Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 14, No. 1, 2025

Study on Coalbed Methane Well Production Forecasting Based on VMD and Hybrid Time Series Models

Qianyu Zheng[1, *]

[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, China
[*]Correspondence: Qianyu Zheng

Abstract: To address the challenges of non-stationarity and nonlinearity in forecasting daily gas production for coalbed methane (CBM) wells, a hybrid prediction model incorporating Variational Mode Decomposition (VMD), Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM), and Sparrow Search Algorithm-optimized Support Vector Regression (SSA-SVR) is proposed. Initially, VMD adaptively decomposes the original time series into a series of intrinsic mode functions, followed by assessing each mode's complexity via permutation entropy for precise feature delineation. CNN-LSTM is then applied to capture deep spatiotemporal features of highly complex components, while SSA-SVR effectively predicts lower-complexity components through regression. Ultimately, the predictions are linearly combined to determine CBM wells' daily gas production. Comparative experiments reveal that the proposed model attains a mean absolute error (MAE) of 1.5158, root mean square error (RMSE) of 2.0608, and coefficient of determination (R^2) of 0.9998, outperforming other commonly used models in prediction accuracy and generalization capacity. This framework presents an enhanced model structure for CBM daily production forecasting.

Keywords: CBM well daily gas production forecasting; hybrid model; time series; variational mode decomposition.

1. Introduction

The prediction of gas production in coalbed methane (CBM) wells serves as a fundamental basis for comprehensively evaluating the productive capacity of coal reservoirs and the efficacy of extraction techniques. High-precision forecasting not only underpins the scientific development of extraction protocols but also facilitates outcome predictions for low-yield wells following modification, thus offering valuable guidance for the effective management of CBM well production[1].

The intricate characteristics of CBM reservoirs, coupled with the influence of extraction protocols, pose challenges for traditional mechanistic models[2][3], such as modeling difficulties and limited stability, which significantly restrict their practical application[4]. In recent years, the use of artificial intelligence models for predicting CBM well production has gained considerable attention. Wu Caifang et al. [5] applied a BP neural network to forecast CBM well productivity, accurately predicting gas production for the subsequent seven days and optimizing extraction protocols accordingly. Dong Weiqiang et al.[6] utilized a Long Short-Term Memory (LSTM) network to uncover the latent dynamic relationships within extraction data, achieving favorable fitting results. Given the highly nonlinear and non-stationary characteristics of CBM extraction data series, the predictive performance of single models is often inadequate. Consequently, some researchers have introduced hybrid models that integrate multiple methods to enhance prediction accuracy. Ma X et al. [7]proposed a hybrid GRU-MLP neural network model, which accurately predicts shale gas production by capturing the sequential dependencies in production data and the nonlinear relationships with physical constraints. Zhao Haifeng et al. [8] implemented an attention-based Convolutional Neural Network-Gated Recurrent Unit (CNN-GRU) model to overcome the limitations of traditional methods, which struggle with nonlinearity, temporal dependency, and information loss in CBM productivity forecasting.

To further improve prediction accuracy, modal decomposition techniques have been applied to handle non-stationary and nonlinear time series. Data decomposition techniques such as Empirical Mode Decomposition (EMD), Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), and Variational Mode Decomposition (VMD) enabling the extraction of critical features from the original sequence and thus optimizing hybrid time series models. Liu et al. [9]utilized the EEMD to decompose the original oil and gas production series into multiple subsequences, effectively addressing the mode mixing issue inherent in the EMD method. By combining EEMD with LSTM, this approach provides rapid and accurate oil and gas production forecasts while avoiding the reliance of traditional methods on high-quality historical matching and precise geological models. CEEMDAN and EEMD are refined EMD methods, with CEEMDAN offering superior noise suppression and improved mode decomposition. Zhao Xingyu et al. [10] applied Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to simplify raw data complexity, and established a short-term electricity load forecasting model through a Temporal Convolutional Network-Long Short-Term Memory (TCN-LSTM) model, achieving more precise load data predictions.

Nevertheless, existing decomposition-based forecasting models commonly rely on conventional decomposition methods, which produce an unregulated number of decomposition modes, potentially resulting in mode mixing and pseudo-mode issues that compromise prediction accuracy[11]. In contrast, VMD, as a non-recursive and adaptive decomposition method, overcomes the limitations of the recursive solving process in EMD and offers better harmonic separation[12]. Furthermore, VMD exhibits enhanced robustness in handling noise and sampling errors. As a result, VMD is better suited to process the complex,

multimodal oscillatory signals present in coalbed methane production data, effectively mitigating the mode mixing issues that can arise with EMD. In scenarios characterized by high noise levels and intricate signal patterns, VMD demonstrates greater adaptability and stability, thereby enabling more accurate signal decomposition.

This study proposes a method for forecasting daily gas production in CBM wells, employing Variational Mode Decomposition (VMD) combined with a hybrid time series model. Initially, VMD is utilized to decompose the original daily production data into multiple Intrinsic Mode Functions (IMFs). Subsequently, an optimal predictive model for each component is selected based on its Permutation Entropy (PE) value. Finally, the individual predictions for each component are integrated to produce a more accurate forecast of daily CBM well production. The model is applied to a dataset of CBM well production from a specific block in the Qin shui Basin, Shanxi, and its effectiveness is validated by comparing it with several other models.

2. Basic Theory

2.1. VMD

Variational Mode Decomposition (VMD) is a signal processing technique[13]. Its fundamental principle is to formulate and solve a variational problem that iteratively and adaptively adjusts the central frequency and bandwidth of each mode, thereby enabling effective separation of modal components and ultimately achieving the optimal solution to the variational problem[14].

The variational problem is first formulated, with the original signal $f(t)$ decomposed into K modal components. The constraint expression of the variational model is as follows:

$$\begin{cases} \min\limits_{\{\mu_k\},\{\omega_k\}} \left\{ \sum\limits_k \left\| \partial_t \left[\left(\delta(t) + \dfrac{j}{\pi t} \right) * \mu_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \\ s.t. \sum\limits_{k=1}^{K} \mu_k = f(t) \end{cases} \tag{1}$$

where $\{\mu_k\}$ and $\{\omega_k\}$ represent the $k-th$ modal component and its central frequency, $\delta(t)$ denotes the Dirac delta function, acting as a unit impulse function, * indicates the convolution operator, and $s.t. \sum\limits_{k=1}^{K} \mu_k = f(t)$ defines the constraint, where the sum of all modal components equals the original signal.

The constrained variational problem is subsequently converted into an unconstrained variational problem through the introduction of the Lagrange multiplier λ ,and the augmented Lagrange expression is defined as follows:

$$L(\{\mu_K\},\{\omega_K\},\lambda) = \alpha \sum_K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mu_K(t) \right] e^{-j\omega_K t} \right\|_2^2 + \tag{2}$$

Where the penalty factor α mitigates Gaussian noise interference, the optimal solution of the variational model is derived by locating the saddle point of the augmented Lagrange function, and iteratively updating the parameters through alternating optimization. The iterative procedure for solving the variational problem is outlined as follows[15]:

Step 1: Parameter initialization. The modal components $\mu_k^0 (k = 1,2,\ldots, K)$,central frequencies ω_k^0 ,Lagrange multiplier λ^0 ,penalty factor α ,and maximum iteration count N are initialized.

Step 2: Parameter Update. The parameters μ_k and ω_k are alternately updated using equations (3) and (4), while the Lagrange multiplier λ is updated according to equation (5), ensuring that the sum of all modal components accurately reconstructs the original signal.

$$\hat{\mu}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq K} \hat{\mu}_i(\omega) + \dfrac{\hat{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{3}$$

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{\mu}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{\mu}_k^{n+1}(\omega) \right|^2 d\omega} \tag{4}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^{n}(\omega) + \gamma \left(\hat{f}(\omega) - \sum_K \hat{\mu}_k^{n+1}(\omega) \right) \tag{5}$$

Step 3: Accuracy Evaluation and Result Output. A threshold ε is established, and if equation (6) is satisfied, the iteration is terminated. Otherwise, the process returns to Step 2 and continues until either the maximum iteration limit is reached or the desired accuracy is attained.

$$\frac{\sum_K \left\| \hat{\mu}_K^{n+1} - \hat{\mu}_K^{n} \right\|_2^2}{\left\| \hat{\mu}_K^{n} \right\|_2^2} < \varepsilon \tag{6}$$

2.2. PE

Entropy serves as a measure of the disorder within a time series, while the permutation entropy algorithm is capable of detecting noise in the sequence. This approach is particularly suited for analyzing the complexity of coalbed methane well production data after VMD decomposition, The computational procedure is outlined as follows [13][14]:

Step 1: Conduct phase space reconstruction on the components $v(i)$ derived from the VMD decomposition to obtain the matrix X, expressed as follows:

$$X=\begin{bmatrix} v(1) & v(1+t) & \cdots & v[1+(m-1)t] \\ v(2) & v(2+t) & \cdots & v[2+(m-1)t] \\ \vdots & \vdots & & \vdots \\ v(j) & v(j+t) & \cdots & v[j+(m-1)t] \\ \vdots & \vdots & & \vdots \\ v(K) & v(K+t) & \cdots & v[K+(m-1)t] \end{bmatrix} \quad (7)$$

where t and m represent the time delay and embedding dimension.

Step 2: The row vectors of matrix X are ordered in increasing are ordered in ascending sequence to define the reconstructed vector $x(l)$, thereby yielding the position index sequence.

$$x(l)=(j_1,j_2,\cdots,j_m) \quad (8)$$

Step 3: The probability $P_1,P_2,\cdots,P_k$ for each permutation is computed.

Step 4: The permutation entropy of the time series $v(i)$ is defined according to the Shannon entropy formula as follows:

$$H_p=-\sum_{l=1}^{m!}P_l\log P_l \quad (9)$$

Step 5: The permutation entropy values are normalized, mapping each component's value to the range [0,1], as follows:

$$H=\frac{H_P}{\log(m!)} \quad (10)$$

A smaller H value indicates a less complex sequence with reduced randomness, while a larger H signifies a more intricate sequence with higher randomness.

3. Materials and Methods

3.1. Construction of a Hybrid Time Series Forecasting Model

Based on the aforementioned description, the flowchart of the proposed VMD-based hybrid time series forecasting model for addressing the non-stationary and nonlinear characteristics of coalbed methane well production data can be summarized in Figure 1.

The main steps include:

Step 1: Decompose the original data into a limited number of Intrinsic Mode Functions (IMFs) using VMD, capturing essential characteristics of the original data while reducing noise and redundant information to enhance forecast accuracy.

Step 2: Measure the complexity information of each sequence using permutation entropy values, then classify the IMFs by assigning components 1 through k to the High-complexity category and components k+1 through n to the Medium-to-Low complexity category.

Step 3: Develop distinct forecasting models tailored to each IMF. The CNN-LSTM model is employed to forecast components 1 through k, while the SSA-SVR model is applied to components k+1 through n.

Step 4: Upon obtaining the predicted values from each model, the forecasts for each component are linearly synthesized in the time domain to yield the final prediction.

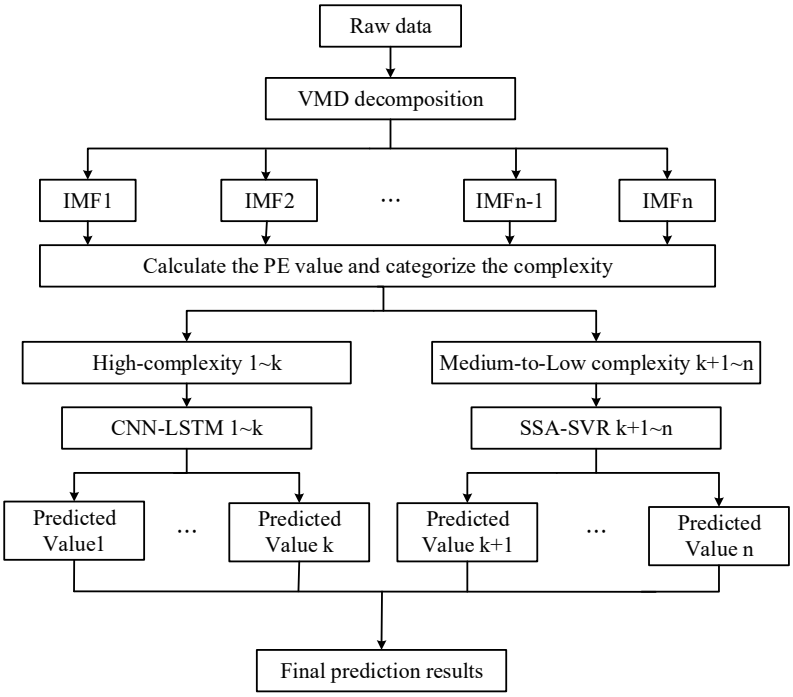

Figure 1. Hybrid model prediction flowchart based on VMD

3.1.1. CNN-LSTM

CNN[18] is capable of extracting latent spatial features through convolution operations , whereas LSTM[19] possesses the capability of long-term and short-term memory capabilities to effectively capture and maintain temporal dependencies in sequence data. By combining CNN and LSTM, the former is used for spatial feature extraction,

followed by the fusion of the extracted spatial features with the raw time series data. Then, LSTM's powerful capabilities are leveraged to mine the temporal dependencies in the time series[17][18]. This architecture effectively integrates the extraction and learning of both spatial and temporal features, making it particularly suitable for predicting the trends of highly complex IMFs. In this paper, a CNN-LSTM model for coalbed methane well production prediction is constructed, with its structure is shown in Fig. 2. The model primarily includes input layer, CNN layer, LSTM layer and output layer. The specific functions of each layer are described as follows:

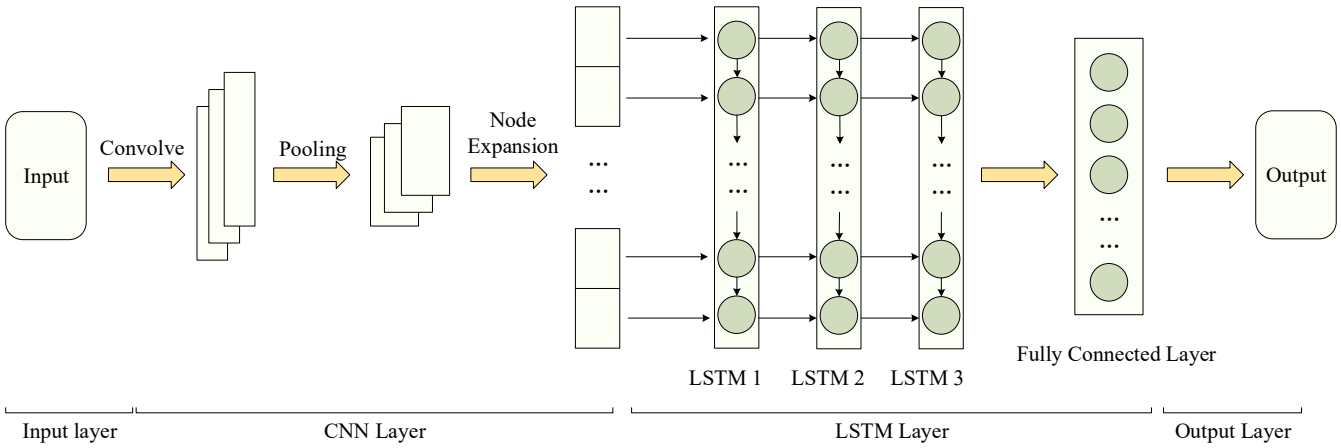

Figure 2. Architecture of CNN-LSTM predictive model

1) Input layer: determine the input training dataset, and the data is normalized using the min-max scaling method, with the normalized data serving as the input to the model.

2) CNN Layer: The CNN layer is utilized to extract features from the time series data. The IMF sequence data is passed through the convolutional layer, where one-dimensional convolution is performed using three convolutional kernels, followed by dimensionality reduction through a max pooling layer.

3) LSTM Layer: The features extracted in the previous layer are input into the LSTM network, which learns the temporal dependencies inherent in the sequence data.

4) Output Layer: A fully connected layer is incorporated at the end of the network structure. This layer generates predictions based on the features extracted and learned in the previous layers, ultimately producing the forecasted values.

3.1.2. SSA-SVR

SVR is a regression technique derived from Support Vector Machines (SVM), which maps input parameters to a high-dimensional linear feature space through a nonlinear transformation. By utilizing a kernel function, SVR replaces the complex computations in traditional high-dimensional spaces, By employing a kernel function, it substitutes the complex calculations in traditional high-dimensional spaces, thereby reducing computational complexity and constructing a fitting function between variables to achieve precise predictions of the sample dataset[22]. By transforming the nonlinear problem into a high-dimensional linear regression model, SVR can effectively deal with complex nonlinear relationships, as expressed by:

$$f(x) = \omega^T \varphi(x) + b \tag{11}$$

where ω is the weight vector, $\varphi(x)$ is the nonlinear mapping function with respect to x, and b is the defined bias term.

The performance of SVR relies heavily on the selection of parameters, particularly the configuration of the penalty coefficient C and the RBF kernel function. Therefore, parameter optimization is a critical factor in SVR applications. SSA[23], an optimization algorithm based on collective intelligence, globally searches for the optimal solution by simulating the foraging behavior of sparrows[24]. The integration of SVR and SSA leverages the exceptional global search optimization capabilities of SSA and the robust modeling power of SVR for nonlinear problems. SSA maps the positions of sparrows to the key parameters of the SVR model that require optimization. Through SSA's global optimization capacity, the optimal sparrow position is determined, corresponding to the globally optimal parameters, thus effectively merging and complementing the strengths of both methods.

The SSA-SVR model exhibits remarkable efficacy in processing low-frequency and trending sequences, effectively extracting both trend and nonlinear features from the data. Its flexibility and robustness make it an optimal choice for handling sequences of lower complexity. The implementation steps of the SSA-SVR model are shown in Figure 3:

1) Determine the training sample data set and initialize the sparrow population parameters. This involves determining the population size, setting the algorithm's maximum number of iterations, defining the spatial boundaries for sparrow activity, and establishing the fitness evaluation function. The mean squared error (MSE) is chosen as the fitness metric to evaluate the quality of the sparrow positions, with its formula expressed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \tilde{y}_i\right) \tag{12}$$

where y_i represents the true values of the training set, and $\tilde{y}_i$ represents the predicted values of the training set.

2) The positional information of each sparrow in the current population is utilized to configure the parameters of the SVR model, which is subsequently trained.

3) The SVR model derived from Step 2 is employed for prediction. The fitness value of each sparrow's position is

computed based on the fitness metric established in Step 1, and the sparrow positions are ranked accordingly to assess the optimal position for the current SVR model.

4) Judge whether the termination condition is satisfied. If satisfied, the optimal parameters are output, and the algorithm concludes with the optimal model. If not, the sparrow positions are updated, and their fitness values recalculated to distinguish the best and worst individuals in the population.

5) The optimal parameter combination is generated and integrated into the SVR model to derive the optimal model.

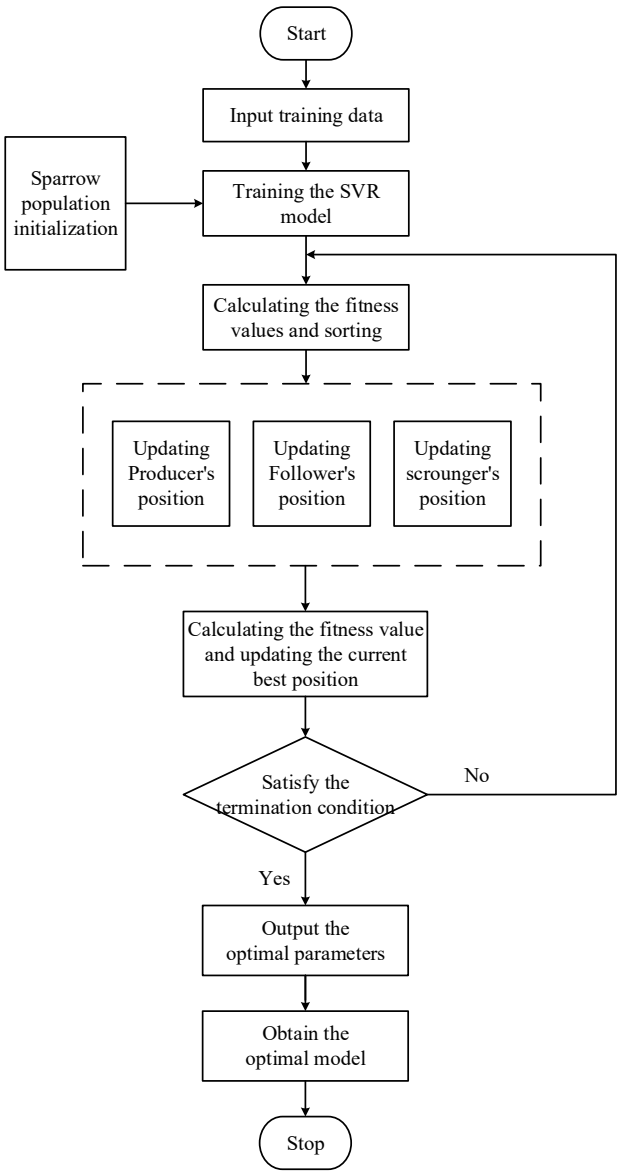

Figure 3. Flowchart of SSA-SVR predictive model

3.2. Evaluation indexes

To objectively evaluate the predictive performance, this paper employs the following metrics to assess algorithm accuracy: mean absolute error (MAE), root mean square error (RMSE), and goodness-of-fit (R^2). MAE and RMSE respectively measure the average deviation and error between the predicted and actual values, with lower values indicating better predictive accuracy. R^2 reflects the proportion of variance between the predicted and true values, ranging from 0 to 1, where values closer to 1 indicate a better fit. The formulas for each metric are as follows:

$$MAE(y,\tilde{y}) = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \tilde{y}_i\right| \tag{13}$$

$$RMSE(y,\tilde{y}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \tilde{y}_i\right)^2} \tag{14}$$

$$R^2(y,\tilde{y}) = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \tilde{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \tilde{y}\right)^2} \tag{15}$$

where y represents the actual production data, $\tilde{y}$ represents the predicted production data, and n is the number of prediction samples.

4. Experiments and Analysis

4.1. Data Source

The Shi Zhuang, South Block, located in the southeastern part of the Qin shui Basin, is abundant in coalbed methane resources and ranks among China's earliest and primary blocks for coalbed methane exploration and development. The main coal seams in the study area are the #3 and #15 seams. The #3 seam, characterized by high regional stability, lies at depths of 689.41 to 781.15 m, with a thickness of 5 to 6.52 m and a gas content ranging from 10.24 to 19.84 m³/t. This seam offers a promising target for coalbed methane exploration and is the focal coal seam in this study.

The research dataset is sourced from the Shi Zhuang South Block, covering daily gas production data from November 2016 to September 2022. This dataset includes nearly six years of production data from wells in the study area. Due to potential influences such as site maintenance, unexpected failures, or other uncontrollable external factors during the coalbed methane extraction phase, the production data may contain missing values or exhibit extreme variations, such as sudden drops or increases in production. To maintain data integrity and ensure the effectiveness of the analysis, this study adopts a dual strategy: outlier removal and mean imputation for preprocessing. This approach ensures both the completeness of the dataset and the stability of the model. Taking Well 299 in the target area as an example, after outlier removal and mean imputation, the adjusted dataset contains 1,958 valid daily gas production data points. Figure 4 shows the time-series trend of the imputed data for some wells in the target study area.

Since the model is highly sensitive to the scale of input data, the data is normalized using min-max normalization to compress the time series data into the [0, 1] range before being fed into the model. This normalization enhances the efficiency and stability of model training, enabling faster convergence and improved accuracy. After the prediction, the data is reverse normalized to its original scale for performance evaluation. Regarding dataset splitting, an 80:20 ratio was used to divide the dataset into a training set and a test set. The training set contains 1,566 data points, while the test set contains 392 data points. This method ensures both data integrity and model stability, providing a reliable

foundation for subsequent model training and evaluation. Figure 4 presents the temporal trends of the completed data for selected wells within the target study area.

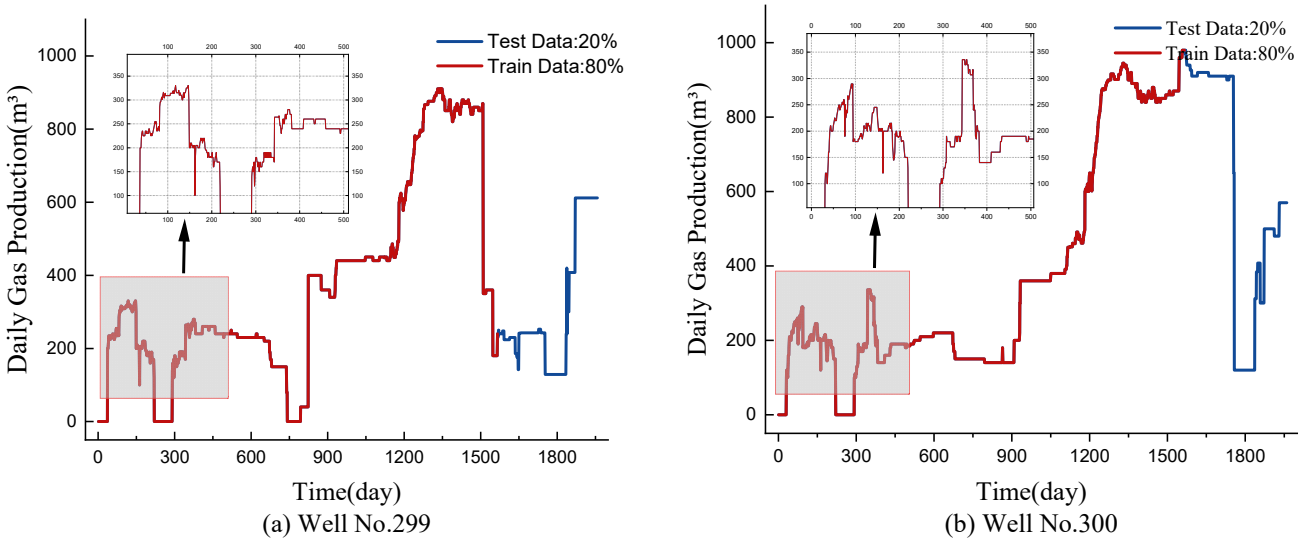

Figure 4. Daily gas production data of coalbed methane wells in the target area

4.2. VMD Decomposition

The selection of the decomposition number K is critical when applying VMD to the original sequence. When the chosen number of modes is too small, important information from the original signal may be filtered out, which negatively impacts the accuracy of subsequent predictions. On the other hand, when the number of modes is too large, the center frequencies of adjacent modes may become very close to each other, leading to mode redundancy or the introduction of additional noise. In this study, the selection of K is primarily determined by combining the center frequency method with frequency aliasing considerations.

The primary distinction between different modes lies in their center frequencies. Therefore, by analyzing the variation of center frequencies under different numbers of modes, the optimal value of K can be reasonably determined. In this study, we select K values ranging from 2 to 8 for pre-decomposing the signal and calculate the center frequency of each mode. When the center frequencies of adjacent modes become close, it is considered an indication of over-decomposition. In this case, the optimal number of decomposition layers corresponds to the K value of the previous mode when the center frequencies are close. Table 1 presents the center frequency values of each mode component.

Table 1. Center frequencies for different values of K

K	Center frequencies /Hz							
2	0.000143	0.010697						
3	0.00012	0.00708	0.02788					
4	0.000112	0.006238	0.022873	0.053907				
5	0.000111	0.006114	0.02204	0.049143	0.087932			
6	0.00011	0.006013	0.021903	0.049035	0.0877	0.411148		
7	0.00011	0.005969	0.02174	0.048479	0.084346	0.150367	0.413322	
8	0.00011	0.005961	0.02172	0.048427	0.084089	0.148192	0.355827	0.417711

The data presented in Table 1 reveals that when K is set to 8, When the number of decomposition layers is 8, the center frequencies of IMF7 and IMF8 are found to be very close, with the growth rate not exceeding 20%, and the rate of increase significantly drops. This indicates that the frequency variation introduced by the new mode is minimal, potentially introducing redundant information. In addition, the frequency spectrum obtained by performing the Fast Fourier Transform (FFT) with K = 8 is shown in Figure 5. It can be observed that mode mixing occurs when the decomposition level is set to 8. Therefore, the optimal decomposition level for VMD is determined to be 7.

The penalty parameter (α) in the VMD algorithm also significantly influences the decomposition results. Research has shown that as the penalty parameter α decreases, the bandwidth of the obtained IMF components increases. Conversely, as α increases, the bandwidth of the components decreases. After determining the number of decomposition modes (K), multiple experiments were conducted to adjust the value of α and verify the decomposition performance. Several common α values (100, 500, 1000, 2000) were selected, and the mean square error (MSE) was used to evaluate the error between the original signal and the signal obtained after VMD decomposition and reconstruction. The smaller the error, the higher the correlation between the reconstructed result and the original signal. The signal-to-noise ratio (SNR) was used to evaluate the ratio between the signal and noise; a higher SNR indicates better reconstruction performance. The experimental results are shown in the table 2. The experimental results show that the decomposition performs best when α = 500, as this value provides the optimal balance between denoising effectiveness and the preservation of details during the VMD decomposition process.

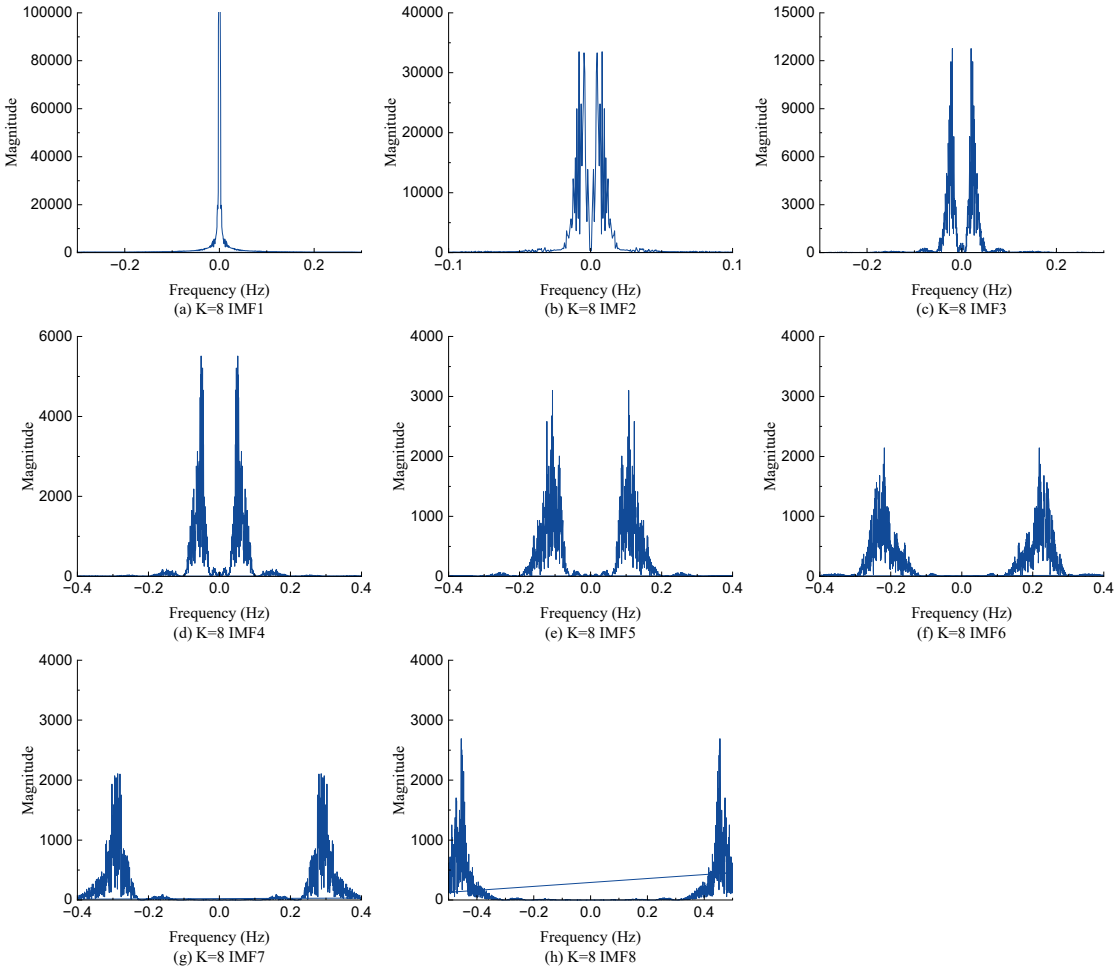

Figure 5. The frequency spectrum of the mode components at K = 8.

Table 2. Performance Evaluation of VMD Decomposition for Varying α Values

α	MSE	MAE	SNR
100	0.00122455	0.03497085	82.1935281
500	0.00122331	0.03496892	82.19790499
1000	7.92863078	1.27994684	44.08132178
2000	26.13934977	2.46786865	38.90035592

Therefore, the optimal value for K is determined to be 7, and the penalty parameter α is set to 500.The VMD decomposition results for well 299 are illustrated in Figure 6.

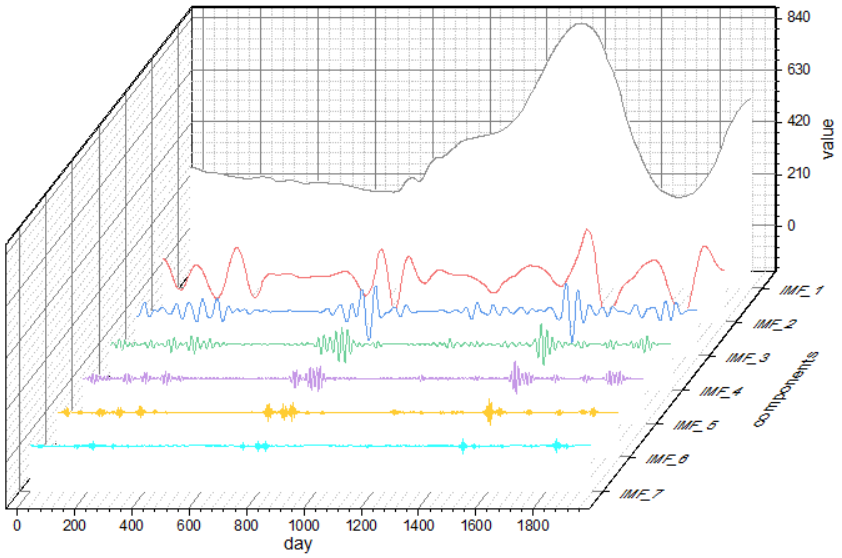

Figure 6. VMD modal decomposition results

4.3. PE Value

After performing VMD, the permutation entropy values for each sequence were calculated to quantify their complexity and dynamic characteristics. Table 3 displays the computed permutation entropy values, and Figure 7 depicts the variation in entropy values over time and across sequences.

Table 3. Permutation entropy values for each IMFs

IMFs	PE Value	IMFs	PE Value
IMF1	6.7972	IMF4	5.3954
IMF2	6.8090	IMF5	4.5515
IMF3	6.1355	IMF6	3.9122
-	-	IMF7	2.3486

As shown in Figure 6, the complexity of each modal component progressively diminishes after decomposition, signifying a decrease in the overall sequence complexity. To improve prediction accuracy, to improve prediction accuracy, the IMF components are classified into different levels based on their PE average values. According to the data in Table 3, the average PE for IMF1-7 is 5.1356. Consequently, IMF1-3, with values substantially exceeding the average, are designated as high-complexity sequences; IMF5-7, with values falling below the average, are categorized as low-complexity sequences; and IMF4, with a value near the average, is assigned to the medium-complexity category.

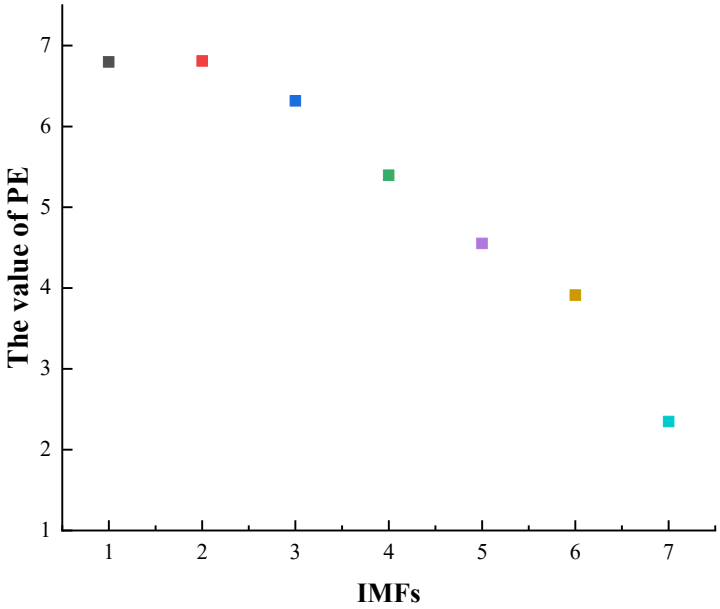

Figure 7. Trend of permutation entropy values for each IMFs

4.4. Hybrid Model Prediction

Each IMF component was input into its respective hybrid time series model for training and forecasting. Specifically, High-complexity sequences were processed using the CNN-LSTM model, which combines the advantages of CNN and LSTM. CNN automatically extracts important information from local spatiotemporal features, while LSTM captures long-term temporal dependencies, making it particularly suitable for handling complex time series with nonlinear and non-stationary components. Therefore, CNN-LSTM is well-suited for processing high-complexity, nonlinear data with long-term dependencies. For medium- and low-complexity sequences, the SSA-SVR model was employed. The Sparrow Search Algorithm (SSA) optimizes the hyperparameters of the Support Vector Regression (SVR) model, thereby enhancing its prediction accuracy for stable, low-complexity time series. SVR has demonstrated strong generalization ability and accuracy in handling time series with fewer noises or linear trends. As a result, SSA-SVR is effective in noise reduction and suitable for modeling and forecasting relatively stable time series.

In this study, the hyperparameter selection for CNN-LSTM was initially guided by existing literature and domain-specific experience to establish reasonable parameter ranges. Subsequently, the Optuna library was used to further refine the selection by automatically exploring various parameter combinations to identify the optimal configuration. For the SVR, the parameters were optimized through SSA, which searches for the best parameter combination within a specified range, evaluating the impact of each set of parameters on the SVR model, thereby determining the parameter configuration that maximizes model performance. The final parameter settings are presented in Tables 4 and 5.

Table 4. Hyperparameters of CNN-LSTM

Description	IMF1: CNN-LSTM 1	IMF2: CNN-LSTM 2	IMF3: CNN-LSTM 3
Number of Filters	64	64	64
Kernel Size	2	2	2
Stride	1	1	1
Hidden Layer Size	64	64	64
Number of LSTM Layers	2	3	3
Input Sequence Length	2	2	2
Output Sequence Length	1	1	1
Batch Size	64	64	64
Learning Rate	0.001	0.001	0.001
Number of Epochs	700	500	600
Optimizer	Adam	Adam	Adam
Loss Function	MSE Loss	MSE Loss	MSE Loss

Table 5. Hyperparameters of SSA-SVR

Description	IMF4: SSA-SVR 1	IMF5: SSA-SVR 2	IMF6: SSA-SVR 3	IMF7: SSA-SVR 4
Num Sparrows	50	50	50	50
Max iter	60	60	60	60
lb	[0.1, 0.001, 0.001]	[0.1, 0.001, 0.001]	[0.1, 0.001, 0.001]	[0.1, 0.01, 0.001]
ub	[10, 0.05, 10]	[10, 0.05, 10]	[10, 0.1, 10]	[15, 0.1, 15]
C	5.5468	5.1556	8.7821	13.1235
Epsilon	0.001	0.01	0.001	0.01
Gamma	8.0695	9.6023	1.6951	1.6951

This selective model matching approach based on the complexity of the IMF components enables each component's forecast to fully leverage the strengths of different models tailored to specific data characteristics, thereby enhancing the overall prediction accuracy and stability. Ultimately, by temporally stacking and integrating the forecasts of the individual components, a more accurate gas production prediction is obtained. The forecasting results for each component are presented in Figure 8, and the performance metrics are summarized in Table 6. By temporally stacking and integrating the individual component forecasts, the final gas production prediction is derived.

Table 6. Evaluation metrics for the prediction performance of each IMFs

IMFs	Evaluation Metrics		
	MAE	RMSE	R^2
IMF1	1.24952	1.49771	0.99986
IMF2	0.34557	0.40946	0.99997
IMF3	0.31623	0.38441	0.99947
IMF4	0.22744	0.34378	0.99861
IMF5	0.32934	0.49940	0.99630
IMF6	0.43084	0.74934	0.96342
IMF7	0.23382	0.48875	0.98376

4.5. Prediction results analysis

To assess the performance of the proposed VMD-based hybrid time series forecasting model, comparative experiments were conducted using gas production data from well 299 in the target study area. Both comparative and ablation experimental results are presented and discussed. The analysis is organized into four scenarios: (1) forecasting the original gas production data using various individual prediction models, as illustrated by Models 1 and 2; (2) forecasting the original gas production data using hybrid network models, as shown by Models 3 and 4; (3) constructing feature sequences using the VMD method applied in this study, followed by forecasting with different models, as exemplified by Models 5 and 6; (4) applying different mode decomposition methods to the original data, followed by forecasting using the hybrid time series model established in this study, as represented by Models 7 and 8. The specific descriptions of the models are as follows:

Model 1: LSTM – The sequence is forecasted using an LSTM neural network.

Model 2: SVR – The sequence is forecasted using an SVR neural network.

Model 3: CNN-LSTM – The sequence is forecasted using an LSTM model enhanced by a one-dimensional convolutional neural network.

Model 4: SSA-SVR – The sequence is forecasted using an LSTM model optimized through the Sparrow Search Algorithm.

Model 5: VMD-CL – The sequence is first decomposed using VMD, after which each component is independently forecasted using the CNN-LSTM model. The final prediction is derived by aggregating the individual forecasts.

Model 6: VMD-SS – The sequence is first decomposed using VMD, after which each component is independently forecasted using the SSA-SVR model. The final prediction is derived by aggregating the individual forecasts.

Model 7: EMD-CL-SS – The sequence is first decomposed using EMD, and based on the permutation entropy values of each component, high-complexity components are forecasted

using the CNN-LSTM model, while the remaining components are forecasted using the SSA-SVR model. The final prediction is obtained by aggregating the individual forecasts.

Model 8: CEEMDAN-CL-SS – The sequence is first decomposed using CEEMDAN, and based on the permutation entropy values of each component, high-complexity components are forecasted using the CNN-LSTM model, while the remaining components are forecasted using the SSA-SVR model. The final prediction is obtained by aggregating the individual forecasts.

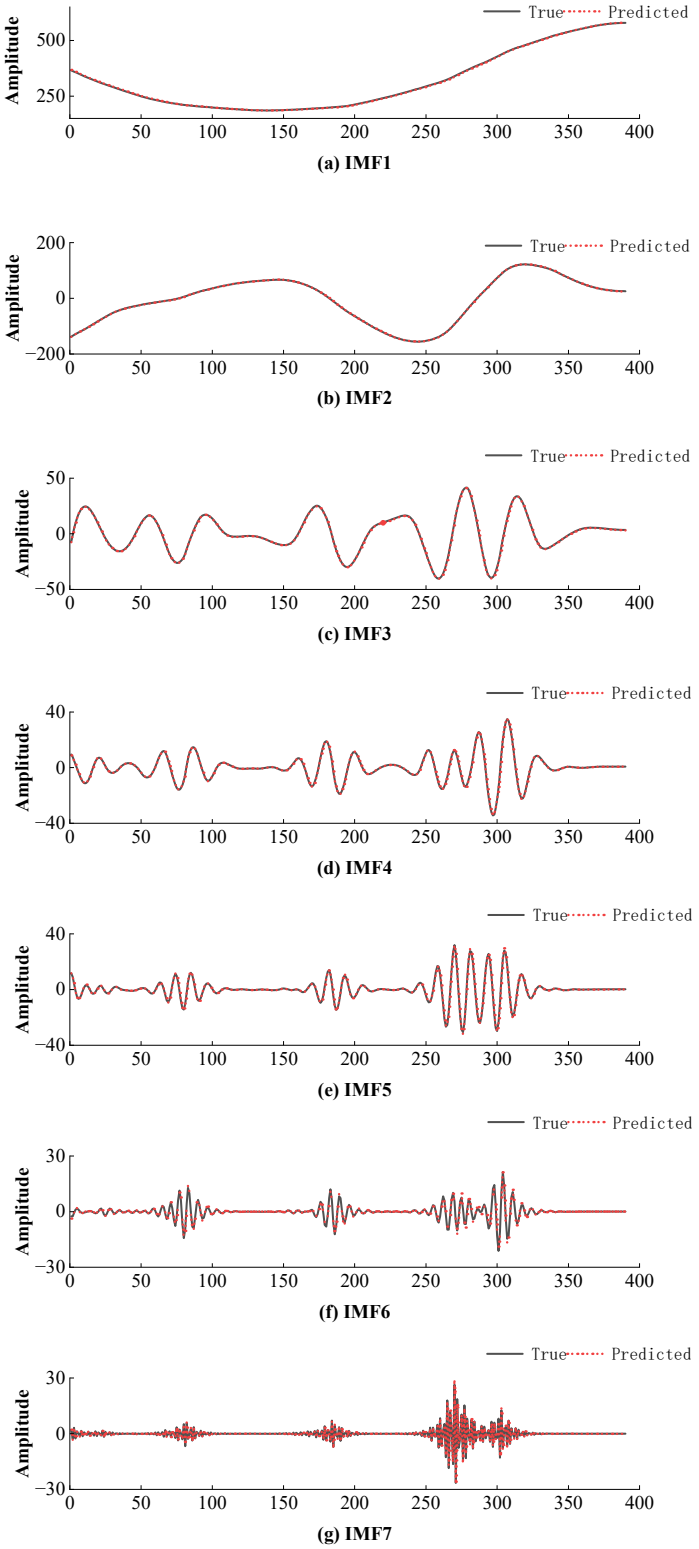

Figure 8. Predict results for each IMFs

The model proposed in this study was thoroughly evaluated against the eight previously mentioned models based on their fitting performance on the test set. The results of the comparison are illustrated in Figure 8, and the evaluation metrics are detailed in Table 7.

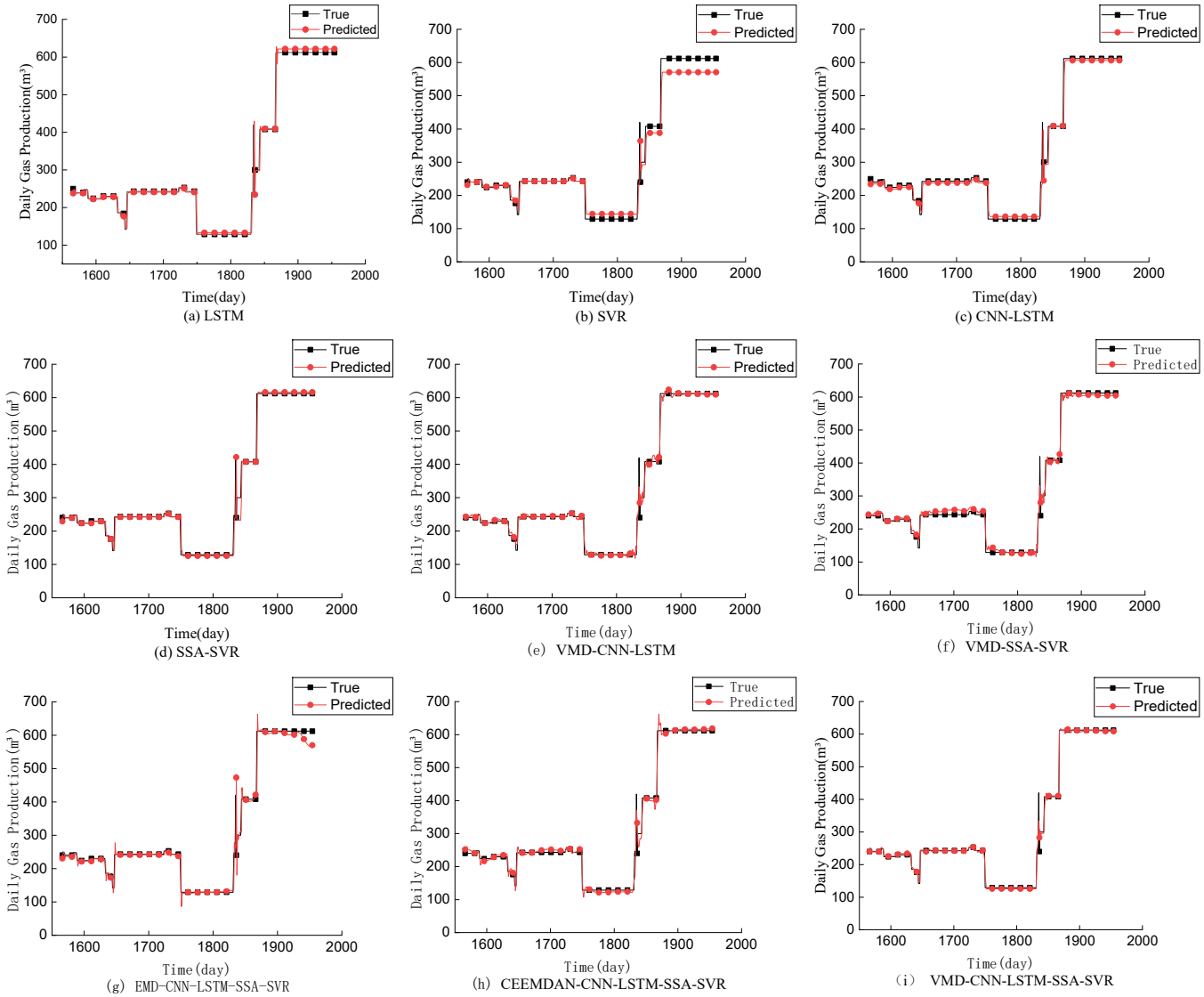

Figure 9. Comparison of prediction results for various modes

Table 7. Evaluation metrics for the prediction performance of various modes

Model	Evaluation Metrics		
	MAE	RMSE	R^2
LSTM	7.1574	21.2494	0.9854
SVR	17.5175	29.1886	0.9725
CNN-LSTM	6.8370	20.6678	0.9862
SSA-SVR	4.9957	20.6067	0.9863
VMD-CL	2.7330	4.1677	0.9994
VMD-SS	5.7358	6.7892	0.9984
EMD-CL-SS	8.2525	23.4347	0.9822
CEEMDAN-CL-SS	7.3870	12.4426	0.9950
Proposed	1.5158	2.0608	0.9998

The comparative experiments show that the model proposed in this paper improves in various evaluation metrics. Compared to LSTM, CNN-LSTM reduces the MAE and RMSE by 4.48% and 2.74%, respectively. SSA-SVR reduces the MAE by 12.5218 as well as the RMSE by 8.5819 compared with SVR, which suggests that prediction with a single LSTM or SVR only is susceptible to the interference of the intrinsic characteristics of the data, which affects the stability and accuracy of the prediction results.

As shown in Table 4, it can be observed that the prediction accuracy significantly improves when the production data is decomposed prior to prediction. Compared to CNN-LSTM and SSA-SVR, the prediction performance of models built after VMD decomposition, such as VMD-CL and VMD-SS, shows varying degrees of improvement. Specifically, the MAE and RMSE of VMD-CL decreased by 4.104 and 16.5001, respectively, while the R^2 increased by approximately 1.34%. For VMD-SS, the RMSE decreased by 13.8175, and the R^2 improved by around 1.23%, which strongly demonstrates the importance of incorporating decomposition techniques. Furthermore, to highlight the superiority of the VMD decomposition method, the error metrics of the VMD-CL-SS model were compared with those of models using EMD-CL-SS and CEEMDAN-CL-SS

decompositions. The MAE and RMSE of the VMD-CL-SS model were reduced by 6.7367 and 21.3739, respectively, as well as 5.8712 and 10.3818, indicating that VMD is more effective in signal decomposition and better at extracting frequency components that aid in prediction.

In conclusion, the proposed method in this study demonstrates better alignment with the real values and higher accuracy, whether applied to single models, combined models, or various decomposition techniques.

Figure 10 shows the distribution of the absolute prediction errors for our method compared to other methods. As can be seen from Figure 10, The hybrid model exhibits smaller error values compared to the single models. Specifically, the proposed method demonstrates the smallest absolute error values across various model combinations and decomposition methods, with a more concentrated distribution. This further emphasizes the superior predictive accuracy of the proposed approach.

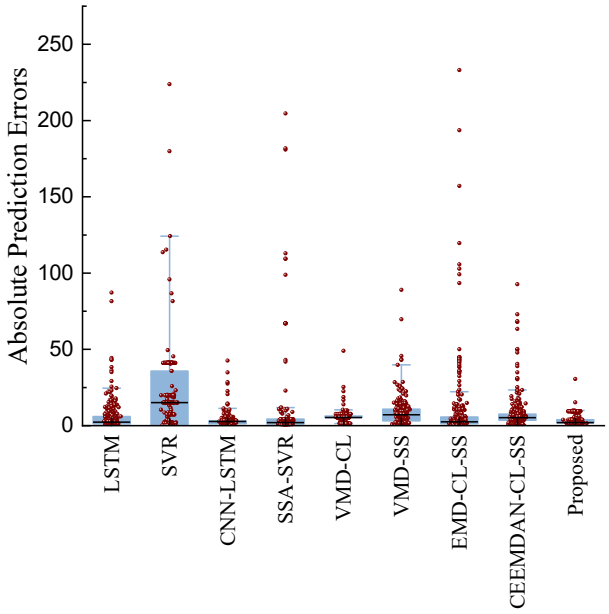

Figure 10. Box-plots of absolute forecasting error for each model

4.6. Other Experimental Analyses

To further validate the accuracy and generalization ability of the proposed method, we conducted experiments using actual gas production data from different wells within the study area, covering various time periods. Specifically, data from Well No. 300, spanning from September 2017 to November 2021, was selected for this purpose. In each case, 80% of the data was used for training, while the remaining 20% served as the test set. The original production data for each well, along with the prediction results generated by the various models, are presented in Figure 11. The corresponding error metrics are summarized in Table 8.

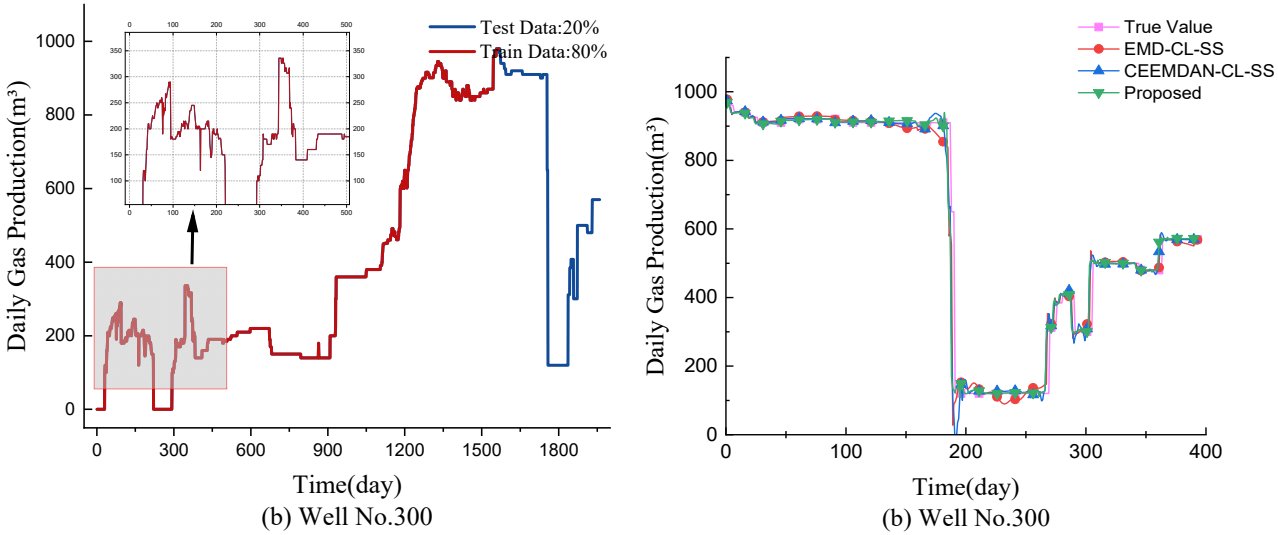

Figure 11. Data of No.300

Table 8. Error index values of each mode

Dataset	Model	MAE	RMSE	R^2
No.300	EMD-CL-SS	13.1546	31.8002	0.9900
	CEEMDAN-CL-SS	7.7721	19.3873	0.9963
	Proposed	3.7967	6.6265	0.9995

5. Conclusions

the VMD-based hybrid forecasting model demonstrates a high degree of accuracy in predicting the daily gas production of coalbed methane wells. Through both theoretical analysis and empirical validation, the following conclusions have been derived:

1) By applying Variational Mode Decomposition (VMD) to the original gas production data, it is decomposed into multiple intrinsic mode functions (IMFs), effectively reducing the inherent instability of the data. Each IMF exhibits corresponding frequency and amplitude characteristics, which aids in extracting latent information from the original data, thereby enhancing the quality of input data for subsequent models.

2) The integration of the permutation entropy algorithm enables the classification of modal components according to their permutation entropy values, thereby streamlining the training of various models in subsequent phases.

3) IMFs classified by permutation entropy values are separated into high-frequency and mid-to-low frequency components. The high-frequency components typically capture detailed information and rapid variations, while the mid-to-low frequency components represent trends and slow changes in the signal. Predictive modeling is then performed using a combination of CNN-LSTM and SSA-SVR time series models, which aids in more accurately capturing the data's underlying characteristics, enhancing model flexibility, and improving prediction accuracy.

4) The prediction results indicate that the proposed algorithm outperforms other commonly used forecasting methods in terms of various evaluation metrics. It exhibits superior accuracy in forecasting variations in coalbed methane well gas production and shows robust generalizability, providing a novel model framework for coalbed methane production forecasting.

Although the proposed method demonstrates strong performance in this case study, the dataset is limited to coalbed methane production data from a specific region and a subset of wells over a particular time period, which introduces inherent regional and temporal constraints. The method is particularly effective for datasets characterized by significant nonlinearity and seasonality, and it performs well in predicting relatively regular production patterns. However, in the presence of extreme climatic events or irregular fluctuations, the model's robustness and prediction accuracy may be impacted. To address these limitations, future research will focus on expanding the dataset to include data from a wider range of regions and time periods, thereby enhancing the generalizability of the method. Additionally, further work will investigate the optimization of the algorithm, with particular attention to fine-tuning model parameters to improve its adaptability in more complex and variable environments. This will help increase the model's stability and prediction accuracy, particularly in situations characterized by high noise levels or irregular production patterns. Through these improvements, we aim to increase the practical applicability of the proposed method and broaden its potential applications in various fields.

References

[1] Y. D. T. S. C. Ming and R. P. Shida, ‘Study on productivity prediction model of horizontal coalbed methane well and its applicability’, Coal Science and Technology, no. 12, 2016, Accessed: Nov. 27, 2024. [Online]. Available: https://www.mtkxjs.com.cn/en/article/id/6fda586e-9d2f-4891-b75b-59784e529747

[2] T. Xia, F. Gao, J. Kang, and X. Wang, ‘A fully coupling coal–gas model associated with inertia and slip effects for CBM migration’, Environ Earth Sci, vol. 75, no. 7, p. 582, Apr. 2016, doi: 10.1007/s12665-016-5378-y.

[3] C. R. Clarkson and F. Qanbari, ‘A semi-analytical method for forecasting wells completed in low permeability, undersaturated CBM reservoirs’, in SPE Asia Pacific Unconventional Resources Conference and Exhibition, SPE, 2015, p. SPE-176869. Accessed: Nov. 28, 2024. [Online]. Available: https://onepetro.org/SPEURCE/proceedings-abstract/15URCE/All-15URCE/183897

[4] D. LIU, Q. JIA, and Y. CAI, ‘Research progress on coalbed methane reservoir geology and characterization technology in China’, Coal science and technology, vol. 50, no. 1, pp. 196–203, 2022.

[5] C. F. Wu, S. Yao, and Y. F. Du, ‘Production systems optimization of a CBM well based on a time series BP neural network’, Journal of China University of Mining and Technology, vol. 44, no. 1, 2015.

[6] D. Weiqiang, M. Zhaoping, S. Zhen, Z. Zhimin, and C. Tao, ‘Research on coalbed methane well gas production forecast methodbased on cyclic neural network’, Coal Science and Technology, vol. 49, no. 9, pp. 176–183, 2021.

[7] Ma X, Hou M, Zhan J, et al. Enhancing production prediction in shale gas reservoirs using a hybrid gated recurrent unit and multilayer perceptron (GRU-MLP) model[J]. Applied Sciences, 2023, 13(17): 9827.

[8] ZHAO Haifeng, ZHU Likai, LIU Changsong, et al. Prediction of coalbed methane well productivity based on attention mechanism of CNN-GRU[J]. Safety in Coal Mines, 2023, 54（12）: 11−17.

[9] Liu W, Liu W D, Gu J. Forecasting oil production using ensemble empirical model decomposition based Long Short-Term Memory neural network[J]. Journal of Petroleum Science and Engineering, 2020, 189: 107013.

[10] G. Zheng, L. Kong, Z. Su, M. Hu, and G. Wang, ‘Approach for Short-Term Power Load Prediction Utilizing the ICEEMDAN–LSTM–TCN–Bagging Model’, J. Electr. Eng. Technol., Sep. 2024, doi: 10.1007/s42835-024-02040-1.

[11] X. Zhang, K. Yang, Q. Lu, J. Wu, L. Yu, and Y. Lin, ‘Predicting carbon futures prices based on a new hybrid machine learning: Comparative study of carbon prices in different periods’, Journal of Environmental Management, vol. 346, p. 118962, 2023.

[12] Lahmiri S. Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices[J]. IEEE Systems Journal, 2015, 11(3): 1907-1910.

[13] K. Dragomiretskiy and D. Zosso, ‘Variational mode decomposition’, IEEE transactions on signal processing, vol. 62, no. 3, pp. 531–544, 2013.

[14] X. PANG, C. WANG, and W. SUN, ‘Study on removing noise effect of magneto telluric signals based on multi-resolution VMD algorithm’, COAL SCIENCE AND TECHNOLOGY, vol. 49, no. 5, pp. 227–233, 2021.

[15] L. U. Tieding, L. I. Zhen, H. E. Xiaoxing, and Z. Shijian, ‘GNSS vertical time series prediction method integrating VMD and XGBoost algorithms’, Acta Geodaetica et Cartographica Sinica, vol. 52, no. 8, p. 1235, 2023.

[16] B. Zhang, C. Ding, W. Yan, L. Guo, J. Wang, and F. Hou, ‘Analysis of Magnetoencephalography based on symbolic

transfer entropy', in 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, 2017, pp. 1–5. Accessed: Nov. 27, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8302087/

[17] C. QIN et al., 'A method for predicting the time series of microseismic events in coal mines based on modal decomposition and deep learning', Journal of China Coal Society, vol. 49, no. 9, pp. 3781–3797, 2024.

[18] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, 'Review on Convolutional Neural Networks (CNN) in vegetation remote sensing', ISPRS journal of photogrammetry and remote sensing, vol. 173, pp. 24–49, 2021.

[19] A. Graves, 'Long Short-Term Memory', in Supervised Sequence Labelling with Recurrent Neural Networks, vol. 385, in Studies in Computational Intelligence, vol. 385. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 37–45. doi: 10.1007/978-3-642-24797-2_4.

[20] H. Wei-Jian, L. Yong-Tao, and H. Yuan, 'Prediction of chaotic time series using hybrid neural network and attention mechanism', Acta Physica Sinica, vol. 70, no. 1, 2021.

[21] K. Wang, C. Ma, Y. Qiao, X. Lu, W. Hao, and S. Dong, 'A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction', Physica A: Statistical Mechanics and its Applications, vol. 583, p. 126293, 2021.

[22] H. Tang, J. Cheng, and S. Wang, 'Support vector machine regression model of CBM content and application', in 2009 IEEE International Conference on Intelligent Computing and Intelligent Systems, IEEE, 2009, pp. 99–102. Accessed: Nov. 28, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/5357929/

[23] J. Xue and B. Shen, 'A novel swarm intelligence optimization approach: sparrow search algorithm', Systems Science & Control Engineering, vol. 8, no. 1, pp. 22–34, Jan. 2020, doi: 10.1080/21642583.2019.1708830.

[24] L. DENG, J. YUAN, J. LIU, and W. SHANG, 'Detection method of wind speed anomaly fluctuation based on SSA-LSTM', COAL SCIENCE AND TECHNOLOGY, vol. 52, no. 3, pp. 139–147, 2024.