Article

Enhancing Production Prediction in Shale Gas Reservoirs Using a Hybrid Gated Recurrent Unit and Multilayer Perceptron (GRU-MLP) Model

Xianlin Ma *, Mengyao Hou, Jie Zhan * and Rong Zhong

College of Petroleum Engineering, Xi'an Shiyou University, Xi'an 710065, China;
22212010144@stumail.xsyu.edu.cn (R.Z.)
* Correspondence: xianlinm@xsyu.edu.cn (X.M.); zhanjie@xsyu.edu.cn (J.Z.)

Abstract: Shale gas has revolutionized the global energy supply, underscoring the importance of robust production forecasting for the effective management of well operations and gas field development. Nonetheless, the intricate and nonlinear relationship between gas production dynamics and physical constraints like shale formation properties and engineering parameters poses significant challenges. This investigation introduces a hybrid neural network model, GRU-MLP, to proficiently predict shale gas production. The GRU-MLP architecture can capture sequential dependencies within production data as well as the intricate nonlinear correlations between production and the governing constraints. The proposed model was evaluated employing production data extracted from two adjacent horizontal wells situated within the Marcellus Shale. The comparative analysis highlights the superior performance of the GRU-MLP model over the LSTM and GRU models in both short-term and long-term forecasting. Specifically, the GRU model's mean absolute percentage error of 4.7% and root mean squared error of 120.03 are notably 66% and 80% larger than the GRU-MLP model's performance in short-term forecasting. The accuracy and reliability of the GRU-MLP model make it a promising tool for shale gas production forecasting. By providing dependable production forecasts, the GRU-MLP model serves to enhance decision-making and optimize well operations.

Keywords: GRU; hybrid neural network; LSTM; shale gas; well production

1. Introduction

Shale gas has emerged as a promising resource to fulfill global energy demand, primarily attributed to notable discoveries of reserves and the combined application of horizontal drilling techniques and multi-stage hydraulic fracturing methods [1,2]. According to data from the U.S. Energy Information Administration (EIA), the United States possesses technically recoverable shale gas resources amounting to 862 trillion cubic feet [3]. These resources are distributed across diverse shale formations, including the Marcellus, Utica, Barnett, and Haynesville.

These shale formations are recognized for their low porosity and ultra-low permeability. Nevertheless, advancements in horizontal drilling and hydraulic fracturing technologies have rendered the extraction of natural gas from these formations economically feasible [4,5]. This entails the initial vertical drilling of a well, followed by its subsequent horizontal orientation within the shale formation. Sequential stimulations of multiple stages are then implemented along the horizontal wellbore through the injection of high-pressure fluids, typically comprising water, sand, and additives. This process leads to the creation of a stimulated reservoir volume, thereby amplifying the contact areas between the wellbore and the shale formation. Consequently, the horizontal wells with multistage hydraulic fracturing enhance gas drainage efficiency and augment overall gas production. In the year 2022, dry natural gas production from the shale formations in the United States was estimated at 28.5 trillion cubic feet due to the combination of horizontal drilling

with multistage hydraulic fracturing, equating to roughly 80% of the total dry natural gas production [6].

To enhance shale gas recovery from hydraulically fractured reservoirs, the prediction of production at a specified time is imperative for selecting the optimal completion method and well operational parameters, as well as for refining gas field development strategies [7–9]. Conventional methods for production forecasting, such as decline curve analysis (DCA) and numerical reservoir simulation, are extensively employed in the oil and gas sector to foresee well production performances from hydrocarbon reservoirs.

DCA, a prevalent empirical approach, relies on historical production data to anticipate future production outcomes [10,11]. This method assumes that the production decline of a well or a reservoir follows a certain mathematical function. Through fitting the decline curve to the available production data, essential parameters like the initial production rate, decline rate, and ultimate recovery can be estimated. DCA offers a straightforward and expedient estimation of forthcoming production; however, it operates under the assumption of boundary-dominated flow and may not accurately capture intricate reservoir behavior or shifts in production mechanisms over time.

Conversely, numerical reservoir simulation considers reservoir heterogeneities, well configurations, fluid properties, and other factors exerting influence on hydrocarbon production [12–14]. This approach permits the integration of intricate reservoir dynamics, encompassing fluid flow dynamics, pressure variations, and reservoir geometry. For instance, the reservoir simulations adeptly incorporate various intricate mechanisms governing shale gas transport in the production prediction, including Fick diffusion and non-Darcy flow, alongside the geometry of the simulated reservoir volume [15,16]. Nonetheless, the reservoir simulations require substantial data input encompassing reservoir properties, well specifics, and fluid characteristics. The accuracy of production predictions heavily relies on the quality of the input data and the reliability of the reservoir models. Moreover, the simulations usually demand considerable computational resources and entail time-intensive model construction and execution processes.

In recent years, there has been an increased focus on incorporating data-driven and machine learning-based approaches to forecast hydrocarbon production [17,18]. These approaches view such production prediction as a time-series forecasting problem [19]. By utilizing time series production as training data, these methods can effectively predict well production by capturing the trends and characteristics present in the hydrocarbon production data. Deep neural networks have been extensively studied for the time series forecasting of hydrocarbon production due to their ability to handle nonlinear data. Previous studies on hydrocarbon production prediction using deep learning models can be categorized based on the count of input and forecasted features, as illustrated in Table 1. These prediction categories are classified into univariate, covariate, and multivariate scenarios [19], as visually depicted in Figure 1.

In Figure 1a, the univariate prediction entails the projection of the future values of a single variable based on its historical data. In certain instances, when predicting the values of a single variable, it is necessary to consider the relationships among multiple variables and to model their associations to forecast the forthcoming value of the single variable. This form of prediction is termed covariate forecasting.

Within the context of covariate forecasting, both one or multiple time series data and additional physical factors governing production are considered. While the number of features predicted in the covariate forecasting remains single, akin to the univariate prediction, it incorporates multiple input features, as depicted in Figure 1b. Beyond the aforementioned predictive paradigms, there are multivariate predictions that focus on capturing the associations among multiple variables to forecast the values of the variables. These prediction scenarios are depicted in Figure 1c.

These investigations have demonstrated the impressive effectiveness and accuracy of deep neural networks in the prediction of hydrocarbon production, as evidenced by the results presented in Table 1.

**Table 1.** Summary of recent studies using deep neural networks in hydrocarbon production forecasting.

Categories	Author (Year)	Well Production	Neural Network	Description
Univariate	Lee et al. (2019) [20]	Monthly shale gas production	LSTM	LSTM outperforms DCA methods.
	Sagheer et al. (2019) [21]	Daily oil production	Deep LSTM (DLSTM)	DLSTM outperformed RNN and GRU.
	Ning et al. (2022) [22]	Oil production rate	LSTM	LSTM outperformed traditional prediction methods
	Yang et al. (2022) [23]	Daily shale gas	LSTM	LSTM outperforms ARIMA and DCA
Covariate	Le et al. (2021) [24]	Oil production rate	Physics-guided model	Physics-guided model outperforms LSTM
	Li et al. (2022) [25]	Daily oil rate	Bidirectional GRU and Hybrid network	BiGRU-DHNN outperforms RNN, GRU, BiGRU and LSTM
Multivariate	Zhang et al. (2019) [26]	Water saturation, formation pressure and oil production	LSTM, GRU, RNN	LSTM is better than GRU and standard RNN
	Kim et al. (2021) [27]	Oil and water rates	LSTM	RNN-based proxy model for well-control optimization
	Huang et al. (2022) [28]	Daily oil, water, gas rates gas-oil ratio	LSTM	LSTM outperforms numerical reservoir simulation
	Yang et al. (2022) [29]	Gas, water flow rates	GRU-MLP combined neural network	GRU-MLP is superior to RNN, GRU and LSTM

Recent scholarly pursuits have delved into the exploration of the potential of physics-constrained machine learning (PCML) methodologies for projecting hydrocarbon production. These methods excel at capturing the inherent physical dependencies between variables through the design of neural network architectures. In the realm of PCML, physical insights are integrated into the neural network framework, wherein designated neurons or modules are endowed with explicit physical interpretations [25]. Notably, Shi et al. [30] put forth an integrated neural network comprising LSTM and MLP architectures for the prediction of geothermal productivity in multilateral wells, incorporating the physical constraints into the model to enhance its predictive capability. A hybrid network was devised, wherein the integration of physical constraints and LSTM outputs serves as input for the MLP. Through this integration, the intricate physical dependencies and geothermal production were systematically explored.

Similarly, Li et al. [25] introduced a composite neural network encompassing Bidirectional Gated Recurrent Unit (BiGRU) and Deep Hybrid Neural Network (DHNN) components. This approach leverages the complementary strengths of physics knowledge and machine learning, effectively capturing the dynamics of time-series production. By considering the static and dynamic, temporal and spatial aspects of the fractured wells as the physical constraints shaping long-term production predictions, this PCML technique facilitated the generation of oil production forecasts that were more precise and dependable.

Motivated by the recent advancements in the application of PCML to the realm of time-series production forecasting, this study introduces a hybrid neural network named GRU-MLP, which seamlessly integrates the architectures of GRU and MLP. The GRU-MLP model is designed to forecast well productivity within shale gas reservoirs while simultaneously accounting for the physical constraints associated with shale gas production.

Embedded within the GRU-MLP model, the MLP component is trained to apprehend the intricate nonlinear correlation existing between the production data and the accompanying physical constraints, including formation properties and engineering parameters. On the other hand, the GRU component is harnessed to capture the sequential relationships innate to the production data.
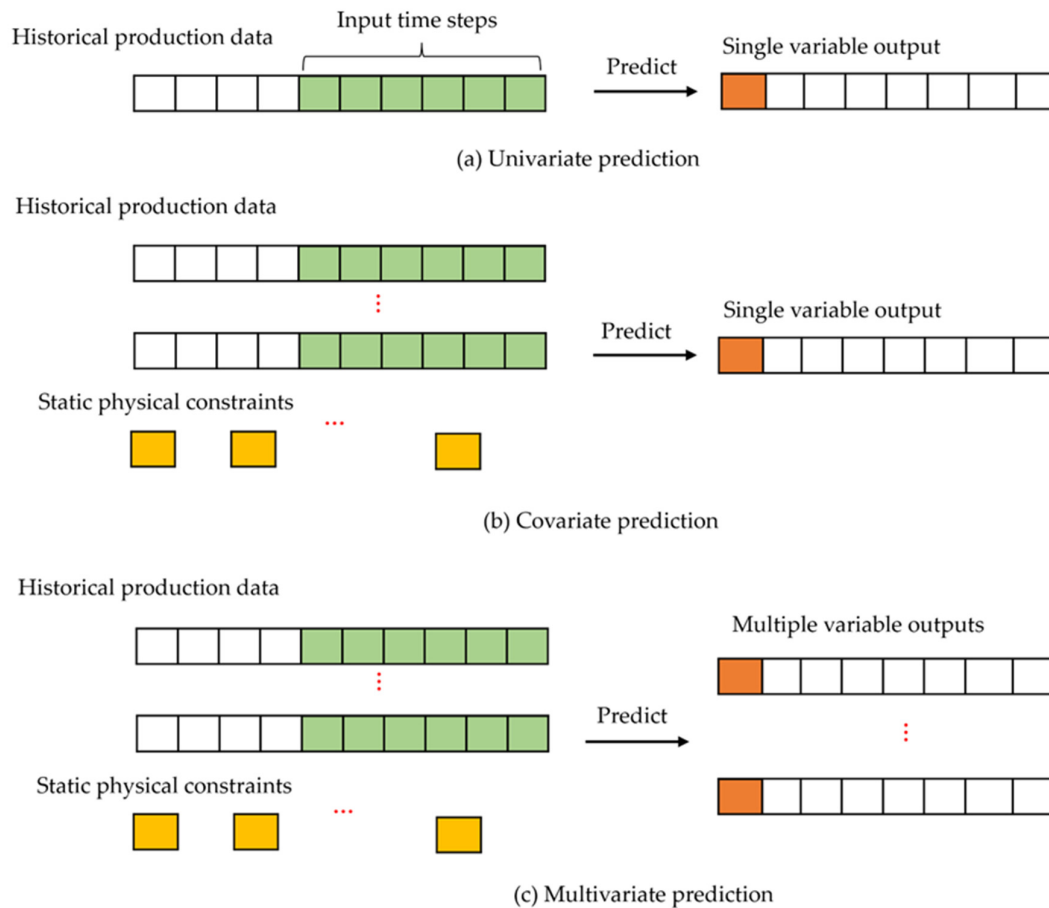
Figure 1. Categories of hydrocarbon production forecasting.

This paper offers three key contributions: (i) the creation of a novel hybrid GRU-MLP architecture designed for well production prediction within shale gas reservoirs; (ii) a comprehensive comparative evaluation of the proposed GRU-MLP model with established LSTM and GRU models, demonstrating its improved predictive performance; and (iii) the integration of geological attributes, fracture geometry, and fracturing treatment parameters as inputs, enabling a holistic assessment of intricate interactions impacting production outcomes. Collectively, these contributions introduce innovative methodologies and pragmatic advancements to the realm of production prediction in shale gas reservoirs. By incorporating the physical constraints, the model not only enhances the predictive accuracy but also aligns the prognostications with the physical principles governing shale gas reservoir dynamics.

2. Description of Deep Learning Model

2.1. RNN-Based Models

Recurrent neural networks (RNNs) are a class of artificial neural networks that are engineered to effectively model sequential and time-evolving data, such as time series datasets. In contrast to conventional feedforward neural networks, RNNs possess hidden states that acquire the information from preceding steps and pass it along to the subsequent step, which allows the networks to preserve the information and learn long-term dependencies within the data [31].

2.1.1. Long Short-Term Memory

The standard RNN model suffers from short-term memory, gradient explosion, and gradient disappearing problems [32]. LSTM is a variation of the RNN architecture that addresses the gradient disappearance problem by incorporating a gating mechanism [33]. The LSTM architecture lies in its memory cell, which is responsible for storing and passing pertinent information across different time steps. The memory cell comprises distinct components, including an input gate, a forget gate, and an output gate. These gates engage with the input (x_t) at the current time step (t) and the hidden state (h_{t-1}) from the previous time step $(t-1)$. Subsequently, by means of a sequence of nonlinear operations performed by these gates, the pertinent information contained within (x_t) and (h_{t-1}) is transferred to the cell state (c_t) and output (h_t). A visual representation of a LSTM cell's structure is depicted in Figure 2.
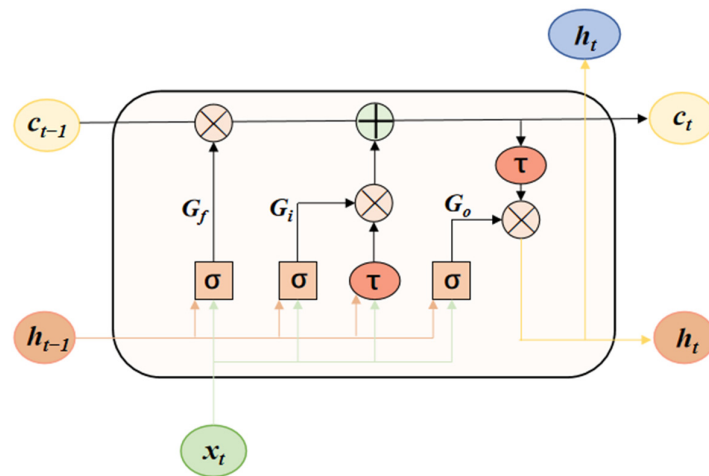
Figure 2. Structure of a LSTM cell with forget gate (G_f), input gate (G_i), and output gate (G_o).

The forget gate (G_f) controls how much information from the preceding memory cell should be retained or discarded. It takes the previous hidden state (h_{t-1}) and the current input (x_t) and produces an output ranging between 0 and 1. This computation is expressed as:

$$G_f = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{1}$$

Next, the input gate (G_i) determines how much new information should be stored in the memory cell. By considering both the previous hidden state (h_{t-1}) and the current input (x_t) as inputs, the input gate generates an output confined within the interval of 0 and 1. This gate includes two operations. A sigmoid layer first updates the values. Then, a hyperbolic tangent (tanh) layer constructs a vector of new candidate values (c_t'). The formulation of the input gate is articulated as follows:

$$G_i = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{2}$$

$$c_t' = \tau\left(W_{xg}x_t + W_{hg}h_{t-1} + b_g\right) \tag{3}$$

Subsequently, the cell state (c_{t-1}) needs to be updated to (c_t), which takes the previous memory cell, the forget gate, and the input gate as inputs and produces the current cell state, denoted by c_t.

$$c_t = G_f c_{t-1} + G_i c_t' \tag{4}$$

The output gate (G_o) determines how much information from the current memory cell should be outputted as the current hidden state. It takes the previous hidden state and the current input and produces an output between 0 and 1. It is composed of a sigmoid layer

and a tanh layer. Finally, based on the inputs (x_t) and (h_{t-1}) as well as the cell state (c_t), the output gate decides what information will be output.

$$G_o = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{5}$$

$$h_t = G_o\tau(c_t) \tag{6}$$

In these equations, W and b represent the weight matrices and bias vectors, respectively. The sigmoid function is mathematically represented as $\sigma(x)$, and it transforms an input x into a value that falls within the range of 0 and 1. On the other hand, the tanh function is symbolized as $\tau(x)$, and it operates by mapping an input x onto a value that lies within the span of -1 and 1.

The LSTM architecture allows for the learning of long-term dependencies by selectively retaining or discarding information in the memory cell through the input and forget gates. This enables the LSTM networks to effectively process and model sequential data with complex temporal dependencies, making them widely used in tasks such as natural language processing, speech recognition, and time series prediction.

2.1.2. Gated Recurrent Unit

Though proven effective, the LSTM networks have a high computational cost. As a simplified version of LSTM, GRU retains cell units analogous to the forget gate mechanism, albeit omitting the output gate. Consequently, this design choice leads to a reduction in the overall parameters [34], as visually depicted in Figure 3.
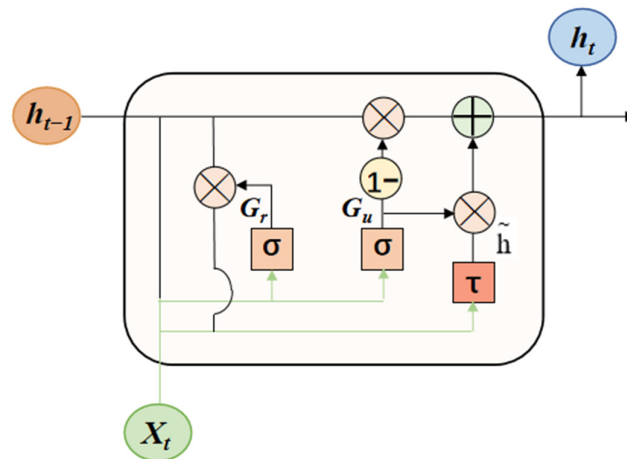

Figure 3. Structure of a GRU cell with reset gate (G_r) and update gate (G_u).

The GRU network is like the LSTM network but has a simplified structure with two gates: an update gate and a reset gate. The role of the update gate determines the amount of historical information to be carried forward to the current time step. The update gate incorporates the previous hidden state (h_{t-1}) and the current input (x_t) as input parameters, generating an output that lies within the interval of 0 and 1, denoted as G_u.

$$G_u = \sigma(W_{xu}x_t + W_{hu}h_{t-1} + b_u) \tag{7}$$

The reset gate controls how much of the historical information should be forgotten. The reset gate takes the previous hidden state (h_{t-1}) and the current input (x_t), generating an output confined within the range of 0 and 1, symbolized as G_r.

$$G_r = \sigma(W_{xr}x_t + W_{hr}h_{t-1} + b_r) \tag{8}$$

The candidate activation process calculates a fresh candidate activation value, which is a fusion of the previous hidden state and the current input. This operation involves

taking the previous hidden state, the present input, and the reset gate as input parameters, yielding the candidate activation denoted by $\widetilde{h}_t$.

$$\widetilde{h}_t = \tau(W_{xh}x_t + W_{hh}(G_r h_{t-1} + b_h)) \tag{9}$$

The hidden state combines the previous hidden state and the candidate activation to produce the current hidden state, denoted by h_t.

$$h_t = (1 - G_u)h_{t-1} + G_u\widetilde{h}_t \tag{10}$$

2.2. Hybrid GRU-MLP Model

The prediction of gas production from stimulated shale reservoirs is inherently a multivariate problem, as it depends not only on historical production data but also on various physical constraints. These constraints include a wide spectrum of factors, ranging from geological attributes to engineering parameters such as formation properties, fracture geometries, and fracturing treatment specifics. Consequently, the development of forecasting models for well gas production requires the incorporation of numerous input parameters derived from diverse domains.

Given that the selected parameters in this study are static, MLP networks are robust models for addressing such non-sequential data. The hybrid GRU-MLP network capitalizes on the advantages of two distinct neural network architectures. The GRU component can capture long-term dependencies and temporal patterns present in the historical production data, while the MLP component can handle non-sequential features and learn intricate nonlinear relationships within the dataset. This hybrid approach allows for a more comprehensive representation and understanding of the input data, leading to improved accuracy in the prediction of well gas production.

Figure 4 illustrates the architectural composition of the hybrid GRU-MLP neural network [29]. Initially, a multi-layer GRU neural network is employed to capture the relationship between historical production data (x_1, x_2,..., x_t) and projected forthcoming value (h_t). The output of the GRU network is then passed through a linear activation layer and subsequently combined with the physical constraints (c_1, c_2,..., c_n) as inputs to the MLP component, capturing the non-linear relationship between the production data and the constraints. Finally, the hybrid neural network generates the production prediction (x_{t+1}) at the time step ($t + 1$).
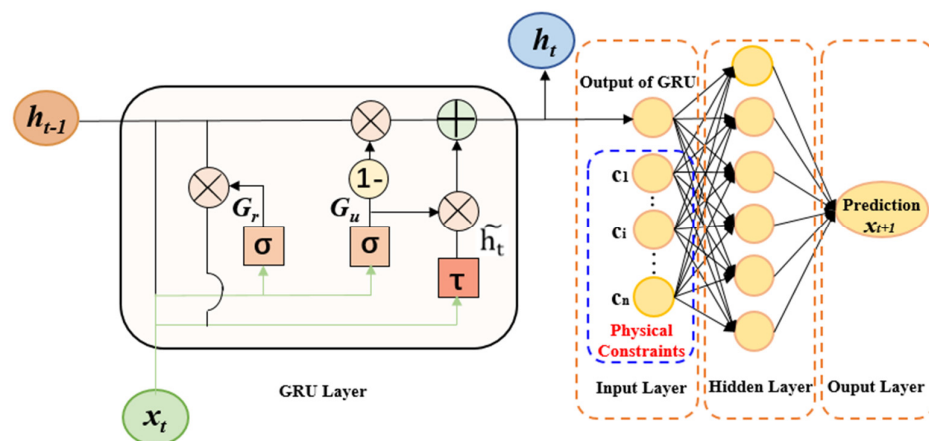

Figure 4. Structure of a hybrid GRU and MLP neural network (modified from [29]).

3. Data Preparation

The data employed in this study were sourced from the Marcellus Shale Energy and Environment Laboratory (MSEEL) project [35,36], accessible online through the website http://www.mseel.org (accessed on 12 April 2023). The MSEEL project, sponsored by the US Department of Energy, was geared towards enhancing the comprehension of shale

resources and ensuring the extraction of shale gas that is both efficient and environmentally responsible. Administered by Northeast Natural Energy, the MSEEL represents one of the most expansive shale gas research initiatives on a global scale.

The MSEEL field laboratory is comprised of two legacy wells (MIP-4H and MIP-6H) that were drilled in 2011, in addition to two more recent horizontal wells (MIP-3H and MIP-5H) that were drilled and completed in 2015 (as depicted in Figure 5). These horizontal wells span an average lateral length of 6000 feet, with a well spacing of 1700 feet. Natural gas production from two horizontal wells commenced in December 2015.
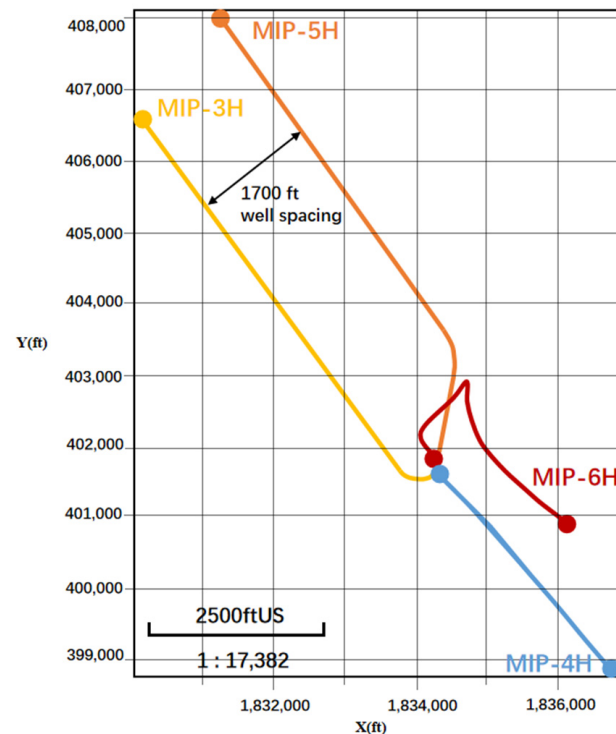
Figure 5. Well configurations in the MSEEL project.

The MIP-3H and MIP-5H fractured horizontal wells, featuring their multi-stage fractures, were both drilled and completed within the Marcellus Shale formation [35]. The Marcellus Shale, positioned beneath the Appalachian Basin, holds paramount significance as the largest natural gas-producing formation in the United States. The U.S. EIA has estimated the presence of approximately 11.33 trillion cubic meters of technically recoverable natural gas reserves within the Marcellus Shale [37].

3.1. Geological and Engineering Factors

The production of shale gas from a fractured horizontal well is closely related to a combination of factors encompassing geological characteristics, the geometry of induced fractures, and parameters pertaining to the fracturing treatment. The interplay of these elements significantly influences shale gas production. Table 2 succinctly presents an overview of the physical factors, along with their respective correlations with shale gas production, as examined in the study.

3.2. Historical Production Data

In addition to the aforementioned static constraints, the well production prediction in the shale gas reservoirs is a time-series forecasting problem when considering the historical production data. Figure 6 displays the daily shale gas production profiles for the MIP-3H and MIP-5H wells.

Table 2. The physical constraints for shale gas production.

Type	Control Factors	Relevance with Gas Production
Geological	Formation thickness (ft)	affects the volume of gas available for production.
	True vertical depth (ft)	affects the distribution of shale gas reservoir
Fracture	Fracture half-length (ft) Fracture height (ft)	A longer and taller fracture allows for a larger contact area with enhanced gas production.
Fracturing treatment	Lateral length (ft)	Longer lateral length increases surface area available for gas production.
	Total fracturing fluid injected (gal)	A larger volume of fluid can create more extensive and interconnected fractures, resulting in potentially higher gas production.
	Total proppant mass (lbm)	Increasing the amount of proppant used can enhance fracture conductivity, allowing for improved gas production.
	Injection rate (bbl/min)	Higher injection rates can lead to the creation of longer and more extensive fractures, increasing the surface area available for gas flow.
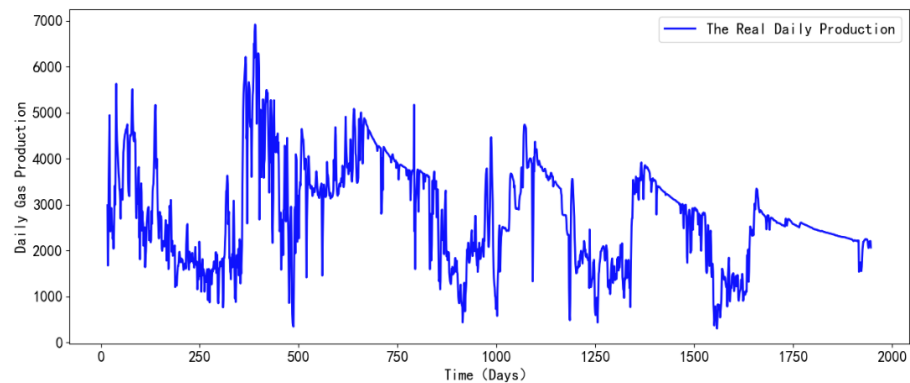
(**a**) Production data of well MIP-3H.
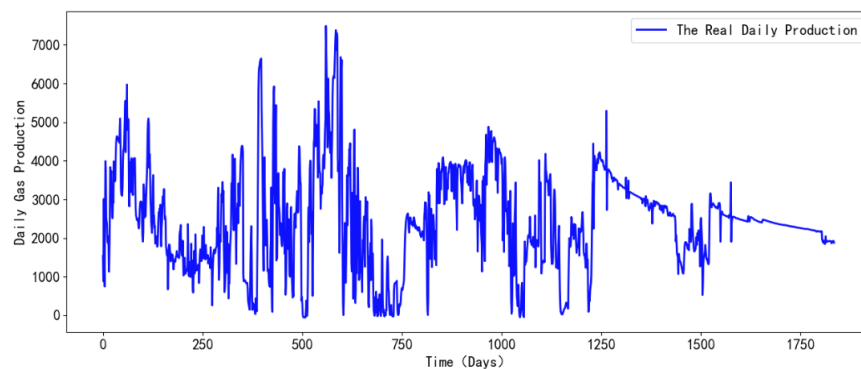
(**b**) Production data of well MIP-5H.

**Figure 6.** Production profiles of the two wells.

3.3. Data Preprocessing

3.3.1. Data Smoothing

　　Given the large oscillations and noise observed in the raw on-site data, it becomes imperative to undertake a smoothing process to mitigate the extent of fluctuation. This process makes the data more stable and conducive to later analysis and modeling. Exponential smoothing is a widely used method for time-series forecasting that assigns exponentially decreasing weights to past observations. It assumes that recent observations bear more relevance in the prediction of future values.

The exponential smoothing method computes the forecasts by combining a weighted average derived from previous observations with a smoothing factor. The smoothing factor determines the weight given to the most recent observation and governs the pace at which the relevance of preceding observations wanes. This phenomenon is articulated through the subsequent formula [38]:

$$y_t = \alpha x_t + (1 - \alpha) y_{t-1} \tag{11}$$

where y_t is the smoothing value at time (t), x_t is the actual value at time t, y_{t-1} is the smoothing value at time (t − 1), and α is the smoothing factor ranging from 0 to 1. Figure 7 shows a comparison of the daily gas production of well MIP-3H pre- and post-smoothing. The smoothed data will be utilized for the subsequent study.
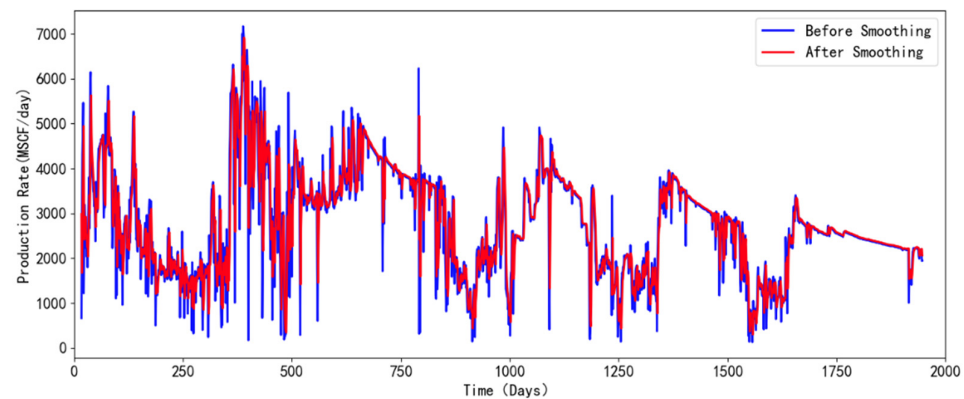
Figure 7. Contrast between daily gas production of well MIP-3H pre- and post-smoothing.

3.3.2. Data Normalization

Normalization is a widely used data preprocessing technique that aims to standardize the scale of data within a common range. It helps eliminate differences in features and bring the data to a unified scale. Several normalization methods are commonly employed, including Min-Max Scaling, Standardization, and Norm Normalization, among others. In this study, the data were normalized using Min-Max Scaling. By applying this scaling technique, the original data is transformed proportionally within the range of [0, 1] to remove any biases that might arise from the original data scale.

3.4. Prediction Accuracy Evaluation

To assess the performance and accuracy of the forecasting models, it is essential to select appropriate evaluation metrics, as some are scale-dependent while others are scale-independent [19]. Scale-dependent evaluation metrics, such as root mean squared error (*RMSE*) and mean absolute error (*MAE*), are commonly used. *RMSE* quantifies the overall fit of the model by measuring how the errors are distributed. *MAE*, on the other hand, captures the average deviation between the predicted values and the actual values. These metrics are advantageous as they utilize the same scale as the original data, enabling straightforward comparisons.

Additionally, representing the error in percentage form provides a clearer understanding of the model's performance. For this purpose, the mean absolute percentage error (*MAPE*) is utilized. *MAPE* calculates the average of the absolute values of the relative errors, offering insights into the magnitude of the error relative to the actual values.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(y_i^{pred} - y_i^{act} \right)^2}{N}} \tag{12}$$

$$MAE = \frac{\sum_{i=1}^{N} \left| y_i^{pred} - y_i^{act} \right|}{N} \tag{13}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{\left| y_i^{pred} - y_i^{act} \right|}{y_i^{act}} \right] \times 100 \qquad (14)$$

In the above equations, N is the total number of observations, y_i^{act} is the observed value at the ith position, and y_i^{pred} is the predicted value at the ith position.

The *RMSE*, *MAE*, and *MAPE* metrics are utilized to measure the accuracy of the model in forecasting the time-series production of the well. A lower value for these errors, including *RMSE*, *MAE*, and *MAPE*, indicates better model performance and greater precision in the prediction of the production data. Conversely, higher values of these errors suggest a less efficient forecasting model.

4. Results and Discussion

This section focuses on the verification and evaluation results of the LSTM, GRU, and GRU-MLP models for shale gas production prediction. To demonstrate the performance of these models, we specifically analyze two wells: MIP-3H and MIP-5H. By examining the results obtained from these wells, we can assess the accuracy and effectiveness of the LSTM, GRU, and GRU-MLP models in forecasting shale gas production.

4.1. Hyperparameter Tuning

Optimizing the architecture of neural networks represents a fundamental facet of deep learning methodologies to enhance model performance. The selection of hyperparameters within the deep learning model bears a direct influence on both the accuracy and efficiency of the ultimate model. In this study, a combination of grid search and the Particle Swarm Optimization (PSO) technique [39] was employed for the hyperparameter tuning within a 10-fold cross-validation framework. The aim was to balance prediction accuracy/convergence with potential overfitting.

For the GRU network, varying the number of hidden layers was investigated first. Increasing the number of layers improved the predictive performance but came at an additional computational cost. Fewer than three hidden layers yielded unsatisfactory results, while more than three provided similar performances on the validation datasets. Consequently, three hidden layers were chosen. The PSO was then used to determine the number of neurons in each hidden layer, dropout rate, and batch size, with predefined ranges set at [50, 300], [0.1, 0.5], and [32, 128], respectively. The PSO algorithm converged to a set of hyperparameters that resulted in the lowest average validation loss. The specific values of these hyperparameters were then utilized to configure the final neural network model. Furthermore, a series of experiments were conducted to progressively increase the number of epochs from 100 to 500 while monitoring the convergence of both training and validation losses. The most optimal performance emerged at 100 epochs.

In the MLP network, the number of hidden layers was selected through trial and error. Opting for more than one hidden layer produced subpar results, so a single hidden layer was chosen. The number of neurons within this layer was tested with 32, 64, and 128, and 32 neurons generated the best performance on the cross-validation. In addition, the choice of activation function favored ReLU due to its simpler derivative compared to Tanh, which facilitates the training process.

The fine-tuned hyperparameters for the GRU-MLP model were itemized in Table 3. In light of the possibility of suboptimal configurations of the model, forthcoming research will delve into hyperparameter tuning techniques like Bayesian optimization and nondominated sorting genetic algorithms II [40,41]. This systematic exploration aims to achieve a balance between model convergence and the prevention of overfitting.

Table 3. Fine-tuned hyperparameters for the GRU-MLP model.

Neural Network	Hyperparameter	Value
	No. of hidden layers	3
	No. of neurons in the hidden layers	[251, 192, 102]
	Dropout rate	0.1
GRU	Batch size	64
	Epochs	100
	Loss function	MSE
	Optimizer	Adam
	No. of hidden layer	1
MLP	No. of neurons in the hidden layer	32
	Activation function	Relu

4.2. Comparisons of Different Deep Learning Models

A detailed and comprehensive comparison of different methods was conducted using field shale gas production data. The outcomes indicated similar performances for both wells, thereby prompting the designation of well MIP-3H (referred to as "well1") as the illustrative case for the principal discourse. Well1 serves as an exemplar, demonstrating the approach's efficacy. In this context, 10% of the dataset (comprising the final 140 days) was allocated for testing the short-term production prediction, 20% (encompassing the last 280 days) for the medium-term prediction evaluation, and 30% (covering the last 420 days) for the long-term prediction assessment. Detailed production particulars for both wells, as well as the dataset divisions pertaining to the respective testing intervals, are outlined in Table 4.

Table 4. Production data for two wells and dataset partitioning.

Well	Well Name	Start Date	End Date	Production Time (Days)	Short-Term Prediction (Days)	Medium-Term Prediction (Days)	Long-Term Prediction (Days)
Well1	MIP-3H	12/12/2015	7/8/2021	1931	194	387	580
Well2	MIP-5H	12/11/2015	7/8/2021	1835	184	367	551

The fine-tuned GRU-MLP model was employed to predict daily gas production for well 1 (MIP-3H), leveraging both the daily gas production data and pertinent geological and engineering variables delineated in Table 2 for the predictive endeavor. Additionally, two other recurrent neural networks, specifically GRU and LSTM, were selected as benchmarks for comparison with the hybrid GRU-MLP model. To ensure parity in the comparison, the hyperparameters of two neural networks were set to align with the GRU component of the GRU-MLP model.

Table 5, accompanied by Figure 8, presents the mean prediction errors of the three neural networks with regards to gas production. The observations derived from this analysis reveal that among the considered networks, the hybrid GRU-MLP network consistently showcases diminished prediction errors in comparison to both the LSTM and GRU networks. The observations underscore the hybrid GRU-MLP model's capacity to merge the inherent strengths of the GRU and MLP models, fostering heightened accuracy in forecasting daily gas production.

Table 5 in conjunction with Figure 8 highlights the models exhibit better performances, particularly in the context of short-term predictions. This trend is in line with the common tendency in deep-learning-based models, where predictive accuracy often hinges on the volume of input samples. Figures 9–11 juxtapose the daily gas production forecasts engendered by the GRU and GRU-MLP models for well 1 (MIP-3H).

Table 5. Assessment metrics for various methods in gas production prediction.

Well	Model		*RMSE*	*MAE*	*MAPE*
Well1	LSTM	LSTM-L	181.75	134.33	7.4
		LSTM-M	134.40	95.16	5.6
		LSTM-S	102.31	85.97	3.8
	GRU	GRU-L	165.91	112.83	6.3
		GRU-M	133.87	99.30	5.2
		GRU-S	120.03	105.30	4.7
	GRU-MLP	GRU-MLP-L	151.35	81.93	5.4
		GRU-MLP-M	125.12	83.67	4.8
		GRU-MLP-S	66.73	33.59	1.6
Well2	LSTM	LSTM-L	182.70	114.54	6.0
		LSTM-M	181.46	82.84	5.0
		LSTM-S	64.30	58.23	2.7
	GRU	GRU-L	204.26	145.77	7.0
		GRU-M	178.14	80.88	4.0
		GRU-S	61.00	52.97	2.5
	GRU-MLP	GRU-MLP-L	175.06	105.15	5.0
		GRU-MLP-M	174.73	63.03	3.6
		GRU-MLP-S	59.80	50.43	2.0

L—Long-term prediction; M—Medium-term prediction; S—Short-term prediction.

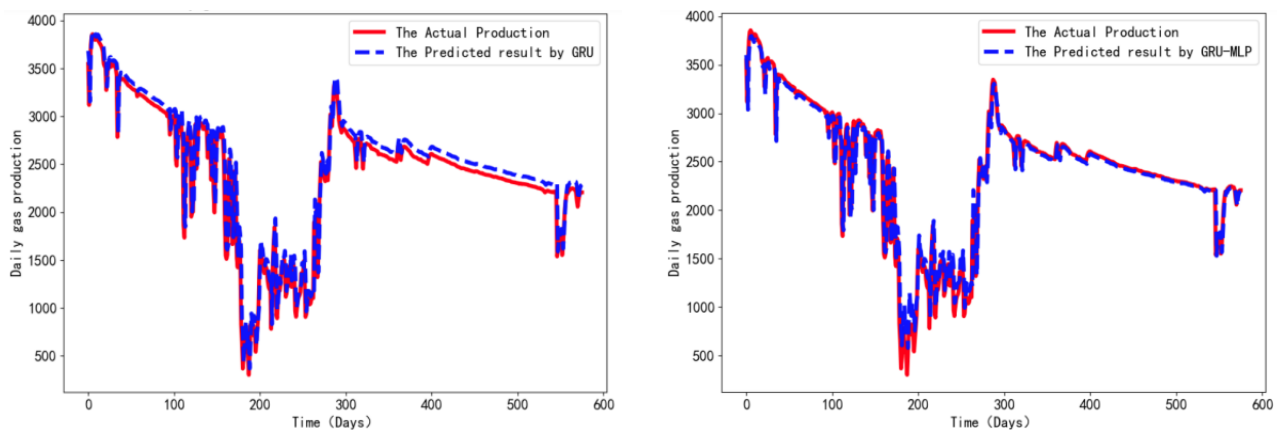(**a**)　　　　　　　　　　　(**b**)　　　　　　　　　　　(**c**)
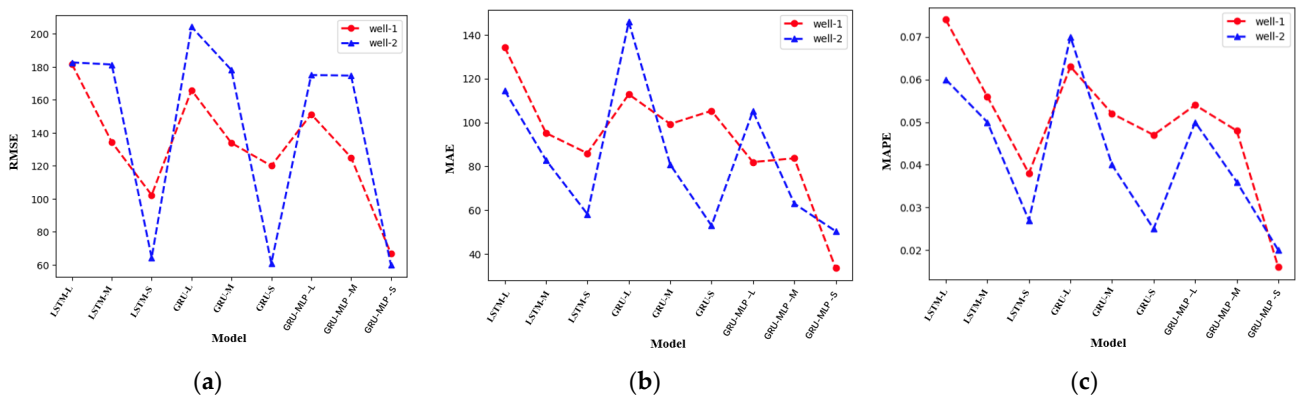
Figure 8. Evaluation metrics for diverse model performances. (**a**) RMSE values; (**b**) MAE values; (**c**) MAPE values.
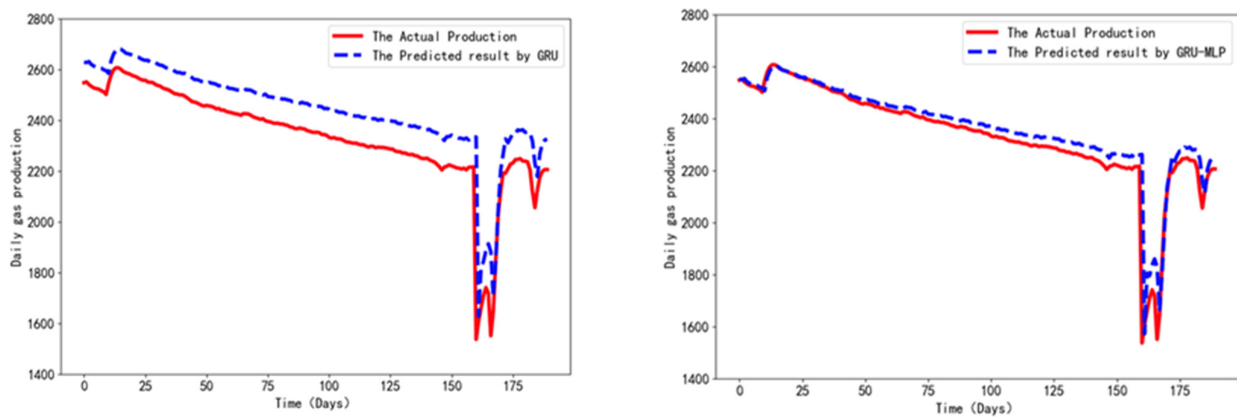
Figure 9. Comparisons of the final 194 days of production (short-term) for well 1: real data vs. GRU and GRU-MLP predictions.
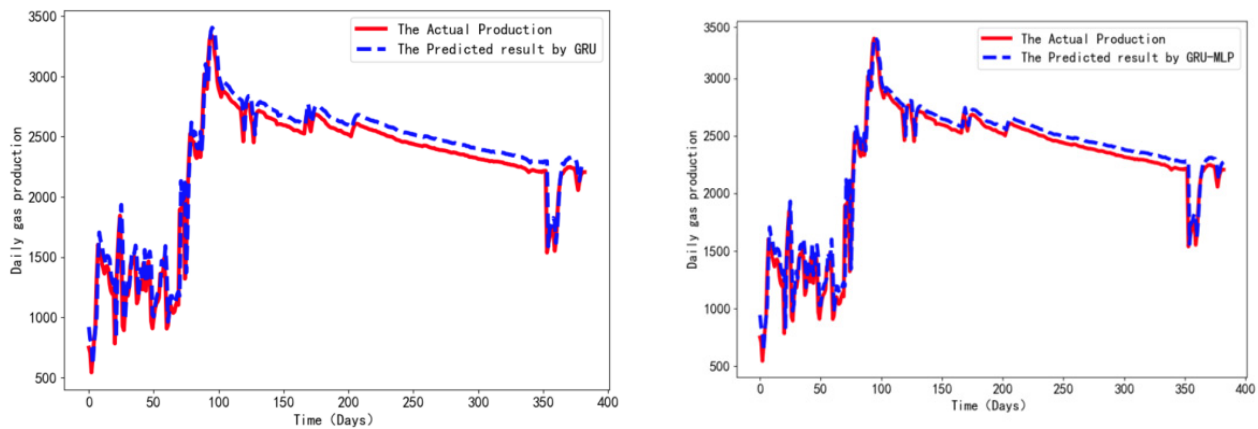
Figure 10. Comparisons of the final 387 days of production (medium-term) for well 1: real data vs. GRU and GRU-MLP predictions.

Figure 11. Comparisons of the final 580 days of production (long-term) for well 1: real data vs. GRU and GRU-MLP predictions.

Upon scrutiny of Figure 9, it becomes palpable that the short-term predicted production trajectory derived from the GRU-MLP model aligns better with the factual shale gas production. Moving further, Figures 10 and 11 spotlight the outcomes for medium-term and long-term predictions, as rendered by both the GRU and GRU-MLP models. These visual comparisons underscore the consistency between the model predictions and the in situ data, which are notably evident in the long-term predictions produced by the GRU-MLP model. This comprehensive scrutiny substantiates the efficacy of the GRU-MLP model in both short-term and long-term forecasting endeavors.

4.3. Production Prediction Using Adjacent Well

Given the proximity of well 1 (MIP-3H) and well 2 (MIP-5H), lying a mere 1700 feet apart, this analysis was undertaken to discern whether the historical production data from well 2 could effectively forecast well 1's production over a 150-day interval. The comparative assessment between well 1's actual production data and the projected outcomes derived from well 2's historical production data is elucidated in Figure 12.

When comparing the short-term production prediction using well 1's own historical production data with the predictions stemming from well 2's data through the GRU network, the accuracy of forecasting diminishes when leveraging data from the adjacent well. The prediction outcomes drawn from well 2 do not mirror the performance achieved through predictions on well 1's data.
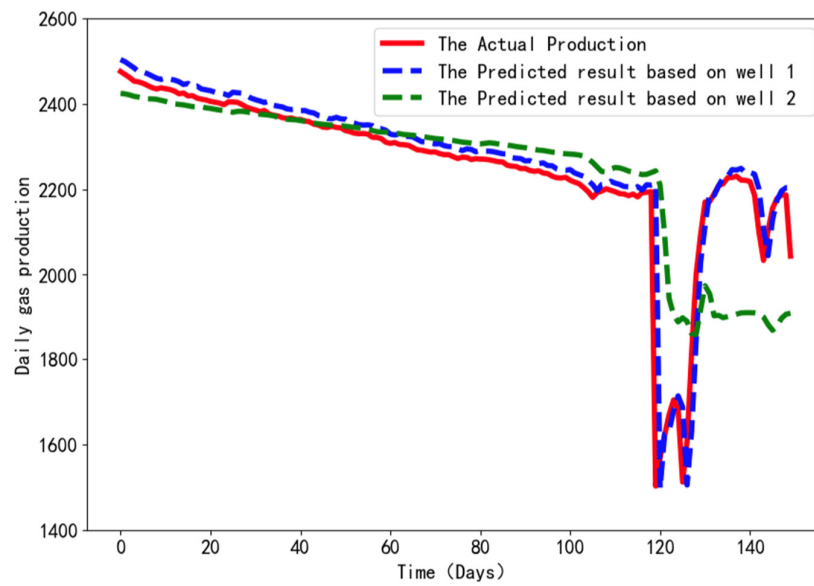
Figure 12. Comparisons of 150-day predictions using well 1 and well 2 historical data (GRU).

This distinction is particularly pronounced when one examines the *RMSE*, *MAE*, and *MAPE* metrics associated with the GRU-based prediction of the neighboring well's production. As outlined in Table 6, the *RMSE*, *MAE*, and *MAPE* values for the adjacent well's prediction are 156.97, 99.26, and 0.048, respectively. These values significantly surpass the corresponding metrics obtained through the utilization of well 1's own production data.

Table 6. Evaluation metrics using different well production data.

Predicted Well	Model	Input Variable	*RMSE*	*MAE*	*MAPE*
Well1	GRU	Well1	72.39	38.82	0.019
		Well2	156.97	99.26	0.048

5. Conclusions

In this paper, a hybrid GRU-MLP deep learning model was developed to facilitate the prediction of shale gas production. This model seamlessly integrates both the historical production data and the inherent physical constraints into its framework. The GRU network was harnessed to capture the nonstationary patterns of the production data, exploiting its long-term memory capability. Conversely, the MLP component was enlisted to unveil the intricate and multifaceted nonlinear relationships existing between the production data and the prevailing physical constraints.

Through a comprehensive analysis conducted on the field data, the proposed physics-constrained GRU-MLP model demonstrated its ability to effectively capture the intricate and dynamic patterns that characterize production sequences. Furthermore, the model successfully considered the nonlinear dependencies between geological properties, fracture geometry, fracturing treatment parameters, and production outcomes. In comparison to the original LSTM and GRU models that did not incorporate such constraints, the GRU-MLP model showcased better performances in both short-term and long-term forecasting tasks.

Extending the findings of the research to other shale reservoirs holds significant potential for advancing production forecasting. Despite originating from a specific dataset, the inherent adaptability and generalizability of the methodology make it suitable for such expansion. Exploring diverse shale reservoirs in a similar manner aims to enhance predictive precision and foster a broader comprehension of reservoir dynamics. This research endeavor aligns with the dedication to provide a versatile and transferable tool to optimize production across diverse shale reservoirs.

References

1. Middleton, R.S.; Gupta, R.; Hyman, J.D.; Viswanathan, H.S. The shale gas revolution: Barriers, sustainability, and emerging opportunities. *Appl. Energy* **2017**, *199*, 88–95. [CrossRef]
2. Wang, Q.; Li, R. Research status of shale gas: A review. *Renew. Sustain. Energy Rev.* **2017**, *74*, 715–720. [CrossRef]
3. EIA. 2011. Available online: https://www.eia.gov/analysis/studies/worldshalegas/archive/2011/pdf/fullreport.pdf (accessed on 6 March 2023).
4. Yuan, J.; Luo, D.; Feng, L. A review of the technical and economic evaluation techniques for shale gas development. *Appl. Energy* **2015**, *148*, 49–65. [CrossRef]
5. Vishkai, M.; Ian Gates, I. On multistage hydraulic fracturing in tight gas reservoirs: Montney Formation, Alberta, Canada. *J. Pet. Sci. Eng.* **2019**, *174*, 1127–1141. [CrossRef]
6. EIA. 2023. Available online: https://www.eia.gov/tools/faqs/faq.php?id=907&t=8 (accessed on 2 April 2023).
7. Wang, H. What Factors Control Shale-Gas Production and Production-Decline Trend in Fractured Systems: A Comprehensive Analysis and Investigation. *SPE J.* **2016**, *22*, 562–581. [CrossRef]
8. Liang, H.B.; Zhang, L.H.; Zhao, Y.L.; Zhang, B.N.; Chang, C.; Chen, M.; Bai, M.X. Empirical Methods of Decline-Curve Analysis for Shale Gas Reservoirs: Review, Evaluation, and Application. *J. Nat. Gas Sci. Eng.* **2020**, *83*, 103531. [CrossRef]
9. Zhao, Y.; Lu, G.; Zhang, L.; Wei, Y.; Guo, J.; Chang, C. Numerical simulation of shale gas reservoirs considering discrete fracture network using a coupled multiple transport mechanisms and geomechanics model. *J. Pet. Sci. Eng.* **2020**, *195*, 107588. [CrossRef]
10. Arps, J.J. Analysis of decline curves. *Trans* **1945**, *160*, 228–247. [CrossRef]
11. Fetkovich, M.J.; Fetkovich, E.J.; Fetkovich, M.D. Useful concepts for decline-curve forecasting, reserve estimation, and analysis. *SPE Res Eng.* **1996**, *11*, 13–22. [CrossRef]
12. Wang, L.; Wang, S.; Zhang, R.; Wang, C.; Xiong, Y.; Zheng, X.; Li, S.; Jin, K.; Rui, Z. Review of multi-scale and multi-physical simulation technologies for shale and tight gas reservoirs. *J. Nat. Gas Sci. Eng.* **2017**, *37*, 560–578. [CrossRef]
13. Cipolla, C.L.; Lolon, E.P.; Erdle, J.C.; Rubin, B. Reservoir Modeling in Shale-Gas Reservoirs. *SPE Reserv. Eval. Eng.* **2010**, *13*, 638–653. [CrossRef]
14. Wu, Y.; Li, J.; Ding, D.; Wang, C.; Di, Y. A generalized framework model for the simulation of gas production in unconventional gas reservoirs. *SPE J.* **2014**, *19*, 845–857. [CrossRef]
15. Chen, Y.; Ma, G.; Jin, Y.; Wang, H.; Wang, Y. Productivity evaluation of unconventional reservoir development with three-dimensional fracture networks. *Fuel* **2019**, *244*, 304–313. [CrossRef]
16. Berawala, D.S.; Andersen, P. Numerical investigation of Non-Darcy flow regime transitions in shale gas production. *J. Pet. Sci. Eng.* **2020**, *190*, 107114. [CrossRef]
17. Cao, C.; Jia, P.; Cheng, L.; Jin, Q.; Qi, S. A review on application of data-driven models in hydrocarbon production forecast. *J. Pet. Sci. Eng.* **2022**, *212*, 110296. [CrossRef]
18. Liang, B.; Liu, J.; You, J.; Jia, J.; Pan, Y.; Jeong, H. Hydrocarbon production dynamics forecasting using machine learning: A state-of-the-art review. *Fuel* **2023**, *337*, 127067. [CrossRef]
19. Chen, Z.; Ma, M.; Li, T.; Wang, H.; Li, C. Long sequence time-series forecasting with deep learning: A survey. *Inf. Fusion* **2023**, *97*, 101819. [CrossRef]
20. Lee, K.; Lim, J.; Yoon, D.; Hyungsik, J. Prediction of Shale-Gas Production at Duvernay Formation Using Deep-Learning Algorithm. *SPE J.* **2019**, *24*, 2423–2437. [CrossRef]
21. Sagheer, A.; Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **2019**, *323*, 203–213. [CrossRef]

22. 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]

23. Yang, R.; Liu, X.; Yu, R.; Hu, Z.; Duan, X. Long short-term memory suggests a model for predicting shale gas production. *Appl. Energy* **2022**, *322*, 119415. [CrossRef]

24. Le, N.T.; Shor, R.J.; Chen, Z. Physics-constrained deep learning for production forecast in tight reservoirs. In Proceedings of the Asia Pacific Unconventional Resources Technology Conference, Virtual, 16–18 November 2021. URTEC-208394-MS. [CrossRef]

25. Li, X.; Ma, X.; Xiao, F.; Xiao, C.; Wang, F.; Zhang, S. A physics-constrained long-term production prediction method for multiple fractured wells using deep learning. *J. Pet. Sci. Eng.* **2022**, *217*, 110844. [CrossRef]

26. Zhang, Q.; Wei, C.; Wang, Y.; Du, S.; Zhou, Y.; Song, H. Potential for Prediction of Water Saturation Distribution in Reservoirs Utilizing Machine Learning Methods. *Energies* **2019**, *12*, 3597. [CrossRef]

27. Kim, Y.D.; Durlofsky, L.J. A Recurrent Neural Network–Based Proxy Model for Well-Control Optimization with Nonlinear Output Constraints. *SPE J.* **2021**, *26*, 1837–1857. [CrossRef]

28. Huang, R.; Wei, C.; Wang, B.; Yang, J.; Xu, X.; Wu, S.; Huang, S. Well performance prediction based on Long Short-Term Memory (LSTM) neural network. *J. Pet. Sci. Eng.* **2022**, *208*, 109686. [CrossRef]

29. Yang, R.; Qin, X.; Liu, W.; Huang, Z.; Shi, Y.; Pang, Z.; Zhang, Y.; Li, J.; Wang, T. A Physics-Constrained Data-Driven Workflow for Predicting Coalbed Methane Well Production Using Artificial Neural Network. *SPE J.* **2022**, *27*, 1531–1552. [CrossRef]

30. Shi, Y.; Song, X.; Song, G. Productivity prediction of a multilateral-well geothermal system based on a long short-term memory and multi-layer perceptron combinational neural network. *Appl. Energy* **2021**, *282*, 116046. [CrossRef]

31. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef]

32. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training recurrent neural networks. In Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 16–21 June 2013. Available online: https://dl.acm.org/doi/10.5555/3042817.3043083 (accessed on 8 May 2023).

33. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

34. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [CrossRef]

35. Carr, T.R.; Wilson, T.H.; Kavousi, P.; Amini, S.; Sharma, S.; Hewitt, J.; Costello, I.; Carney, B.J.; Jordan, E.; Yates, M.; et al. Insights from the Marcellus shale energy and environment laboratory (MSEEL). In Proceedings of the Unconventional Resources Technology Conference, Austin, TX, USA, 24–26 July 2017; pp. 1130–1142. [CrossRef]

36. Bohn, R.; Hull, R.; Trujillo, K.; Wygal, B.; Parsegov, S.G.; Carr, T.; Carney, B.J. Learnings from the Marcellus Shale Energy and Environmental Lab (MSEEL) using fiber optic tools and Geomechanical modeling. In Proceedings of the Paper presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference, Virtual, 20–22 July 2020. [CrossRef]

37. EIA. Marcellus Shale Play Geology Review. U.S. Energy Information Administration, U.S. Department of Energy. 2017. Available online: https://www.eia.gov/maps/pdf/MarcellusPlayUpdate_Jan2017.pdf (accessed on 2 May 2023).

38. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Lander, E.S.; Mitzenmacher, M.; Sabeti, P.C. Detecting Novel Associations in Large Data Sets. *Science* **2011**, *334*, 1518–1524. [CrossRef]

39. Shami, T.M.; El-Saleh, A.A.; Alswaitti, M.; Al-Tashi, Q.; Summakieh, M.A.; Mirjalili, S. Particle swarm optimization: A comprehensive survey. *IEEE Access* **2022**, *10*, 10031–10061. [CrossRef]

40. White, C.; Neiswanger, W.; Savani, Y. BANANAS: Bayesian optimization with neural architectures for neural architecture search. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 18 May 2021; pp. 10293–10301. [CrossRef]

41. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [CrossRef]