energies

MDPI

Article

Synergizing Machine Learning and Physical Models for Enhanced Gas Production Forecasting: A Comparative Study of Short- and Long-Term Feasibility

Bafren K. Raoof [1,*], Ali Rabia [2], Usama Alameedy [3], Pshtiwan Shakor [1,*] and Moses Karakouzian [4]

[1] Technical College of Engineering, Sulaimani Polytechnic University, Sulaymaniyah 46001, Iraq
[2] Wireline Logging Engineer, COSL Middle East Iraq Branch, Baghdad 10001, Iraq
[3] Petroleum Engineering Department, University of Baghdad, Baghdad 10001, Iraq
[4] Civil Engineering Department, University of Nevada, Las Vegas, NV 89154, USA
* Correspondence: bafren.kamil@spu.edu.iq (B.K.R.); pshtiwan.shakor@spu.edu.iq (P.S.)

Abstract: Advanced strategies for production forecasting, operational optimization, and decision-making enhancement have been employed through reservoir management and machine learning (ML) techniques. A hybrid model is established to predict future gas output in a gas reservoir through historical production data, including reservoir pressure, cumulative gas production, and cumulative water production for 67 months. The procedure starts with data preprocessing and applies seasonal exponential smoothing (SES) to capture seasonality and trends in production data, while an Artificial Neural Network (ANN) captures complicated spatiotemporal connections. The history replication in the models is quantified for accuracy through metric keys such as mean absolute error (MAE), root mean square error (RMSE), and R-squared. The future forecast is compared with an outcome of a previous physical model that integrates wells and reservoir properties to simulate gas production using regressions and forecasts based on empirical and theoretical relationships. Regression analysis ensures alignment between historical data and model predictions, forming a baseline for hybrid model performance evaluation. The results reveal the complementary attributes of these methodologies, providing insights into integrating data-driven and physics-based approaches for optimal reservoir management. The hybrid model captured the production rate conservatively with an extra margin of three years in favor of the physical model.

Keywords: reservoir simulation; material balance equation; IPR; VLP; water influx; history matching; predication; forecast; ML; ANN; SES

Academic Editor: Dameng Liu

Received: 28 December 2024
Revised: 8 February 2025
Accepted: 10 February 2025
Published: 28 February 2025

Citation: Raoof, B.K.; Rabia, A.; Alameedy, U.; Shakor, P.; Karakouzian, M. Synergizing Machine Learning and Physical Models for Enhanced Gas Production Forecasting: A Comparative Study of Short- and Long-Term Feasibility. *Energies* **2025**, *18*, 1187. https://doi.org/10.3390/en18051187

1. Introduction

Efficient reservoir management, operation strategy, and hydrocarbon recovery, particularly hydrocarbon forecasting, is essential for the petroleum industry. The evolution of the hydrocarbon forecasting first started with traditional reservoir-forecasting approaches like material balance equation and decline curve analysis, and both have been frequently employed in programs such as Mbal that provide a reliable farmwork for future hydrocarbon production by estimating the reservoir drive mechanisms, reservoir pressure, and production data. However, these approaches often have limitations in handling complex, non-linear reservoir behavior or when data are sparse or incomplete [1]. Following years of relying on numerical simulators, a new, intreseting era has been declared, linking the vacant points by transcending the limitations in the model. To leverage the strength of the

prediction and the accuracy of the forecast, machine learning algorithms and statistical methods have been employed in recent years to mitigate simplified aspects and bypass limitations in representing the non-linearity of physical models such as the Artificial Neural Network (ANN) and Long-Short Term Memory (LSTM). Machine learning and artificial intelligence should make field development faster and cheaper while keeping the production margins higher on a longer-term basis [2], as machine learning could be enforced to counteract various difficulties found in the oil and gas industries and help in maturing profitable strategies [3]. For instance, when it comes to predicting the future, such as the yearly gas supply, employing the ANN can effectively capture the trend [4]. Others are intreated by the same neural network to discern an interactive pattern and connection in the base by analyzing production and historical parameters [5]. A study, through employing the production history of 11 producer wells' and 3 injectors' data, predicted the production rate with an ARE of 12% using ANN [6]. The ability of ANNs to model gas-production dynamics has been effectively proven [7]. ANN can learn from data points without assuming a pre-determined model and subsequently predict a well's performance [8], and through experimenting on Multi-Linear Regression (MLR) and ANN, with the latter providing higher accuracy [9].

While machine learning-based models demonstrated strong predictive capabilities, they remained sensitive to data quality; thus, a hybrid model throughout a multi-layering system is an attractive predictive tool [10]. The introduction of a hybrid model combines Convolutional Networks (CNNs) and Long Short-Term Memory (LSTM) to predict production trends and achieve an accuracy of up to 73.63% for multiyear forecasting [11]. Multilayer predication studies utilize various constraints in the production system, such as targeting the cumulative gas production as a function of completion designs, depths, well spacing, orientations, and liquid yield. Other studies have explored the application of machine learning methods, such as Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM). An adaptive hyperband algorithm was employed to fine-tune the ANN model for hydraulic fracturing purposes with two outputs; the cumulative produced gas and the produced water resulted in reduced freshwater consumption during fracturing [12]. As mentioned previously, machine learning approaches are employed to predict different outputs in the petroleum industry. However, deep learning models may suffer significant errors when used for long-term forecasts, because the predictions are computed sequentially and depend on past predictions appended to the data. This causes a gradual accumulation of errors over time [13]. To reduce the margin of error, effective development and maintenance plans must be created using precise analysis and diagnosis of the targeted field [14]. Standalone AI models often lack physical interpretability, which can limit their application in reservoir forecasting; therefore, several investigations have explored integrating a physical model with machine learning methodologies. For instance, to optimize gas production, researchers have developed a multilayer regression model based on variables obtained from nodal analysis and sensitivity tests, providing a time-efficient analysis for the production system [15]. The decline curve analysis and reservoir properties served as a base for the hybrid model to leverage both machine learning and simulation data to predict the flow rate of a new well [16].

Unlike previous studies that focus solely on either material balance or nodal analysis as a benchmark to accurately represent fluid dynamics within the reservoir pore volume, the current research uniquely combines the strength of both approaches specifically when the hybrid model and seasonal smoothing are implemented. The constraints are overcome by the conventional regression and implication of the geological stock. These machine learning forecasts are directly compared to the forecasts from the physical simulator. The comparative approach is further enhanced by our demonstration between the physical

model and machine learning models, enabling a seamless exchange of information. Our paper evaluates the variability for both short- and long-term forecast feasibility, provides a comprehensive analysis, bridges the gap between traditional reservoir simulators and advanced data analysis, and unfolds the merit of AI and the effectiveness of a numerical simulator.

2. Materials and Methods

This study evaluates the current production facility and modifies the configurations to achieve stability in production naturally, along with future forecasting; the required data are gathered and prepared for the simulation process. Lift curves originated from the Prosper program after the assessment of the models, and the data were matched with available correlations to employ the one with the lowest degree of deviation. The process applies to tubing-fluid flow modeling for three wells (D7, D8, D9). The calculated results match the well tests. For modeling the reservoir, the MBAL program is used to create the structural and material balance model describing its operation. After setting the material balance model, the production history was run by matching based on the data recorded during production (cumulative gas produced and measured reservoir pressures). The importance of history matching is that the model data, calculated through simulation, fits the input data. This process helps track reservoir pressure decline and drive mechanisms over time, ensuring that the production forecast yields more realistic results. After elaborating the existing reservoir's model and its production wells, a production forecast can be run. During the production forecast, various modification parameters can be defined, which can influence the forecast result, such as the maximum and minimum production recovery values, periodicity or continuity of production, and other factors. The generalized steps are illustrated in Figure 1.

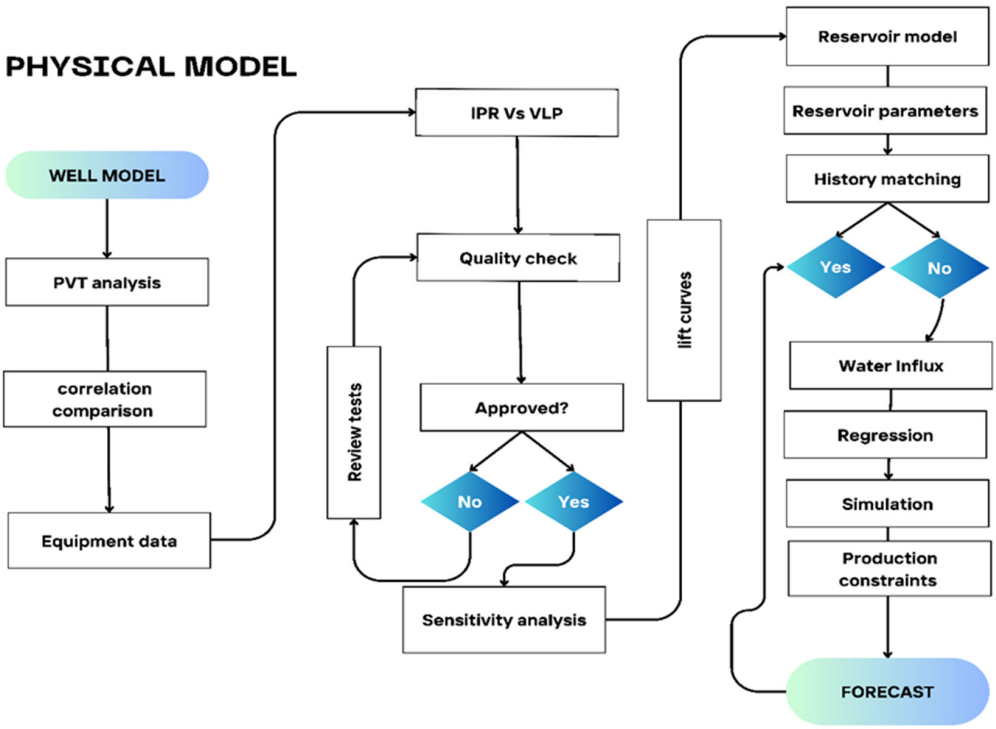

Figure 1. The generalized steps of the physical model.

At many crucial stages, data analysis uses a hybrid model to estimate future cumulative gas production. The analysis begins with historical information on cumulative gas production and related variables, including reservoir pressure and cumulative water pro-

duction, which is gathered over a set time, precisely 67 monthly dates over eight years. The dataset provides a cushion for the enhancement of deep learning, as large amounts of high-quality data are essential to create more accurate AI algorithms [17]. The forecasting system depends on this information. Secondly, data preparation is initiated through cleaning the dataset by employing outlier and missing value correction, thereby ensuring the integrity of the data. Moreover, data are standardized to ensure that all input data equally affect the model training process. The model architecture consists of input and output layers. After preprocessing, the dataset is divided into training and test sets, with 70–80% allocated for model training and 20–30% used for performance assessment. The hybrid model initialization is shown in Figure 2.

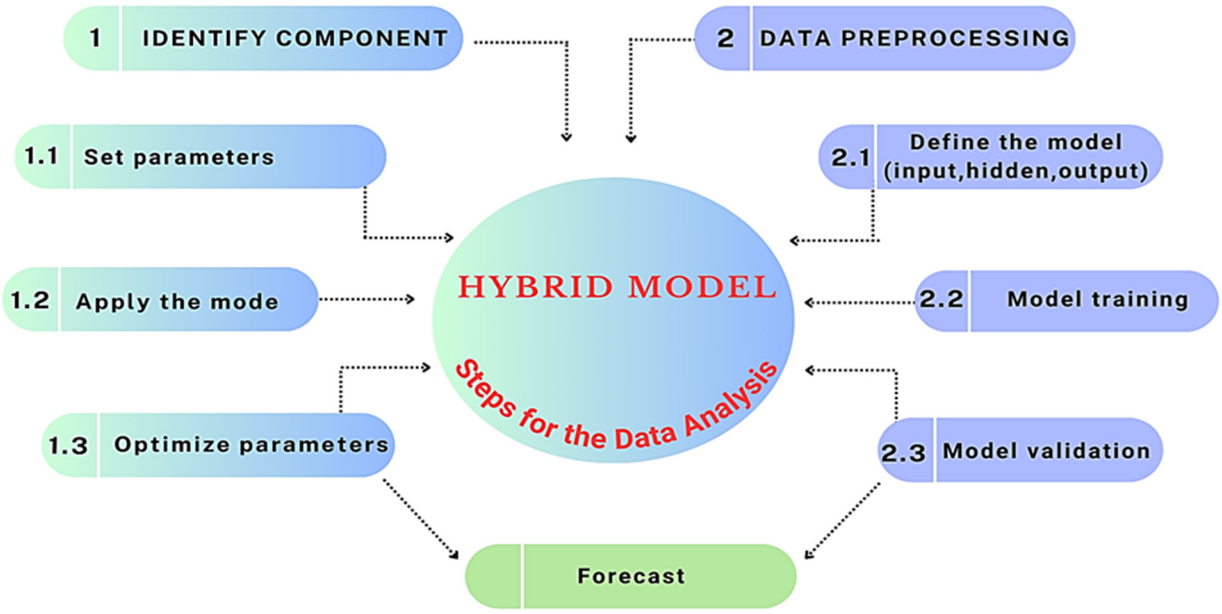

Figure 2. Steps for the data-analysis hybrid model.

Neurons and their connection strengths (weights) are the fundamental components of a neural network. A typical multi-layer neural network consists of an input layer, one or more hidden layers as shown in Figure 3, and an output layer using JMP pro 18.0.0 AI software. The "topology" or structure of the network defines how the neurons in different layers are connected.

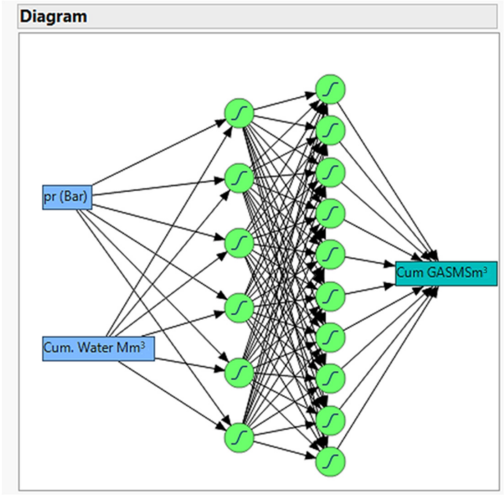

Figure 3. The neural architecture on the left is the input data for the layer, those in the middle are hidden layers, and the green button on the right is the output of the neural network.

3. Results

3.1. Wells' Model

Even in the same gas field, flow behavior may vary from one well to another due to reservoir heterogeneity and fluid properties [18]. Prosper serves to evaluate the wells' ability to produce under different conditions, and later, lift curves are integrated into the Mbal program. The production path is stimulated within the active wells in the reservoir. It creates the production and transport curves characteristic of the well based on the appropriate data definition, and the curves enable one to determine the work point of the well. Moreover, well performance analysis identifies early problems, such as water loading, which can lead to the premature evacuation of the well [19]. Using the program, the production pathway can be followed, up from entering the produced reservoir to reaching the surface equipment. The fluid characteristics are required, and it is necessary to have the operational parameters in the PVT section. Furthermore, the production type, trajectory, and configuration are submitted to the equipment data section. The fluid properties and operational parameters are considered similar for all wells in the reservoir, as shown in Table 1.

Table 1. Reservoir fluid and operational parameters for well D7.

Parameter	Input	Unit
Gas Gravity	0.64	
Separator Pressure	4	Bara
Condensate to Gas Ratio	8.3×10^{-8}	m^3/m^3
Condensate Gravity	750	Kg/m^3
Water to Gas Ratio	1×10^{-5}	m^3/m^3
Water Salinity	1480	Ppm
Co2	8.53	%
N2	5.82	%

Data collection for Table 1 is as follows:

- Relative density is the ratio of the density of a specific gas and that of air under identical conditions.
- The value of the condensate and gas volume ratio was determined by the obtained production data. For the determination of condensate/gas ratio, the volume of produced condensate was divided by the volume of gas for each production data by well, and the average of the results was taken.
- The value of water salinity and the quantity of hydrogen sulfide, carbon dioxide, and nitrogen gases expressed in percentage for the specific gas were defined first, based on the values from water analyses gathered from the well log.

Afterward, a regression was made to modify the correlations to best fit the laboratory-measured data. In the last step, the program requires the definition of the gas correlation process to consider the gas viscosity; Lee et al.'s correlation [20] is employed, as it is considered a sufficient tool to calculate the viscosity of natural gases in most hydrocarbon field software and, usually, the matched one is used to stimulate the fluid behavior in the tubing.

3.1.1. Generation of IPR Curves

The IPR curve is to be generated in the program's next step, which follows. Multiphase C and N models are chosen based on the available data. The recoveries arising during choke capacity measurements and the corresponding pressure values are defined in Table 2. The IPR curves were obtained in Figure 4 for wells D7, D8, and D9, respectively.

Table 2. Multi-rate flow data.

Well	D 7			D8				D9		
Pressure (Bar)	122.4	116.4	111.98	182.84	168.09	145.59	124.14	200	194.4	186.05
Gas Rate (1000 m^3/day)	158	176.1	180	62.2	85.6	134.2	157.7	90	149.6	207.6

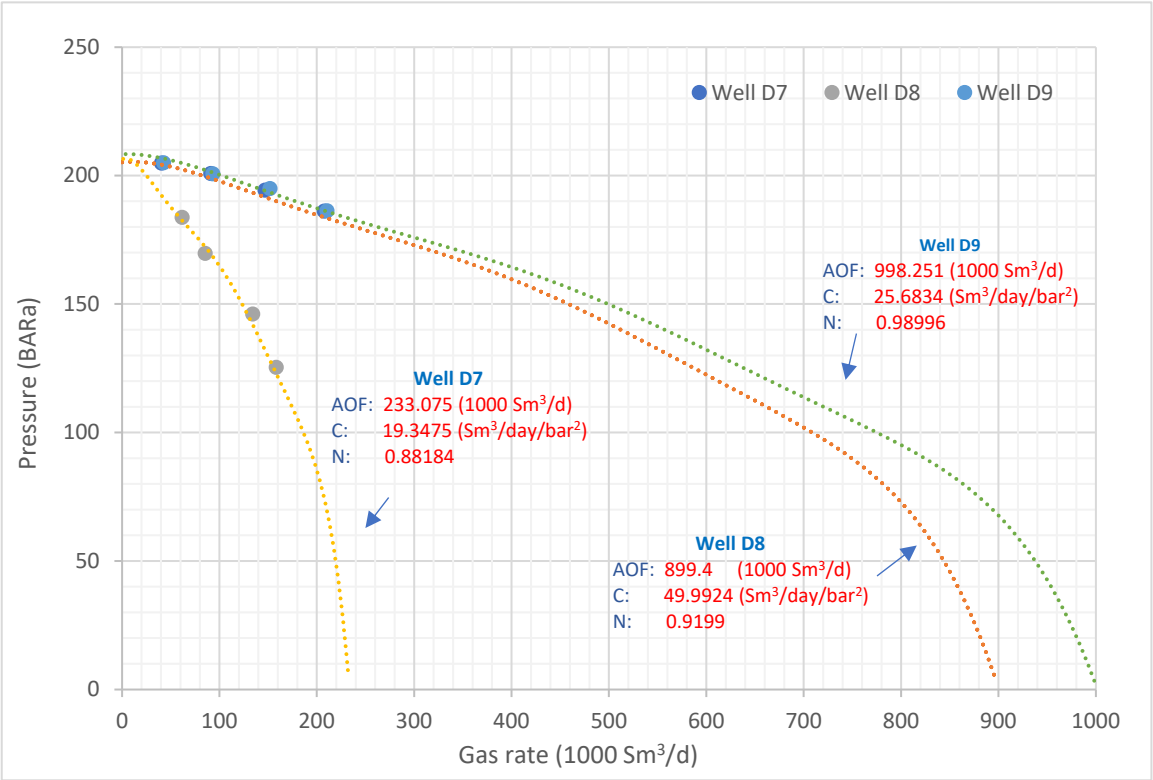

Figure 4. IPR curve for wells D7, D8, and D9.

3.1.2. Data Quality Check

The production rate is largely dependable on the difference between bottom-well pressure (BWP) and the wellhead pressure (WHP) [21]. Even though flowmeters are attached to each flowline in the gas field, certain errors might arise due to several factors [22]. The test data are required to be validated and tested against the IPR and utilize the fittest correlation to predict the flow behavior in the tubing. However, there is no best correlation. The wells' characteristics fit a specific correlation and can reproduce the actual field data, such as Duns and Ros, which are applied for deviated wells [23]. Orkiszewski is applied with significant gas–liquid [24] and others [25–28] through correlation comparison and matching the capacity calculation of the selected correlation to the actual tubing flow capacity. After rendering the image of the generated curve, the result of the comparison and calibration between the correlation and the actual tubing performance is run and the best-fit test data are chosen. The Petroleum Expert method has proved to be the most accurate correlation, specifically in the case of gas wells. There is a sequence of numbers behind the method's name, as this method can still be refined more so that it fits. If a correlation is matched, it has two parameters whose values must be close to unity; otherwise, they will distort the result. The quality check assessment is achieved as illustrated in Figure 5 for all three wells through field test data.

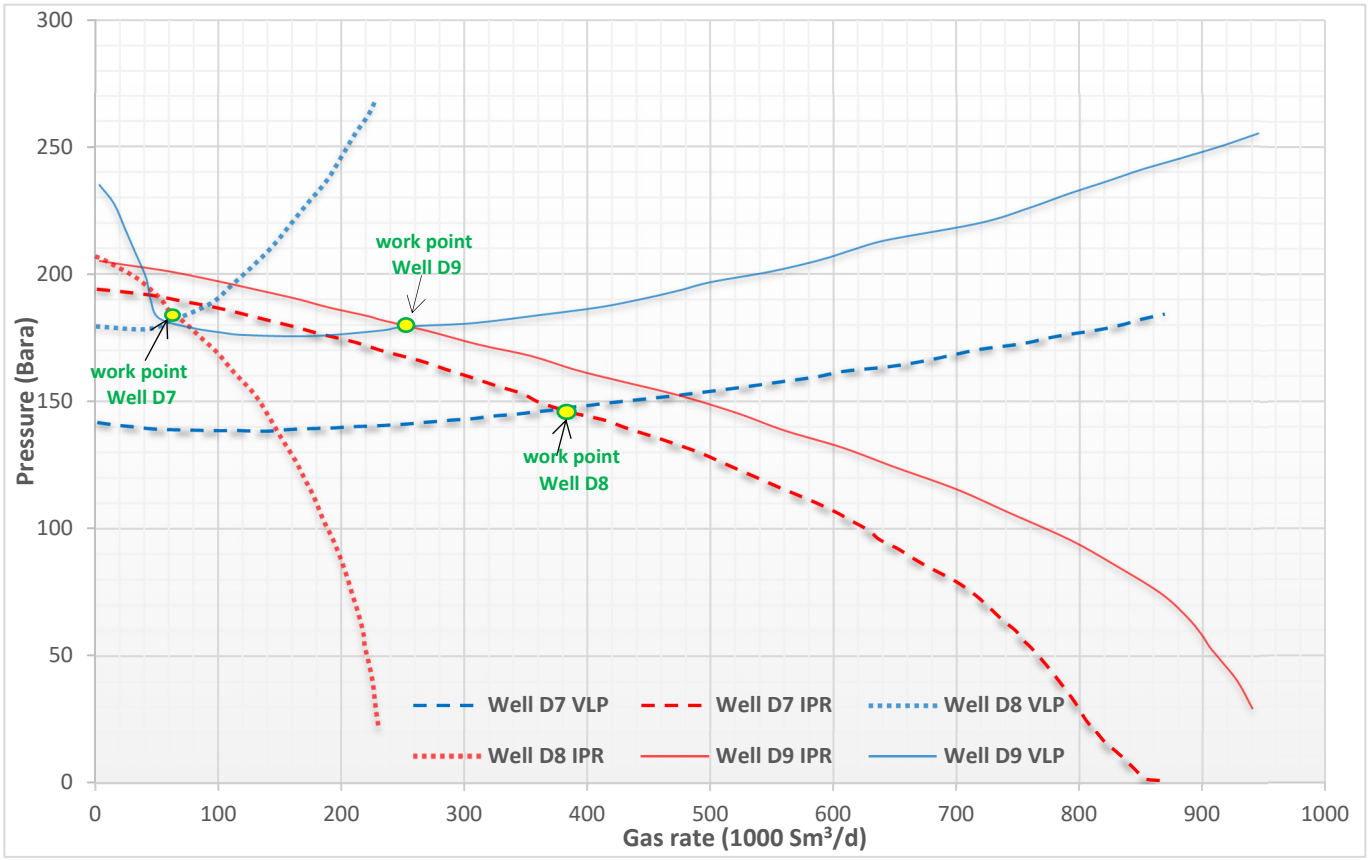

Figure 5. The VLP/IPR matching D7, D8, and D9.

A quality check to the test data is achieved after excluding unfit tests and activating the best-fit one, modifying the reservoir pressure slightly based on the test to capture the actual flow behavior in the simulator. Then, the stable producing flow rate must be estimated under current conditions and set scenarios for future management. The intersection between the red IPR and the blue (VLP) indicated the possible gas rate for wells (7, 8, 9) under current conditions. The lift curves are then generated and prepared to be imported to Mbal. This step is of the utmost importance, as it serves as a connection between the reservoir and the production capabilities, examines the reservoir pressure, drive mechanism, and fluid behavior impacts on the flow rate, and thus limits overestimation during forecasting and addresses economic feasibility for long-term field planning.

3.2. MBAL

The application of the material balance equation (MBE) in contemporary reservoir engineering has been employed by MBAL. In addition, it contains the classical reservoir engineering tool. Realistic reservoir production profiles can be interpreted with or without history-matching MBAL. The logical framework offers a dependable model that enables a reservoir engineer to develop production forecasts under various conditions [29]. To create the model, the reservoir's structural model had to be generated first. The model comprises four elements: a hydrocarbon reservoir and three active wells. With the proper adaptation of the material balance equation (MBE), it is possible to integrate production modeling through Petroleum Expert software 8.2 to predict the hydrocarbon behavior in the reservoir, as well as its behavior under different driving mechanisms.

3.2.1. Generation of the Reservoir Fluid

The fluid composition varies along the production path as the pressure and temperature drops. This approach is applied to predict the effect of distinctive variations in the pressure and temperature on the flow in the tubing. Table 3 lists values defined for MBAL for the reservoir's PVT generation. The program is calculated by employing the parameters defined here. In this case, the program needs the definition of the parameters to create the necessary sample. These parameters are listed in Table 3.

Table 3. Reservoir input for production forecast.

Description	Input	Unit
Reservoir Type	Gas	
Temperature	111	Deg C
Initial Reservoir Pressure	206.85	Bara
Water Saturation	25	%
Water Compressibility	Correlation	1/bar
Initial Gas in Place	850	MSm^3

3.2.2. Reservoir Modeling Through Material Balance Generation

The most straightforward way to explain mass conversion is through the material-balance equation (MBE). The equation mathematically provides a seamless link between the reservoir fluid and rock expansion to the succeeding extracted fluid and quantitatively characterizes the producing mechanism. The MBE treats the reservoir as a single tank and analyzes the average pressure, reserves, and productivity of the reservoir based on the conservation of mass for fluids entering and leaving the tank [30]. The impact of the relative permeability plays a vital role in reservoir simulation, where they impact the phase flow in the reservoir. Table 4 shows that water flows at less than half of its potential and is hindered by the presence of gas and formation properties. It does not flow easily until a sufficient amount is filled in the pores. Conversely, the gas's ability to flow is much easier and it can achieve 80% of the absolute permeability.

Table 4. Relative permeability parameters.

Description	Residual Saturation	Endpoint	Exponent
Krw	0.45	0.64	3
Sgr	0.2	0.8	1

The rock compressibility was set as default. The next step is to import the production history to detect the presence of the driving force in the reservoir. The MBAL requires the definition of cumulated gas production (in this case, the cumulated gross gas production) and reservoir pressures measured during production in the monthly breakdown along with the cumulated monthly values of water and gas. Through history matching, the model was not aligned with the provided data, so introducing the water influx parameters comes as the next step.

3.2.3. Drive Mechanism

Water influx parameters are introduced to the program. This value determines how much water flows into the system as a result of a given pressure variation from the water body. Here, one can attribute the water body to the elements, and the required parameters can be set. In the first step, the program involves the definition of what model to use for the calculations. Throughout the interpretation of the material balance as a straight line, with the initial hydrocarbon in place, the water's invasion into the hydrocarbon system

and the original volume of the gas cap are estimated when compared to the hydrocarbon zone and the driving force [31]. The pre-evaluation of the initial reserve is essential and estimates the effect of water influx affinity into the reserve through a mathematical model matched to the field history. There have been numerous attempts to predict and evaluate the MBE, such as Van Everdingen and Hurst [32] and Fetkovich [33]. The way Campbell's plot works is by adapting the material balance equation to adapt the presence of the aquifer and its strength to find out whether it is weak, medium, or strong [34]. It also provides an insight into the presence of the aquifer. Building on the discussion of aquifer strength, further parameters should be introduced to capture the driving force. As the graphical method assumes a steady state or a semi-state, the revalidation of the data from the aquifer via regression is a prime factor in the simulation to recompensate the idealized conditions.

The initial reservoir modeling has been found to be inaccurate, requiring proper history matching and multiple regressions to eliminate the impedance between the correlations and the tank history. The analytical method employed for the regression option adjustment for an IGIP of 838 Msm^3 leads to more accurate fitting during history matching, as illustrated in Figure 6. The dominant drive is shown in Figure 7.

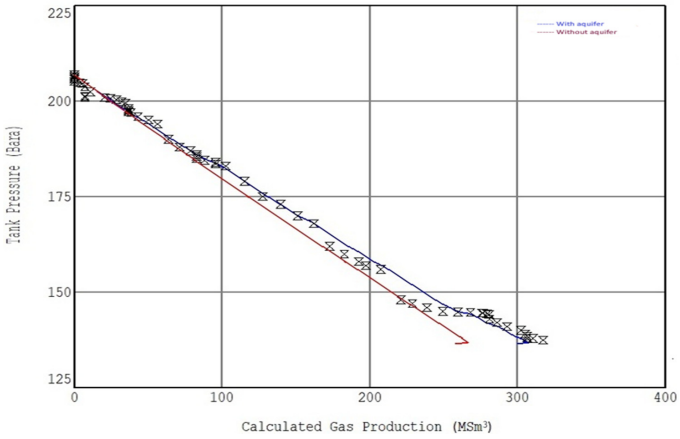

Figure 6. The analytical method for the reservoir (where the black markers ⧖ represent the measured tank pressure at various cumulative-gas-production volumes within the provided history).

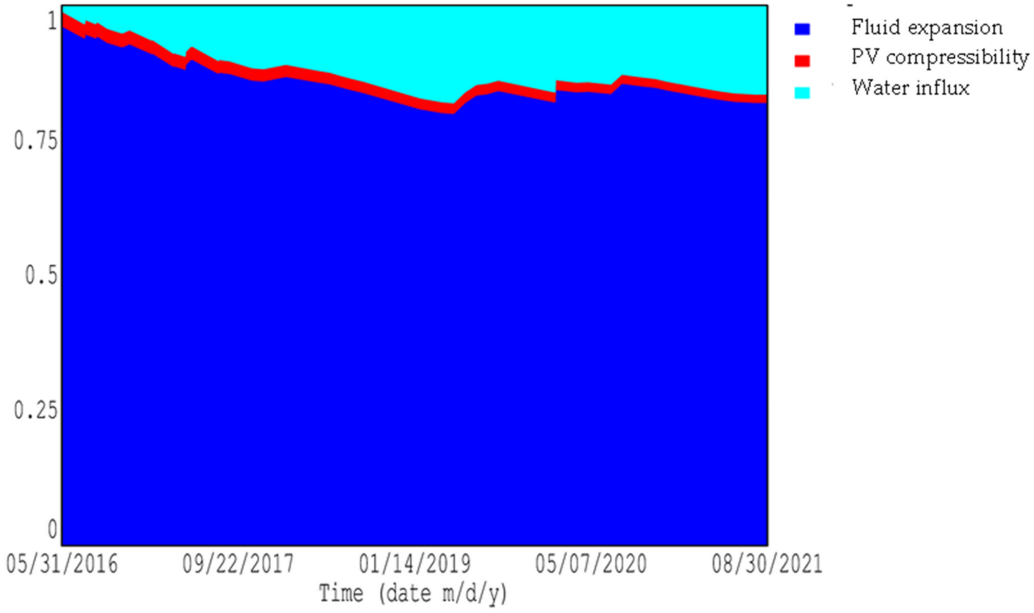

Figure 7. Energy plot for the drive mechanisms in D reservoir.

The pressure decline is steady, and during shutdown periods, stability is witnessed but no significant pressure rise is observed. This can also confirm the weak communication through the water influx. The communication and the effect of the aquifer varies from one well to another, and the reasons behind it are twofold. Firstly, geological formation effects, such as the depth of perforation intervals, formation fractures, and thin high permeability streaks, influence cumulative water production differently from one perforation zone to another. Secondly, mechanical consideration through a casing integrity test is required to be conducted. The dominance of the fluid expansion aligns the aquifer to a weak aquifer and provides minimal support. A discrepancy may arise between the energy plot and graphical method, and the possible reasons for this may be attributed to the fact that the aquifer model in the graphical method idealizes the conditions, and the energy plot depends on more precise details such as reservoir characteristics and production history. The delay in aquifer response is best demonstrated under the analytical method. However, the simplicity of detecting the major trend in the graphical method cannot be denied.

3.2.4. Stimulation and Forecast

The vital part is that the reservoir pressure simulation matches the defined values. Squares designate reservoir pressure data defined in the program's production history; the pink curve represents the pressure curve calculated by the simulation. They follow each other quite well; there is a deviation of 1.5 and 2.5 bar between the values at some points. Out of the simulations ran, this had the most accurate result. Figure 8 illustrates the simulation results as the curve started to fit the given pressure values.

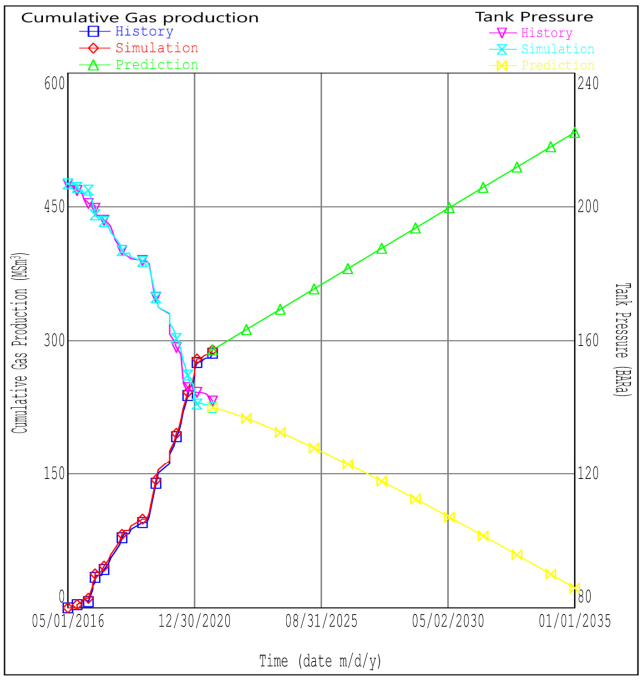

Figure 8. Production history fitting with the simulation results and future forecast.

The objective of the production forecast was to determine the most feasible gas rate to produce under certain adjustments. The program defines certain constraints, which it considers in the calculation processes. Higher well capacity is required to evaluate maximum reservoir depletion. The manifold pressure is 28 Bara. Next, the well-type definition had to be completed by importing the tubing performance curve that was generated in the prosper program and setting a minimum value to the following bottom hole pressure to be 83 Bara. The history is available until mid-2022 with a tank pressure of 141 Bara cumulative gas production of 295.77 and a recovery of 36.7%. However, the

production forecast starts at the beginning of 2021 until the end of the production. The reason for initiating a forecast at that date and leaving behind one year is to validate the simulator forecast. The field is currently considered a mature field and the active wells are declining. Execute the predication in two runs first, with the minimum value of the gas rate set at 50,000 Sm^3/day (long-term production). The forecast results are illustrated in Figure 8. The second forecast is run with a maximum value of 100,000 Sm^3/day (Short-term). A comparison between the two runs is illustrated in Figure 9.

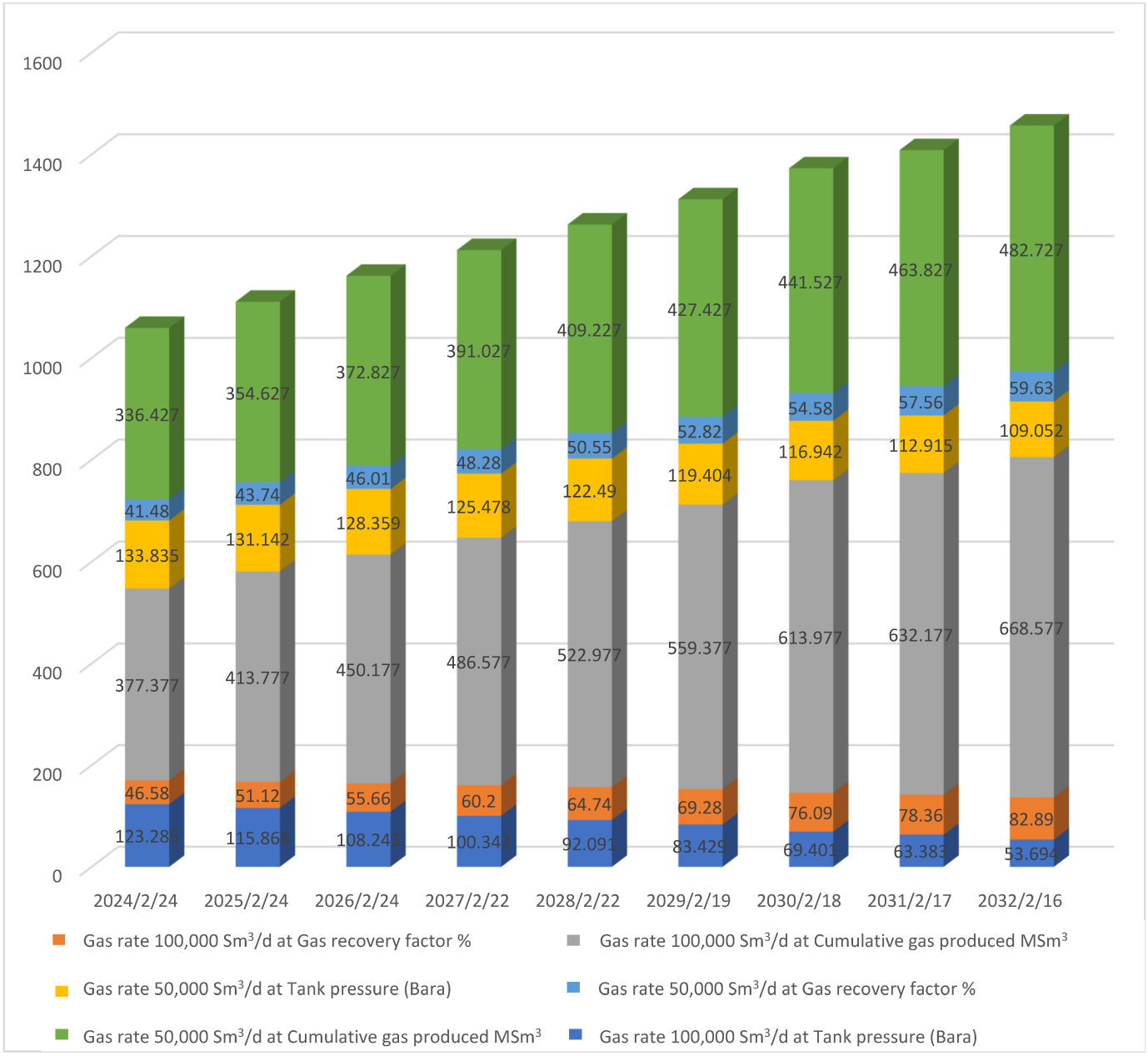

Figure 9. Comparison between long-term forecast and short-term forecast in a physical model.

3.3. Data Analysis

The development of models forms the foundation of the investigation. After compilation, the Artificial Neural Network (ANN) learns using the training data, changing the weights among its neurons depending on the input data and the desired output. Backpropagation network training is divided into three main stages. The forward pass is used to activate the network that receives the input vector during the first phase. This results in a

discrepancy (error) between the network's input and the intended output. The calculated output error spreads back over the network (error backward pass) in the second phase. By feeding the sum of the squared errors from the output layer back to the input layer via the hidden layers, connection weights are adjusted in the third phase. Until the connection weights generate an output that falls within a certain tolerance of the intended output, this process is repeated.

To predict Cumulative GAS MSm^3 using an Artificial Neural Network (ANN) with Pr (Bar) and Cumulative Water (Mm^3) as input features, the general equation for the ANN model can be represented in Equation (1):

$$Cummulative\ gas\ Mm^3 = f\left(Pr,\ Cummulative\ water\ \mathrm{Mm}^3\right) \tag{1}$$

Hidden layers: For a simple case with one hidden layer, the output from the hidden layer (for each neuron) would be:

$$h_i = \sigma(w_{i1} \cdot x_1 + w_{i2}x_h + bi) \tag{2}$$

Output layer: The output layer computes the predicted Cumulative GAS MSm^3, denoted as Yi, as a weighted sum of the hidden layer outputs:

$$yi = \sum_i w_{Oi} \cdot h_i + bo \tag{3}$$

The weights and biases are learned through training. The process is repeated until the final layer is calculated, minimizing the prediction error and finding the optimal values for the weights and biases. After that, we simultaneously use the Seasonal Exponential Smoothing (SES) model using previous cumulative-gas-production data. This method provides an additional perspective on Seasonal Exponential Smoothing (SES). The general form is represented in Equation (4).

$$F_t = (L_t + S_{t-m}) + \varepsilon_t \tag{4}$$

$$L_t = \alpha(y_t - S_{t-m}) + (1 - \alpha)L_{t-1} \tag{5}$$

St-m: The seasonal component at time t-m (where m is the season length) and is illustrated in Equation (6).

$$S_t = \beta(y_t - L_t) + (1 - \beta)S_{t-m} \tag{6}$$

The integration of both models occurs once they have evolved. We must combine the SES models' data and ANN to obtain the last prediction. For this integration, we might emphasize the more accurate model employing a weighted average or average of the projections. Even now, evaluation of the hybrid model is essential even during development. Realizing the accuracy of the forecasts requires the Artificial Neural Network (ANN) model with two hidden layers, each containing 10 and 6 neurons. The hidden layers employ the ReLU activation function, whereas the output layer uses the linear activation function. We used the Adam optimizer with a learning rate of 0.001 to reduce the mean squared error (MSE) loss function. We trained the model for 500 epochs with a batch size of 32. Before training, input features (pr (bar) and Cum. Water (mm)) were standardized to [0, 1]. We divided the dataset into two parts: training (42 data points, 63%) and testing (25 data points, 27%). The desired neural network is developed using the training set, where the network adjusts the weights between its neurons based on the intended output (supervised training). Once the network has "converged" and learned from the training

data, the test set is introduced for validation. It is important to note that the network has not seen the test set outputs during training, ensuring the robustness and integrity of the model.

3.4. Hyperparameter Optimization Procedure

We utilized Bayesian optimization to refine the hyperparameters for enhanced model performance. This method enabled us to effectively investigate the hyperparameter space and determine the ideal configuration that enhanced predicted accuracy. The selection criterion was predicated on attaining the maximum R^2 value on the validation dataset. The tuning process examined differences in the quantity of neurons within each hidden layer, learning rate, batch size, and methods of weight initialization.

3.5. The Impact of Hyperparameters on Model Efficacy

Learning Rate: Various learning rates were evaluated to determine their effect on convergence. Lower learning rates (e.g., 0.0001) caused sluggish convergence, but extremely high learning rates (e.g., 0.01) resulted in instability. A learning rate of 0.001 achieved an equilibrium between convergence velocity and model stability.

Neuron Count: Augmenting the neuron count beyond 10 and 6 did not markedly enhance model accuracy, but diminishing it resulted in underfitting.

A batch size of 32 was determined to yield consistent training and efficient generalization. Increased batch sizes (e.g., 64) expedited training but diminished generalization, whereas decreased batch sizes (e.g., 16) amplified noise in weight updates.

The ReLU activation function in hidden layers shows efficacy in capturing intricate linkages within the data. Alternative activation functions, including sigmoid and tanh, were evaluated but yielded diminished accuracy and prolonged training times due to vanishing gradient problems.

After training and testing the Artificial Neural Network (ANN) model, the prediction process consists of various phases. To process input data (pr (Bar) and Cum. Water (mm)), the following steps are taken:

1. Data preprocessing: involves normalizing input values (min–max scaling) to guarantee consistency and compliance with the trained model.
2. Prediction: The trained ANN model receives normalized input data. The model predicts the output value for cumulative gas production.
3. After processing: the predicted output may need to be changed back to its original scale (by reversing normalization) to match the target variable's units.

The average (mean) square error between the target and network outputs helps one evaluate the model. With an R-squared value of 0.983 and a mean absolute error (MAE) of 0.91, the error is less than a certain tolerance value. These measures show a minor error percentage, therefore verifying the model's excellent training quality.

Frequent field monitoring and validation methods are employed to ensure the data's dependability. We used cleaning techniques to remove any inconsistencies or irregularities in the data, ensuring that the dataset accurately reflects the well's activity.

Time-series data cross-valuation is performed to evaluate the model's durability and accuracy. This involved defining a sliding window across the previous data to forecast the upcoming period. Our approach preserves the continuity of the training set data, which is crucial for our models. It enables a more realistic projection of the model's predictive power across a wider temporal horizon.

To capture non-linearity in the data, we combined an Artificial Neural Network (ANN) with seasonal exponential smoothing (SES) and hybrid model. This method assesses the data to enhance its accuracy and reliability, thereby optimizing the pressure forecasts for

the oil field. With the residuals, Figure 10 shows the historical data (black dots) and the projected values (green line). This figure's steady growth aligns with the estimated increase in cumulative gas production. The model fits the historical data well, and the residuals fairly capture the seasonality as well as the underlying trend. By evaluating unseen data using k-fold cross-validation, we have also improved the model's dependability. Figure 11 illustrates the quantitative evaluation of the prediction results using several criteria. These contrasts show how well the model forecasts future cumulative gas production based on past data.

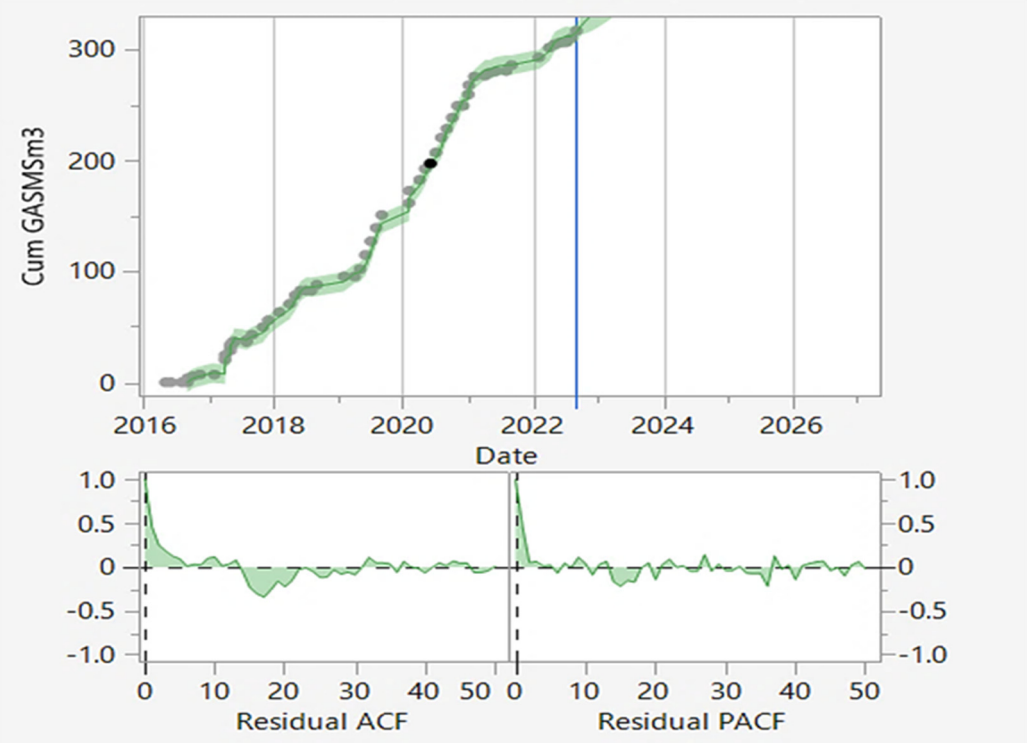

Figure 10. Quantitative evaluation for cumulative gas production.

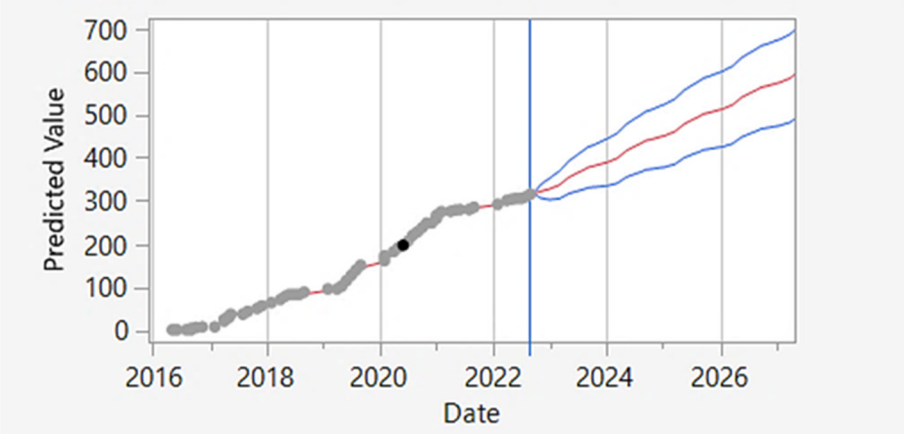

Figure 11. Future forecast through the hybrid model.

This study shows that the neural network model predicts cumulative gas production well. The model explains a lot of data variance, and the MAE (0.91) suggests that predictions are close to reality. SES and ANN help capture non-linear correlations in data, which is essential to effectively model gas reservoir behavior. Continuous cumulative gas production increases match the gas reservoirs' natural increase, proving the model's prediction ability.

Forecasting the future production is shown in Figure 10; The upper graph illustrates in Figure 10 cumulative gas (cum GAS) over time, with black dots representing field data and green shading indicating software predictions. The blue vertical line marks the transition from observed data to forecasted values. The lower graph depicts the percentage error of these predictions. As the lines spread, they indicate greater uncertainties. The upper blue line represents the best-case scenario and the red one represents the worst-case scenario. However, as the forecast moves into the future, it loses its uncertainty.

Data prediction calls for a characterization of the interaction between the actual and predicted variables. Determining the expected value is made possible by aggregating the values of several members and giving weights based on their membership condition. Numerical Equations (7) and (8) are often used in several disciplines to assess the accuracy of forecasts, evaluating the mean absolute error (MAE) and the root mean squared error (RMSE), respectively. Optimal models are selected depending on their capacity to forecast values with the lowest possible error.

$$RMSE = \sqrt{\frac{\Sigma(y_i - y_p)^2}{n}} \tag{7}$$

$$MAE = \frac{|(y_i - y_p)|}{n} \tag{8}$$

y_i For actual value; y_p is predicted value.

The model was trained and assessed to predict Cum GASMSm^3 for new values of pr (Bar) and Cum. Water (mm), respectively. Table 5 shows an example of the input attributes and anticipated output for clarity.

Table 5. Hybrid model and (SES) and (ANN) predication results sample.

Input Feature: pr (Bar)	Input Feature: Cum. Water (mm)	Actual Output: Cum GASMSm^3	Predicted (Hybrid Model): Cum GASMSm^3	Predicted (SES): Cum GASMSm^3	Predicted (ANN): Cum GASMSm^3
144.7	0.0039201	267.876	265.326	270.521	262.268
144.6	0.0040586	275.723	274.754	279.657	271.372
144.5	0.0041051	276.491	276.725	278.351	273.152
144.4	0.0042613	279.279	278.945	283.485	274.719
144.2	0.0043046	280.472	279.671	285.346	275.671

To recompensate for the uncertainties as the future forecast proceeds, a comparison with the forecast between the physical model and the hybrid is established. Figure 11 illustrates the comparison. The red line represents the predicted values from the hybrid model and the blue curve above it is (SES), and the blue curve below the red curve is the prediction using ANN, while the black dots correspond to the values derived from the physical model. The alignment of these data points demonstrates a strong correlation between the two datasets. The coefficient of determination (R^2) of 0.967 highlights the hybrid model's accuracy and its ability to replicate the physical model's results. The coefficient of determination (R^2) of 0.942 highlights the (SES) model's accuracy and its ability to replicate the physical model's results. The coefficient of determination (R^2) of 0.935 highlights the ANN model's accuracy and its ability to replicate the physical model's results. This high R^2 value indicates that 96.7% of the variance in the physical model's predictions is explained by the hybrid model. The blue line in Figure 12 is the actual data from the physical model, while the black dots represent the predicted value using the ANN;

the narrow spread of data points around the blue line suggests minimal deviation, further underscoring the hybrid model's reliability in capturing cumulative gas production trends.

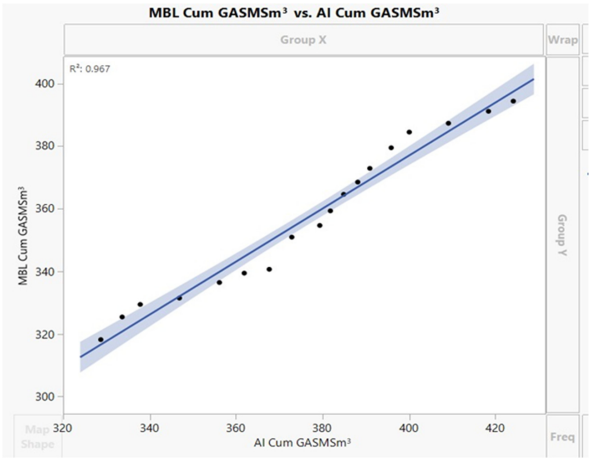

Figure 12. A comparison between future forecasts for the physical model and the hybrid.

4. Discussion

4.1. Economical Evaluation for the Production Forecast

The forecast was carried out in session to evaluate the permittable flow rate. The first run with 100,000 Sm^3/day delivered substantial gas recovery with a cumulative gas production of 668.577 MSm^3/d; however, it comes at a higher cost in the form of the reservoir depletion, where the pressure dropped significantly to 53.694. The second run was with a lower flow rate of 50,000 Sm^3/day; the pressure dropped to 109.052 Bara with a cumulative production of 482.727 MSm^3.

The assessment of the production for the two-flow rate is based on serval factors:

- Operational cost and gas prices: An exponential decline curve is suitable for a steady decline and a mature gas field when comparing the revenue from gas sales with the operation costs. The gas price is assumed to be \$0.213 per Sm^3, which will generate a total revenue of 102,820,851 USD in 2032.
- To verify the net revenue, the capital costs considered 20 million set as a CAPEX value, and the operating cost is assumed to be \$0.05/m^3, resulting in a total operational cost at the end of the predication period equal to \$24,136,350. The taxes and loyalty, 10% of the revenue, equate to \$10,282,085, and the total costs are about \$54,418,435, resulting in net profit of \$48,402,415 in 2032. Based on the calculations, it would be more economical to continue producing at 50,000 Sm^3/day rather than increasing the rate to 100,000, considering the long-term recovery and pressure decline. Thus, the forecast prediction of 50,000 m^3/day is regarded as a benchmark for the hybrid model.

4.2. Comparison Between the Hybrid and Physical Models

Forecasting the cumulative gas production through the hybrid and physical models creates the following results:

- The hybrid model captures the higher trend of production while the physical model is more conserved and avoids over-estimation. This can be attributed to assumptions grounded on the reservoir characteristics, fluid properties, and flow dynamics that may be not able to capture real-world non-linearity fully.
- The physical model predication for the reservoir pressure is more reliable, as it depends on the laws of physics. While the hybrid model accurately predicts the future cumulative gas production, it suffers in capturing trends and predicting future pressures.

- Though the results validate the hybrid model's robustness and its potential as a reliable forecasting tool for cumulative gas production, it loses its certainty for an extended period of future casting.
- A successful k-fold cross-validation strengthens the model's robustness, allowing it to generalize to new data. This is crucial in reservoir management, where precise projections may guide operational decisions and optimize production.
- The present study utilized a dataset that includes 42 training data points and 25 test data points from several wells in a gas reservoir. Although the dataset is smaller in comparison to other studies, it is justified by the data's homogeneity, the simplicity of the ANN design, and the stringent assessment criteria that show the model's dependability. Practical limits in data collecting, especially in reservoirs containing bottom water, limit the availability of large datasets. Nonetheless, the findings in this work give a proof of concept for the use of ANNs in gas production prediction. Future studies will focus on extending the dataset to incorporate data from injection wells, which will increase the model's generalizability.
- The amount of bottom water in the reservoir affects the gas well's production characteristics. The study uses cumulative water production as a significant input feature in the ANN model to account for its influence on cumulative gas output, and the model captured the strength of the aquifer efficiently.

5. Conclusions

- From the tank pressure data, the pressure will decline from 136 Bara to 109 Bara. The decline will take place steadily over the years. There is a higher possibility to produce for an extended period through a reduced rate. Wells are uneconomical when the production rate drops and the revenue can no longer cover the costs.
- Through the cost, the time-effective technology of artificial intelligence is employed to capture the degree of uncertainty and represent the non-linearity in the reservoir with a valuable connection. While the physical model requires substantial data, this model can forecast prediction through the training of production history and capture hidden trends that the physical model probably would overpass.
- This investigation shows that the neural network approach predicts cumulative gas production and that advanced modeling techniques improve reservoir management accuracy and reliability. However, further attempts are anticipated for future studies to accurately predict the reservoir pressure by incorporating different variables and expand the certainty of cumulative prediction for an extended time.

Author Contributions: B.K.R. Conceptualization, methodology, writing—original draft, preparation, resources; A.R. data curation, writing, AI model; U.A. supervision, visualization, and review; P.S. writing—review, editing, and investigation; M.K. writing—review, editing, software, validation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors extend their sincere appreciation to Shelby Smith as a proofreader from the University of Nevada, Las Vegas, for her meticulous review of the manuscript. We are grateful for her time, expertise, and dedication in refining our research study.

Conflicts of Interest: Author Ali Rabia was employed by the Wireline Logging Engineer, COSL Middle East Iraq Branch. All the 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.

Abbreviations

The following abbreviations are used in this manuscript

Pr	The input pressure in Bar; the input layer is represented as x_1.
Cumulative. Water	The accumulated water in Mm^3; the second input layer is represented as x_2.
F	The neural network function consists of a series of weighted connections, activation functions, and layers.
hi	The output of the itch neuron in the hidden layer.
σ	Function.
w_{i1} and w_{i2}	The weights associated with the inputs x_1 and x_2 represent the pressure and cumulative water production, respectively.
bi	The bias term for the ith neuron.
Yi	The predicated cum gas (Mm^3).
w_{Oi}	The weight from the hidden layer output to the output.
bo	The output bias term.
ε_t	The error or the noise, the difference between the future forecast and the observed value y_t.
α	Smoothing parameters for the level.
β	Smoothing parameters for the seasonal component.
m	Seasonal period (67 months).

References

1. Ashour, M.A.H. Optimizing Gas Production Forecasting in Iraq Using a Hybrid Artificial Intelligence Model. In Proceedings of the 2023 IEEE 14th Control and System Graduate Research Colloquium (ICSGRC), Shah Alam, Malaysia, 5 August 2023.
2. Koroteev, D.; Tekic, Z. Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future. *Energy AI* **2021**, *3*, 100041. [CrossRef]
3. Sircar, A.; Yadav, K.; Rayavarapu, K.; Bist, N.; Oza, H. Application of machine learning and artificial intelligence in oil and gas industry. *Pet. Res.* **2021**, *6*, 379–391. [CrossRef]
4. Al-Fattah, S.M.; Startzman, R.A. Predicting Natural Gas Production Using Artificial Neural Network. In Proceedings of the SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, TX, USA, 2–3 April 2001.
5. Nande, S. Application of Machine Learning for Closure Pressure Determination. In Proceedings of the SPE Annual Technical Conference and Exhibition, Dallas, TX, USA, 24–26 September 2018.
6. Elmabrouk, S.; Shirif, E.; Mayorga, R. Artificial Neural Network Modeling for the Prediction of Oil Production. *Pet. Sci. Technol.* **2014**, *32*, 1123–1130. [CrossRef]
7. Al-Ghamdi, A.A.; Gajbhiye, R.N. Development of an AI Model to Estimate Flowing Bottom-Hole Pressure in High-Pressure High-Temperature Gas Well. In Proceedings of the International Petroleum Technology Conference, Dhahran, Saudi Arabia, 12–14 February 2024.
8. Cao, Q.; Banerjee, R.; Gupta, S.; Li, J.; Zhou, W.; Jeyachandra, B. Data Driven Production Forecasting Using Machine Learning. In Proceedings of the SPE Argentina Exploration and Production of Unconventional Resources Symposium, Buenos Aires, Argentina, 1–3 June 2016.
9. Bowie, B. Machine Learning Applied to Optimize Duvernay Well Performance. In Proceedings of the SPE Canada Unconventional Resources Conference, Calgary, AB, Canada, 13–14 March 2018.
10. Manda, P.; Nkazi, D. Accuracy Assessment of Single and Hybrid Models for Predicting Shale Gas Production. *Energy Fuels* **2021**, *35*, 6068–6080. [CrossRef]
11. Li, W.; Wang, L.; Dong, Z.; Wang, R.; Qu, B. Reservoir production prediction with optimized artificial neural network and time series approaches. *J. Pet. Sci. Eng.* **2022**, *215*, 110586. [CrossRef]
12. López-Flores, F.J.; Lira-Barragán, L.F.; Rubio-Castro, E.; El-Halwagi, M.M.; Ponce-Ortega, J.M. Development and Evaluation of Deep Learning Models for Forecasting Gas Production and Flowback Water in Shale Gas Reservoirs. *Ind. Eng. Chem. Res.* **2023**, *62*, 6434–6447. [CrossRef]
13. Tadjer, A.; Hong, A.; Bratvold, R.B. Machine learning based decline curve analysis for short-term oil production forecast. *Energy Explor. Exploit.* **2021**, *39*, 1747–1769. [CrossRef]
14. Rostamian, A.; Miranda, M.V.d.S.; Mirzaei-Paiaman, A.; Botechia, V.E.; Schiozer, D.J. Analysis of different objective functions in petroleum field development optimization. *J. Pet. Explor. Prod. Technol.* **2024**, *14*, 2785–2805. [CrossRef]
15. Hyoung, J.; Lee, Y.; Han, S. Development of Machine Learning-Based Production Forecasting for Offshore Gas Fields Using a Dynamic Material Balance Equation. *Energies* **2024**, *17*, 5268. [CrossRef]

16. Li, Y.; Han, Y. Decline Curve Analysis for Production Forecasting Based on Machine Learning. In Proceedings of the SPE Symposium: Production Enhancement and Cost Optimisation, Kuala Lumpur, Malaysia, 7–8 November 2017.
17. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [CrossRef] [PubMed]
18. Ahmed, T.; McKinney, P.D. (Eds.) 4-Performance of Oil Reservoirs. In *Advanced Reservoir Engineering*; Gulf Professional Publishing: Burlington, MA, USA, 2005; pp. 291–325.
19. Meng, F.; He, D.; Yan, H.; Zhao, H.; Zhang, H.; Li, C. Production performance analysis for slanted well in multilayer commingled carbonate gas reservoir. *J. Pet. Sci. Eng.* **2021**, *204*, 108769. [CrossRef]
20. Lee, A.L.; Gonzalez, M.H.; Eakin, B.E. The Viscosity of Natural Gases. *J. Pet Technol.* **1996**, *18*, 997–1000. [CrossRef]
21. Wiggins, M.L.; Lake, L.W.; Clegg, J.D. Inflow and Outflow Performance. In *Production Operations Engineering*; Society of Petroleum Engineers: Richardson, TX, USA, 2007; pp. 1–40.
22. Hansen, L.S.; Pedersen, S.; Durdevic, P. Multi-Phase Flow Metering in Offshore Oil and Gas Transportation Pipelines: Trends and Perspectives. *Sensors* **2019**, *19*, 2184. [CrossRef]
23. Duns, H., Jr.; Ros, N.C.J. Vertical flow of gas and liquid mixtures in wells. In Proceedings of the 6th World Petroleum Congress, Frankfurt am Main, Germany, 19–26 June 1963.
24. Orkiszewski, J. Predicting Two-Phase Pressure Drops in Vertical Pipe. *J. Pet. Technol.* **1967**, *19*, 829–838. [CrossRef]
25. Aziz, K.; Govier, G.W. Pressure Drop In Wells Producing Oil And Gas. *J. Can. Pet. Technol.* **1972**, *11*, PETSOC-72-03-04. [CrossRef]
26. Hagedorn, A.R.; Brown, K.E. Experimental Study of Pressure Gradients Occurring During Continuous Two-Phase Flow in Small-Diameter Vertical Conduits. *J. Pet. Technol.* **1965**, *17*, 475–484. [CrossRef]
27. Alameedy, U.; Al-Sarkhi, A.; Abdul-Majeed, G. Comprehensive review of severe slugging phenomena and innovative mitigation techniques in oil and gas systems. *Chem. Eng. Res. Des.* **2025**, *213*, 78–94. [CrossRef]
28. Gray, H.E. Vertical Flow Correlation in Gas Wells. In *User's Manual for API 14B Surface Controlled Subsurface Safety Valve Sizing Computer Program*, 2nd ed.; American Petroleum Institute: Washington, DC, USA, 1978.
29. Ahmed, T.; McKinney, P.D. (Eds.) 5-Predicting Oil Reservoir Performance. In *Advanced Reservoir Engineering*; Gulf Professional Publishing: Burlington, MA, USA, 2005; pp. 327–363.
30. Okotie, S.; Ikporo, B. *Reservoir Engineering*; Springer: Berlin/Heidelberg, Germany, 2019.
31. Havlena, D.; Odeh, A. The Material Balance as an Equation of a Straight Line. *J. Pet. Technol.* **1963**, *15*, 896–900. [CrossRef]
32. Van Everdingen, A.; Hurst, W. The Application of the Laplace Transformation to Flow Problems in Reservoirs. *J. Pet. Technol.* **1949**, *1*, 305–324. [CrossRef]
33. Fetkovich, M.J. A Simplified Approach to Water Influx Calculations-Finite Aquifer Systems. *J. Pet. Technol.* **1971**, *23*, 814–828.
34. Ahmed, T.; McKinney, P.D. (Eds.) 2-Water Influx. In *Advanced Reservoir Engineering*; Gulf Professional Publishing: Burlington, MA, USA, 2005; pp. 149–185.