energies

MDPI

Article

Physics-Guided Machine Learning with Flowing Material Balance Integration: A Novel Approach for Reliable Production Forecasting and Well Performance Analytics

Eghbal Motaei [1], Tarek Ganat [2,*] and Hai T. Nguyen [3,4,*]

[1] Petroleum Engineering Department, Petronas Carigali Sdn. Bhd, Kuala Lumpur 50088, Malaysia; eghbal.motaei@petronas.com.my
[2] Department of Petroleum Chemical Engineering, Sultan Qaboos University, P.O. Box 33, Al-Khod, Muscat 123, Oman
[3] Laboratory of Applied Physics, Science and Technology Advanced Institute, Van Lang University, Ho Chi Minh City, Vietnam
[4] Faculty of Applied Technology, Van Lang School of Technology, Van Lang University, Ho Chi Minh City, Vietnam
* Correspondence: t.ganat@squ.edu.om (T.G.); nguyenthihai@vlu.edu.vn (H.T.N.)

Abstract

Reliable production forecasting is a critical task for evaluating asset valuation and commercial performance in oil and gas reservoirs. Conventional short-term forecasting methods, such as Arps' decline curve analysis, rely on simple mathematical curve fitting and often oversimplify reservoir performance. On the other hand, long-term forecasting requires complex multidisciplinary models that integrate geophysics, reservoir engineering, and production engineering, but these approaches are time-consuming and have high turnaround times. To bridge the gap between long and short-term production forecasts, reduced-physics models such as Blasingame type curves have been developed, incorporating transient well behaviour derived from diffusivity equations and Darcy's law. These models assume homogeneity and uniform reservoir properties, enabling faster results while honouring pressure performance. However, despite their efficiency, they still face limitations in reliability, particularly when extended to long-term forecasts. This paper proposes a hybrid modelling approach that integrates flowing material balance (FMB) concepts into physics-informed neural networks (PiNNs) and machine learning models to improve the accuracy and reliability of production forecasting. The proposed methodology introduces two hybrid strategies: physics-informed models enriched with FMB feature, and PiNNs. The first proposed hybrid model uses a created FMB-derived feature as input to neural networks. The second PiNN model embeds data-driven loss functions with a physics-based envelope to reflect reservoir response into the machine learning model. The primary loss function is mean squared error, ensuring minimization of data misfit between predicted and observed production rates. The study validates both proposed physically informed neural network models through performance metrics such as RMSE, MAE, MAPE, and R^2. Results application on field data shows that the integration of FMB into neural network models using the PiNN concept guides the neural network models to predict the production rates with higher reliability over the full span of the tested data period, which was the last year of unseen production data. Additionally, the proposed PiNN model is able to predict the well productivity index via hyper-tuning of the PiNN model. Furthermore, the PiNN is not improving the metric performance of conventional neural networks, as it has to satisfy an additional material balance equation. This is due to a lower degree of freedom in the PiNN models.

Check for updates

Received: 7 March 2026
Revised: 13 April 2026
Accepted: 17 April 2026
Published: 22 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: physics-informed neural networks (PiNNs); flowing material balance (FMB); machine learning

1. Introduction

Reliable Production forecasting is a critical step in the evaluation of asset valuation and commercial performance of the well and reservoir [1]. The production forecasting is conventionally performed on short-term (less than two years) using basic well performance data, such as well production data; for long-term production forecasting, comprehensive well and reservoir models are used. The simplest form of short-term production forecast was introduced using only a mathematical curve fitting approach [2], which oversimplifies the nature of oil or gas production rate to a decline function. On the other hand, the full physics that is used in long-term production forecast slows down the forecasting process due to complex multi-disciplinary involvement, including geophysics, geoscience, reservoir engineering, production engineering, and facilities engineering. Attempts are performed to fill the gap in the reliability of short-term approaches by [3] to bring more physics into short-term production forecasting and make it reliable through rate decline curve analysis. These models and workflows are data-driven and are physics-based models. For instance, Blasingame type curves are outlined from the transient behaviour of wells derived from the diffusivity equation, arising from conservation of mass, fluid properties, and principal fluid flow in porous media known as Darcy's Law. The governing equation is derived analytically and assumes the reservoir is a single cell tank with uniform properties. The assumption of homogeneity in analytical models is reduced; physics models deliver fast results for the DCA level while capturing more details and honouring the pressure performance of the well.

The Blasingame type curves can be applied to one well or a group of wells [4,5]. An alternate approach, introduced as Capacitance Resistance Model (CRM), is an efficient approach to address the multi-well production forecasting while honouring material balance, and reservoir heterogeneity between the wells, yet in an analytical way and without three-dimensional geological gridding [6]. The CRM models implement flowing pressure behaviour and vary in their derivation; however, the variation is considered negligible and dismissed from the calculation, which is acceptable in some cases.

CRM enhances the production forecasting from a short-term to be reliable for the longer term by bringing the flow mobility among multi-wells into consideration, and it bridges the gap between SPTF and LPTF. Similar integrated approaches have been utilized in analytical models to cater for pressure inclusion in the production forecasting [7].

Recent advancements in machine learning have provided another approach for production forecasting with a data-driven engine that utilizes all of the available features such as production rates, flowing pressure, choke size and other operating well conditions [8,9]. Different engines, such as neural networks [10] and Long-Term Short Memory (LTSM) models [11], are used to build models that can provide efficient model training and verification that can be used for production forecasting, production optimization with a wide range of applicability in conventional and unconventional reservoirs [12,13]. However, these models also have several challenges, such as data availability long enough to build the model confidence and also difficulty in interpreting them physically as they are physically unaware and they are black boxes in nature [14,15].

The challenge of model explainability of machine learning models and black box models brought the physics-informed models that are called hybrid models, which guide the machine learning models to follow physical principles through additional inputs,

feature inclusion [16,17], or Physically Informed Neural Networks (PiNN) models [18]. PiNN models are combining measured data, such as well production rate, well-head pressure, choke size, etc., with the governing physical equation, such as the diffusivity or material balance equation, as a physics principle into a single machine learning model, and the model will be trained to history match the data while honouring the main governing physical model. This advancement of hybrid models allows the model to be trained with higher confidence and avoid the risk of overfitting while it transforms the machine learning model into a physically explainable model. Table 1 summarizes the advantages and limitations of existing models and their evolution on production forecasting, starting from simple mathematical models such as DCA to the most recent hybrid models.

Table 1. Summary of advantages and limitations of existing approaches and methods for production forecasting.

Existing Methods	Applications	Advantages	Limitations
Decline Curve Analysis (DCA)	Short-term production forecasting	Fast model	Does not reflect the actual fluid flow in reservoirs
Full-physics simulation Dynamic Models	Short-term and Long-term production forecasting	Consideration of complex physics	• High computational cost • Lengthy forecasting process
Reduced physics model: • Rate Transient Analysis (RTA) • Capacitance Resistance Model (CRM)	Short-term production forecasting Reservoir connectivity	• Computational efficiency • Considers the flow behaviour and material balance	Over-simplification of fluid flow and a lack of integration of operating conditions in the workflow
Pure machine learning	Short-term production forecasting	Computational efficiency	• Large amount of data is required • No model explainability
Hybrid Models • Physics informed models (Features Engineering) • Physically-informed Neural Networks (PiNN)	Short-term production forecasting to long-term production forecasting	• Reliable Models • Computational efficiency • Model interpretability and explainability	

The aim of this paper is to propose a general approach of a hybrid model that fills the gap of the model's reliability in production forecasting by incorporating the flowing material balance concept in different models as a subset of PiNN and physically informed machine learning models. In the following sections of this paper, the methodology is presented for the proposed hybrid model. Next, the model verification is presented, and it is applied to field cases.

2. Methodology

Two different hybrid modelling strategies are proposed: a physics-informed model, with a flowing material balance feature, and a PiNN combined with a physically informed machine learning model. In this section, the problem of interest is fully stated and discussed. Next, the two approaches in machine learning (ML) models, including pure ML models and hybrid ML models, are discussed. Finally, the proposed hybrid model by the authors is described.

2.1. Problem Statement and Assumptions

The main area of interest is predicting the production rate using the available historical data that has enough reliability to be considered as a long-term production forecast. The producing wells have three-phase fluid production, including oil, gas and water production rates as main features, as well as multiphase production rate. Additional features, such as well head pressure and choke size, are key features that reflect the well operating conditions. Additionally, in the case that direct measurement of flowing bottom-hole pressure exists, it is used as a feature; otherwise, well-head pressure is transformed to flowing bottom-hole pressure using well models to have the sub-surface well control points.

Two proposed hybrid machine learning models are developed to correlate the production rate to all features using the historical well measurements. These hybrid models are trained to learn the well performance to estimate the well production rate and predict reservoir pressure. The ability to predict reservoir pressure from production data allows us to validate the machine learning models and enhance their reliability using the introduced physics-informed features. This allows for assessing the reservoir connectivity across multiple wells by having reservoir pressure to reservoir performance optimization and improving reservoir management. As the hybrid machine learning models are built based on historical well data, numerical or analytical models are not required to be history-matched to be utilized as model verification. This enforces the requirement for data measurement sufficiency to build the models.

2.2. Flowing Material Balance (FMB)

A flowing material balance is developed based on the simple concept that the shut-in reservoir pressure can be estimated using the flowing pressure by embedding the well productivity index into the material balance equation [19]. The flowing material balance (FMB) is used mostly to avoid the shut-in of wells to minimize the well downtime, as the wells are the cash generators for assets. Additionally, it is quick to run. The main governing equation for FMB is derived from the mass conservation principle. The mass conservation principle can be integrated into the total fluid compressibility equation:

$$C_t = \frac{-1}{v}\frac{\partial v}{\partial p} \tag{1}$$

where c_t is the total compressibility, v is the reservoir pore volume, and P is the average reservoir pressure. Equation can be integrated with the production rate definition of $q = \frac{\partial v}{\partial t}$ and can be rewritten as follows:

$$C_t = \frac{-1}{v}\frac{\partial v}{\partial p} = \frac{-1}{v}\frac{\partial v}{\partial t}\frac{\partial t}{\partial p} = \frac{-q}{v}\frac{\partial t}{\partial p} \tag{2}$$

The pressure gradient in Equation (2) can be expanded as follows:

$$\frac{\partial t}{\partial p} = \frac{0 - t}{(P_R - P_i)} = \frac{t}{(P_i - P_R)} \tag{3}$$

In which t is the time of production of the well, P_i is the initial reservoir pressure, and P_R is the reservoir pressure at time t.

The total production rate can be estimated using the well productivity index as follows:

$$q = J * \left(P_R - P_{wf}\right) \tag{4}$$

where J is the well productivity index, and P_{wf} is the flowing bottom hole pressure. Using Equations (3) and (4), and solving for P_Rt, this can be written as follows:

$$P_R = \frac{\tau P_i + tP_{wf}}{\tau + t} \tag{5}$$

In which τ is defined as follows:

$$\tau = \frac{C_t \cdot v}{J} \tag{6}$$

The aim of developing Equation (5) is to estimate the reservoir pressure at any time using flowing pressure data. This derived reservoir pressure from the flowing material balance will bring the physical principles and boundaries into the machine learning models.

2.3. Original Machine Learning Models for Production Forecasting

Machine learning models have been widely used to build models with the ability of production forecasting using production data [20]. Machine learning models that only use historical data as input data and production rate as target are considered Original or Pure machine learning models [18]. These models can be illustrated as a simple workflow without the inclusion of any additional features other than the historical data. A schematic of original machine learning models with a neural network model is depicted in Figure 1. The model inputs are usually time series data, such as production rate and flowing pressure data, and certainly time, with a target value of production rate; as defined by the target, the output of these models is production rate. The majority of proxy well models that use machine learning models are constructed by original machine learning models [21].

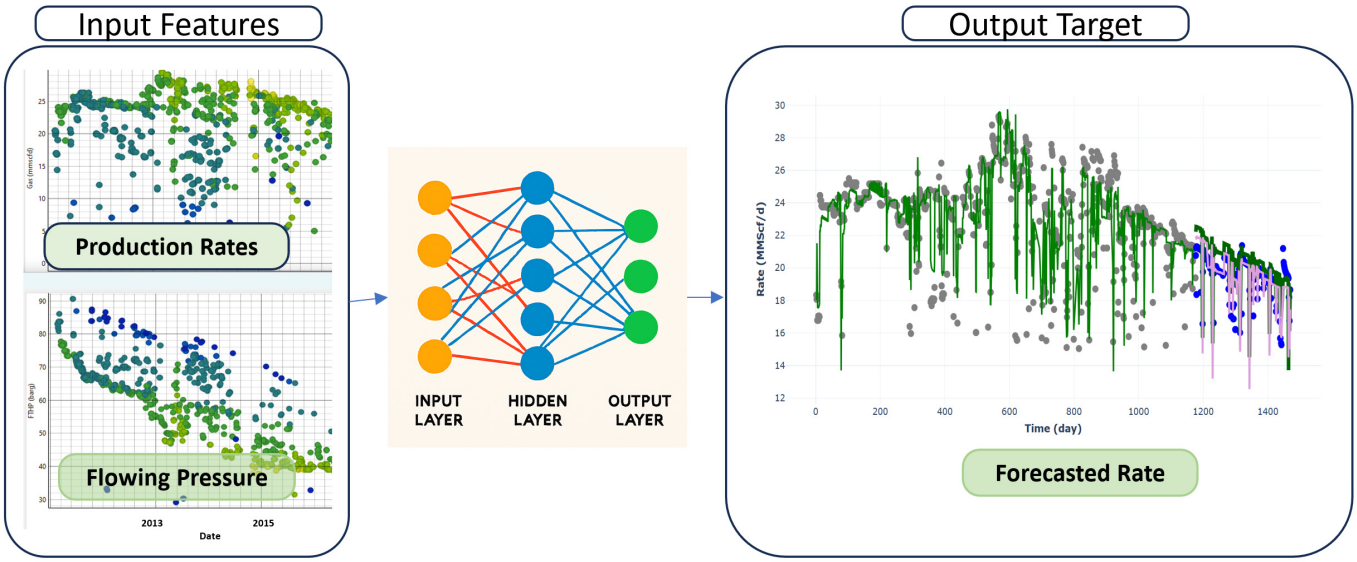


Figure 1. Schematic of original machine learning model using neural networks. The output target is production rate with green lines of the trained model, purple line the forecast of the model for test data while gray dots are the measured data.

2.4. Hybrid Model Types-1: Feature Creation for Physics-Informed Models

To inform the neural network model about the flowing material balance principle, a new feature is developed based on Equation (5) and fed into the neural network as an input feature. The new feature in this model enriches the neural network model, as shown in Figure 2. This model is using the analytical model based on flowing material balance, which in essence is the reduced physics model to obtain the production forecast. The machine learning model forecasts the production rate based on the physically derived input feature, reservoir pressure and original features such as time, flowing pressure and production rate.

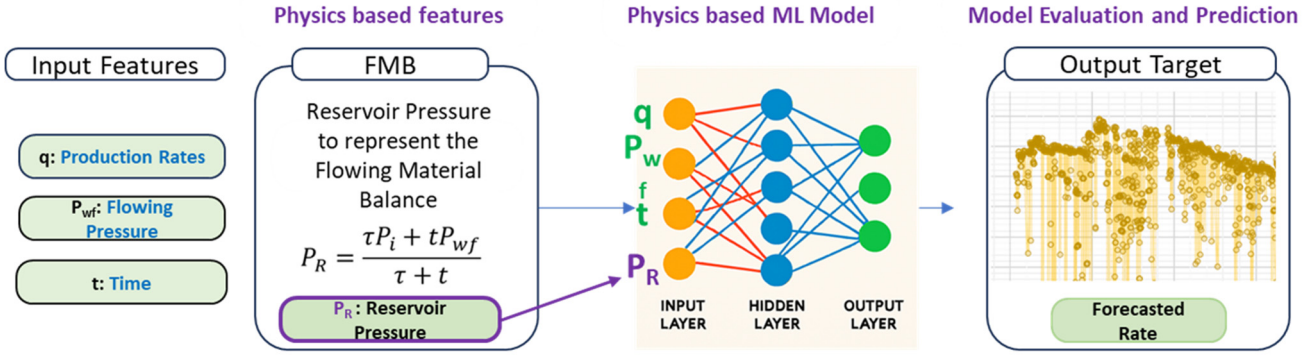


Figure 2. Framework of the physics-informed feature engineering for production rate estimation.

2.5. Hybrid Model Type-2: PiNN Model with Feature Creation and Loss Function

In this section, the PINN framework is developed. The PINN uses a physics-driven term in the loss function of the neural network model. In this application, Equation (4) is used to achieve a material balance check against features and targets. The target feature is rate, q, while P_R, and P_{wf} are input features. The governing loss function can be written as:

$$FMB\ loss = \ L_{FMB} \ = q - J * \left(P_R - P_{wf}\right) \tag{7}$$

In which L_{FMB} is the flowing material balance loss. The neural network model itself has a loss function called $L_{NN\ model}$ due to model misfit on the predicted results. The PiNN model combines both loss functions to reduce the satisfaction of the machine learning model while satisfying the physical model and flowing material time at the same time, using the following:

$$L_{PiNN} = a\ .L_{NN\ Model} + (1 - a)L_{FMB} \tag{8}$$

where the $L_{NN\ Model}$ represents the original NN model misfit and L_{FMB} is representing the FMB misfit level. Standard data misfit errors, such as RMSE, MAE, MAPE, . . . can be used as model performance metrics. This additional check keeps the model informed on the extra deviation and keeps the machine learning model within a physical forecast and a more reliable forecast. The weighting factor of a is used in Equation 8 as an impact of the loss introduced by FMB. It can be any number between 0 and 1. Using $a = 1$ makes the model simplified to a neural network, while any lower number gives a weight to FMB.

The total loss can be subject to weighting in many solutions and can be used as a hyper-tuning parameter for the model [22].

The estimated value of total loss will be the criterion for model performance. If the model is not optimized, it will be imposed on a new learner setup using gradient-based optimizers. The new model will be created based on the optimizer engines (such as Adam and LBFGS). Figure 3 shows the framework of the methodology to build the hybrid model with PiNN engine on top of additional feature creation (reservoir pressure).

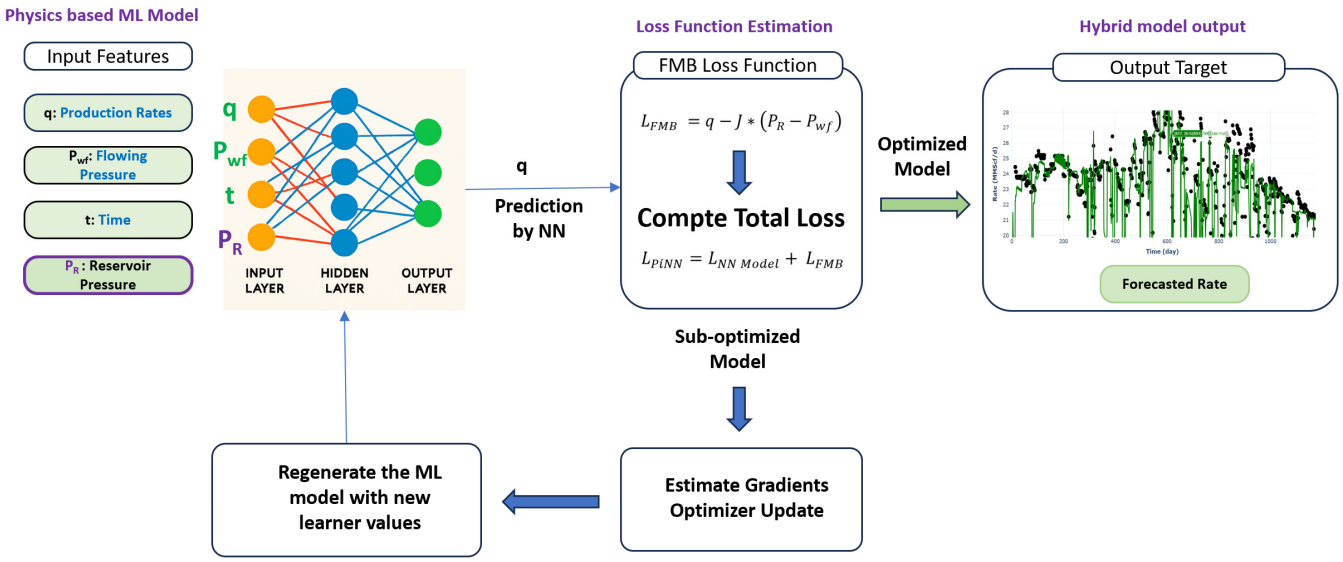


Figure 3. Framework of PiNN model with feature creation and loss function.

L_{FMB} is expected to be minimized while building the machine learning model; however, due to a lack of model awareness on a machine learning model on the well productivity index (J value) in Equation (7), the model needs hyper-tuning to find the right value of J that can represent the best well potential for the modelled period.

3. Model Validation

The prediction performance of two models, Type-1 and Type-2 models, is compared against the original machine learning model:

- Physics-based feature data preprocessing based on FMB (Reservoir Pressure);
- Physics-informed neural network (PINN) with an FMB-driven feature (Reservoir Pressure).

For applying the above approaches, an open-source deep learning library, TensorFlow version 2.21.0, is used [23]. Due to the high efficiency of the models, all calculations were implemented with an Intel(R) Core(TM) CPU i7-14700F @ 2.10 GHz (Intel Corporation, Santa Clara, CA, USA) using eight cores.

3.1. Model Description

In this study, a feed-forward neural network is used in TensorFlow for production rate forecasting. The model architecture is defined by the SimpleNN class, which is fed with four input features, including time, rate, flowing bottom-hole pressure, and estimated reservoir pressure from the flowing material balance. It propagates them through two hidden layers with 32 neurons each, initiated by hyperbolic tangent functions, before producing a single output. The setup of hidden layers is with default values and can be sensitized to enhance the model performance computationally or to achieve better model quality. The optimization process uses the Adam algorithm as a default approach, which is available in the TensorFlow library for its adaptive learning rate and robustness in handling noisy gradients, such as production rates. The primary loss function for neural networks is the mean squared error (MSE), which calculates the difference between predicted and observed production rates, ensuring the model learns to minimize data misfit over consecutive epochs.

A function is defined to distinguish the models Type-1 (Feature Engineering only) and Type-2 (PiNN loss function and Feature Engineering). The function is executed with the use_PiNN flag enabled as a Type-2 model, which includes an additional physics-based

residual term incorporated into the loss function. This term enforces consistency with governing physical equations by penalizing deviations between the network's predictions and the expected pressure–rate relationships (Equation (5)), scaled by the hyperparameter lambda_phys. The total loss allows the model to balance NN model accuracy with the FMB check in place. Within 50,000 training iterations, the optimizer iteratively updates the network parameters to minimize this total composite loss, yielding a physics-informed neural network that integrates observational data FMB for improved model reliability.

3.2. Model Training

Firstly, the database has been loaded with about five years of data at a daily frequency. The dataset is filtered to remove unreliable measurements, such as outliers, either high rates or low rates, then split sequentially, with 80% of the rows used for training and the remaining 20% reserved for testing. This reflects that the last production year is not used in the training process, and is held for testing and validation later. The input features include time, flowing pressure, reservoir pressure, and choke size, which are converted into tensors for use in TensorFlow. The target variable is the production rate, which the network aims to predict.

The input features are normalized and scaled based on the maximum value of each feature. For instance, rates are normalized by dividing the rate by 28 (maximum measured rate in MMScf/d) and a new feature called q_n to represent the normalized rate. The same has been performed for all other features. The data list is presented in Table 2. Data refers to a gas well with more than 1400 days of production (t is a time feature with units in days) with daily frequency.

Table 2. Summary of input parameters in the machine learning model.

Parameter	**Original Features**					**Normalized Features**			
index	t	P_{wf}	P_r	q	CS	t_n	q_n	P_{wfn}	P_{rn}
count	1468	1468	1468	1468	1468	1468	1468	1468	1468
mean	734	773	1148	21	0.308	0.37	0.77	0.43	0.64
std	424	192	218	4	0.084	0.21	0.16	0.11	0.12
min	0	461	838	0	0.100	0.00	0.01	0.26	0.47
25%	367	573	917	20	0.250	0.18	0.70	0.32	0.51
50%	734	742	1156	23	0.280	0.37	0.81	0.41	0.64
75%	1100	966	1356	24	0.388	0.55	0.87	0.54	0.77
max	1469	1267	1515	29	0.700	0.73	1.05	0.70	0.90

The constructed model is a standard feed-forward neural network with several hidden layers and nonlinear activation functions. Adam optimizer is chosen to perform training with a small learning rate to ensure stable convergence. The loss function is defined to be the mean squared error, which measures the difference between predicted flow rates from the model and actual production rates. Over 5000 iterations, the optimizer updates the network weights to minimize this error. RMSE, MAE, MAPE, and R^2 metric performance are used to address the model quality and reliability.

To improve model reliability, the process is extended to a physics-informed neural network (PINN) by defining an additional physics-based loss alongside the data loss with the aim of making the predictions more interpretable.

The J value on the loss function is not known, so the model is run against different values of J to hyper-tune the models. A range of 1 to 50 is used to find the closest possible value on both models. The Type-1 model, which uses physics-based feature data preprocessing based on FMB, is called the NN model, while the other model is called the PINN model for simplicity to evaluate the training quality. The results of hyper-tuning are

presented in Table 3. As shown, the PINN with a J value of 15 shows the best performance for the PINN model, which translates to the best model performance metrics based on all parameters, such as an RMSE of 0.05, MAE = 0.03, MAPE = 4.3, and R^2 = 0.68. Individual metrics could be better in some cases; however, considering all performance metrics, J = 40 gives the best training performance.

Table 3. Performance metrics for first batch of PiNN model hyper-tuning.

Model Type	J-Value	RMSE	MAE	MAPE	R^2
PiNN	5	0.088136	0.081588	9.753807	0.091295
PiNN	10	0.052049	0.034876	4.355432	0.683092
PiNN	15	0.077629	0.058531	7.234559	0.295042

The model prediction quality is shown in Figure 4. As shown, the model's ability to reproduce the production for the training dataset is becoming closer to the J value of 15. The higher values of J and lower values affect the model quality.

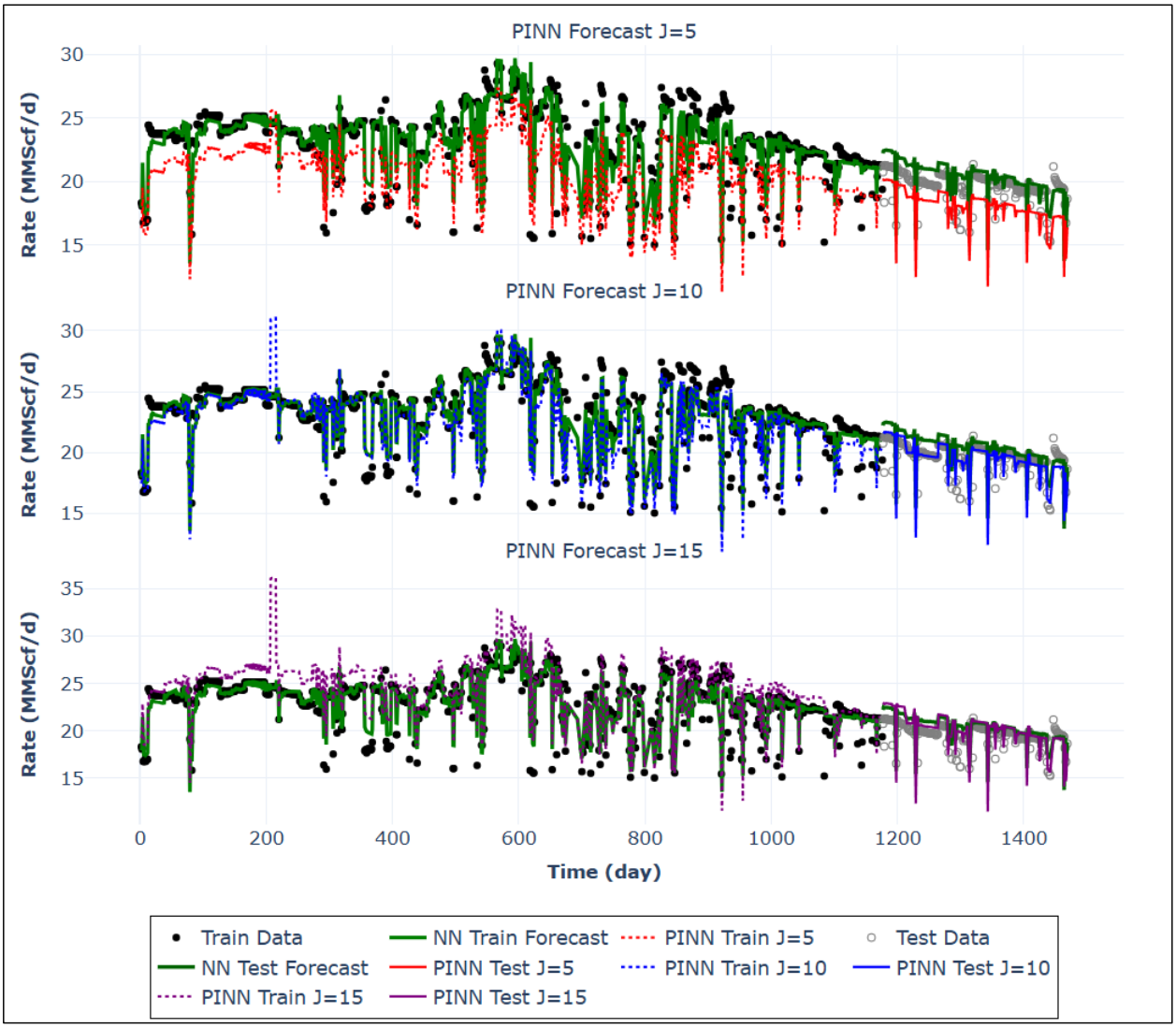


Figure 4. Training quality check for refined hyper-tuning of well productivity index, J.

A better refinement on the J value is performed to find a better estimate of the well productivity index using PiNN, and also to enhance the PiNN model reliability. As shown

in Table 4, the refined ranges of J value from 5 to 17 are run, and the best model performance metrics are achieved at a J of 11, which is highlighted in blue.

Table 4. Performance metrics for refined PiNN model hyper-tuning.

Model Type	J	RMSE	MAE	MAPE	R^2
PiNN	5	0.084995	0.080376	9.971696	0.283387
PiNN	7	0.066183	0.059666	7.474061	0.565499
PiNN	9	0.052666	0.04091	5.231935	0.724849
PiNN	**11**	**0.048494**	**0.031626**	**4.102491**	**0.766722**
PiNN	13	0.055016	0.03551	4.49736	0.699755
PiNN	15	0.07	0.048464	6.0046	0.513928
PiNN	17	0.089908	0.067808	8.361886	0.198136

As a cross-validation of the workflow, another production dataset from the second well in the same field is tested. The same model is built for another well to ensure the model's robustness. The resulting model performance metrics are presented in Table 5, and the prediction quality is visualized in Figure 5. As shown, the model has better prediction quality with PiNN, with a zoomed view of the testing section, as shown in Figure 6.

The nature of the well decline imposes a variation in J value. As shown in Figure 7, the well PI reduces and, therefore, the J value becomes a variable to sensitize and tune for the best well model performance.

Table 5. Performance metrics for refined PiNN model hyper-tuning for second well as a validation check.

Model Type	J	RMSE	MAE	MAPE	R^2
NN_Train	4.5	0.008004	0.005805	1.081721	0.997134
PINN Train	4.5	0.030975	0.02696	4.469941	0.957084
NN_Test	**4.5**	**0.008871**	**0.005337**	**1.268326**	**0.876655**
PINN Test	**4.5**	0.010079	0.008693	2.028308	0.910778

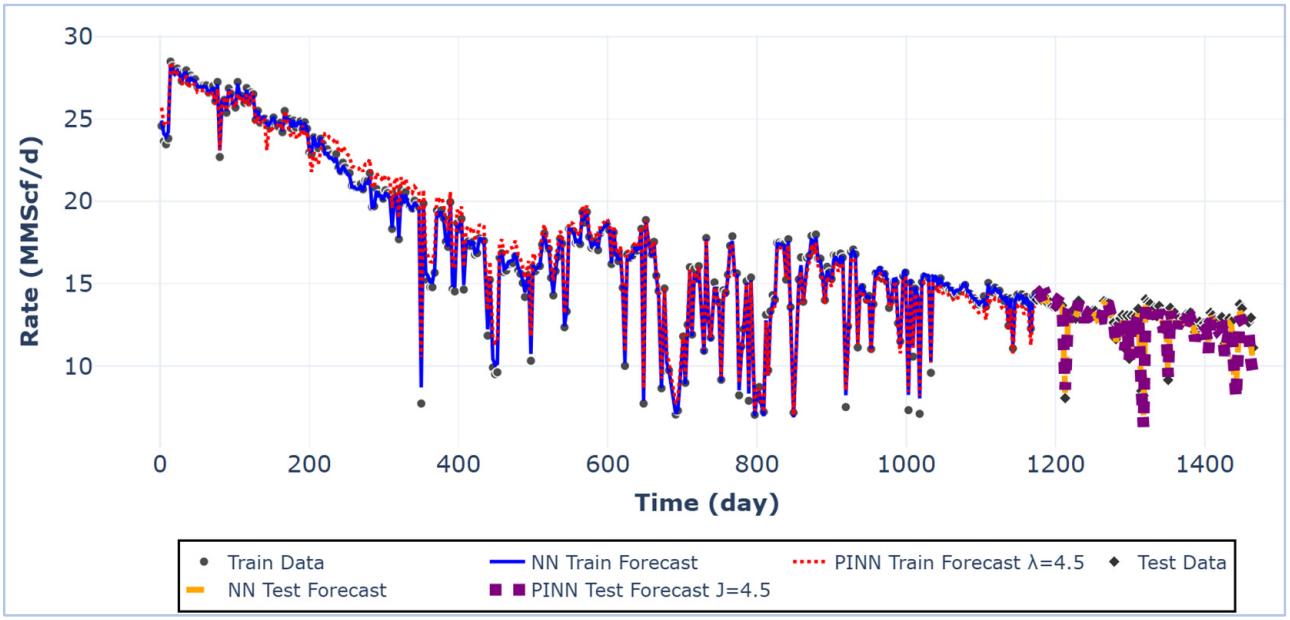


Figure 5. Model performance for second well to test the model robustness and validity.

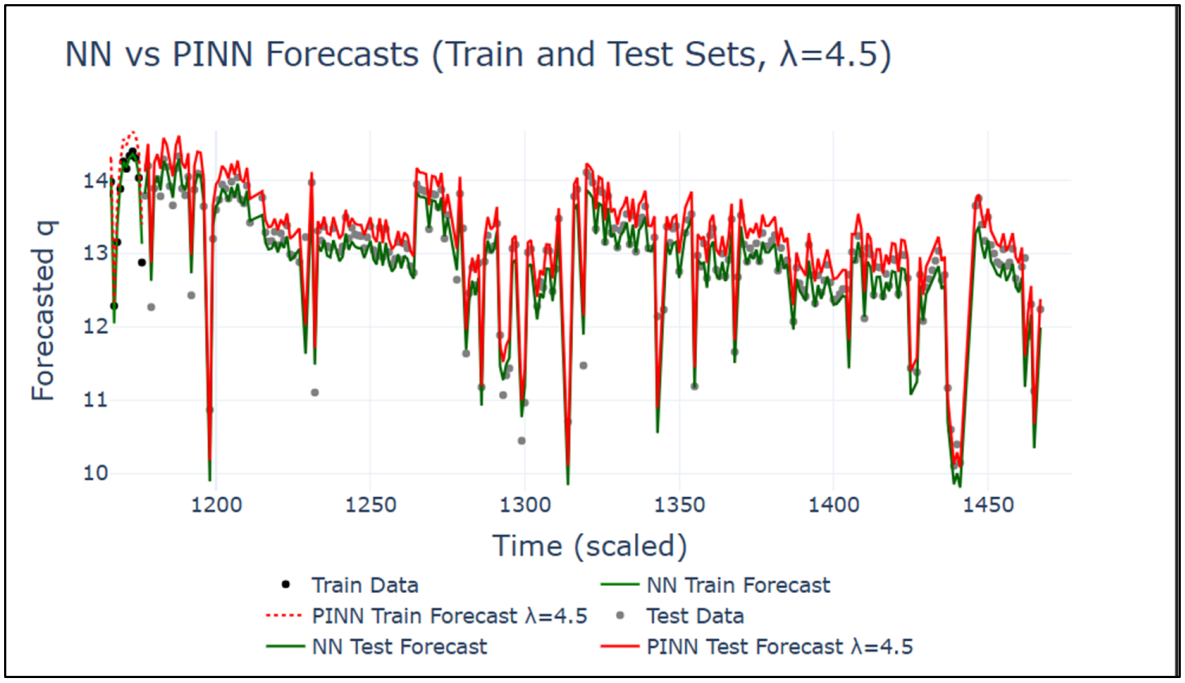


Figure 6. Model prediction zoomed for the tested section. PiNN show better prediction quality.

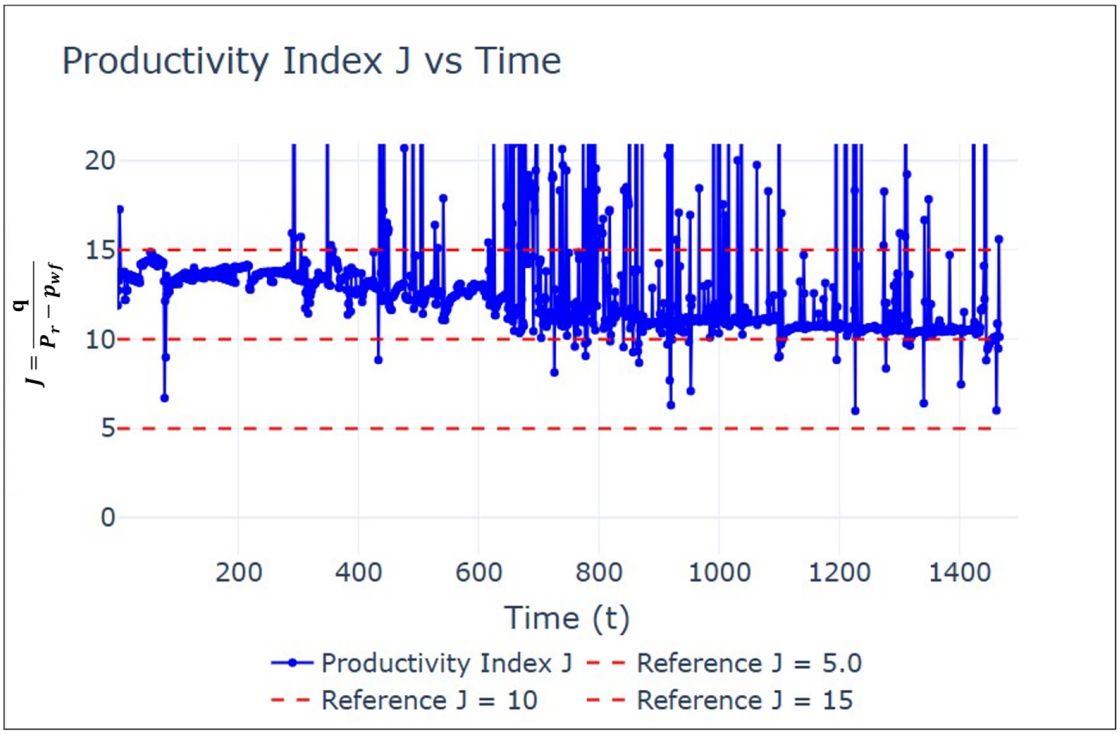


Figure 7. Well PI variable analysis.

This cross-check can be performed once the reservoir pressure is calculated as a new feature.

3.3. Prediction Result

Figure 8 shows the prediction performance of the proposed PINN models for the last year's production, which was not provided to the model. The model has been run with the refined J value around 11, with 3 values as per the hyper-tuned model from the training exercise. A narrow range of 10 to 12 is considered for prediction as a confidence range. As

shown, the prediction from the PiNN matched reasonably well for the whole blind test section, which is reflected in red, purple, and orange lines for all J values. However, still, the best match comes from the J value, which is reflected in Figure 8 as a zoomed window to focus on the prediction part of the model.

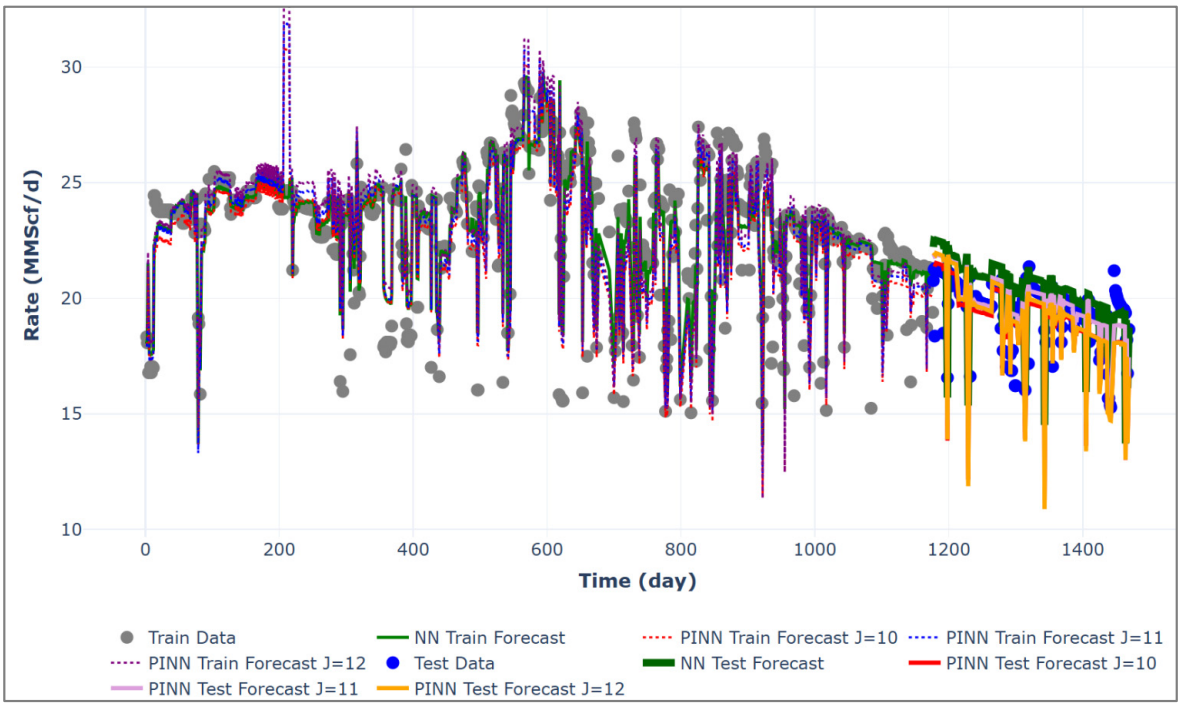


Figure 8. Model Prediction quality check for the tested period of last 20% of data or last year's prediction.

The Type-1 model, which has additional feature engineering as reservoir pressure derived from the material balance equation estimated, is also built the same way; the results of the model for the training section are better than the PiNN, as presented in Table 6. The physically informed model, the Type-1 model, is used to predict the production rate for the same test period and is visualized as a green curve in Figure 9. As shown for some periods of time, it overestimates the rate and goes beyond the acceptable range of 5% error.

Despite better performance metrics, the PiNN has more reliability over the testing period to predict more accurately and maintains consistency for the whole test period, from the beginning of the forecast to the end of the forecast. In other words, despite a lower R^2, the model is consistent to keep the prediction of the rate more close to the measured data, while NN model tends to be very accurate in some part of the prediction while is off at some other section of the production forecast This is due to nature of the NN models which have more reliability within seen data or called over-fitting. The physics-based model, PiNN, keeps the model more physical even for unseen conditions, such as a lower reservoir pressure, a lower flowing pressure, or a higher choke size.

Table 6. Performance metrics of physically informed model by feature engineering vs. physically informed loss function driven model (PiNN).

Model Type	RMSE	MAE	MAPE	R^2
Feature Engineering	0.041101	0.024725	3.216178	0.802383
PiNN	0.048494	0.031626	4.102491	0.766722

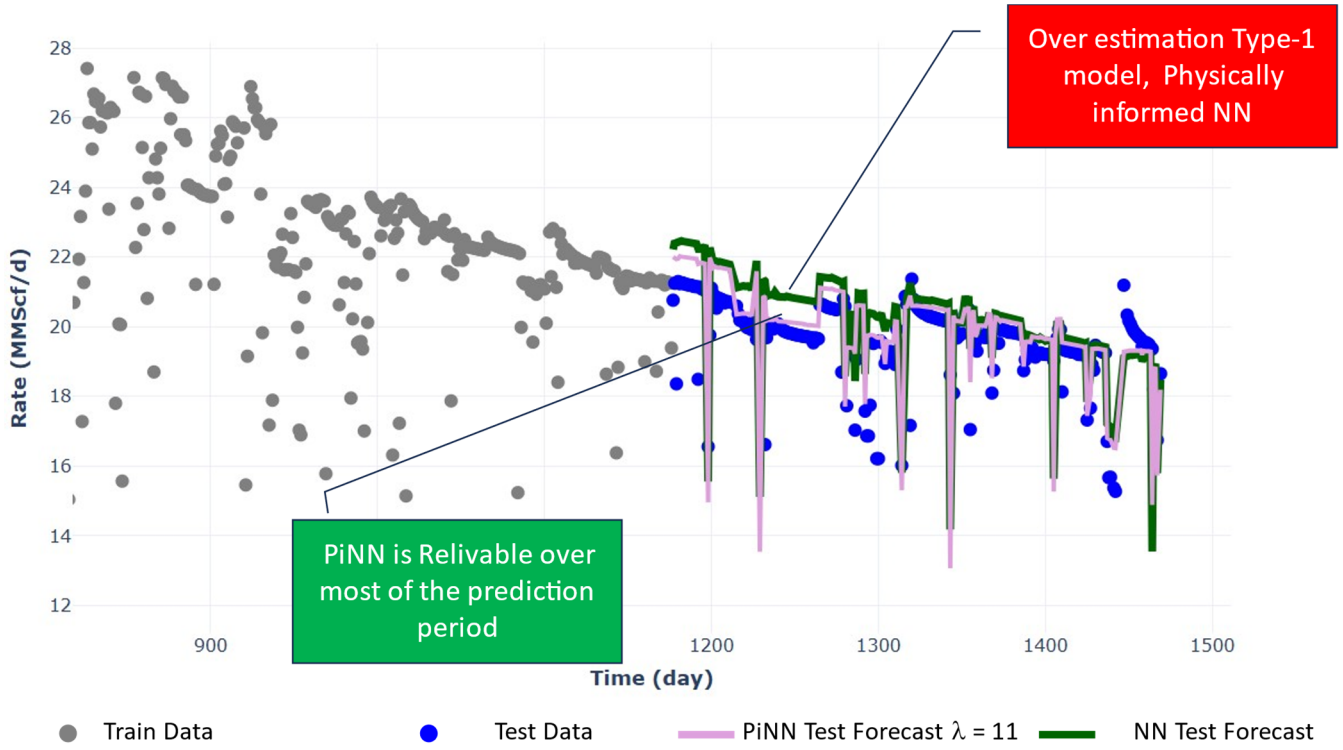


Figure 9. Model Prediction quality check for the last 20% of data, at last year's prediction.

In terms of quantitative performance, the pure FMB method achieved moderate accuracy with RMSE around 0.065, MAE near 0.045, and a MAPE of about 6.8%, corresponding to an R^2 of 0.62. The feature-engineered NN model improved these metrics significantly, reaching RMSE ≈ 0.041, MAE ≈ 0.025, MAPE ≈ 3.2%, and R^2 ≈ 0.80. The PiNN–FMB hybrid model delivered RMSE ≈ 0.048, MAE ≈ 0.032, MAPE ≈ 4.1%, and R^2 ≈ 0.77. Although the PiNN metrics are slightly less favourable than the feature-engineered NN, they outperform pure FMB and, more importantly, maintain prediction consistency across unseen data in the test series.

For completeness, the hybrid PiNN–FMB workflow was benchmarked against the pure FMB analytical method, as shown earlier in Figure 9. The pure FMB approach, which estimates reservoir pressure directly from flowing pressure and productivity index, provides rapid forecasts and avoids shut-in requirements. However, its predictive capability is limited when operating conditions vary, such as when the choke varies or multiphase flow interactions happen. In contrast, the PiNN–FMB hybrid consistently reproduced production trends across the blind test period, maintaining forecast stability even under unseen operating conditions with the aid of the governing material balance. While the pure FMB method delivered fast analytical estimates, the hybrid PiNN–FMB approach demonstrated superior reliability by embedding physical constraints into the learning process, thereby reducing divergence and improving robustness of long-term forecasts.

4. Conclusions

The integration of Flowing Material Balance is introduced into neural network models using the PiNN concept to guide the neural network models to predict the production rates within the physically possible and avoid over-estimation or under-estimation beyond the well's technical potential. The proposed PiNN model is able to predict the well productivity index via hyper-tuning of the PiNN model. It is able to take the value directly as well from the input data. However, the well productivity index is estimated with an approach to be used externally for well performance modelling.

The PiNN is not outperforming the conventional neural networks, as it has to satisfy the additional material balance equation. This is due to a lower degree of freedom in the PiNN models.

Multi-well analysis might be used to address the reservoir connectivity as future work through a material balance check across multiple wells. Well PI tracking can be performed by implementing multi-section neural network integration in PiNN to track the well productivity change (J-value) at different time windows. In future work, this will bring insight into well performance evaluation at different times of production.

An important outcome of this study is that the insertion of FMB constraints within the PiNN framework acts as a physical regulator. By penalizing the model that fits the training data but violates reservoir fluid material balance, the PiNN maintains generalization and avoids the overfitting commonly observed in conventional neural networks through the loss function. This double guidance ensures that production forecasts remain both statistically accurate and physically meaningful. In practice, this principle strengthens production matching, as the PiNN can reproduce historical trends while simultaneously validating reservoir pressure performance, providing a more reliable basis for reservoir management decisions. Beyond the immediate forecasting integrity, this approach also opens the door to practical extensions such as multi-well connectivity analysis, dynamic productivity index tracking, and integration with reservoir surveillance workflows. These applications highlight the broader utility of PiNN–FMB models, not only as forecasting tools but as frameworks for embedding physics into machine learning to support long-term reservoir management strategies.

In reservoir engineering, two principal categories of PiNN are typically recognized: the diffusivity equation in Partial differential equation (PDE) form governing fluid flow in porous media, and the material balance equation. Conventional material balance approaches rely on static pressure or build-up data; in contrast, this study introduces the utilization of flowing data without the need for well shut-in building PiNN models. This enables the construction of more reliable machine learning models that remain applicable under continuous production conditions, and enhances their utility for long-term forecasting.

Author Contributions: Methodology, E.M. and T.G.; Software, H.T.N.; Validation, E.M., T.G. and H.T.N.; Formal analysis, H.T.N.; Investigation, T.G.; Resources, H.T.N.; Data curation, E.M. and H.T.N.; Writing—original draft, T.G.; Writing—review & editing, T.G.; Visualization, E.M. and H.T.N.; Supervision, T.G.; Project administration, T.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: Author Eghbal Motaei was employed by the company Petronas Carigali Sdn. Bhd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Le Quang, D.; Ngoc, D.H.; Rangponsumrit, M.; Supalasate, P.; Ngo, K.D.; Nguyen, D.H.; Tran, M.D.; Ho, V.T.; Van, C.L.; Ali, S.; et al. Production Optimisation and Surface Network Debottlenecking Using an Integrated Asset Model for Vietnam Offshore Field Development. In Proceedings of the International Petroleum Technology Conference, Bangkok, Thailand, 1–3 March 2023; p. D022S024R001. [CrossRef]
2. Arps, J. Analysis of Decline Curves. *Trans. AIME* **1945**, *160*, 228–247. [CrossRef]

3. Blasingame, T.A.; Lee, W.J. Properties of Homogeneous Reservoirs, Naturally Fractured Reservoirs, and Hydraulically Fractured Reservoirs from Decline Curve Analysis. In Proceedings of the Permian Basin Oil and Gas Recovery Conference, Midland, TX, USA, 13–15 March 1986; p. SPE–15018–MS. [CrossRef]
4. Eghbal, M.; Abdolrahim, A.; Ebrahim, Y.M. Production Analysis PA and Application of Rate Transient Analysis RTA in Carbonate Oil Reservoirs for Reservoir Characterization and Connectivity. In Proceedings of the Offshore Technology Conference, Kuala Lumpur, Malaysia, 20–23 March 2018. [CrossRef]
5. Ataei, A.; Motaei, E.; Yazdi, M.E.; Masoudi, R.; Bashir, A. Rate Transient Analysis RTA and Its Application for Well Connectivity Analysis: An Integrated Production Driven Reservoir Characterization and a Case Study. In Proceedings of the SPE Asia Pacific Oil and Gas Conference and Exhibition, Brisbane, Australia, 23–25 October 2018. [CrossRef]
6. Yousef, A.A.; Gentil, P. A Capacitance Model To Infer Interwell Connectivity From Production-and Injection-Rate Fluctuations. *SPE Reserv. Eval. Eng.* **2006**, *9*, 630–646. [CrossRef]
7. Ameer, A. Bridging Short and Long-Term Forecasting: Pressure-Based Modeling Unification for Integrated Production Forecasting in Kuwait Jurassic Gas Fields. In Proceedings of the Kuwait Oil & Gas Show, Kuwait City, Kuwait, 3–5 February 2026. [CrossRef]
8. Chahar, J.; Verma, J.; Vyas, D.; Goyal, M. Data-driven approach for hydrocarbon production forecasting using machine learning techniques. *J. Pet. Sci. Eng.* **2022**, *217*, 110757. [CrossRef]
9. Tong, S.; Wang, F.; Gao, H.; Zhu, W. A machine learning-based method for analyzing factors influencing production capacity and production forecasting in fractured tight oil reservoirs. *Int. J. Hydrogen Energy* **2024**, *70*, 136–145. [CrossRef]
10. Al-Shabandar, R.; Jaddoa, A.; Liatsis, P.; Hussain, A.J. A deep gated recurrent neural network for petroleum production forecasting. *Mach. Learn. Appl.* **2021**, *3*, 100013. [CrossRef]
11. Ning, Y.; Kazemi, H.; Tahmasebi, P. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Comput. Geosci.* **2022**, *164*, 105126. [CrossRef]
12. Fulford, D.S.; Bowie, B.; Berry, M.E.; Bowen, B.; Turk, D.W. Machine learning as a reliable technology for evaluating time/rate performance of unconventional wells. *SPE Econ. Manag.* **2016**, *8*, 23–39. [CrossRef]
13. Sarkar, P.; Yoon, S.; Kim, J.; Baek, S.; Sun, A.; Yoon, H. Surrogate models for development of unconventional shale reservoirs by an integrated numerical approach of hydraulic fracturing, flow and geomechanics, and machine learning. *Geomech. Energy Environ.* **2025**, *43*, 100691. [CrossRef]
14. Manta-Costa, A.; Araujo, S.O.; Peres, R.S.; Barata, J. Machine Learning Applications in Manufacturing—Challenges, Trends, and Future Directions. *IEEE Open J. Ind. Electron. Soc.* **2024**, *5*, 1085–1103. [CrossRef]
15. Rathore, M.M.; Shah, S.A.; Shukla, D.; Bentafat, E.; Bakiras, S. The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities. *IEEE Access* **2021**, *9*, 32030–32052. [CrossRef]
16. Yeomans, C.M.; Shail, R.K.; Grebby, S.; Nykänen, V.; Middleton, M.; Lusty, P.A.J. A machine learning approach to tungsten prospectivity modelling using knowledge-driven feature extraction and model confidence. *Geosci. Front.* **2020**, *11*, 2067–2081. [CrossRef]
17. Motaei, E.; Tabatabai, S.M.; Ganat, T.; Khanifar, A.; Dzaiy, S.; Chis, T. Physics-driven feature creation to improve machine learning models performance for oil production rate prediction. *Rom. J. Pet. Gas Technol.* **2024**, *5*, 291–306. [CrossRef]
18. Nagao, M.; Datta-Gupta, A. Physics-informed machine learning for identification of preferential flow paths and performance forecasting of subsurface carbon storage and utilization. *Geoenergy Sci. Eng.* **2026**, *259*, 214357. [CrossRef]
19. Han, G.; Liu, Y.; Sun, L.; Liu, M.; Gao, D.; Li, Q. Determination of pore compressibility and geological reserves using a new form of the flowing material balance method. *J. Pet. Sci. Eng.* **2019**, *172*, 1025–1033. [CrossRef]
20. Mask, G.M.; Wu, X.; Nicholson, C. Enhanced hydrocarbon production forecasting combining machine learning, transfer learning, and decline curve analysis. *Gas Sci. Eng.* **2025**, *134*, 205522. [CrossRef]
21. Kim, J.; Choe, J.; Lee, K. Sequential field development plan through robust optimization coupling with CNN and LSTM-based proxy models. *J. Pet. Sci. Eng.* **2022**, *209*, 109887. [CrossRef]
22. Escapil-Inchauspé, P.; Ruz, G.A. Hyper-parameter tuning of physics-informed neural networks: Application to Helmholtz problems. *Neurocomputing* **2023**, *561*, 126826. [CrossRef]
23. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. Large-Scale Machine Learning on Heterogeneous Systems. Available online: https://www.tensorflow.org/ (accessed on 1 March 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.