electronics

MDPI

Article

Advanced Predictive Modeling of Tight Gas Production Leveraging Transfer Learning Techniques

Xianlin Ma [1,2,*], Shilong Chang [1], Jie Zhan [1,2,*] and Long Zhang [1]

[1] College of Petroleum Engineering, Xi'an Shiyou University, Xi'an 710065, China
[2] Engineering Research Center of Development and Management for Low to Extra-Low Permeability Oil & Gas Reservoirs in Western China, Ministry of Education, Xi'an Shiyou University, Xi'an 710065, China
* Correspondence: xianlinm@xsyu.edu.cn (X.M.); zhanjie@xsyu.edu.cn (J.Z.)

Abstract: Accurate production forecasting of tight gas reservoirs plays a critical role in effective gas field development and management. Recurrent-based deep learning models typically require extensive historical production data to achieve robust forecasting performance. This paper presents a novel approach that integrates transfer learning with the neural basis expansion analysis time series (N-BEATS) model to forecast gas well production, thereby addressing the limitations of traditional models and reducing the reliance on large historical datasets. The N-BEATS model was pre-trained on the M4 competition dataset, which consists of 100,000 time series spanning multiple domains. Subsequently, the pre-trained model was transferred to forecast the daily production rates of two gas wells over short-term, medium-term, and long-term horizons in the S block of the Sulige gas field, China's largest tight gas field. Comparative analysis demonstrates that the N-BEATS transfer model consistently outperforms the attention-based LSTM (A-LSTM) model, exhibiting greater accuracy across all forecast periods, with root mean square error improvements of 19.5%, 19.8%, and 26.8% of Well A1 for short-, medium-, and long-term horizons, respectively. The results indicate that the pre-trained N-BEATS model effectively mitigates the data scarcity challenges that hinder the predictive performance of LSTM-based models. This study highlights the potential of the N-BEATS transfer learning framework in the petroleum industry, particularly for production forecasting in tight gas reservoirs with limited historical data.

Keywords: attention-based LSTM; N-BEATS; production prediction; tight gas; transfer learning

Citation: Ma, X.; Chang, S.; Zhan, J.; Zhang, L. Advanced Predictive Modeling of Tight Gas Production Leveraging Transfer Learning Techniques. *Electronics* **2024**, *13*, 4750. https://doi.org/10.3390/electronics13234750

Academic Editor: Ping-Feng Pai

Received: 4 November 2024
Revised: 24 November 2024
Accepted: 29 November 2024
Published: 30 November 2024

1. Introduction

Tight gas, a form of unconventional natural gas, is trapped within rock formations characterized by matrix permeability of less than one millidarcy (mD) and porosity below 10%, often found in sandstone or limestone [1–3]. Unlike conventional reservoirs, where gas moves freely through porous media, the low permeability of tight gas reservoirs restricts flow. Hydraulic fracturing, a method involving high-pressure fluid injection to create fractures held open by proppants, is commonly employed to enhance gas flow [4–6]. With proven reserves exceeding 20 trillion cubic meters in China as of 2021, tight gas has become a critical energy resource, necessitating accurate production forecasting to optimize field development and exploitation strategies [7,8]. In addition, reliable forecasts are crucial for optimizing production techniques such as hydraulic fracturing or infill well drilling.

Production forecasting methods are typically classified into three categories: empirical production decline-curve analysis (EDCA), numerical simulation, and data-driven approaches [9–14]. Each method offers distinct advantages and limitations due to their underlying theoretical assumptions. EDCA, for example, focuses on the relationship between production rates and time without accounting for critical reservoir properties like pressure, permeability, or porosity. These empirical models are derived from historical production trends and used to estimate future production rates. EDCA is simple to apply and requires

limited production data, but its accuracy depends heavily on long-term production history and the quality of available data. Numerical simulation techniques, on the other hand, involve constructing detailed reservoir models that incorporate geological characteristics, fluid flow dynamics, and production mechanisms. For example, Zhang et al. developed an embedded discrete fracture model (EDFM) for tight reservoirs, accounting for gravity and stress sensitivity [15]. Their results showed that fracture distribution and morphology significantly impact well productivity. Fujita et al. proposed a rapid flow simulator based on the concept of diffusive time of flight (DTOF) for shale gas reservoirs, enhancing flow characterization through efficient calculation of drainage volumes [16]. Li et al. introduced a nonlinear flow model for tight sandstone reservoirs, demonstrating improved production accuracy compared to traditional methods [17]. Despite offering high accuracy and detailed insights, numerical simulation methods often require extensive data and computational resources, which may not always be feasible.

Deep learning approaches, particularly recurrent neural networks (RNNs) like long short-term memory (LSTM) networks and gated recurrent units (GRU), have emerged as promising tools for production forecasting [18–20]. These methods treat forecasting as a time series problem, capturing intricate temporal dependencies and patterns in historical data. Recently, the integration of attention mechanisms with RNNs has shown promising results, allowing for more accurate long-term forecasts by focusing on the most relevant segments of the time series data. An LSTM model enhanced with an attention mechanism has been specifically developed for petroleum production forecasting [21]. This fusion has demonstrated substantial improvements in prediction accuracy and robustness. Comparative studies reveal that attention-based networks outperform traditional models in prediction accuracy for petroleum production forecasts. However, the performance of these methods heavily relies on the availability of extensive historical data, presenting challenges in data-scarce environments like tight gas reservoirs.

To address data limitations, transfer learning has been proposed as a viable solution [22–25]. By leveraging pre-trained models from related domains, transfer learning enables the application of learned knowledge to new tasks with limited data. The N-BEATS model, introduced by Oreshkin et al. (2019), stands out as a state-of-the-art deep learning architecture for time series forecasting, excelling in benchmark datasets like M4 and tourism datasets by effectively capturing temporal patterns without requiring feature engineering [26]. Prior research, such as Alali and Horne's work on oil rate forecasting in the Volve field, demonstrated the N-BEATS model's ability to transfer knowledge from the M4 dataset, outperforming LSTM models even under sparse data conditions [27]. Their study emphasized the model's capability to produce accurate forecasts without input variables like reservoir pressure, making it especially useful for green fields with limited historical data.

While this earlier work focused on oil production, our study extends the N-BEATS framework to tight gas reservoirs, which present distinct challenges, such as ultralow permeability, sparse production history, and pronounced seasonal fluctuations. Gas production is subject to greater seasonal variability than oil production, influenced by demand changes and operational constraints during extreme weather. By adapting the N-BEATS model through transfer learning for forecasting tight gas production in the S block of the Sulige gas field, we address these unique complexities. This work pioneers the application of N-BEATS to gas production, showcasing its ability to capture both long-term trends and short-term variations in a challenging domain, thus advancing data-driven methods for tight gas reservoir forecasting.

The remainder of this paper is organized as follows: Section 2 provides an overview of the study area and details the data preparation process. In Section 3, we introduce the two deep learning models used in this study: the attention-based LSTM and the N-BEATS model, while also outlining the production prediction framework incorporating transfer learning. Section 4 presents an in-depth analysis of the results for short-, medium-, and long-term forecasting. Finally, Section 5 summarizes the key findings and conclusions.

2. Data Preparation

2.1. Background of the Study Area

The study area, known as S block, is situated in the central region of the Sulige gas field, within the Ordos Basin, and is adjacent to Wushenqi, as shown in Figure 1. Structurally, S block lies in the middle of the Yishan slope. The average matrix porosity of S block is 6.2%, while the porosity of the gas-bearing intervals ranges from 5% to 15%. The average permeability is 0.14 mD. As of 2023, in the S block there are over 1500 fractured gas producers, with more than 80% of them being vertical wells [28].

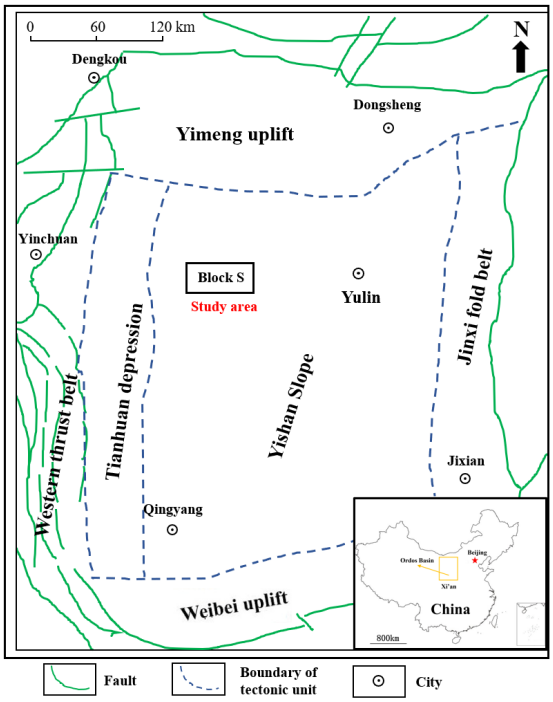

Figure 1. The study area in the Sulige gas field.

For this study, two vertically fractured wells, A1 and A2, were selected from the S block. Both wells have daily production records from 2011 to 2016. The key reservoir properties of these wells are summarized in Table 1.

Table 1. Reservoir properties of gas wells.

Well Name	Vertical Depth (m)	Reservoir Thickness (m)	Permeability (mD)	Porosity (%)	Gas Saturation (%)
A1	3668	2.9	0.22	8.1	72
A2	3649	3.5	0.49	10.6	55

Due to confidentiality restrictions concerning actual production data, we normalized the production data using Min-Max normalization. The formula is as follows:

$$y_{norm} = \frac{y - y_{min}}{y_{max} - y_{min}} \quad (1)$$

where y represents the actual data, y_{norm} denotes the normalized value, and y_{min} and y_{max} are the minimum and maximum values of the production data, respectively. The chosen normalization method was selected for its simplicity and effectiveness in preserving the distribution shape of the original data while making it suitable for models that rely on numerical stability.

Figure 2 illustrates the normalized daily gas production over time for Wells A1 and A2, with the production data scaled linearly to the range [0, 1].

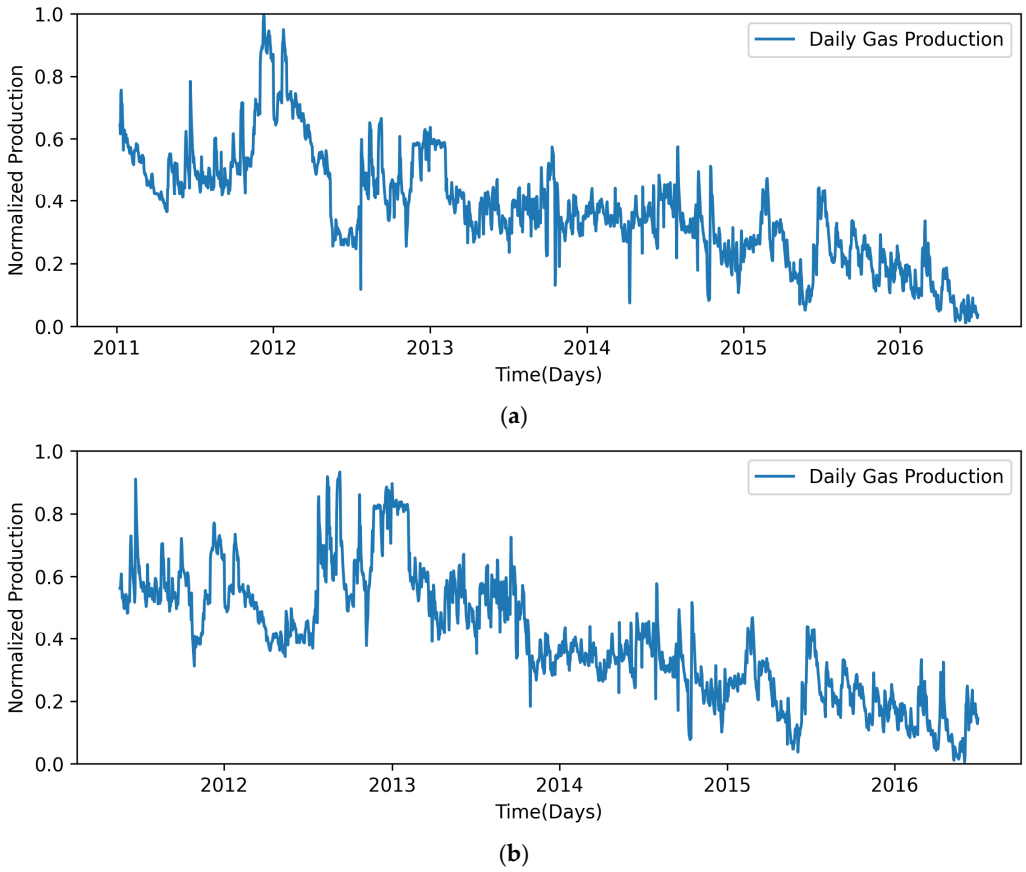

Figure 2. Scaled production profiles of the two gas wells. (**a**) Daily production data of Well A1. (**b**) Daily production data of Well A2.

Figure 2 illustrates the seasonal fluctuations in daily gas production for both wells, showing a decline during the summer followed by an increase in the winter. This pattern is attributed to lower natural gas demand in the summer, leading to reduced production and a subsequent rise in production as demand increases during the winter months.

2.2. Data Smoothing

Production data from gas wells often contain random noise and seasonal fluctuations, which can compromise the accuracy of prediction models. To mitigate these issues, data smoothing techniques are employed to reduce noise and stabilize the time series. One effective method is exponential smoothing, which assigns exponentially decreasing weights to older data points [29]. The method is particularly well-suited for time series data exhibiting trends and seasonal variations, such as tight gas production, where smoothing facilitates better identification of underlying patterns. By assigning exponentially decreasing weights to older observations, the method ensures that recent data points have a greater influence, which is crucial for capturing real-time production trends and reducing the impact of abrupt fluctuations or outliers. The formula for exponential smoothing is as follows:

$$S_t = \alpha y_t + (1 - \alpha) S_{t-1} \tag{2}$$

where S_t is the smoothed value at time t, y_t is the actual value at time t, S_{t-1} is the smoothed value at the previous time step, and α is the smoothing factor, ranging between 0 and 1. The smoothing factor α, which determines the weight assigned to recent data points versus historical values, was selected based on its ability to balance short-term trends and

long-term patterns. A smaller α gives more weight to historical data, reducing noise but potentially overlooking rapid changes, while a larger α prioritizes recent data, capturing fluctuations but possibly amplifying noise. In this study, α was determined empirically by testing multiple values during the preprocessing stage. The final selection was guided by aligning smoothed data with actual production trends without over-smoothing critical transitions. This approach ensures that key production patterns are preserved while irrelevant noise is filtered out.

Figure 3 compares the daily gas production data before and after smoothing for Wells A1 and A2, demonstrating the impact of smoothing on data stability.

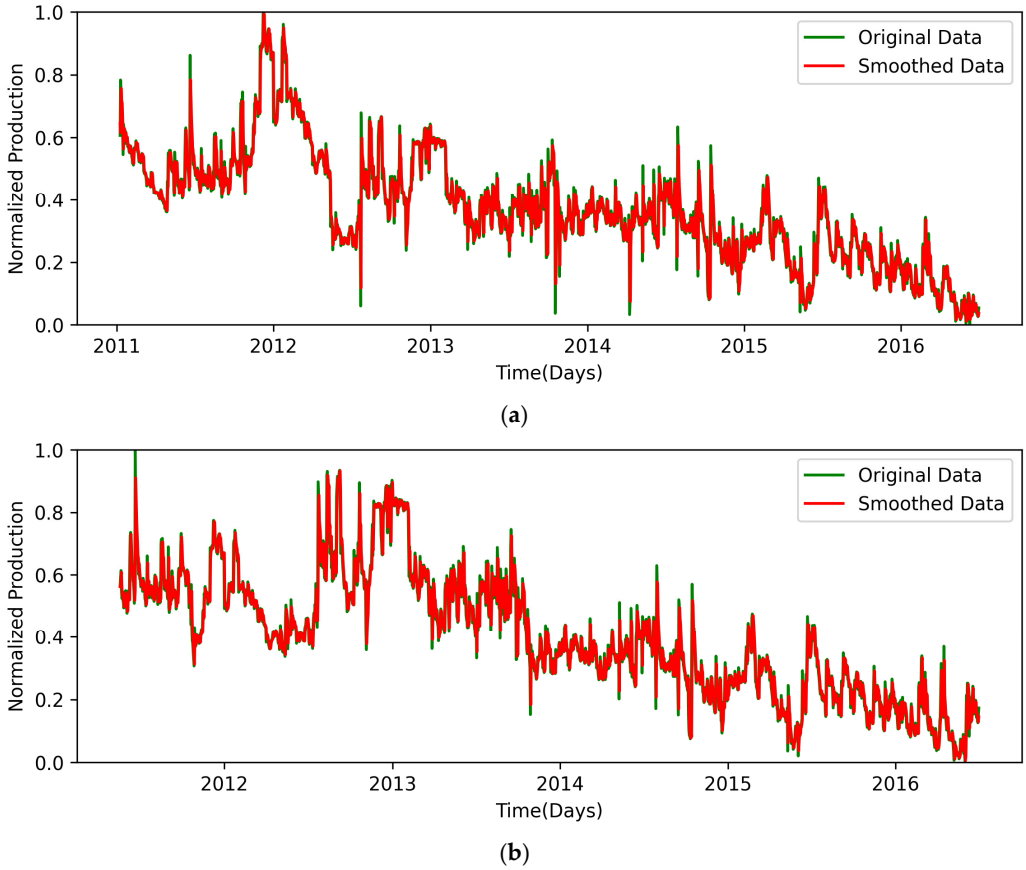

Figure 3. Normalized daily gas production rate of Wells A1 and A2 before and after smoothing. (**a**) Daily gas production rate of Well A1 before and after smoothing. (**b**) Daily gas production rate of Well A2 before and after smoothing.

Other potential smoothing techniques, such as moving average smoothing or Gaussian smoothing, were considered but not applied. Moving average smoothing, while effective in reducing high-frequency noise, often distorts trend information in data with seasonal components. Similarly, Gaussian smoothing, although useful for localized data smoothing, might over-smooth abrupt changes in production, which are vital for accurate forecasting. Exponential smoothing strikes a balance by retaining critical trends and seasonal variations while minimizing noise, thereby serving as the most suitable technique for the dataset used in this study.

3. Deep Learning Model for Production Prediction

In this study, two deep learning models, an attention-based long short-term memory (A-LSTM) model and a neural basis expansion analysis for time series (N-BEATS) transfer model, were employed for production prediction. The primary distinction between the two models lies in their architecture and mechanisms for handling sequential data. The

LSTM model leverages hidden states to retain information from previous steps, enabling the network to learn long-term dependencies. The integration of attention mechanisms enhances this capability by dynamically weighting past data, allowing the model to focus on the most relevant points within the input sequence, which improves performance over vanilla LSTMs in capturing long-range dependencies.

Conversely, the N-BEATS model employs a fully connected architecture with blocks designed to capture different patterns within the data. Each block contributes to the final prediction, enabling a robust, pattern-focused approach. The N-BEATS model used in this study was pre-trained on the M4 competition dataset, which consists of over 100,000 time series from various domains. After pre-training, the model was transferred to forecast daily gas production rates, demonstrating the advantages of transfer learning.

3.1. Long Short-Term Memory

The LSTM, a variant of the recurrent neural network (RNN), addresses the vanishing gradient problem encountered in traditional RNNs, making it suitable for time series forecasting [30]. LSTMs feature three main components: the forget gate, the memory selection phase, and the output phase. During the forget phase, the model selectively decides which information from the previous node should be discarded by reading the current input and previous neuron information, with a forget gate determining what to ignore. In the memory selection phase, the network decides what information to retain and update, using a sigmoid layer to regulate new information and a tanh layer to generate a candidate value to add to the memory state. Finally, in the output phase, the network decides the current output by scaling the memory state using the tanh function, and the transformed value is passed as the output of the current time step [31].

In the forget phase of LSTM, the model selectively discards information from the previous time step. It reads both the current input and the previous state, determining which aspects of the past state should be ignored using a forget gate f_t. In the memory selection phase, the input is processed, with a sigmoid layer acting as an input gate to determine what new information should be stored. A tanh layer generates a new candidate value to be added to the state. The output phase decides which information from the current state will contribute to the next time step, scaling the cell state via the tanh function and producing the final output.

$$f_t = \sigma\Big(W_f \cdot [h_{t-1}, x_t] + b_f\Big) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C}_t \tag{6}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{8}$$

where W represents the weight matrices and b represents the bias vectors. The variables f, i, C, and o correspond to the forget gate, input gate, memory cell, and output gate, respectively. The function σ denotes the sigmoid activation function.

3.2. Attention-Based LSTM

Standard LSTMs can struggle with long-range dependencies, especially in long sequences, as all inputs are treated with equal importance over time. The attention-based LSTM (A-LSTM) enhances time series forecasting by selectively focusing on the most important elements of the input sequence. After the LSTM processes the input, an attention mechanism calculates attention scores by comparing the current hidden state with previous ones. The attention layer, comprising an encoder and decoder, calculates similarity scores

to assign attention weights [32,33]. These weights determine which parts of the input sequence are most relevant, improving the model's ability to capture long-term dependencies and variability, as shown in Figure 4.

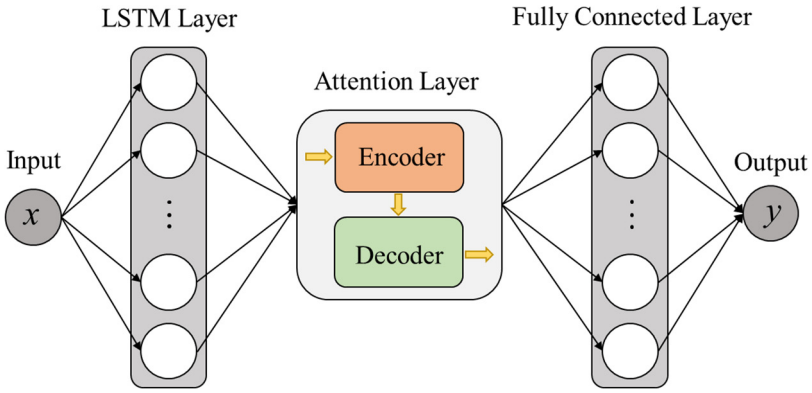

Figure 4. Structure of the A-LSTM model.

Internally, the attention mechanism operates with an encoder-decoder structure. The encoder processes the input sequence and generates feature representations, while the decoder produces outputs based on these representations and the current hidden state, as shown in Figure 5. Attention weights are calculated by comparing the encoder's hidden states with the decoder's current state, allowing the model to prioritize relevant data points. These weights are then applied to the feature representations to generate attention values, which serve as inputs to the decoder. By focusing on key elements of the sequence, A-LSTM improves the model's ability to capture long-term dependencies, enhancing prediction accuracy, especially in complex or highly variable time series data.

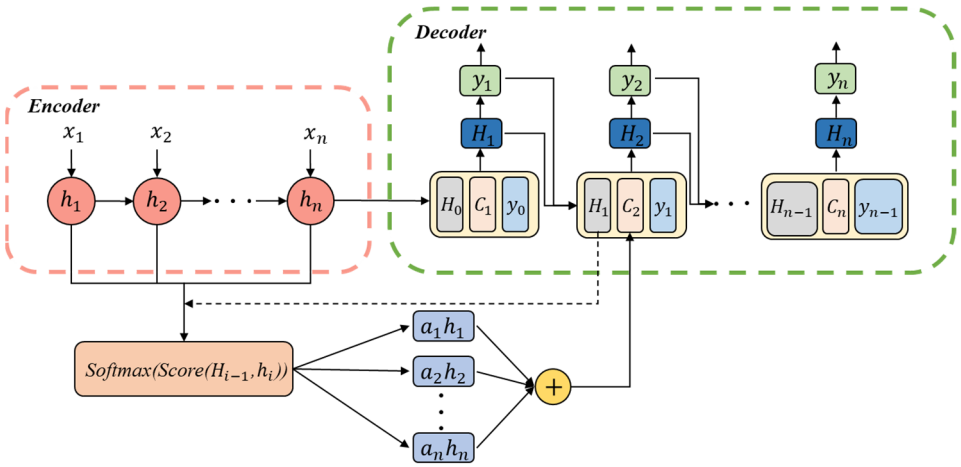

Figure 5. Structure of the attention mechanism.

The calculations for the attention mechanism are defined by the following equations:

$$\alpha_i = Softmax(Sim_i) = \frac{\exp(Sim_i)}{\sum_{j=1}^{Lx} \exp(Sim_j)} \tag{9}$$

$$Sim_i = Score(H_{i-1}, h_i) = H_{i-1} W h_i \tag{10}$$

$$C_i = \sum_{i=1}^{Lx} \alpha_i h_i \tag{11}$$

$$H_i = f(H_{i-1}, y_{i-1}, C_i) \tag{12}$$

where α_i represents the attention weight at the ith input position, Sim_i represents the similarity score at the i-th input position, H_{i-1} is the decoder hidden state from the previous time step, h_i is the encoder hidden state at the ith input position, W is the trainable weight matrix, y_{i-1} is the decoder output from the previous time step, and C_i is the context vector at the current time step.

By explicitly modeling which parts of the sequence are most relevant, A-LSTM improves accuracy in forecasting tasks, particularly when the sequence contains significant variability or long-term dependencies.

3.3. N-BEATS

N-BEATS is a deep learning model tailored for time series forecasting, utilizing a block-wise architecture to handle sequential data, as shown in Figure 6. Each block in the N-BEATS structure performs two main functions: backcasting and forecasting. The backcast aims to reconstruct the historical portion of the time series, while the forecast predicts future values. After each block makes its backcast, the residual error between the original input and the backcast is calculated, and this residual is passed to the next block. The iterative process allows successive blocks to refine the residuals and improve the overall prediction.

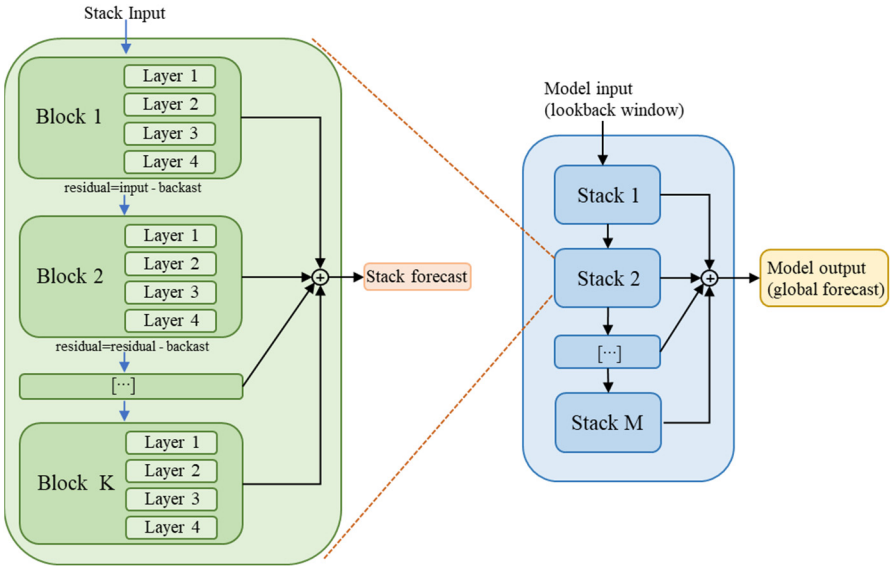

Figure 6. N-BEATS model architecture.

Unlike traditional recurrent neural networks, N-BEATS does not use recurrent layers. Instead, it relies on fully connected layers, followed by a basis expansion layer, which is responsible for learning different components like trends and seasonality. This flexible, non-parametric approach allows the model to capture a wide range of patterns without assuming a specific structure in the data. The simplicity of its design, combined with its ability to model non-linear trends, makes N-BEATS highly effective for time series forecasting.

3.4. N-BEATS Transfer Model

As shown in Figure 7, the N-BEATS model was pre-trained on the M4 dataset, which comprises over 100,000 time series from various domains and frequencies, including data commonly encountered in business, finance, and economic forecasts, sampled from hourly to yearly frequencies. It was designed for the M4 competition, a significant event in the forecasting field aimed at advancing and comparing prediction methods. Known for its scale and diversity, this dataset serves as a primary resource for developing and testing both traditional and modern forecasting models. By leveraging transfer learning through

shared N-BEATS model parameters and weights learned from the M4 sequences, we apply this approach to predict production of the gas wells in the Sulige gas field.

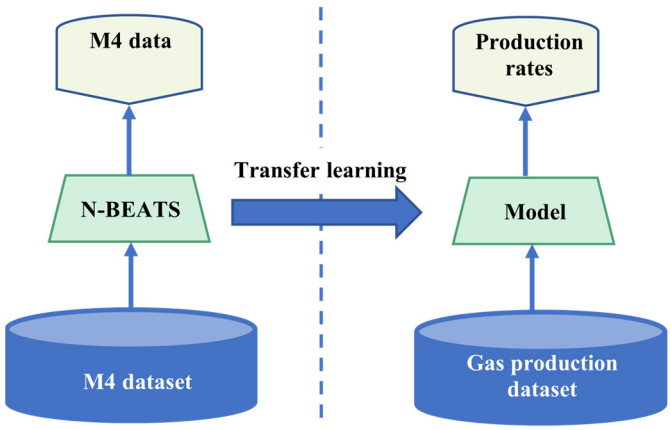

Figure 7. Production prediction using the pretrained N-BEATS transfer model.

3.5. Sliding Window Technique

In this study, both models utilize sliding window techniques, which involve two key parameters: window size and prediction horizon [34]. The window size determines how many previous time steps are considered, while the prediction horizon specifies the number of future time steps to be forecasted. For instance, as illustrated in Figure 8, if the window size is set to 3 and the prediction horizon is 1, production data from the previous 3 days is used to predict the production for the 4th day. Afterward, the sliding window shifts forward by one day, using data from days 2 to 4 to predict production on the 5th day. This iterative process continues, with the sliding window advancing one step at a time until the entire time series data is segmented, and predictions are made.

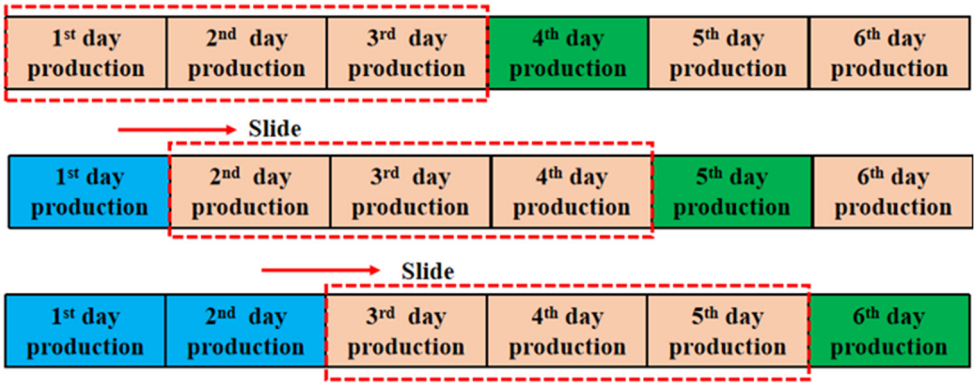

Figure 8. Sliding window technique.

When applying sliding window techniques, it is essential to determine the appropriate window size and sliding step length according to the specific requirements of the problem. The window size must be large enough to capture sufficient historical information without being overly large, which could lead to information dilution. To identify the optimal window size for the N-BEATS and A-LSTM models, we employed a sliding (or rolling) 10-fold cross-validation strategy for the gas production data. The training dataset, comprising 1600 days of production data for Well A1, was divided into 10 folds, each containing 160 days. For each fold, the data was further split into a training set and a validation set. Various window sizes (3, 5, 7, 9) were systematically tested, and the corresponding RMSE values were calculated for each window size using the A-LSTM model. As shown in Figure 9, a window size of 5 delivered the lowest RMSE, indicating superior predictive performance compared to other tested sizes.

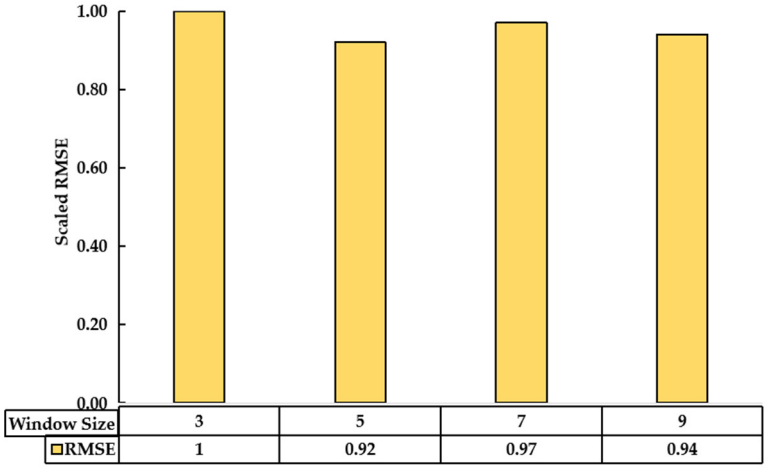

Figure 9. Comparison of scaled RMSE across different window sizes.

Although the optimal window size of 5 was determined using the A-LSTM model, this size was also applied to the N-BEATS model. The decision to use the same window size was motivated by the need for consistency in evaluating and comparing the two models under identical input configurations. This approach ensures a fair and meaningful comparison of their forecasting performances, eliminating potential biases arising from differing window size settings. Additionally, the chosen window size aligns well with the dataset's characteristics, which feature moderate variability and seasonal fluctuations, further supporting its suitability for both models.

3.6. Evaluation Metrics

To evaluate the performance and accuracy of prediction models, selecting appropriate evaluation metrics is crucial. Some metrics are scale-dependent, while others are not. In this study, we employ both scale-dependent metrics and percentage-based error metrics, as they share the same units as the original data, allowing for convenient and meaningful comparisons. The most commonly used scale-dependent evaluation metrics are mean absolute error (MAE) and root mean square error (RMSE) [35].

MAE measures the average deviation between predicted and actual values by taking the arithmetic mean of absolute errors, thus reflecting prediction accuracy on the same scale as the original data. RMSE, on the other hand, provides a more sensitive measure of error by giving greater weight to larger deviations, which helps assess the model's overall fit or error distribution. Errors expressed in percentage terms can offer clearer insight into model performance. Mean absolute percentage error (MAPE), for instance, calculates the average absolute error as a percentage of actual values, providing an intuitive interpretation of the model's relative accuracy. The coefficient of determination (R^2) is a widely used metric that quantifies how well the independent variables explain the variance in the dependent variable.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{14}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N} \left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\% \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y})^2} \tag{16}$$

where N represents the total number of data samples, y_i denotes the true value at the ith position, $\hat{y}_i$ represents the predicted value at the ith position, and $\overline{y}$ is the mean of the true values.

In this study, we assess the accuracy of gas well production forecasts using RMSE, MAE, MAPE, and R^2. Lower values of RMSE, MAE, and MAPE suggest superior model performance and higher precision in forecasting production. Conversely, higher error values indicate reduced predictive accuracy. The R^2 value, ranging from 0 to 1, serves as an additional indicator of model quality, with values closer to 1 denoting a stronger fit to the observed data.

4. Results and Discussion

In this section, we conducted a detailed analysis of two gas wells, A1 and A2, located in the S block of the Sulige gas field. We applied the A-LSTM model and the N-BEATS transfer model to forecast the natural gas production for these wells across three-time horizons: short-term, medium-term, and long-term. The last 10% of production data were reserved for short-term prediction, the next 20% for medium-term, and the following 30% for long-term forecasting. Table 2 provides a detailed breakdown of production specifics for these wells and the corresponding dataset partitions used for each evaluation period. By comparing these forecasts, we assessed the accuracy and performance of both models in predicting well production.

Table 2. Production data for two wells and dataset partitioning.

Well Name	Production Time (Days)	Short-Term Prediction (Days)	Medium-Term Prediction (Days)	Long-Term Prediction (Days)
A1	2000	200	400	600
A2	1870	187	347	561

4.1. Hyperparameter Tuning

Hyperparameters define the operational characteristics of a machine learning model and must be set before training begins. Each ML algorithm involves a range of hyperparameters that need to be optimized for the best performance. Initially, we used grid search to determine optimal values for key hyperparameters in the A-LSTM model, such as learning rate, number of neurons, layers, and training epochs. However, the initial fit to the production data of the two wells was suboptimal. As a result, we resorted to manual fine-tuning, iteratively adjusting network parameters. Several experiments were conducted to determine the optimal number of iterations, gradually increasing from 100 to 500, while monitoring the convergence of both training and cross-validation losses. The model exhibited its best performance after 100 iterations. The final hyperparameters are listed in Table 3. The N-BEATS transfer model was pre-trained on the M4 time series dataset using the hyperparameters listed in Table 4. The model featured 4 layers and 20 stacks, and training was conducted on a GPU machine. For the loss function, we used symmetric mean absolute percentage error (sMAPE), which calculates the average symmetrical difference between predicted and actual values, making it effective for time series data.

The selection of MSE for A-LSTM and sMAPE for N-BEATS reflects their alignment with the data characteristics and model objectives. MSE is ideal for A-LSTM as it calculates errors based on absolute values, which is appropriate for normalized data within the [0, 1] range. This normalization ensures consistent loss computation without being influenced by scale differences, making MSE particularly effective for datasets with uniform units and reducing the impact of outliers. On the other hand, sMAPE evaluates relative differences between predicted and actual values, making it better suited for the diverse value ranges and seasonal fluctuations inherent in tight gas production data. This relative error measure allows N-BEATS to excel in capturing subtle variations and abrupt changes, leveraging its architectural strengths in both global and local temporal pattern recognition.

Table 3. Key parameters of the A-LSTM model.

Parameter	Values
Hidden nodes in in a single LSTM layer	96
Attention layer	1
Learning rate	0.0005
Batch size	128
Optimizer	Adam
Activation function	ReLU
Loss function	MSE

Table 4. Key parameters of the N-BEATS transfer model.

Parameter	Values
Number of stacks	20
Number of blocks	3
Number of layers in a block	4
Layer width	64
Expansion coefficient	11
Learning rate	0.0001
Optimizer	Adam
Loss function	sMAPE
Activation function	ReLU

Note: sMAPE $= \frac{200}{N}\sum_{i=1}^{N}\frac{|y_i-\hat{y}_i|}{|y_i|+|\hat{y}_i|}$.

After pretraining, we employed a transfer learning approach to forecast the gas rates. By transferring the learned model parameters and weights from the M4 series, we applied the pretrained model to predict the gas rates of Wells A1 and A2, demonstrating the transfer learning model's robustness and adaptability.

4.2. Comparative Analysis

To perform long-term, medium-term, and short-term production forecasts for Wells A1 and A2 in the S block of the Sulige gas field, we utilized the A-LSTM and N-BEATS transfer models. To ensure experimental consistency, we standardized the number of iterations, batch sizes, and learning rates for both models. The performance metrics for predicting natural gas production in these wells are summarized in Table 5.

Table 5. Comparison of the N-BEATS transfer model with A-LSTM across three forecast periods.

Well	Model	RMSE	MAE	MAPE (%)	R^2
A1	A-LSTM-L	235	144	20.6	0.86
	A-LSTM-M	258	161	23.4	0.83
	A-LSTM-S	266	175	26.1	0.81
	N-BEATS-L	172	106	5.9	0.92
	N-BEATS-M	207	137	6.1	0.91
	N-BEATS-S	214	142	7.1	0.86
A2	A-LSTM-L	231	132	17.1	0.86
	A-LSTM-M	247	157	22.3	0.85
	A-LSTM-S	272	183	28.4	0.78
	N-BEATS-L	225	120	6.5	0.89
	N-BEATS-M	238	132	6.9	0.88
	N-BEATS-S	257	173	8.2	0.81

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

As shown in Table 5 and Figures 10–13, the N-BEATS transfer model consistently outperforms the A-LSTM model across long-term, medium-term, and short-term forecasts for both Wells A1 and A2. The N-BEATS transfer model demonstrates lower MAPE and higher R^2 values, indicating greater accuracy in predicting natural gas production.

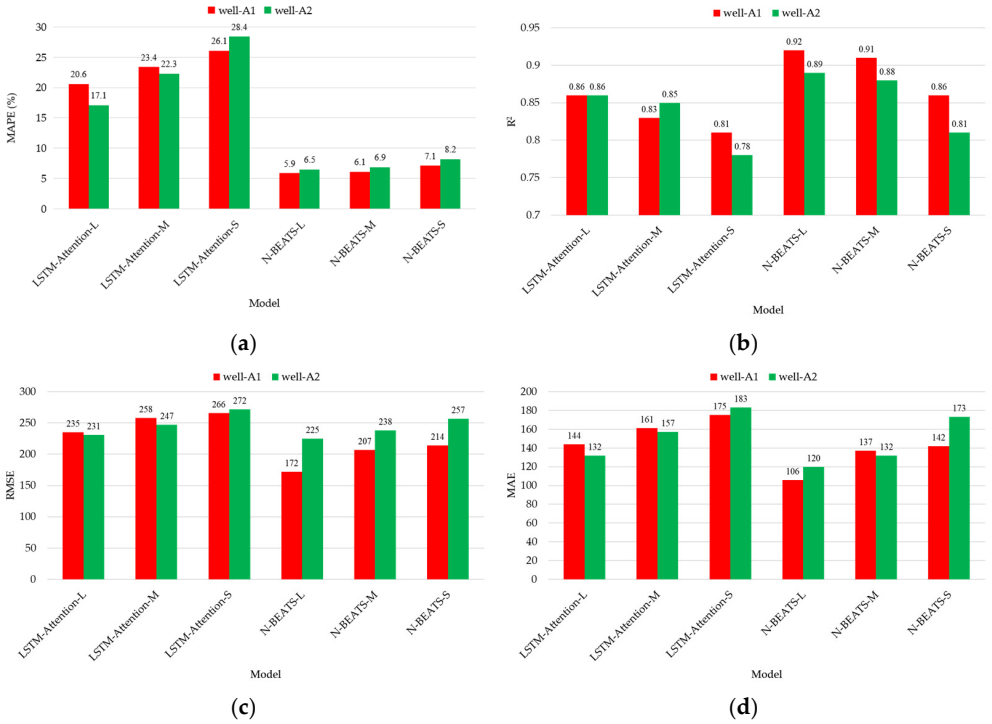

Figure 10. Various metrics across three forecast periods: (**a**) MAPE; (**b**) R^2; (**c**) RMSE; (**d**) MAE.

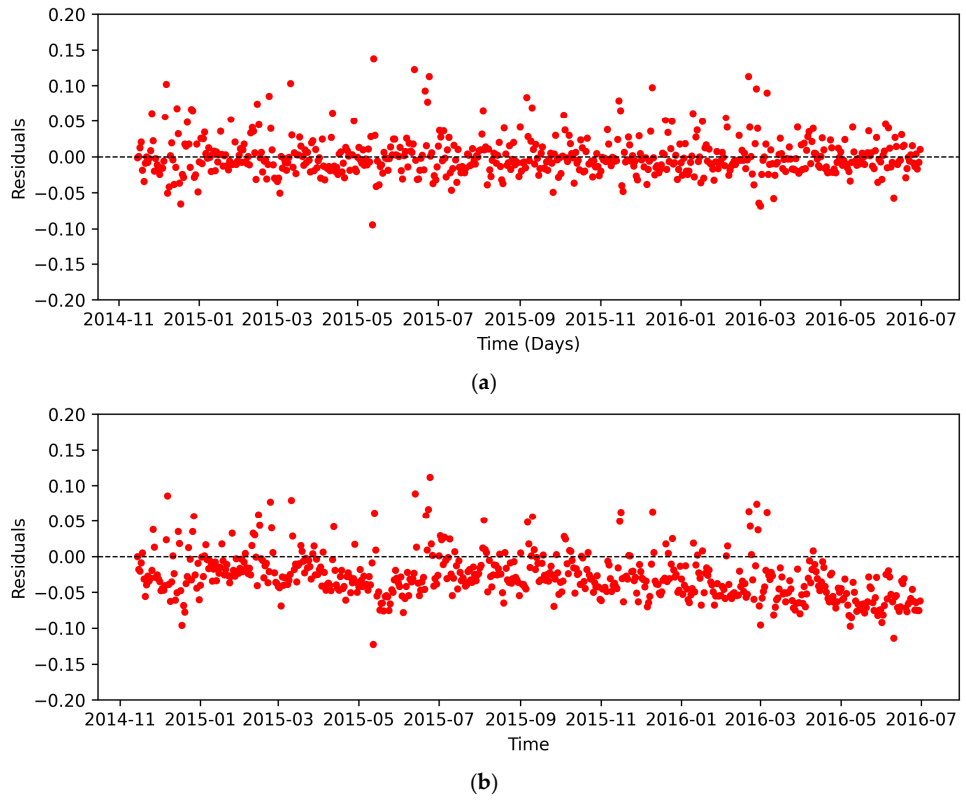

Figure 11. Comparison of residuals for long-term forecasts of Well A1. (**a**) Residuals generated by the N-BEATS transfer model. (**b**) Residuals generated by the A-LSTM model.

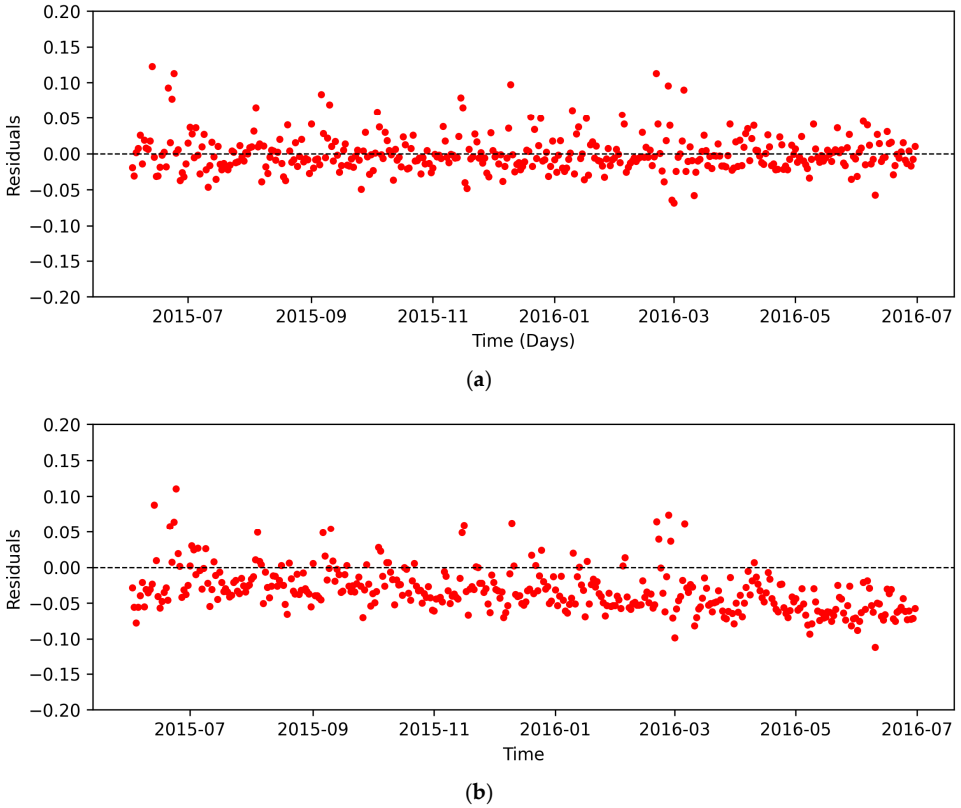

Figure 12. Comparison of residuals for medium-term forecasts of Well A1. (**a**) Residuals generated by the N-BEATS transfer model. (**b**) Residuals generated by the A-LSTM model.

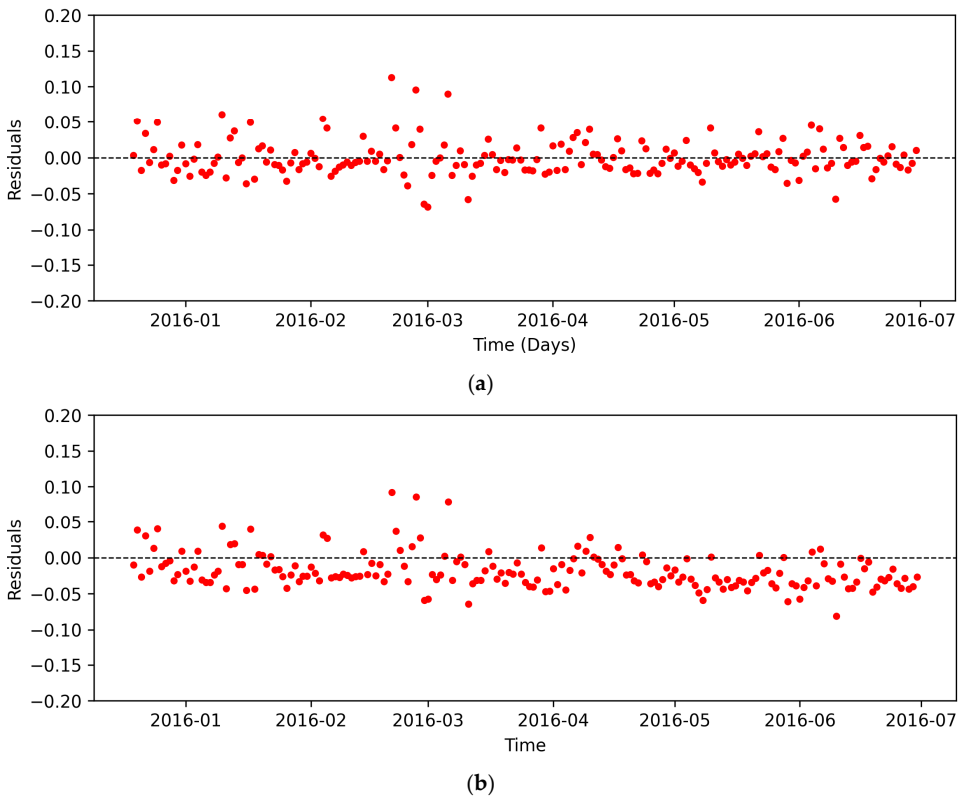

Figure 13. Comparison of residuals for short-term forecasts of Well A1. (**a**) Residuals generated by the N-BEATS transfer model. (**b**) Residuals generated by the A-LSTM model.

Figures 11–13 show the residuals between the actual daily gas production rates of Well A1 and the predicted values over the three forecast periods. The trends observed in these figures are consistent with the results in Table 5, confirming that the N-BEATS transfer model provides superior accuracy in production forecasting compared to the A-LSTM model.

For the long-term production forecast of Well A1, Figure 11 presents a comparison of residuals produced by the N-BEATS transfer model and the A-LSTM model. The visual analysis highlights the consistently lower residuals achieved by the N-BEATS transfer model, indicating its superior capability in capturing the dynamics of daily gas production. This comparative view reinforces the quantitative findings, showcasing the enhanced accuracy and robustness of the N-BEATS model over the A-LSTM model in production forecasting. This performance is largely due to its ability to capture non-linear patterns and seasonality in the time series, which are particularly important in long-term forecasts, where variability tends to be more pronounced.

These comparative analyses demonstrated that the superior performance of the N-BEATS transfer model over the A-LSTM arises from both the unique characteristics of the tight gas production data and the architectural advantages of N-BEATS. Tight gas production data often exhibit seasonal fluctuations and non-linear trends due to factors such as changing seasonal demand and operational dynamics. N-BEATS excels in addressing these complexities by explicitly separating trend and seasonal components through specialized blocks for backcasting and forecasting. This contrasts with A-LSTM, which relies on implicit feature extraction and sequential processing, making it less effective in disentangling intricate seasonal patterns. Another critical advantage of N-BEATS lies in its ability to transfer knowledge from large external datasets, as demonstrated by its pre-training on the M4 dataset. This transfer learning capability allows N-BEATS to generalize effectively to tight gas production data, addressing challenges of data sparsity, a scenario where A-LSTM's dependency on extensive historical data often limits its performance.

The findings of this study have significant industrial implications, particularly in enhancing gas field management practices. By demonstrating the superiority of the transfer learning-based N-BEATS model over traditional recurrent models, the study provides a robust framework for accurate production forecasting in tight gas reservoirs. This capability enables operators to optimize resource allocation and operational planning, such as managing well shut ins. The model's ability to perform effectively with limited historical data addresses a critical challenge in newly developed reservoirs, reducing the dependency on extensive field data collection. Furthermore, accurate forecasting across short-, medium-, and long-term horizons aids in financial planning and risk assessment, empowering stakeholders to make informed decisions about investment and field development strategies. These advancements can significantly improve production efficiency, reduce operational costs, and enhance the sustainability of gas field operations.

However, the proposed work has several limitations that could be addressed in future research. One limitation is the reliance on a high-quality pre-training dataset, such as the M4 competition dataset, which may affect the generalizability of the model to diverse geological formations and gas fields. Future efforts could focus on improving model transferability by incorporating gas production data sources. While the exponential smoothing method handles noise well, extreme production fluctuations or anomalies (e.g., maintenance disruptions) remain challenging; advanced anomaly detection techniques will be explored. Another limitation is model interpretability, as deep learning models are often perceived as black boxes. Explainable AI techniques could provide valuable insights into the model's decision-making process for better industry adoption.

5. Conclusions

In this study, a N-BEATS transfer model was developed to improve the prediction of tight gas production by leveraging transfer learning within the N-BEATS framework. This model was specifically designed to capture both trend and seasonality in production

data. Through an in-depth analysis of two production wells in the S block of the Sulige gas field, the proposed model successfully captured the complex and dynamic patterns of gas production sequences. Compared to the A-LSTM model, the N-BEATS transfer model consistently demonstrated superior accuracy across short-term, medium-term, and long-term forecasting tasks, with improvements in RMSE and MAE scores by 19.5%, and 18.9% of Well A1, respectively, for short-term predictions. Additionally, for medium-term and long-term forecasts, the N-BEATS model consistently maintained a lower error margin, achieving up to 20% greater accuracy across all forecast periods in MAPE compared to A-LSTM. In addition, the proposed N-BEATS transfer learning model offers transformative potential for decision-making in real-world tight gas production scenarios. Its ability to generate accurate production forecasts equips operators with critical insights for optimizing operations such as infill drilling and resource allocation. The model's effectiveness with limited historical data is particularly advantageous for newly developed fields, enabling earlier and more reliable predictions than traditional approaches. Additionally, by accommodating seasonal variations and abrupt production changes, the model supports more adaptive field management. These capabilities enhance investment decision-making and production optimization, ultimately contributing to more sustainable and efficient development of tight gas resources.

Author Contributions: Conceptualization, X.M. and J.Z.; methodology, X.M.; validation, S.C. and L.Z.; formal analysis, X.M.; investigation, X.M. and J.Z.; data curation, S.C. and L.Z.; writing—original draft preparation, X.M.; writing—review and editing, X.M. and J.Z.; visualization, S.C. and L.Z.; funding acquisition, X.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (51934005).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, Z.; Nie, X.; Zhang, C.; Wang, M.; Zhao, J.; Jin, L. Lithology Classification and Porosity Estimation of Tight Gas Reservoirs With Well Logs Based on an Equivalent Multi-Component Model. *Front. Earth Sci.* **2022**, *10*, 850023. [CrossRef]
2. Mao, G.; Lai, F.; Li, Z.; Wei, H.; Zhou, A. Characteristics of Pore Structure of Tight Gas Reservoir and Its Influence on Fluid Distribution during Fracturing. *J. Pet. Sci. Eng.* **2020**, *193*, 107360. [CrossRef]
3. Jin, H.; Liu, C.; Guo, Z. Characterization of Tight Gas Sandstone Properties Based on Rock Physical Modeling and Seismic Inversion Methods. *Energies* **2023**, *16*, 7642. [CrossRef]
4. Cao, P.; Liu, J.; Leong, Y.-K. General Gas Permeability Model for Porous Media: Bridging the Gaps Between Conventional and Unconventional Natural Gas Reservoirs. *Energy Fuels* **2016**, *30*, 5492–5505. [CrossRef]
5. Tang, Y.; Ranjith, P.G.; Perera, M.S.A. Major Factors Influencing Proppant Behaviour and Proppant-Associated Damage Mechanisms during Hydraulic Fracturing. *Acta Geotech.* **2018**, *13*, 757–780. [CrossRef]
6. Suboyin, A.; Rahman, M.M.; Haroun, M. Hydraulic Fracturing Design Considerations and Optimal Usage of Water Resources for Middle Eastern Tight Gas Reservoirs. *ACS Omega* **2021**, *6*, 13433–13446. [CrossRef]
7. Li, L. Development of Natural Gas Industry in China: Review and Prospect. *Nat. Gas Ind. B* **2022**, *9*, 187–196. [CrossRef]
8. Jia, A.; Wei, Y.; Guo, Z.; Wang, G.; Meng, D.; Huang, S. Development Status and Prospect of Tight Sandstone Gas in China. *Nat. Gas Ind. B* **2022**, *9*, 467–476. [CrossRef]
9. Ma, X.; Hou, M.; Zhan, J.; Liu, Z. Interpretable Predictive Modeling of Tight Gas Well Productivity with SHAP and LIME Techniques. *Energies* **2023**, *16*, 3653. [CrossRef]
10. 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]
11. Dutta, R.; Meyet, M.; Burns, C.; Cauter, F.V. Comparison of Empirical and Analytical Methods for Production Forecasting in Unconventional Reservoirs: Lessons from North America. In Proceedings of the European Association of Geoscientists & Engineers, Vienna, Austria, 25–27 February 2014; Volume 2014, pp. 1–26.
12. Wilson, A. Comparison of Empirical and Analytical Methods for Production Forecasting. *J. Pet. Technol.* **2015**, *67*, 133–136. [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. Chen, Z.; Yu, W.; Liang, J.-T.; Wang, S.; Liang, H. Application of Statistical Machine Learning Clustering Algorithms to Improve EUR Predictions Using Decline Curve Analysis in Shale-Gas Reservoirs. *J. Pet. Sci. Eng.* **2022**, *208*, 109216. [CrossRef]

15. Zhang, L.; Liu, S.; Yong, R.; Li, B.; Zhao, Y. EDFM-Based Numerical Simulation of Horizontal Wells with Multi-Stage Hydraulic Fracturing in Tight Reservoirs. *J. Southwest Pet. Univ. (Sci. Technol. Ed.)* **2019**, *41*, 1–11. [CrossRef]
16. Fujita, Y.; Datta-Gupta, A.; King, M.J. A Comprehensive Reservoir Simulator for Unconventional Reservoirs That Is Based on the Fast Marching Method and Diffusive Time of Flight. *SPE J.* **2016**, *21*, 2276–2288. [CrossRef]
17. Li, Y.; Duan, B.; Shang, L.; Sun, Y.; Gong, L.; Xin, C.; Xing, J. Numerical simulation method and its application for nonlinear flow in ultra-low permeability and tight oil reservoir in Jidong Oilfield. *Pet. Geol. Oilfield Dev. Daqing* **2022**, *41*, 153–158. [CrossRef]
18. Martínez, V.; Rocha, A. The Golem: A General Data-Driven Model for Oil & Gas Forecasting Based on Recurrent Neural Networks. *IEEE Access* **2023**, *11*, 41105–41132. [CrossRef]
19. Dettori, S.; Matino, I.; Colla, V.; Speets, R. A Deep Learning-Based Approach for Forecasting off-Gas Production and Consumption in the Blast Furnace. *Neural Comput. Appl.* **2022**, *34*, 911–923. [CrossRef]
20. Ma, X.; Hou, M.; Zhan, J.; Zhong, R. Enhancing Production Prediction in Shale Gas Reservoirs Using a Hybrid Gated Recurrent Unit and Multilayer Perceptron (GRU-MLP) Model. *Appl. Sci.* **2023**, *13*, 9827. [CrossRef]
21. Kumar, I.; Tripathi, B.K.; Singh, A. Attention-Based LSTM Network-Assisted Time Series Forecasting Models for Petroleum Production. *Eng. Appl. Artif. Intell.* **2023**, *123*, 106440. [CrossRef]
22. Iman, M.; Arabnia, H.R.; Rasheed, K. A Review of Deep Transfer Learning and Recent Advancements. *Technologies* **2023**, *11*, 40. [CrossRef]
23. Alolayan, O.S.; Raymond, S.J.; Montgomery, J.B.; Williams, J.R. Towards Better Shale Gas Production Forecasting Using Transfer Learning. *Upstream Oil Gas Technol.* **2022**, *9*, 100072. [CrossRef]
24. Mohd Razak, S.; Cornelio, J.; Cho, Y.; Liu, H.-H.; Vaidya, R.; Jafarpour, B. Transfer Learning with Recurrent Neural Networks for Long-Term Production Forecasting in Unconventional Reservoirs. *SPE J.* **2022**, *27*, 2425–2442. [CrossRef]
25. Weber, M.; Auch, M.; Doblander, C.; Mandl, P.; Jacobsen, H.-A. Transfer Learning With Time Series Data: A Systematic Mapping Study. *IEEE Access* **2021**, *9*, 165409–165432. [CrossRef]
26. Oreshkin, B.N.; Carpov, D.; Chapados, N.; Bengio, Y. N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting. *arXiv* **2019**, arXiv:1905.10437. [CrossRef]
27. Alali, Z.H.; Horne, R.N. A Comparative Study of Deep Learning Models and Traditional Methods in Forecasting Oil Production in the Volve Field. Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, TX, USA, 6–8 October 2023. [CrossRef]
28. Cai, W.; Zhang, H.; Huang, Z.; Mo, X.; Zhang, K.; Liu, S. Development and Analysis of Mathematical Plunger Lift Models of the Low-Permeability Sulige Gas Field. *Energies* **2023**, *16*, 1359. [CrossRef]
29. Rubio, L.; Gutiérrez-Rodríguez, A.J.; Forero, M.G. EBITDA Index Prediction Using Exponential Smoothing and ARIMA Model. *Mathematics* **2021**, *9*, 2538. [CrossRef]
30. Cho, K.; Merriënboer, B.V.; 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, Doha, Qatar, 25–29 October 2014. [CrossRef]
31. Staudemeyer, R.; Morris, E. Understanding LSTM—A tutorial into Long Short-Term Memory Recurrent Neural Networks. *arXiv* **2019**, arXiv:1909.09586. [CrossRef]
32. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 5998–6008.
33. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
34. Chibani, N.; Sebbak, F.; Cherifi, W.; Belmessous, K. Road Anomaly Detection Using a Dynamic Sliding Window Technique. *Neural Comput. Appl.* **2022**, *34*, 19015–19033. [CrossRef]
35. Hodson, T.O. Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not. *Geosci. Model Dev.* **2022**, *15*, 5481–5487. [CrossRef]