American Journal of Modern Energy
2026, Vol. 12, No. 3, pp. 49–58
https://doi.org/10.11648/j.ajme.20261203.12

Research Article

Hybrid Physics-Informed Dynamic Productivity Index (HPD-PI) Model for Oil Reservoir Production Forecasting

Abdul Qoyum Adegoke Olowookere[1, *], **Adewale Usman Oguntola**[2], **Ebenezer Leke Odekanle**[1]

[1]Department of Chemical and Petroleum Engineering, Abiola Ajimobi Technical University, Ibadan, Nigeria
[2]Department of Computer Science, Abiola Ajimobi Technical University, Ibadan, Nigeria

Abstract

Accurate prediction of reservoir production remains a critical challenge due to the complex interplay of multiphase flow behavior, temporal dynamics, and reservoir heterogeneity. This study proposes a Hybrid Physics-Informed Dynamic Productivity Index (HPD-PI) model that integrates classical inflow performance relationships with data-driven temporal features to enhance production forecasting. The model extends the conventional productivity index formulation by incorporating water-cut and gas-oil ratio corrections, dynamic production trends, and temporal memory through lagged production terms. A comprehensive dataset comprising production, pressure, and multiphase-flow parameters was used to evaluate the proposed approach. A baseline Random Forest model was first developed for comparison, achieving a coefficient of determination (R^2) of 0.8138. However, analysis revealed that the model was predominantly driven by temporal dependencies with limited physical interpretability. In contrast, the proposed Hybrid HPD-PI model achieved superior performance, with an R^2 of 0.9466, root mean square error (RMSE) of 71.01, and mean absolute error (MAE) of 38.78. The improvement is attributed to the integration of temporal memory, multiphase flow corrections, and dynamic behavior within a physically consistent framework. The results demonstrate that the hybrid model effectively captures both short-term fluctuations and long-term production trends while maintaining interpretability. The study highlights the importance of combining physical principles with data-driven techniques for robust reservoir modeling. The proposed HPD-PI framework provides a reliable tool for production forecasting, reservoir management, and decision-making, with potential applicability to a wide range of reservoir systems.

Keywords

Hybrid Productivity Index, Reservoir Modeling, Production Forecasting, Multiphase Flow, Water Cut, Gas-Oil Ratio, Physics-Informed Modeling, Time-Series Analysis

1. Introduction

Reservoir performance prediction is a fundamental aspect of petroleum engineering, playing a critical role in production optimization, reservoir management, and economic decision-making. Accurate estimation of production rate enables operators to optimize recovery strategies, minimize operational risks, and improve overall field performance. Traditionally,

*Correspondence: Abdul Qoyum Adegoke Olowookere (abdulqoyumadegoke@gmail.com)

Received: 6 April 2026; **Accepted:** 15 April 2026; **Published:** 27 July 2026

Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

reservoir engineers have relied on analytical models such as the productivity index (PI) and inflow performance relationships (IPR) to describe the relationship between production rate and pressure drawdown [18-20]. These models provide a simplified representation of reservoir behavior and have been widely adopted due to their ease of implementation and interpretability.

The classical productivity index formulation assumes a linear relationship between production rate and pressure differential, based on steady-state flow conditions and single-phase fluid behavior. While this assumption is valid under ideal conditions, real reservoir systems rarely exhibit such simplicity. In practical scenarios, reservoirs are characterized by multiphase flow, pressure depletion, heterogeneity, and evolving operational conditions. The presence of water production, commonly expressed as water cut, reduces effective oil mobility and alters flow dynamics, while gas interference, represented by the gas-oil ratio (GOR), introduces nonlinear effects that significantly impact production performance. These factors violate the assumptions underlying the classical PI model, leading to reduced predictive accuracy.

In addition to multiphase effects, reservoir performance evolves due to depletion, formation damage, and changes in operational strategies. Consequently, the productivity index cannot be considered a constant parameter but rather a time-dependent variable that reflects the dynamic nature of reservoir systems. Traditional models fail to capture this temporal variability, making them less reliable for long-term forecasting and real-time monitoring applications. Furthermore, production data often exhibit irregularities arising from operational disturbances, sensor noise, and unexpected reservoir behavior. Classical physics-based models cannot detect or adapt to such anomalies, further limiting their applicability in complex reservoir environments.

Recent advances in machine learning have provided powerful tools for modeling complex and nonlinear reservoir behavior. Several studies have demonstrated that data-driven approaches can significantly improve prediction accuracy in reservoir performance analysis [1, 5, 7]. For example, Kang *et al.* [1] and Chen *et al.* [5] showed that machine learning models are capable of capturing intricate relationships between reservoir variables and production rates, outperforming traditional analytical methods. Similarly, Rahmanifard and Gates [6] highlighted the effectiveness of data-driven techniques for production forecasting in unconventional reservoirs. Despite these successes, purely data-driven models often suffer from a lack of physical interpretability and may fail to generalize beyond the conditions represented in the training data.

To address these challenges, hybrid modeling approaches that integrate physical principles with machine learning have emerged as a promising direction. Physics-informed machine learning frameworks incorporate governing equations and domain knowledge into data-driven models, thereby improving both interpretability and predictive robustness [11, 13]. These approaches have shown significant potential in subsurface flow modeling and reservoir simulation. However, existing studies primarily focus on general predictive frameworks and do not explicitly reformulate the productivity index to incorporate multiphase effects, temporal dynamics, and anomaly-aware behavior within a unified mathematical structure.

Therefore, there remains a critical need for a comprehensive formulation that bridges the gap between traditional physics-based models and modern data-driven techniques. Such a formulation should be capable of capturing multiphase flow effects, adapting to time-varying reservoir conditions, and responding to anomalous production behavior while maintaining physical consistency and interpretability.

In this study, a Hybrid Physics-Informed Dynamic Productivity Index (HPD-PI) model is proposed to address these limitations. The proposed model extends the classical productivity index by incorporating multiphase flow corrections through water cut and gas-oil ratio terms, introducing temporal dynamics via production rate evolution, and integrating anomaly-aware intelligence derived from machine learning models. By combining physical principles with data-driven adaptability, the HPD-PI model provides a unified and robust framework for reservoir performance prediction.

The main contributions of this work include the development of a dynamic productivity index formulation that accounts for time-dependent reservoir behavior, the integration of multiphase flow effects into the productivity model, the incorporation of temporal production dynamics, and the inclusion of anomaly detection mechanisms to capture irregular reservoir behavior. The proposed approach enhances predictive accuracy while preserving physical interpretability, making it suitable for real-time reservoir monitoring and intelligent production optimization.

2. Related Studies

Recent advancements in reservoir engineering have demonstrated a progressive shift from traditional analytical models toward data-driven and hybrid modeling approaches. This transition is largely driven by the increasing complexity of reservoir systems and the limitations associated with classical productivity index formulations. Early analytical models, such as decline curve analysis and inflow performance relationships, established fundamental relationships between production rate and pressure drawdown [18, 20]. These models have been widely adopted due to their simplicity and interpretability. However, they are based on restrictive assumptions, including steady-state flow conditions, single-phase fluid behavior, and constant reservoir properties, which limit their applicability in real-world reservoir systems.

In practical scenarios, reservoir flow behavior is inherently multiphase, involving oil, water, and gas interactions. The presence of water production, quantified as water cut, significantly reduces effective oil mobility and alters production efficiency. Similarly, gas interference, typically represented by the gas-oil ratio (GOR), introduces nonlinear effects that disrupt stable flow regimes and reduce oil productivity. Recent studies have highlighted the importance of incorporating multiphase behavior into reservoir modeling. For instance, Sepahvand *et al.* [8] developed machine learning models for predicting gas-oil ratio, demonstrating its strong influence on reservoir performance. Additionally, reservoir conformance and water breakthrough studies have shown that neglecting multiphase effects can lead to substantial prediction errors, particularly in mature and heterogeneous reservoirs.

The emergence of machine learning has provided new opportunities for modeling complex and nonlinear reservoir behavior. Several studies have demonstrated the effectiveness of data-driven approaches in predicting reservoir performance. Kang et al. [1] proposed a stacking ensemble model for productivity prediction in tight oil reservoirs, achieving improved accuracy compared to conventional approaches. Similarly, Zhu et al. [3] developed a machine learning-based framework for early productivity forecasting in multilayered wells, while Chen et al. [5] applied machine learning techniques to estimate production capacity in fractured shale reservoirs. Rahmanifard and Gates [6] further emphasized the potential of data-driven methods in forecasting production from unconventional reservoirs. Recent studies have also explored machine learning applications in reservoir characterization and prediction. Cheddad [14] applied machine learning to permeability prediction in heterogeneous reservoirs, while Risha et al. [15] focused on uncertainty-driven modeling of microporosity. Ivlev [16] further demonstrated the use of machine learning for integrating well and seismic data in complex reservoir environments. Despite their predictive strength, these models are often considered black-box approaches and lack physical interpretability, which limits their reliability and generalization in unseen reservoir conditions. Despite their predictive strength, these models are often considered black-box approaches and lack physical interpretability, which limits their reliability and generalization in unseen reservoir conditions.

To address these challenges, hybrid modeling approaches that integrate physical principles with machine learning have been introduced. Physics-informed machine learning frameworks incorporate governing equations and domain knowledge into data-driven models, thereby improving predictive accuracy while maintaining consistency with physical laws. Karniadakis *et al.* [11] introduced physics-informed neural networks, which embed physical constraints into neural network architectures. Willard *et al.* [13] further explored the integration of physics-based models with machine learning, demonstrating enhanced robustness and generalization. In reservoir engineering applications, Kanin *et al.* [4] combined mechanistic modeling with machine learning to generate permeability maps consistent with well test measurements, highlighting the advantages of hybrid approaches.

Although these advancements represent significant progress, existing studies remain limited in their ability to provide a unified productivity-based formulation that captures multiphase effects, temporal reservoir dynamics, and anomaly-aware behavior simultaneously. Classical models lack adaptability, machine learning models lack interpretability, and hybrid approaches have not explicitly reformulated the productivity index to integrate these critical aspects into a single framework. This gap highlights the need for a comprehensive and adaptive model that combines the strengths of physics-based and data-driven approaches while addressing their individual limitations.

Table 1. *Summary of Related Studies on Reservoir Productivity Modeling and Machine Learning Approaches.*

Author(s)	**Study Purpose & Application Area**	**Methodology**	**Dataset**	**Technique**	**Key Findings**	**Limitations**
Kang et al. [1]	Productivity prediction in tight oil reservoirs	Supervised learning	Field production data	Stacking ensemble ML	High prediction accuracy	Lacks physical interpretability
Nashed et al. [2]	Waterflood optimization	Data-driven modeling	Water injection data	ML regression	Improved injection efficiency	No temporal dynamics
Zhu et al. [3]	Early productivity forecasting	ML-based prediction	Multilayered well data	Deep learning	Accurate early prediction	Ignores multiphase physics
Kanin et al. [4]	Reservoir permeability mapping	Hybrid modeling	Well test data	Mechanistic + ML	Improved permeability estimation	Not PI-based
Chen et al. [5]	Productivity prediction in shale reservoirs	Machine learning	Shale dataset	ML regression	Captures nonlinear relationships	Black-box nature
Rahmanifard & Gates [6]	Production forecasting review	Literature review	Multiple datasets	Comparative analysis	ML improves forecasting accuracy	Limited physical grounding
Sepahvand et al. [8]	GOR prediction	ML modeling	Reservoir dataset	ML regression	Accurate GOR estimation	Not integrated with PI

Author(s)	Study Purpose & Application Area	Methodology	Dataset	Technique	Key Findings	Limitations
Khashman et al. [9]	Productivity evaluation	Analytical modeling	Carbonate reservoir	PI analysis	Useful for field evaluation	Static assumptions
Roustazadeh et al. [10]	Recovery factor prediction	ML modeling	Reservoir dataset	XGBoost	High predictive performance	Limited interpretability
Karniadakis et al. [11]	Physics-informed ML	PINNs	Synthetic/physical systems	Deep learning	Improved accuracy with physics constraints	Not reservoir-specific
Tu et al. [12]	Hybrid physics-ML modeling	Hybrid approach	Energy systems	Physics + ML	Better generalization	Not applied to reservoirs
Willard et al. [13]	Hybrid modeling framework	Survey	Multi-domain	Physics + ML	Robust and interpretable models	No PI reformulation

3. Methodology

This section presents the development of the proposed Hybrid Physics-Informed Dynamic Productivity Index (HPD-PI) model. The formulation begins with the classical productivity index, extends to incorporate multiphase flow effects, introduces dynamic reservoir behavior, and finally integrates temporal memory to improve predictive performance.

3.1. Classical Inflow Formulation

The starting point of the proposed model is the conventional productivity index (PI) relationship, which expresses the production rate as a function of pressure drawdown, as given in (1):

$$q = J(P_r - P_{wf}) \tag{1}$$

where q is the production rate, J is the productivity index, P_r is the reservoir pressure, and P_{wf} is the flowing bottom-hole pressure. The term $(P_r - P_{wf})$ represents the pressure driving force governing fluid flow toward the wellbore (drawdown pressure).

Although (1) provides a simplified representation of reservoir behavior, it is based on assumptions of steady-state flow, single-phase conditions, and constant productivity, which limit its applicability in real reservoir systems.

3.2. Multiphase Flow Correction

To account for multiphase flow effects, the classical inflow equation is generalized by introducing a correction function $f(W_c, GOR)$, as expressed in (2):

$$q = J(P_r - P_{wf}) \cdot f(W_c, GOR) \tag{2}$$

Where W_c is the water cut, and GOR is the gas-oil ratio. The function $f(W_c, GOR)$ modifies the ideal production rate to reflect multiphase flow behavior.

3.2.1. Water-Cut Effect

Water cut is defined as the fraction of produced liquid that is water, as shown in (3):

$$W_c = \frac{q_w}{q_o + q_w} \tag{3}$$

The complementary oil fraction is given in (4):

$$1 - W_c = \frac{q_o}{q_o + q_w} \tag{4}$$

To incorporate reservoir-specific sensitivity, the water-cut effect is generalized using a linear approximation, as expressed in (5):

$$g(W_c) = 1 - \alpha W_c \tag{5}$$

Where α is a water-cut sensitivity coefficient.

3.2.2. Gas-Oil Ratio Effect

The effect of gas-oil ratio is modeled using an exponential decay function, as given in (6):

$$h(GOR) = e^{-\beta GOR} \tag{6}$$

where β is a gas sensitivity parameter controlling the influence of gas interference on production.

3.2.3. Combined Multiphase Correction

The combined multiphase correction function is obtained by multiplying the water-cut and gas effects, as expressed in (7):

$$f(W_c, GOR) = (1 - \alpha W_c)e^{-\beta GOR} \quad (7)$$

Substituting (7) into (2), the modified inflow relationship becomes:

$$q = J(P_r - P_{wf})(1 - \alpha W_c)e^{-\beta GOR} \quad (8)$$

3.3. Dynamic Productivity Index Formulation

To capture time-dependent reservoir behavior, the productivity index is reformulated as a dynamic variable, as shown in (9):

$$J \rightarrow J_d(t) \quad (9)$$

The dynamic productivity index is defined as a function of production trend and anomaly behavior, as given in (10):

$$J_d(t) = J_0 \cdot h(\frac{dq}{dt}, A_t) \quad (10)$$

Where J_0 is the baseline productivity index, $\frac{dq}{dt}$ represents the temporal rate of change of production, and $A_t$is the anomaly score.

3.3.1. First-Order Approximation

To ensure consistency with the baseline condition h (0,0) = 1, a first-order Taylor expansion of the correction function is performed, leading to (11):

$$h\left(\frac{dq}{dt}, A_t\right) \approx 1 + \gamma\frac{dq}{dt} - \delta A_t \quad (11)$$

Substituting (11) into (10), the dynamic productivity index is obtained as:

$$J_d(t) = J_0(1 + \gamma\frac{dq}{dt} - \delta A_t) \quad (12)$$

3.3.2. Numerical Implementation

The temporal derivative of production is approximated using finite difference methods, as shown in (14):

$$\frac{dq}{dt} \approx \frac{q_t - q_{t-1}}{\Delta t} \quad (13)$$

The structure of (12) ensures that the model reduces to the baseline productivity index when $\frac{dq}{dt} = 0$ and $A_t = 0$. The term $\gamma\frac{dq}{dt}$ captures temporal production dynamics and allows the productivity index to increase or decrease depending on production trends. In contrast, the anomaly term appears with a negative sign as $-\delta A_t$, reflecting the assumption that anomaly intensity represents degradation in reservoir performance. Consequently, higher anomaly values reduce the effective productivity index, ensuring physical consistency.

The anomaly score A_t is obtained using a machine learning-based anomaly detection model, such as an LSTM autoencoder trained on historical production data [17]. In this study, a proxy anomaly score derived from historical production deviations is used to approximate abnormal behavior.

Substituting Eq. (8) into the multiphase formulation yields:

$$q = J_d(t)(P_r - P_{wf})(1 - \alpha W_c)e^{-\beta GOR} \quad (14)$$

3.4. Hybrid HPD-PI Formulation

While the initial formulation of the HPD-PI model captures the influence of pressure drawdown, multiphase flow effects, and dynamic reservoir behavior, it does not explicitly account for the strong temporal dependency observed in production data. Reservoir production is inherently autocorrelated, with current production rates strongly influenced by recent production history.

To address this limitation, the proposed model is extended to incorporate temporal memory through lagged production terms. The resulting Hybrid HPD-PI model is expressed as:

$$q_t = \theta_1 q_{t-1} + \theta_2 q_{t-2} + J_0\left(P_r - P_{wf}\right)(1 - \alpha W_c)e^{-\beta GOR} + \gamma\frac{dq}{dt} - \delta A_t \quad (15)$$

where q_{t-1} and q_{t-2} represent production rates at previous time steps, and θ_1 and θ_2 are temporal weighting coefficients.

The first term captures the intrinsic temporal dynamics of reservoir production, reflecting decline trends and operational continuity. The second term represents the physics-based inflow relationship incorporating multiphase flow effects. The third term introduces dynamic corrections based on changes in production rate and anomaly behavior.

3.5. Dataset Description and Preprocessing

The dataset used in this study is derived from the publicly available Volve Field Dataset, provided by Equinor. The Volve dataset represents real production data from an offshore oil field located in the North Sea and includes detailed well-level operational, pressure, and production measurements. It has been widely used as a benchmark dataset for reservoir engineering and production forecasting studies.

The dataset contains key variables such as oil production rate (q_o), gas production rate (q_g), water production rate (q_w), bottom-hole pressure (P_{wf}), wellhead pressure, choke size, and temperature measurements. These variables were used to derive additional features required for modeling.

Data preprocessing involved handling missing values, removing non-producing periods, and ensuring physical consistency of the variables. Missing values in pressure-related features were treated using group-wise interpolation within each well to preserve reservoir continuity and avoid unrealistic discontinuities. Rows with insufficient production data were excluded from the analysis.

Feature engineering was performed to incorporate both physical and temporal information. Water cut was computed in equation (3), and the gas – oil ratio was defined as:

$$GOR = \frac{q_o}{q_o} \tag{16}$$

Temporal features were introduced to capture production dynamics. Lagged production values (q_{t-1}, q_{t-2}) were generated for each well, and the rate of change of production was calculated using historical observations as defined in Eq. (5). Additionally, rolling statistical measures were used to estimate an anomaly score (A_t) based on deviations from recent production trends.

To ensure numerical stability and robustness, extreme values of dynamic variables such as $\frac{dq}{dt}$ were clipped within percentile bounds, and anomaly scores were constrained within a predefined range.

3.6. Model Implementation and Evaluation Strategy

The proposed Hybrid HPD-PI model and the baseline machine learning model were implemented and evaluated using a time-aware validation strategy. The dataset was sorted chronologically and divided into training and testing subsets, with 80% of the data used for training and the remaining 20% reserved for testing. This approach ensures that future information is not used during training, thereby avoiding data leakage and maintaining forecasting realism.

The Hybrid HPD-PI model parameters were estimated using nonlinear least-squares optimization, allowing the model to fit the observed production data while maintaining its physical structure. The inclusion of temporal memory, multiphase corrections, and dynamic terms enables the model to capture both physical and temporal characteristics of reservoir behavior.

For comparison, a Random Forest regression model was developed as a baseline. The model was trained using the same dataset and a combination of physical and temporal features, including pressure drawdown, water cut, gas-oil ratio, and lagged production variables. The Random Forest algorithm was selected for its ability to capture nonlinear relationships and its robustness to complex datasets.

Model performance was evaluated using three standard regression metrics: root-mean-square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). These metrics provide a comprehensive assessment of prediction accuracy, error magnitude, and explanatory power.

4. Results and Discussion

This section presents a comprehensive evaluation of the proposed Hybrid Physics-Informed Dynamic Productivity Index (HPD-PI) model using real reservoir production data. The objective is to assess the model's ability to accurately predict production rates while maintaining physical interpretability. The results are analyzed in terms of data characteristics, multiphase flow behavior, temporal dynamics, and model performance in comparison with a baseline machine learning approach.

4.1. Production Data Characteristics and Temporal Behavior

Figure 1 shows the distribution of production rates reveals significant variability, with a wide spread of values ranging from low to high production regimes. This variability reflects the heterogeneous nature of the reservoir, where production performance is influenced by factors such as reservoir pressure depletion, well conditions, and fluid composition. The presence of a positively skewed distribution indicates that while most production values are moderate, occasional high-production periods occur, likely corresponding to optimal operating conditions or early production stages.

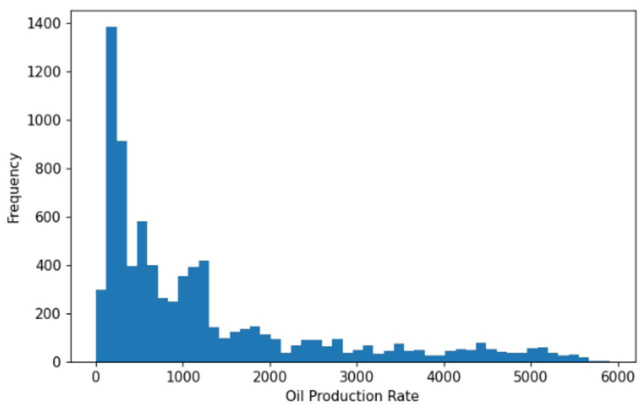

Figure 1. *Distribution of Production Rate in the Reservoir Dataset.*

Beyond the static distribution, the temporal evolution of production provides deeper insight into reservoir dynamics. The production time series exhibits non-monotonic behavior characterized by fluctuations, intermittent increases, and gradual declines. This deviates from the assumptions of classical decline models, which typically assume smooth and continuous decay. The observed irregularities are indicative of complex reservoir processes, including operational adjustments, pressure support mechanisms, and multiphase interactions shown in Figure 2.

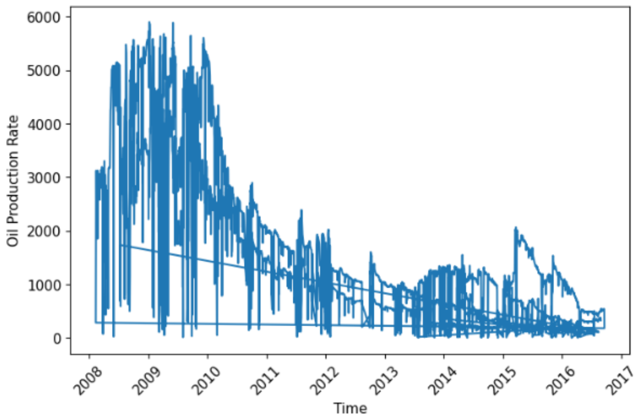

Figure 2. *Temporal Variation of Production Rate Showing Dynamic Reservoir Behavior.*

These observations justify the incorporation of temporal derivatives ($\frac{dq}{dt}$) and lagged production terms in the proposed model, as they enable the capture of both short-term fluctuations and long-term trends.

4.2. Influence of Multiphase Flow on Production

The relationship between water cut and oil production highlights the critical role of multiphase flow in reservoir performance. As water cut increases, a consistent decline in oil production is observed. This behavior can be attributed to the displacement of oil by water, which reduces effective oil saturation and impairs flow efficiency within the reservoir. The nearly monotonic decline confirms that water production is a dominant limiting factor in mature reservoirs shown in Figure 3.

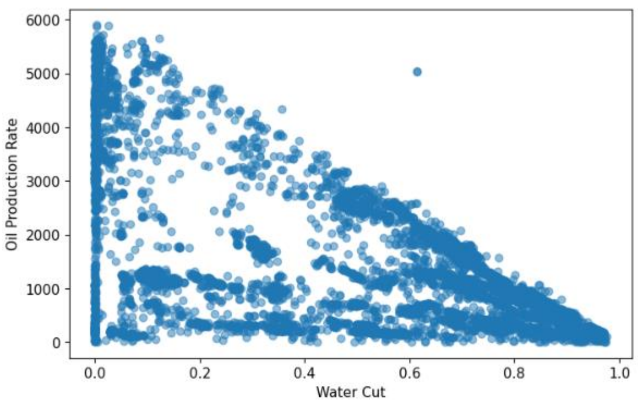

Figure 3. *Relationship Between Water Cut and Oil Production Rate.*

Similarly, the gas-oil ratio exhibits a nonlinear relationship with production displayed in Figure 4. At lower GOR values, the impact on oil production is relatively mild; however, as GOR increases, a more pronounced reduction in oil production is observed. This behavior reflects the increasing dominance of gas in the flow system, which reduces the effective mobility of oil and alters pressure gradients within the reservoir.

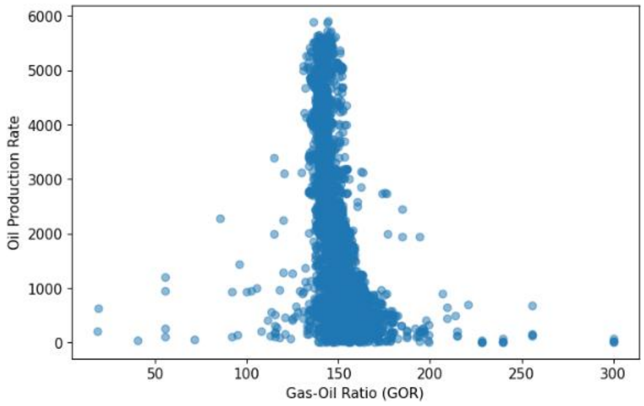

Figure 4. *Nonlinear Impact of Gas-Oil Ratio on Oil Production Performance.*

The observed nonlinear trend validates the use of an exponential decay formulation for GOR in the proposed model. Together, these results confirm that multiphase flow effects must be explicitly incorporated to achieve realistic production predictions.

4.3. Baseline Machine Learning Model Analysis

The baseline Random Forest model achieved an R^2 value of 0.8138, with an RMSE of 132.52 and MAE of 76.90. While these results indicate strong predictive capability, further analysis reveals that the model relies heavily on lagged production features.

Specifically, feature importance analysis shows that the first lagged production term accounts for over 85% of the predictive contribution, indicating that the model primarily learns temporal dependencies rather than underlying physical relationships. This behavior is expected in time-series data, where production values are highly autocorrelated.

However, the limited contribution of physical variables such as pressure drawdown, water cut, and GOR suggests that the model lacks interpretability. Although it captures patterns effectively, it does not provide meaningful insights into reservoir mechanisms, which is a key requirement for engineering applications.

4.4. Performance of the Hybrid HPD-PI Model

The Hybrid HPD-PI model demonstrates a substantial improvement in predictive performance, achieving an R^2 of 0.9466, RMSE of 71.01, and MAE of 38.78. This represents a significant reduction in prediction error compared to the baseline model.

Table 2. Model Performance Comparison.

Model	RMSE	MAE	R^2
Random Forest (Baseline)	132.52	76.90	0.8138
Hybrid HPD-PI (Proposed)	71.01	38.78	0.9466

The improvement can be attributed to the hybrid structure of the model, which integrates temporal memory, physical constraints, and dynamic behavior. Unlike the baseline model, the Hybrid HPD-PI model explicitly combines these components, allowing it to capture both data-driven patterns and physical relationships. The model successfully tracks the overall production trend while also capturing short-term variations, indicating robustness under varying reservoir conditions.

4.5. Physical Interpretation of Model Parameters

The optimized parameters provide meaningful insights into reservoir behavior. The temporal coefficients indicate strong dependence on recent production history, confirming the autoregressive nature of reservoir systems. This aligns with the physical understanding that reservoir depletion and flow continuity influence future production.

The water-cut coefficient demonstrates a significant negative impact on oil production, reinforcing the importance of water management in reservoir operations. The gas-oil ratio coefficient captures the suppressive effect of gas on oil flow, consistent with multiphase flow theory.

The negative value of the dynamic coefficient associated with production rate change indicates that declining production trends lead to further reductions in output, effectively capturing depletion dynamics. The negligible contribution of the anomaly term suggests that production variability in this dataset is primarily governed by temporal and physical factors rather than irregular disturbances.

4.6. Discussion of Key Findings and Implications

The results highlight several important findings. First, temporal dependency plays a dominant role in production prediction, necessitating the inclusion of lagged production terms. Second, multiphase flow effects significantly influence production behavior and must be incorporated for realistic modeling.

The failure of the initial physics-only formulation underscores the limitations of purely mechanistic models in capturing complex reservoir dynamics. By integrating temporal memory with physical relationships, the Hybrid HPD-PI model overcomes these limitations and achieves superior performance.

From a practical perspective, the proposed model provides a valuable tool for real-time production forecasting and reservoir management. Its ability to balance accuracy and interpretability makes it particularly suitable for engineering decision-making, where both predictive performance and physical understanding are essential.

5. Conclusion and Recommendations

This study presented a Hybrid Physics-Informed Dynamic Productivity Index (HPD-PI) model for predicting oil production in reservoir systems by integrating classical inflow relationships, multiphase flow effects, temporal dynamics, and data-driven components within a unified framework. The proposed model was developed to address the limitations of conventional productivity index formulations, which assume steady-state and single-phase flow, as well as purely data-driven approaches that often lack physical interpretability.

The results demonstrated that the baseline machine learning model achieved a reasonable predictive performance with an R^2 value of 0.8138; however, the model relied heavily on temporal features, particularly lagged production values, with minimal contribution from physical parameters. This confirms that while data-driven models can capture patterns in production data, they may fail to provide meaningful insights into the underlying reservoir mechanisms. In contrast, the proposed Hybrid HPD-PI model significantly improved predictive performance, achieving an R^2 of 0.9466, RMSE of 71.01, and MAE of 38.78. This improvement is attributed to the incorporation of temporal memory, multiphase flow corrections, and dynamic production behavior within a physically interpretable structure.

The model parameters provide important insights into reservoir behavior. The strong influence of lagged production terms confirms the inherent autocorrelation in reservoir systems, while the water-cut sensitivity coefficient highlights the substantial negative impact of water production on oil output. The gas-oil ratio parameter reflects the suppressive effect of gas on oil flow, and the dynamic term effectively captures production decline trends. The anomaly component showed negligible contribution in this study, indicating that production variability is largely governed by temporal continuity and physical flow mechanisms rather than irregular disturbances.

Overall, the Hybrid HPD-PI model demonstrates a robust balance between predictive accuracy and physical interpretability, making it a valuable tool for reservoir performance evaluation and production forecasting. Based on these findings, it is recommended that the model be further validated across different reservoir types and geological settings to assess its generalizability. Additionally, future research should explore the integration of more advanced anomaly detection techniques, such as deep learning-based models, to enhance the responsiveness of the anomaly term. The inclusion of additional reservoir properties, including permeability, porosity, and well completion data, may further strengthen the physical component of the model. Furthermore, the proposed framework can be extended for real-time production monitoring and optimization within digital oilfield environments, and its applicability to gas reservoirs and unconventional systems should also be investigated to broaden its scope.

Abbreviations

HPD-PI	Hybrid Physics-Informed Dynamic Productivity Index
PI	Productivity Index
GOR	Gas-Oil Ratio
Wc	Water Cut
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
R^2	Coefficient of Determination
ML	Machine Learning
RF	Random Forest
LSTM	Long Short-Term Memory
IPR	Inflow Performance Relationship

Acknowledgments

The authors acknowledge the use of the Volve Field Dataset provided by Equinor for research purposes. The authors also appreciate the contributions of the open research community whose tools and resources supported the implementation of this study. The authors further acknowledge the support and academic environment provided by Abiola Ajimobi Technical University, Ibadan, as well as the Society of Petroleum Engineers (SPE) Student Chapter of the institution, for fostering research development and professional growth.

Author Contributions

Abdul Qoyum Adegoke Olowookere: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft

Adewale Usman Oguntola: Supervision, Validation, Project administration, Writing – review & editing

Ebenezer Leke Odekanle: Investigation, Methodology, Validation, Resources, Writing – review & editing

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability Statement

The data used in this study are publicly available from the Volve Field Dataset provided by Equinor.

Conflicts of 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.

References

[1] Z. Kang, Y. Zheng, T. Zhang, H. Chen, X. Zhou, Q. Cai, and Y. Sun, "Productivity prediction in tight oil reservoirs: A stacking ensemble approach with hybrid feature selection," Processes, vol. 14, no. 7, p. 1089, Mar. 2026, https://doi.org/10.3390/pr14071089

[2] S. Nashed, O. Ejehu, and R. Moghanloo, "Advanced machine learning approaches for optimizing waterflooding: A focus on injector-producer dynamics," Petroleum Research, in press, corrected proof, Jan. 2026, https://doi.org/10.1016/j.ptlrs.2026.01.001

[3] R. Zhu, N. Li, G. Liu, F. Qu, C. Long, X. Wang, S. Xiu, F. Ling, Q. Liao, and G. Li, "Machine learning-driven early productivity forecasting for post-fracturing multilayered wells," Water, vol. 17, no. 19, p. 2804, Sep. 2025, https://doi.org/10.3390/w17192804

[4] E. Kanin, A. Garipova, S. Boronin, V. Vanovskiy, A. Vainshtein, A. Afanasyev, A. Osiptsov, and E. Burnaev, "Combined mechanistic and machine learning method for construction of oil reservoir permeability map consistent with well test measurements," Petroleum Research, vol. 10, no. 2, pp. 247–265, Jun. 2025, https://doi.org/10.1016/j.ptlrs.2024.09.001

[5] Y. Chen, J. Li, S. Qin, C. Liang, and Y. Chen, "Application of machine learning to predict the capacity of fractured horizontal wells in shale reservoirs," Processes, vol. 12, no. 11, p. 2527, Nov. 2024, https://doi.org/10.3390/pr12112527

[6] H. Rahmanifard and I. Gates, "A comprehensive review of data-driven approaches for forecasting production from unconventional reservoirs: Best practices and future directions," Artificial Intelligence Review, vol. 57, no. 8, p. 213, Jul. 2024, https://doi.org/10.1007/s10462-024-10865-5

[7] M. H. A. Idris, J. M. Cebula, J. Gholinezhad, S. Masum, and H. Ma, "Out-of-sample hydrocarbon production forecasting: Time series machine learning using productivity index-driven features and inductive conformal prediction," arXiv preprint, Aug. 2025. Available: https://arxiv.org/abs/2508.14078

[8] M. Sepahvand, M. Mohammadi, A. Madani, and M. Schaffie, "Advanced machine learning techniques for predicting solution gas-oil ratios in petroleum reservoirs: A comprehensive study and new empirical correlation," Digital Chemical Engineering, vol. 18, p. 100288, Mar. 2026, https://doi.org/10.1016/j.dche.2026.100288

[9] M. A. Khashman, H. Shirazi, and A. N. Al-Dujaili, "Comparative evaluation of productivity indicators in carbonate reservoir modeling: A case study of the Mishrif Formation in the Iraqi Buzurgan Oilfield," Discover Geoscience, vol. 3, p. 17, 2025, https://doi.org/10.1007/s44288-025-00118-5

[10] A. Roustazadeh, F. Male, B. Ghanbarian, M. B. Shadmand, V. Taslimitehrani, and L. W. Lake, "Machine learning-based estimation of oil recovery factor using XGBoost: Insights from classification and data-driven analyses," InterPore Journal, vol. 2, no. 3, Art. no. 53, 2025, https://doi.org/10.69631/ipj.v2i3nr53

[11] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, "Physics-informed machine learning," Nature Reviews Physics, vol. 3, pp. 422–440, May 2021, https://doi.org/10.1038/s42254-021-00314-5

[12] H. Tu, S. Moura, Y. Wang, and H. Fang, "Integrating physics-based modeling with machine learning for lithium-ion batteries," Applied Energy, vol. 329, p. 120289, Jan. 2023, https://doi.org/10.1016/j.apenergy.2022.120289

[13] C. Etienam, Y. Juntao, I. Said, O. Ovcharenko, K. Tangsali, P. Dimitrov, and K. Hester, "A novel AI-enhanced reservoir characterization with a combined mixture of experts: NVIDIA Modulus-based physics-informed neural operator forward model," arXiv preprint, Apr. 2024. Available: https://arxiv.org/abs/2404.14447

[14] F. A. Cheddad, "Enhancing petrophysical studies with machine learning: A field case study on permeability prediction in heterogeneous reservoirs," arXiv preprint, May 2023. Available: https://arxiv.org/abs/2305.07145

[15] M. Risha, M. Elsaadany, and P. Liu, "Uncertainty-driven modeling of microporosity and permeability in clastic reservoirs using random forest," arXiv preprint, Mar. 2025. Available: https://arxiv.org/abs/2503.16957

[16] D. Ivlev, "Reservoir prediction by machine learning methods on well data and seismic attributes for complex coastal conditions," arXiv preprint, Jan. 2023. Available: https://arxiv.org/abs/2301.03216

[17] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder-decoder for multi-sensor anomaly detection," arXiv preprint, Jul. 2016. Available: https://arxiv.org/abs/1607.00148

[18] J. J. Arps, "Analysis of decline curves," Trans. AIME, vol. 160, no. 1, pp. 228–247, Dec. 1945, https://doi.org/10.2118/945228-G

[19] T. Ahmed, Reservoir Engineering Handbook, 3rd ed. Woburn, MA, USA: Butterworth-Heinemann, 2001.

[20] M. Vogel, "Inflow performance relationships for solution-gas drive wells," Journal of Petroleum Technology, vol. 20, no. 1, pp. 83–92, 1968.