPU	Graphics Processing Unit
HSL	Huber-like Asymmetric Loss
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
MSPE	Mean Squared Percentage Error
NMAE	Normalized Mean Absolute Error
QL	Quantile Loss
TCN	Temporal Convolutional Network
WMSE	Weighted Mean Squared Error
WSE	Weighted Squared Error

Acknowledgment. The authors would like to thank the Center for Energy and Petroleum Studies (CEPETRO), the Faculty of Technology - Limeira (FT), the School of Mechanical Engineering (FEM), the Energi Simulation Group, the Research Group in Reservoir Simulation and Management (UNISIM-UNICAMP), and Computer Modeling Group Ltd. (CMG).

Disclosure statement. The authors declare that this study did not involve human participants or animal subjects.

Competing Interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. The authors would like to thank the support of Petrolíam Nasional Berhad (PETRONAS) Petróleo Brazil, through the Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP) Research and Development Levy, as a commitment to research and development investments (Grant Number: 23051-6).

Data Availability. The UNISIM-IV-2026 benchmark dataset used in this work is publicly available at <https://www.unisim.cepetro.unicamp.br/benchmarks/en/unisim-iv/unisim-iv-2026>.

Code availability. The source code of this study will be publicly available on GitHub after the paper’s acceptance.

References

- AlRassas AM, Al-Qaness MA, Ewees AA, et al (2022) Advance artificial time series forecasting model for oil production using neuro fuzzy-based slime mould algorithm. *Journal of Petroleum Exploration and Production Technology* 12(2):383–395. <https://doi.org/10.1007/s13202-021-01405-w>
- Aranha P, Policarpo NA, Sampaio M (2024) Unsupervised machine learning model for predicting anomalies in subsurface safety valves and application in offshore wells during oil production. *Journal of Petroleum Exploration and Production Technology* 14:567–581. <https://doi.org/10.1007/s13202-023-01720-4>
- Botechia V, Correia M, Rios V, et al (2023) Unisim-iv: Benchmark proposal for light oil carbonate reservoir with high co2 content. *Brazilian Journal of petroleum and Gas* 16(4). <https://doi.org/10.5419/bjpg2022-0012>
- Chahar J, Verma J, Vyas D, et al (2022) Data-driven approach for hydrocarbon production forecasting using machine learning techniques. *Journal of Petroleum Science and Engineering* 217:110757. <https://doi.org/10.1016/j.petrol.2022.110757>
- Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* 7:1–30. URL <https://dl.acm.org/doi/10.5555/1248547.1248548/>
- Fu D, Fu Y, Zhang Y, et al (2024) Experimental simulation study on influencing factors of liquid production capacity in heterogeneous water drive reservoirs. *Physics of Fluids* 36(1):016612. <https://doi.org/10.1063/5.0184992>
- Ghazvini A, Sharef NM, Sidi FB (2024) Prediction of course grades in computer science higher education program via a combination of loss functions in lstm model. *IEEE Access* 12:30220–30241. <https://doi.org/10.1109/ACCESS.2024.3351186>
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural computation* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hu Y, Xin X, Yu G, et al (2025) Deep insight: an efficient hybrid model for oil well production forecasting using spatio-temporal convolutional networks and kolmogorov–arnold networks. *Scientific Reports* 15(1):8221. <https://doi.org/10.1038/s41598-025-91412-2>
- Huang Y, Ren R (2024) A garch model selection and estimation method based on neural network with the loss function of mean square error and model confidence set. *Journal of Forecasting* 43(8):3177–3193. <https://doi.org/10.1002/for.3175>
- Jadon A, Patil A, Jadon S (2024) A comprehensive survey of regression-based loss functions for time series forecasting. In: *International Conference on Data Management, Analytics & Innovation*, Springer, pp 117–147, https://doi.org/10.1007/978-98-98-98-98-9_10

- Kim S, Kim TW, Jo S (2025a) Artificial intelligence in geoenergy: bridging petroleum engineering and future-oriented applications. *Journal of Petroleum Exploration and Production Technology* 15(2):35. <https://doi.org/10.1007/s13202-025-01939-3>
- Kim Y, Min B, Sun A, et al (2025b) Learning-based pattern-data-driven forecast approach for predicting future well responses. *Journal of Petroleum Exploration and Production Technology* 15(2):32. <https://doi.org/10.1007/s13202-025-01937-5>
- Lavi B, Bertini Junior JR, Oliveira Pires L, et al (2025) The impact of asymmetric loss functions on data-driven short-term forecasting approaches. In: 2025 12th International Conference on Soft Computing & Machine Intelligence (ISCMI), pp 109–113, <https://doi.org/10.1109/ISCMI67495.2025.11358468>
- Lawrence S, Giles CL, Tsoi AC, et al (1997) Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks* 8(1):98–113. <https://doi.org/10.1109/72.554195>
- Lea C, Flynn MD, Vidal R, et al (2017) Temporal convolutional networks for action segmentation and detection. In: proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 156–165, <https://doi.org/10.1109/cvpr.2017.113>
- Loutfi AA, Sun M, Loutfi I, et al (2022) Empirical study of day-ahead electricity spot-price forecasting: Insights into a novel loss function for training neural networks. *Applied Energy* 319:119182. <https://doi.org/10.1016/j.apenergy.2022.119182>
- Lu S, Cui C, Wu Z, et al (2025) Simultaneous multi-well production forecasting and operational strategy awareness in heterogeneous reservoirs: a spatiotemporal attention-enhanced multi-graph convolutional network. *Scientific Reports* 15(1):42602. <https://doi.org/10.1038/s41598-025-26664-z>
- Olsen F, Schillaci C, Ibrahim M, et al (2022) Borough-level covid-19 forecasting in london using deep learning techniques and a novel mse-moran's i loss function. *Results in Physics* 35:105374. <https://doi.org/10.1016/j.rinp.2022.105374>
- Peng L, Han G, Chen Z, et al (2022) Dynamically coupled reservoir and wellbore simulation research in two-phase flow systems: a critical review. *Processes* 10(9):1778. <https://doi.org/10.3390/pr10091778>
- Pires LO, Lavi B, Bertini JRJ, et al (2026) Short-term reservoir forecasting for operational management: Applicability limits of reservoir simulation and machine learning models. In: SPE Europec featured at EAGE Conference and Exhibition, p D011S001R004, <https://doi.org/10.2118/233181-MS>
- Pires LO, Lavi B, Bertini Junior JR, et al (2026) Data-driven production forecast for oil and gas reservoirs - advantages and limitations. *Geoenergy Science and*

- Engineering 265:214569. <https://doi.org/10.1016/j.geoen.2026.214569>
- Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* 45(11):2673–2681. <https://doi.org/10.1109/78.650093>
- Sturm WL, Belfroid SPC, van Wolfswinkel O, et al (2004) Dynamic reservoir well interaction. In: *SPE Annual Technical Conference and Exhibition*, pp SPE–90108–MS, <https://doi.org/10.2118/90108-MS>
- Wang Q, Ma Y, Zhao K, et al (2022) A comprehensive survey of loss functions in machine learning. *Annals of Data Science* 9(2):187–212. <https://doi.org/10.1007/s40745-020-00253-5>
- Wang Q, Hussein MA, Kanabar B, et al (2025) Precise prediction of choke oil rate in critical flow condition via surface data. *Journal of Petroleum Exploration and Production Technology* 15(7):117. <https://doi.org/10.1007/s13202-025-02013-8>
- Yang J, Li S, Zeng J, et al (2024) Data-driven assessment of well stimulation in unconventional gas reservoirs. *Scientific Reports* 14(1):31133. <https://doi.org/10.1038/s41598-024-82454-z>
- Zhang L, Cheng Z, Hao Z, et al (2025a) Integrating inflow control valve control with lstm networks for oil production forecasting in horizontal intelligent well application. *Journal of Petroleum Exploration and Production Technology* 15(3):51. <https://doi.org/10.1007/s13202-025-01964-2>
- Zhang X, Huang Z, Wu Y, et al (2025b) Multi-period learning for financial time series forecasting. In: *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, pp 2848–2859, <https://doi.org/10.1145/3690624.3709422>