

Article

A VMD-Based Dual-Branch Spatiotemporal Graph Model for Short-Term Gas Concentration Prediction in Coal Mine Return-Air Corners

Shaojie Chen *, Tong Qiao *, Jianing Song, Dongming Li and Zuojin Duan

School of Safety Engineering, University of Emergency Management, Langfang 065201, China; songjianing2026@163.com (J.S.); bzon@163.com (D.L.); duanzuojin2026@163.com (Z.D.)

* Correspondence: chenshaojie@ncist.edu.cn (S.C.); qiaotong2026@163.com (T.Q.)

Abstract

Gas concentration in coal mine return-air corners is affected by ventilation, mining disturbance and gas drainage conditions, and it shows strong nonstationarity, local fluctuation and dynamic multi-point correlations. To improve frequency information separation, monitoring point relationship modeling, and short-term prediction accuracy, a variational mode decomposition (VMD)-based dual-branch spatiotemporal graph method is proposed. Gas concentrations from four key monitoring points are used as inputs, and the return-air corner gas concentration is taken as the output. First, the raw series are decomposed by VMD and reconstructed into low- and high-frequency components. Then, two branches are built for different frequency components. The low-frequency branch combines adaptive graph learning, graph convolution and gated recurrent units to extract global variation features, while the high-frequency branch combines graph attention and gated recurrent units to capture local disturbance features. Finally, a feature-fusion module generates multi-step predictions, and a lightweight short-term warning strategy is developed based on the predicted values. The proposed model achieves MAE, RMSE and R^2 values of 0.0338, 0.0471 and 0.9499 in one-step prediction, respectively, and outperforms GRU, LSTM, GCN-GRU, GAT-GRU, VMD-GRU, Informer and STGCN under three-step and six-step conditions. Cross-dataset validation and inference time analysis indicate good adaptability and online prediction potential.

Keywords: return-air corner; gas concentration prediction; variational mode decomposition; spatiotemporal graph network; dual-branch structure; short-term warning



Academic Editor: Wei Liu

Received: 18 May 2026

Revised: 18 June 2026

Accepted: 2 July 2026

Published: 11 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Gas concentration is a key monitoring variable in coal mine safety production. Its variation is jointly affected by ventilation conditions, gas drainage status, mining-induced disturbances, and air leakage from the goaf, showing strong nonlinearity, nonstationarity, and spatial heterogeneity. The return-air corner is located at the intersection of the goaf and the return airflow, where low local airflow velocity and delayed gas migration can easily lead to gas accumulation. Therefore, accurate short-term prediction of gas concentration at the return-air corner is important for early identification of gas exceedance risk and field safety management.

Existing gas concentration prediction studies have used statistical models, machine learning methods, and deep learning approaches. ARIMA-based models can describe linear

time-series variations, while LSTM, GRU, hybrid neural networks, and ensemble models have improved the ability to extract nonlinear temporal features from gas monitoring data [1–8]. Transformer-based models and their variants have further enhanced long-sequence time series forecasting and multivariate sequence modeling [9–13]. These methods have improved prediction accuracy to different degrees. However, most of them mainly focus on temporal dependence and do not sufficiently describe the dynamic response relationships among multiple monitoring points in coal mine gas monitoring systems.

Gas concentrations at different monitoring points are not independent. They are influenced by gas emission from the working face, return-airflow transport, gas accumulation at the corner, and gas drainage variation. Graph neural networks provide an effective way to model spatial correlations in multivariate sequences. GCN, GAT, STGCN, DCRNN, and adaptive graph models have been widely used in spatiotemporal forecasting tasks [14–18]. Nevertheless, in coal mine gas monitoring scenarios, the relationships among monitoring points may change with ventilation structure, mining disturbance, and drainage conditions. A fixed or unified graph structure may be insufficient to represent these dynamic multi-point response relationships.

In addition, gas concentration sequences contain both slow-varying trends and short-term local disturbances. VMD can decompose nonstationary signals into modal components with different center frequencies and has been used in gas concentration and other complex time series prediction tasks [19–23]. However, most existing decomposition-based methods still feed the reconstructed components into a single prediction model. They do not fully distinguish the different modeling requirements of low-frequency trend evolution and high-frequency disturbance responses. Meanwhile, although prediction models can provide future gas concentration values, their connection with lightweight and field-applicable warning strategies remains insufficient [24].

Based on the above analysis, the existing studies still have limitations in three aspects. First, the dynamic correlations among multiple monitoring points have not been sufficiently represented. Second, low-frequency trend information and high-frequency disturbance information have not been adequately modeled in a differentiated manner. Third, prediction results have not been effectively converted into lightweight short-term warning information for field application. To address these limitations, this study proposes a VMD-based dual-branch spatiotemporal graph model for short-term gas concentration prediction at the return-air corner. The raw multi-point sequences are first decomposed and reconstructed into low- and high-frequency components. A low-frequency branch based on self-adaptive graph learning, graph convolution, and GRU is then used to extract global variation features, while a high-frequency branch based on graph attention and GRU is used to capture local disturbance features. Finally, a feature fusion module outputs multi-step prediction results, and a lightweight warning strategy is constructed based on the predicted values. Comparative experiments, ablation analysis, cross-dataset validation, and inference time analysis are conducted to verify the effectiveness and engineering applicability of the proposed method.

2. Gas Concentration Prediction Model

The overall framework of the proposed model is shown in Figure 1. The model consists of data preprocessing, VMD-based low-/high-frequency reconstruction, a dual-branch spatiotemporal graph network, and feature fusion with multi-step prediction.

2.1. VMD-Based Decomposition Algorithm

To reduce the influence of the superposition of slow trend components and short-term fluctuation components in the raw gas concentration sequences on feature extraction,

variational mode decomposition (VMD) is adopted to decompose the gas concentration sequences from four monitoring points separately. The decomposed results are then reconstructed into low-frequency and high-frequency components, which are used as the inputs of the subsequent dual-branch model. Recent mine gas prediction studies have also shown that decomposition-based modeling can reduce the complexity of nonstationary gas concentration sequences and improve multi-step prediction stability [25].

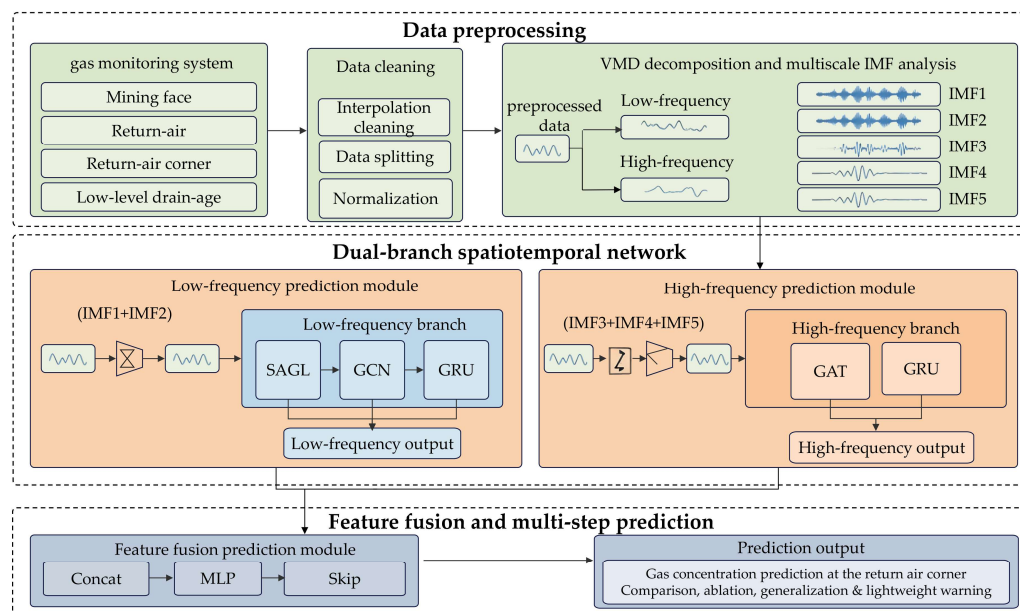


Figure 1. Overall framework of the proposed gas concentration prediction model.

VMD was selected in this study because gas concentration sequences at the return-air corner contain both slowly varying trend components and short-term local disturbances. Compared with empirical decomposition methods such as EEMD and CEEMDAN, VMD formulates signal decomposition as a variational optimization problem and decomposes the original sequence into a limited number of band-limited intrinsic mode functions with specific center frequencies. This mechanism helps reduce mode mixing and provides more stable frequency separation for nonstationary monitoring data. EEMD and CEEMDAN are also effective noise-assisted decomposition methods, but they usually require multiple ensemble operations and may introduce additional computational cost and residual noise uncertainty. Considering the need for short-term prediction and online monitoring in coal mine safety applications, VMD was adopted to obtain compact and stable low- and high-frequency representations before graph spatiotemporal modeling.

At time t , the multi-point input vector is defined in Equation (1), where x_t denotes the input vector, $x_1(t)$, $x_2(t)$, $x_3(t)$, and $x_4(t)$ denote the gas concentration at the mining face, the gas concentration in the return airway, the gas concentration at the return-air corner, and the gas concentration in the low-level gas drainage pipeline, respectively, and t is the time index.

$$x_t = [x_1(t), x_2(t), x_3(t), x_4(t)]^T \quad (1)$$

The objective of VMD is to decompose the raw monitoring sequence into K modal components with different center frequencies while minimizing the sum of their bandwidths. For any input variable $x_i(t)$, the constrained variational problem is formulated in Equation (2), where $x_i(t)$ denotes the raw monitoring value of the i -th input variable at time t , $u_{i,k}(t)$ denotes the k -th modal component decomposed from the i -th input variable, $\omega_{i,k}$ denotes the corresponding center frequency, K is the number of decomposed modes,

$\delta(t)$ is the Dirac delta function, j is the imaginary unit, ∂_t denotes the partial derivative with respect to time t , $*$ denotes the convolution operation, and $\|\cdot\|_2$ denotes the L_2 -norm.

$$\begin{cases} \min_{\{u_{i,k}\}, \{\omega_{i,k}\}} \left\{ \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_{i,k}(t) \right] e^{-j\omega_{i,k}t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{k=1}^K u_{i,k}(t) = x_i(t) \end{cases} \quad (2)$$

To solve the constrained variational problem, the bandwidth of each modal component is first calculated. Equation (3) defines the bandwidth term of the k -th modal component for the i -th input variable, where $J_{i,k}$ denotes the corresponding bandwidth measure.

$$J_{i,k} = \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_{i,k}(t) \right] e^{-j\omega_{i,k}t} \right\|_2^2 \quad (3)$$

On this basis, the penalty factor and the Lagrange multiplier are introduced to transform the constrained optimization problem into an unconstrained form. Equation (4) gives the augmented Lagrangian function, where $\mathcal{L}_i$ denotes the augmented Lagrangian function corresponding to the i -th input variable, α is the penalty factor, $\lambda_i(t)$ is the Lagrange multiplier, and $\langle \cdot, \cdot \rangle$ denotes the inner product operation.

$$\mathcal{L}_i = \alpha \sum_{k=1}^K J_{i,k} + \left\| x_i(t) - \sum_{k=1}^K u_{i,k}(t) \right\|_2^2 + \langle \lambda_i(t), x_i(t) - \sum_{k=1}^K u_{i,k}(t) \rangle \quad (4)$$

The above optimization problem is solved iteratively using the alternating direction method of multipliers, and the modal components corresponding to each input variable are obtained.

Considering the characteristics of the gas concentration data used in this study, the number of decomposed modes is set to $K = 5$. The first two modal components are reconstructed as the low-frequency component, whereas the last three modal components are reconstructed as the high-frequency component. Equation (5) describes the reconstruction process, where $x_i^{\text{low}}(t)$ and $x_i^{\text{high}}(t)$ denote the reconstructed low-frequency and high-frequency components of the i -th input variable, respectively, and $u_{i,k}(t)$ denotes the k -th intrinsic mode function component.

$$\begin{cases} x_i^{\text{low}}(t) = u_{i,1}(t) + u_{i,2}(t) \\ x_i^{\text{high}}(t) = u_{i,3}(t) + u_{i,4}(t) + u_{i,5}(t) \end{cases} \quad (5)$$

After the above processing, the raw monitoring data are divided into low-frequency and high-frequency parts. The low-frequency part mainly reflects the overall variation trend of gas concentration, whereas the high-frequency part mainly represents local disturbances and short-term fluctuations. This treatment enables the subsequent model to extract frequency-specific features separately, thereby improving the prediction accuracy of gas concentration at the return-air corner.

2.2. Low-Frequency Branch Model

After VMD, the low-frequency part mainly reflects the overall variation in gas concentration at the four monitoring points. This component is strongly affected by ventilation conditions, gas drainage status, and production activities at the working face, and usually changes relatively smoothly. To capture the correlations among different monitoring points, the low-frequency branch adopts a spatiotemporal feature extraction structure composed of self-adaptive graph learning (SAGL), graph convolutional network (GCN), and gated

recurrent unit (GRU). This branch first learns the connection relationships among monitoring points, then extracts spatial features at each time step, and finally combines temporal information to generate the low-frequency output feature.

For a prediction sample ending at time n , Equation (6) defines the low-frequency input matrix, where $\mathbf{X}_{\text{low}}^{(n)}$ denotes the low-frequency input matrix ending at time n , $\mathbf{x}_{\text{low}}(t)$ is the low-frequency node feature vector at time t , and L is the input length, which is set to 12 in this study.

$$\mathbf{X}_{\text{low}}^{(n)} = [\mathbf{x}_{\text{low}}(n-L+1)^T, \mathbf{x}_{\text{low}}(n-L+2)^T, \dots, \mathbf{x}_{\text{low}}(n)^T]^T \quad (6)$$

Since the influence relationships among different monitoring points are not fixed, SAGL is used to construct the adjacency matrix of the low-frequency branch. Equation (7) defines the adaptive adjacency matrix $\mathbf{A}_{\text{adp}}$, where a_{pq} denotes the normalized connection weight from node p to node q , $\mathbf{e}_p$ and $\mathbf{f}_q$ are the learnable embedding vectors of nodes p and q , respectively, and $\text{ReLU}(\cdot)$ is the rectified linear activation function. The denominator normalizes the connection weights of node p over the four monitoring nodes.

$$\begin{cases} \mathbf{A}_{\text{adp}} = [a_{pq}]_{4 \times 4} \\ a_{pq} = \frac{\exp(\text{ReLU}(\mathbf{e}_p^T \mathbf{f}_q))}{\sum_{q'=1}^4 \exp(\text{ReLU}(\mathbf{e}_p^T \mathbf{f}_{q'}))} \end{cases} \quad (7)$$

Equation (7) enables the connection relationships among monitoring points to be adaptively adjusted according to the data. After the adaptive adjacency matrix is obtained, graph convolution is used to extract low-frequency spatial features. Equation (8) describes the graph convolution operation at time τ , where $\mathbf{h}_{\text{low}}(\tau)$ denotes the low-frequency spatial feature, $\mathbf{x}_{\text{low}}(\tau)$ denotes the low-frequency node feature vector at time τ , $\tilde{\mathbf{A}} = \mathbf{A}_{\text{adp}} + \mathbf{I}$ is the adjacency matrix with self-connections, $\mathbf{I}$ is the identity matrix, $\tilde{\mathbf{D}}$ is the degree matrix of $\tilde{\mathbf{A}}$, $\mathbf{W}_g$ is the learnable weight matrix, and $\sigma(\cdot)$ is the activation function.

$$\mathbf{h}_{\text{low}}(\tau) = \sigma \left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{x}_{\text{low}}(\tau) \mathbf{W}_g \right) \quad (8)$$

The low-frequency component contains continuous temporal variation information. Therefore, the graph convolution outputs within the historical window are further fed into a GRU for temporal encoding. Equation (9) gives the output of the low-frequency branch, where $\mathbf{c}_{\text{low}}^{(n)}$ denotes the low-frequency branch feature ending at time n , $\mathbf{h}_{\text{low}}(\tau)$ denotes the low-frequency spatial feature at time τ , L is the input window length, and $\text{GRU}(\cdot)$ represents the temporal encoding process.

$$\mathbf{c}_{\text{low}}^{(n)} = \text{GRU} \left([\mathbf{h}_{\text{low}}(n-L+1), \mathbf{h}_{\text{low}}(n-L+2), \dots, \mathbf{h}_{\text{low}}(n)]^T \right) \quad (9)$$

The extracted feature mainly represents the overall variation information of gas concentration across the four monitoring points. Therefore, the low-frequency branch provides a stable trend-level representation for subsequent high- and low-frequency feature fusion and return-air corner gas concentration prediction.

2.3. High-Frequency Branch Spatiotemporal Feature Extraction Model

After VMD, the high-frequency part mainly reflects local disturbances and short-term fluctuations in gas concentration at the four monitoring points. Compared with the low-frequency part, the high-frequency component changes more rapidly and requires stronger sensitivity to key monitoring points and critical time steps. Therefore, a GAT-GRU

structure is adopted in the high-frequency branch. The graph attention mechanism is first used to highlight important node information, and the GRU is then used to extract temporal variation features.

For a prediction sample ending at time n , Equation (10) defines the high-frequency input matrix, where $\mathbf{X}_{\text{high}}^{(n)}$ denotes the high-frequency input matrix ending at time n , $\mathbf{x}_{\text{high}}(t)$ is the high-frequency node feature vector at time t , and L is the input window length.

$$\mathbf{X}_{\text{high}}^{(n)} = [\mathbf{x}_{\text{high}}(n-L+1)^T, \mathbf{x}_{\text{high}}(n-L+2)^T, \dots, \mathbf{x}_{\text{high}}(n)^T]^T \quad (10)$$

To emphasize the different roles of monitoring points in high-frequency variations, a graph attention mechanism is used to assign adaptive weights to node features. Equation (11) defines the unnormalized attention coefficient and the normalized attention weight between nodes p and q , where $e_{pq}^{(\tau)}$ denotes the unnormalized attention coefficient at time τ , $\alpha_{pq}^{(\tau)}$ denotes the normalized attention weight, $\mathbf{W}_a$ is the linear transformation matrix, $\mathbf{a}$ is the attention parameter vector, $\parallel$ denotes vector concatenation, and $\mathcal{N}_p$ is the neighborhood set of node p .

$$\begin{cases} e_{pq}^{(\tau)} = \text{ReLU}(\mathbf{a}^T [\mathbf{W}_a \mathbf{x}_p^{\text{high}}(\tau) \parallel \mathbf{W}_a \mathbf{x}_q^{\text{high}}(\tau)]) \\ \alpha_{pq}^{(\tau)} = \frac{\exp(e_{pq}^{(\tau)})}{\sum_{m \in \mathcal{N}_p} \exp(e_{pm}^{(\tau)})} \end{cases} \quad (11)$$

After the attention weights are obtained, the high-frequency features of neighboring nodes are weighted and aggregated, and the resulting spatial representations are further fed into a GRU for temporal encoding. Equation (12) describes this process, where $\mathbf{h}_p^{\text{high}}(\tau)$ denotes the high-frequency spatial feature of node p at time τ , $\mathbf{h}^{\text{high}}(\tau)$ denotes the high-frequency spatial representation of all nodes at time τ , and $\mathbf{c}_{\text{high}}^{(n)}$ denotes the output feature of the high-frequency branch for the prediction sample ending at time n .

$$\begin{cases} \mathbf{h}_p^{\text{high}}(\tau) = \sigma \left(\sum_{q \in \mathcal{N}_p} \alpha_{pq}^{(\tau)} \mathbf{W}_a \mathbf{x}_q^{\text{high}}(\tau) \right) \\ \mathbf{c}_{\text{high}}^{(n)} = \text{GRU} \left([\mathbf{h}^{\text{high}}(n-L+1), \mathbf{h}^{\text{high}}(n-L+2), \dots, \mathbf{h}^{\text{high}}(n)]^T \right) \end{cases} \quad (12)$$

Through this structure, the high-frequency branch strengthens the representation of key monitoring points and local fluctuation information. The extracted feature $\mathbf{c}_{\text{high}}^{(n)}$ provides high-frequency disturbance information for subsequent feature fusion and gas concentration prediction at the return-air corner.

2.4. Feature Fusion and Multi-Step Prediction Model

After feature extraction by the low-frequency and high-frequency branches, the output features $\mathbf{c}_{\text{low}}^{(n)}$ and $\mathbf{c}_{\text{high}}^{(n)}$ are obtained, respectively. The former mainly represents the overall variation information of gas concentration, while the latter reflects local disturbances and short-term fluctuations. To jointly use these two types of features, a feature fusion and multi-step prediction module is constructed to model the dual-branch outputs and generate the future H -step prediction results of gas concentration at the return-air corner. Multi-step gas concentration prediction studies have shown that combining temporal feature extraction with attention mechanisms can improve prediction robustness under nonlinear fluctuations [26].

The low-frequency and high-frequency features are first concatenated to form a fused feature vector. Equation (13) defines this fusion operation, where $\mathbf{f}^{(n)}$ denotes the fused feature vector at time n .

$$\mathbf{f}^{(n)} = \left[\left(\mathbf{c}_{\text{low}}^{(n)} \right)^T \quad \left(\mathbf{c}_{\text{high}}^{(n)} \right)^T \right]^T \quad (13)$$

On this basis, a multilayer perceptron is used to perform nonlinear mapping on the fused feature vector. Equation (14) gives the main-branch prediction output, where $\hat{\mathbf{y}}_{\text{main}}^{(n)}$ denotes the future H -step prediction vector generated by the main branch, $\mathbf{W}_1$ and $\mathbf{W}_2$ are learnable weight matrices, $\mathbf{b}_1$ and $\mathbf{b}_2$ are bias vectors, and $\phi(\cdot)$ is the nonlinear activation function.

$$\hat{\mathbf{y}}_{\text{main}}^{(n)} = \mathbf{W}_2 \phi \left(\mathbf{W}_1 \mathbf{f}^{(n)} + \mathbf{b}_1 \right) + \mathbf{b}_2 \quad (14)$$

Considering the strong short-term continuity of gas concentration at the return-air corner, a short-window branch is further introduced to map the recent observations of the target point into the prediction space. Equation (15) defines the short-window input and the adaptive compensation coefficient, where $\mathbf{z}^{(n)}$ denotes the short-window input vector at time n , $x_3(t)$ denotes the observed gas concentration at the return-air corner, s is the short-window length, $\alpha^{(n)}$ denotes the step-wise compensation coefficient vector, $\mathbf{W}_\alpha$ and $\mathbf{b}_\alpha$ are the weight matrix and bias vector of the coefficient generation layer, respectively, and $\text{sigmoid}(\cdot)$ is the activation function.

$$\begin{cases} \mathbf{z}^{(n)} = [x_3(n-s+1), x_3(n-s+2), \dots, x_3(n)]^T \\ \alpha^{(n)} = \text{sigmoid} \left(\mathbf{W}_\alpha \mathbf{f}^{(n)} + \mathbf{b}_\alpha \right) \end{cases} \quad (15)$$

The final prediction result is obtained by combining the main-branch output with the short-window compensation branch. Equation (16) gives the final prediction output, where $\hat{\mathbf{y}}^{(n)}$ denotes the final H -step prediction result, $\mathbf{W}_s$ and $\mathbf{b}_s$ are the weight matrix and bias vector of the short-window mapping layer, respectively, and $\odot$ denotes element-wise multiplication.

$$\hat{\mathbf{y}}^{(n)} = \hat{\mathbf{y}}_{\text{main}}^{(n)} + \alpha^{(n)} \odot \left(\mathbf{W}_s \mathbf{z}^{(n)} + \mathbf{b}_s \right) \quad (16)$$

Equation (16) shows that the model performs the main prediction using the fused dual-branch features while incorporating recent observations at the return-air corner for adaptive compensation. This design improves the stability and accuracy of short-term prediction. The proposed fusion prediction module consists of the main-branch mapping, the short-window branch, and the adaptive compensation mechanism. It can jointly use low-frequency overall variation information, high-frequency local disturbance information, and recent target-point observations to achieve multi-step prediction.

In this study, the 1-step, 3-step, and 6-step prediction tasks were implemented using a direct multi-output forecasting strategy rather than a recursive forecasting strategy [27]. For each prediction horizon $H \in \{1, 3, 6\}$, an independent model with the same architecture was trained, and only the output dimension of the final prediction layer was adjusted according to H . Equation (17) defines the direct multi-output forecasting process, where $\mathbf{X}^{(n)}$ denotes the historical multi-point input sequence ending at time n , $f_{\theta_H}(\cdot)$ denotes the direct forecasting model trained for prediction horizon H , and $\hat{\mathbf{y}}^{(n,H)}$ denotes the H -step prediction vector of the return-air corner gas concentration.

$$\hat{\mathbf{y}}^{(n,H)} = f_{\theta_H} \left(\mathbf{X}^{(n)} \right) = [\hat{x}_3(n+1), \hat{x}_3(n+2), \dots, \hat{x}_3(n+H)]^T \quad (17)$$

In Equation (17), the model directly outputs the future return-air corner gas concentration sequence from $n + 1$ to $n + H$ in one forward pass. The predicted value at a previous step is not fed back into the model as an input for the next step. Therefore, the proposed forecasting strategy avoids the explicit step-by-step error accumulation that commonly occurs in recursive multi-step forecasting.

In summary, the proposed model integrates VMD-based frequency decomposition, dual-branch graph spatiotemporal feature extraction, and adaptive feature fusion for multi-step prediction of gas concentration at the return-air corner. By outputting 1-step, 3-step, and 6-step prediction results, the model provides a basis for subsequent short-term risk judgement under different prediction horizons.

3. Data Preprocessing

3.1. Data Source and Feature Analysis

The data used in this study were obtained from the gas monitoring system of the 151306 working face in a high-gas coal mine in Shanxi Province, China. Four key monitoring points were selected as model inputs, including gas concentration at the mining face, gas concentration in the return airway, gas concentration at the return-air corner, and gas concentration in the low-level gas drainage pipeline. Among them, the gas concentration at the return-air corner was used as the only prediction target.

The sampling interval was 2 min, and the data were collected from 00:00 on 1 March 2025 to 00:00 on 1 April 2025. The monitoring records during this period were continuous and complete, reflecting the variation characteristics of gas concentration under actual production conditions. The locations of the four monitoring points are shown in Figure 2.

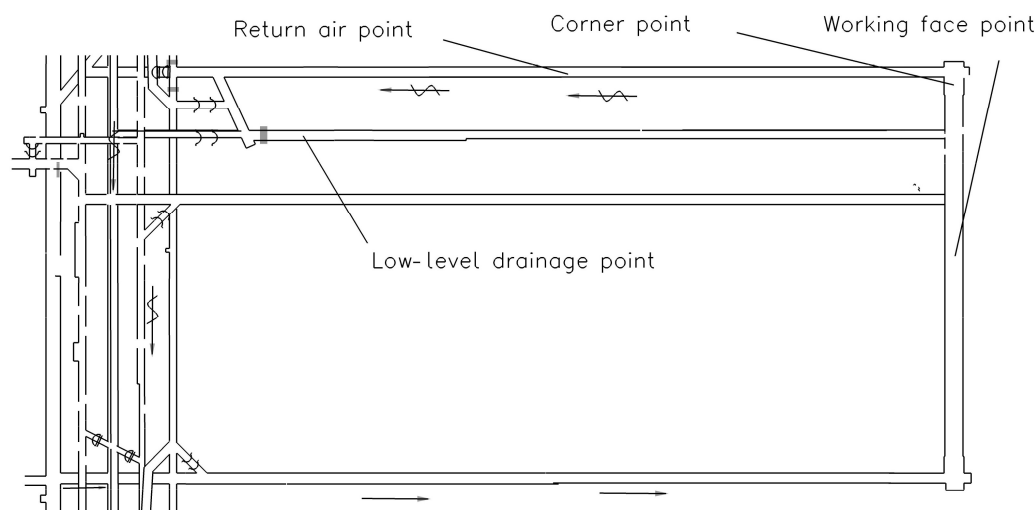


Figure 2. Distribution of monitoring points at the 151306 working face of a high-gas coal mine.

The four monitoring points correspond to different stages of the gas migration process, including gas emission from the working face, return-airflow transport, gas accumulation at the return-air corner, and gas drainage variation. These variables are correlated with each other, but they also differ in fluctuation amplitude and response time. Therefore, a four-input and one-output modeling strategy is adopted to improve the accuracy and practical applicability of gas concentration prediction at the return-air corner.

To further characterize the statistical properties of the monitored gas concentration data, descriptive statistics were calculated for the four monitoring variables before normalization, as shown in Table 1. The statistical indicators include the minimum, maximum, mean, standard deviation, skewness, and kurtosis values. In addition, histograms of the

four gas concentration series are presented in Figure 3 to illustrate their distributional characteristics.

Table 1. Descriptive statistics of gas concentration data from the four monitoring points.

Monitoring Variable	Min (%)	Max (%)	Mean (%)	Std. (%)	Skewness	Kurtosis
Mining face	0.02	1.8563	0.1501	0.1047	1.1462	5.5108
Return-air	0.08	2.0038	0.2313	0.119	1.0837	6.0397
Return-air corner	0.005	1.9023	0.128	0.0916	1.7675	16.2748
Low-level drainage	0	2.524	0.4417	0.4809	0.8985	0.1363

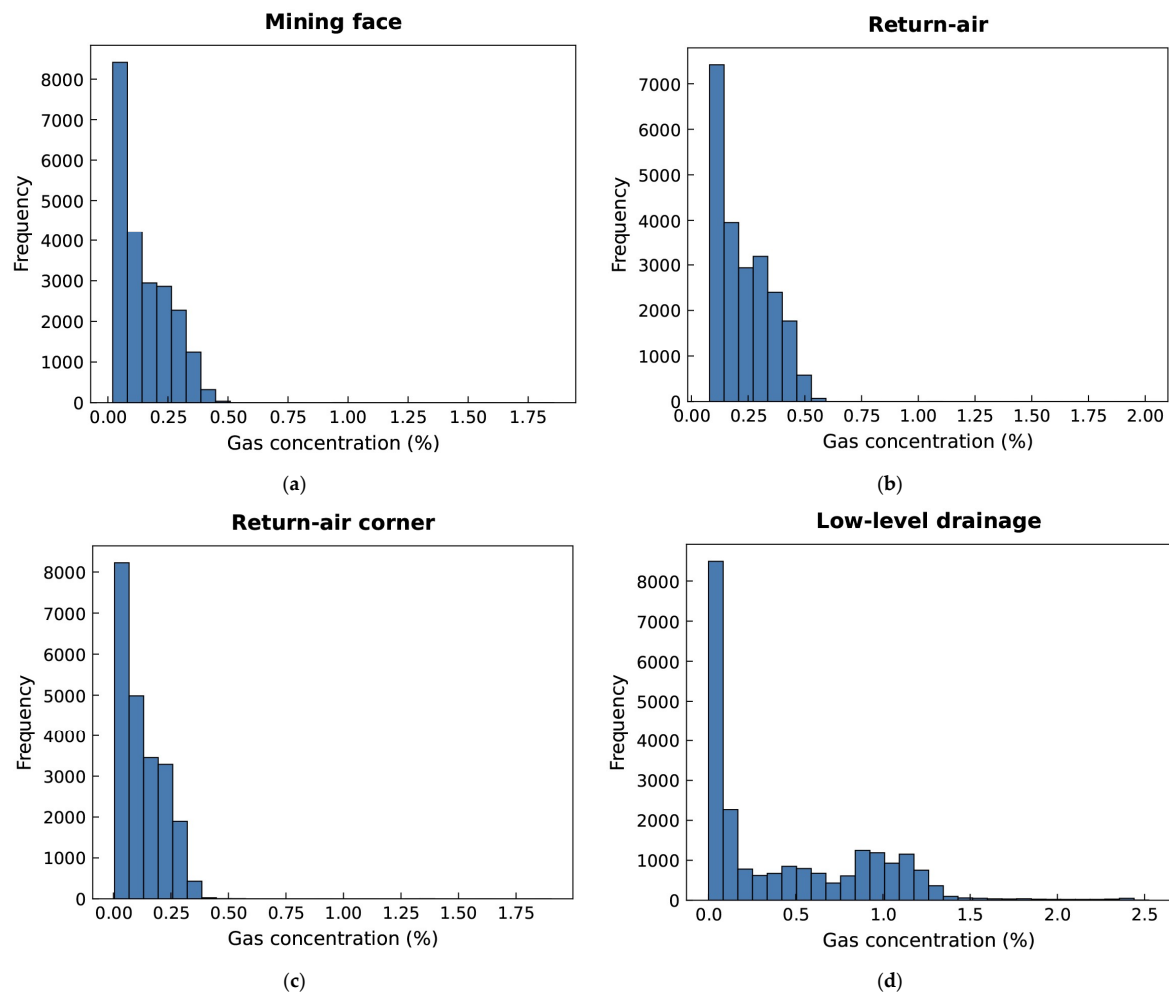


Figure 3. Histograms of gas concentration distributions at the four monitoring points. (a) Mining face; (b) Return-air; (c) Return-air corner; (d) Low-level drainage.

As shown in Table 1 and Figure 3, the gas concentration series at different monitoring points exhibit clear differences in value range and fluctuation intensity. The return-air corner gas concentration, which is used as the prediction target in this study, shows relatively evident variability and non-Gaussian distribution characteristics. The skewness and kurtosis values indicate that the monitored gas concentration series contain asymmetric fluctuations and local peak variations rather than following a simple normal distribution. These characteristics are consistent with the influence of ventilation, gas emission, mining disturbance, and drainage conditions in the working face. Therefore, the dataset contains both slowly varying trend information and short-term local disturbance information, which provides the data basis for the subsequent VMD-based frequency decomposition and dual-branch graph spatiotemporal modeling.

3.2. VMD and Low-/High-Frequency Sequence Reconstruction

Before model construction, missing values were filled by linear interpolation, while missing values at the beginning and end of the sequence were handled by forward and backward filling. Outliers were identified using the 3σ criterion and corrected by threshold truncation to maintain the continuity of the time series.

To avoid information leakage, the dataset was divided in chronological order into training, validation, and test sets at a ratio of 70%, 15%, and 15%, respectively. The Min–Max normalization parameters were fitted only on the training set and then applied to the validation and test sets. For VMD, the hyperparameters, including K , α , τ , DC, initialization mode, and convergence tolerance, were determined based on the training set and then kept fixed for the validation and test sets [28]. VMD decomposition was performed after chronological data splitting under the same parameter configuration. This strategy avoids the use of future test set information during model training while maintaining a consistent decomposition setting across different data subsets.

VMD was performed separately on the four monitoring variables. The parameters were set as follows: penalty factor $\alpha = 2000$, noise tolerance $\tau = 0$, number of modes $K = 5$, DC component $DC = 0$, frequency initialization mode $init = 1$, and convergence tolerance $tol = 1 \times 10^{-7}$.

Under these settings, each monitoring variable was decomposed into five IMF components. The value of $K = 5$ was selected by considering both decomposition completeness and mode redundancy. When K was too small, the trend and disturbance information could not be sufficiently separated, whereas an excessively large K could lead to redundant modes and possible over-decomposition [29]. According to the center-frequency order and fluctuation characteristics of the decomposed modes, IMF1 and IMF2 were reconstructed as the low-frequency component, while IMF3, IMF4, and IMF5 were reconstructed as the high-frequency component. IMF1 and IMF2 mainly preserve the slow variation trend and overall fluctuation level, whereas IMF3–IMF5 mainly reflect rapid oscillations, local disturbances, and intermittent peak responses.

Under these settings, each monitoring variable was decomposed into five IMF components. According to the frequency characteristics of the IMF components, IMF1 and IMF2 were reconstructed as the low-frequency part, while IMF3, IMF4, and IMF5 were reconstructed as the high-frequency part. Taking the gas concentration at the return-air corner as an example, the VMD results are shown in Figure 4.

As shown in Figure 4, the gas concentration sequence at the return-air corner exhibits clear multiscale characteristics after VMD. IMF1 and IMF2 mainly represent low-frequency variation information. Specifically, IMF1 preserves the main profile of the original signal, while IMF2 further reflects slow fluctuations superimposed on the trend variation. During the middle and late stages of the monitoring period, both the overall level and fluctuation intensity of IMF1 and IMF2 increase markedly. In contrast, IMF3, IMF4, and IMF5 have relatively small amplitudes and mainly show local rapid oscillations and intermittent peak responses, indicating short-term disturbance characteristics in the sequence.

As illustrated by the low- and high-frequency components in Figure 4, the low-frequency component effectively retains the main variation process of the raw data and reflects the overall trend of gas concentration at the return-air corner. It clearly captures the sustained increase and intensified fluctuation after approximately mid-March 2025. The high-frequency component fluctuates around zero with a much smaller amplitude than the low-frequency component, mainly representing local disturbances and short-term abnormal variations [30]. After reconstruction, the main variation information and rapid fluctuation information in the raw data are separated, providing a basis for subsequent modeling by the low-frequency and high-frequency branches.

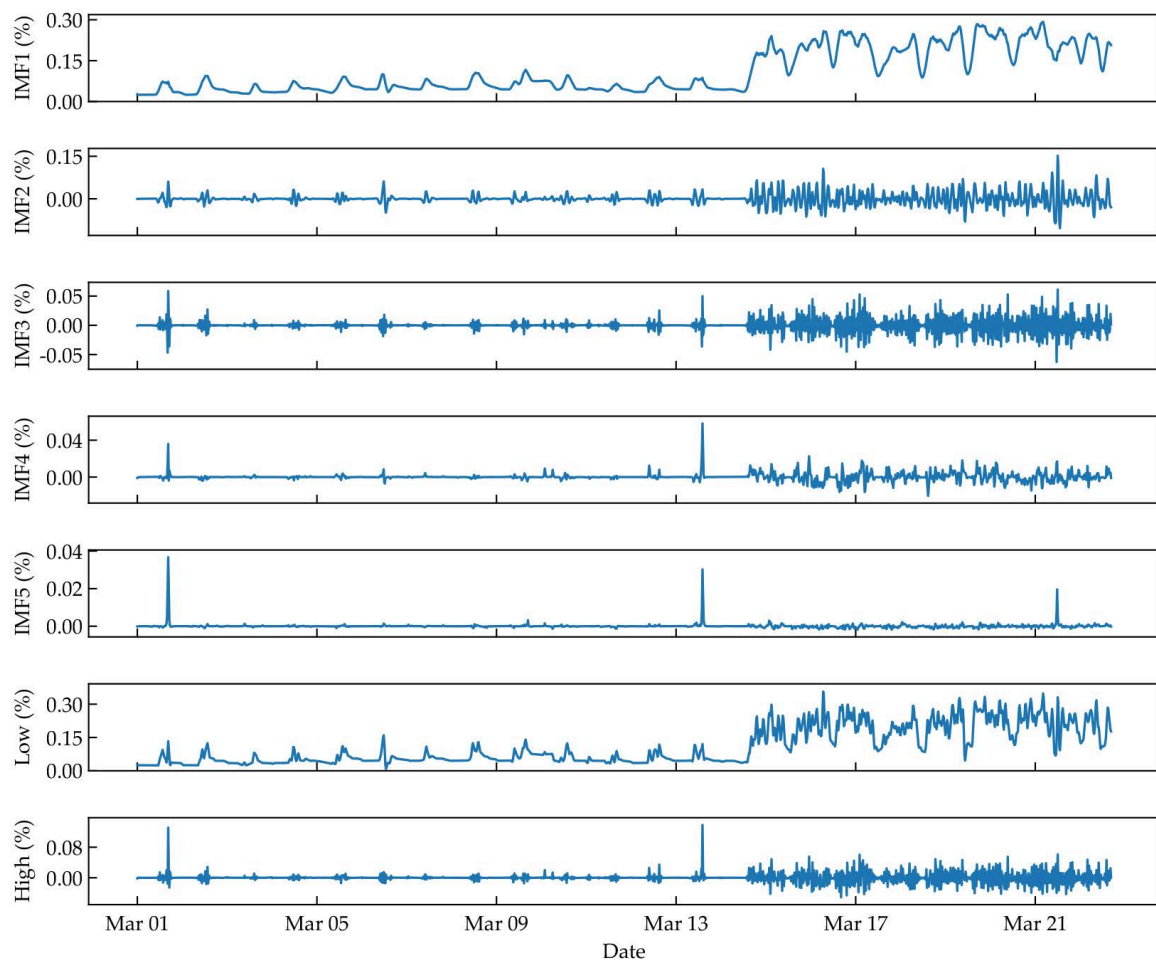


Figure 4. VMD results of gas concentration at the return-air corner with actual monitoring dates.

4. Experiments and Results

4.1. Parameter Settings and Evaluation Metrics

Model training and testing were conducted on a computer running Windows 11, equipped with an Intel i7-8750H processor and 32 GB of RAM. The programming environment was Python 3.13.1, and the deep learning framework was PyTorch 2.9.1.

Since the sampling interval of the monitoring data was 2 min, 1-step prediction corresponds to approximately 2 min ahead forecasting. It was selected as the main prediction task because it is closest to the real-time monitoring interval and directly reflects the model's ability to track short-term gas concentration variations at the return-air corner. The 3-step and 6-step prediction tasks correspond to approximately 6 min and 12 min ahead forecasting, respectively, and were used to evaluate the medium-/short-term forecasting ability of the model and provide additional lead time for short-term warning.

For the three prediction horizons, independent models were trained under the same architecture and hyperparameter settings, with only the output dimension adjusted according to H . The predicted values were generated directly in one forward pass and were not recursively fed back into the input sequence. The batch size was 32, the number of training epochs was 40, the learning rate was 1×10^{-3} , and the random seed was fixed at 42. The Adam optimizer was used for model optimization. The detailed architecture and hyperparameter settings of the proposed model are summarized in Table 2. The same architecture was used for the 1-step, 3-step, and 6-step prediction tasks, and only the output dimension of the final prediction layer was adjusted according to the prediction horizon H .

Table 2. Detailed architecture and hyperparameter settings of the proposed model.

Module	Main Setting	Value
SAGL	Node embedding dimension	16
Low-frequency GCN	Layers/output dimension	1/32
Low-frequency GRU	Layers/hidden units	1/64
High-frequency GAT	Heads/hidden units per head	4/16
High-frequency GAT	Output dimension	32
High-frequency GRU	Layers/hidden units	1/64
Feature fusion MLP	Input dimension/hidden dimension	128/64
Feature fusion MLP	Structure	Linear-ReLU
Short-window compensation	Window length/maximum coefficient	3/0.10
Residual correction	Residual scale	0.10
Dropout	Dropout rate	0.1

The different dimensions in Table 2 correspond to different feature transformation stages rather than inconsistent settings of the same hidden layer. In the low-frequency branch, the SAGL module uses a 16-dimensional node embedding to learn adaptive graph relationships, the GCN maps the input features to a 32-dimensional spatial representation, and the GRU further encodes temporal dependence into a 64-dimensional hidden state. In the high-frequency branch, each GAT head produces a 16-dimensional representation, four attention heads are concatenated and projected to a 32-dimensional output, and the GRU then generates a 64-dimensional temporal hidden state. During feature fusion, the 64-dimensional low-frequency and high-frequency target-node representations are concatenated into a 128-dimensional vector and then mapped by the MLP to a 64-dimensional fused representation. Therefore, these dimensions are stage-specific architecture settings used for graph learning, spatial feature extraction, temporal encoding, and feature fusion, respectively.

The GRU modules in both branches adopt the standard gated recurrent unit structure, including the update gate, reset gate, and candidate hidden state. These internal gate parameters are trainable and are optimized together with the graph learning, graph convolution, graph attention, and fusion module parameters during model training.

During training, the total loss was composed of the main prediction loss, the low-frequency auxiliary loss, and the high-frequency auxiliary loss. Equation (18) defines the total loss, where $\mathcal{L}$ denotes the total loss, $\mathcal{L}_{\text{pred}}$ denotes the main prediction loss, and $\mathcal{L}_{\text{low}}$ and $\mathcal{L}_{\text{high}}$ denote the auxiliary losses of the low-frequency and high-frequency branches, respectively. The two auxiliary losses were assigned weights of 0.3. This setting provides additional supervision for the low-frequency and high-frequency branches while keeping the main prediction loss dominant. A smaller auxiliary weight may provide insufficient branch-level supervision, whereas an excessively large weight may weaken the optimization of the final fused prediction output. Therefore, 0.3 was used as a balanced setting in this study.

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + 0.3\mathcal{L}_{\text{low}} + 0.3\mathcal{L}_{\text{high}} \quad (18)$$

All branch losses were calculated using the mean squared error. For the main prediction branch, Equation (19) gives the loss function, where B is the batch size, H is the prediction horizon, $y_{b,h}$ denotes the true value of the b -th sample at the h -th prediction step, and $\hat{y}_{b,h}$ denotes the corresponding predicted value.

$$\mathcal{L}_{\text{pred}} = \frac{1}{BH} \sum_{b=1}^B \sum_{h=1}^H (y_{b,h} - \hat{y}_{b,h})^2 \quad (19)$$

To evaluate prediction performance, mean absolute error (MAE), root mean square error (RMSE), and the coefficient of determination (R^2) were used. Equations (20)–(22) define these metrics, where M denotes the total number of evaluated samples, y_j and $\hat{y}_j$ are the true and predicted values of the j -th sample, respectively, and $\bar{y}$ is the mean of the true values.

$$\text{MAE} = \frac{1}{M} \sum_{j=1}^M |y_j - \hat{y}_j| \quad (20)$$

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{j=1}^M (y_j - \hat{y}_j)^2} \quad (21)$$

$$R^2 = 1 - \left(\frac{\sum_{j=1}^M (y_j - \hat{y}_j)^2}{\sum_{j=1}^M (y_j - \bar{y})^2} \right) \quad (22)$$

Smaller MAE and RMSE values indicate lower prediction errors, while an R^2 value closer to 1 indicates better fitting performance.

4.2. Prediction Result Analysis

The prediction performance of the proposed model was first evaluated under the 1-step prediction condition. The test results show that the MAE, RMSE, and R^2 for gas concentration prediction at the return-air corner are 0.0338, 0.0471, and 0.9499, respectively, indicating that the model can effectively capture the short-term variation characteristics of the target variable.

Figure 5 shows the single-step prediction results for gas concentration at the return-air corner. Figure 5a presents the global comparison between the predicted and observed values, while Figure 5b provides a local zoomed-in comparison. As shown in Figure 5a, the predicted curve is highly consistent with the observed curve over most test samples, indicating that the model can effectively track the overall variation trend of gas concentration. Figure 5b further shows that the model maintains a close fit in the local interval with obvious fluctuations, including rising, falling, and slight oscillation processes. Although small deviations occur at several peak and turning points, the predicted curve follows the same overall direction as the observed curve, with no obvious lag or distortion.

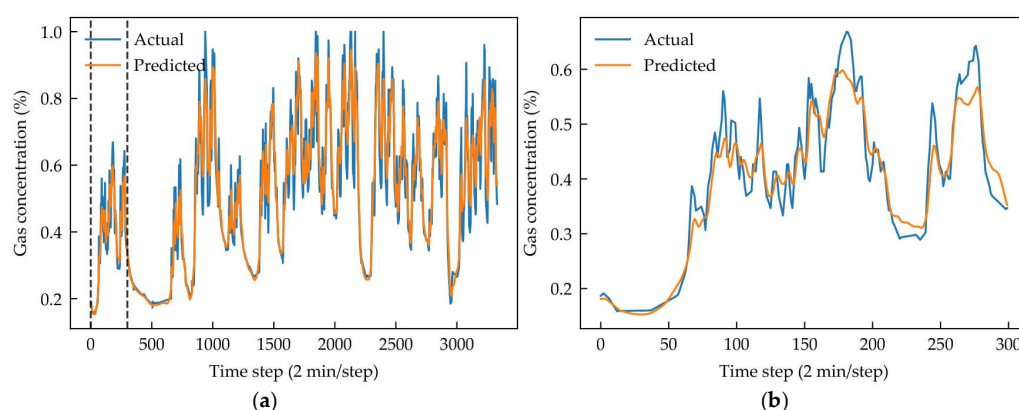


Figure 5. Single-step prediction results. (a) Comparison of global prediction results; (b) Comparison of local zoomed-in results.

From the perspective of field application, accurate short-term prediction is essential for early gas exceedance warning at the return-air corner. The results above demonstrate that the proposed VMD-based dual-branch fusion model can jointly use overall trend information and local disturbance information, thereby achieving reliable 1-step prediction

performance. This provides the basis for further comparative experiments, ablation analysis, and generalization validation under 3-step and 6-step prediction conditions.

To further evaluate the applicability of the proposed model in field monitoring scenarios, the number of parameters, training time, and single-inference time under different prediction horizons were calculated, as shown in Table 3.

Table 3. Complexity and inference efficiency of the proposed model.

Prediction Horizon	Number of Parameters	Training Time (s)	Inference Time per Sample (ms)
1-step	72,257	166.12	0.2035
3-step	72,387	119.41	0.1855
6-step	72,582	159.52	0.2223

When the prediction horizon increases from 1-step to 6-step, the number of model parameters increases only slightly from 72,257 to 72,582. The training time ranges from 119.41 to 166.12 s, and the single-inference time ranges from 0.1855 to 0.2223 ms. Considering that the sampling interval of the monitoring data is 2 min, the inference time is far shorter than the sampling period. These results indicate that the proposed model maintains high prediction accuracy with low computational cost, showing its potential for online short-term prediction in mine safety monitoring systems.

4.3. Comparative Experiment Result Analysis

To ensure the fairness of the comparison, all baseline models were trained and tested under the same experimental conditions. A two-stage training strategy was adopted in the comparative experiments. In the first stage, the skip-connection branch was disabled and the models were trained for 20 epochs. In the second stage, limited compensation was enabled and training continued for another 20 epochs, giving a total of 40 epochs. The learning rate was set to 0.001, the Adam optimizer was used, the loss function was MSELoss, the early stopping patience was set to 6, and the random seed was fixed at 42. Table 4 reports the quantitative comparison results of the proposed model with GRU, LSTM, GCN-GRU, and GAT-GRU under $H = 1$, $H = 3$, and $H = 6$. Figure 6 further compares their prediction performance in terms of MAE, RMSE, and R^2 , where lower MAE and RMSE values indicate smaller prediction errors, and a higher R^2 value indicates better fitting performance.

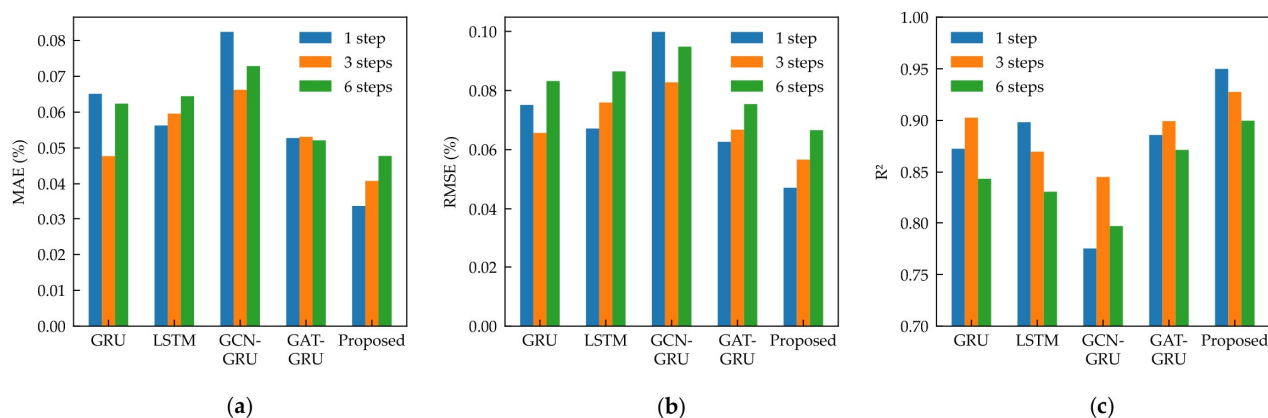


Figure 6. Comparison of prediction performance under different prediction horizons. (a) MAE; (b) RMSE; (c) R^2 .

Table 4. Comparison of experimental results.

Models	MAE (%)			RMSE (%)			R^2		
	1-Step	3-Step	6-Step	1-Step	3-Step	6-Step	1-Step	3-Step	6-Step
GRU	0.0651	0.0477	0.0624	0.0751	0.0657	0.0832	0.8727	0.9027	0.8436
LSTM	0.0563	0.0596	0.0645	0.0671	0.0759	0.0865	0.8983	0.8698	0.8310
GCN-GRU	0.0824	0.0662	0.0729	0.0998	0.0828	0.0948	0.7752	0.8453	0.7968
GAT-GRU	0.0528	0.0531	0.0521	0.0627	0.0668	0.0754	0.8859	0.8994	0.8715
Proposed *	0.0338	0.0408	0.0478	0.0471	0.0566	0.0666	0.9499	0.9276	0.8997

* Proposed denotes the model proposed in this study. Bold values indicate the best performance under each prediction horizon.

At $H = 1$, the proposed model has the lowest error bars and the highest R^2 , showing clearly better overall performance than the other baseline models. Compared with GAT-GRU, the MAE and RMSE decrease by approximately 36% and 25%, respectively. Compared with GCN-GRU, they decrease by approximately 59% and 53%, respectively, while R^2 increases by approximately 23%.

At $H = 3$, GRU is the best-performing baseline model among the compared methods. Even so, the proposed model reduces both MAE and RMSE by approximately 14% compared with GRU, and improves R^2 by approximately 3%. Compared with GCN-GRU, the MAE and RMSE decrease by approximately 38% and 32%, respectively.

At $H = 6$, the prediction errors of all models increase further, but the proposed model still achieves the best results. Compared with the better-performing GAT-GRU, the proposed model reduces MAE and RMSE by approximately 8% and 12%, respectively, and improves R^2 by approximately 3%. Compared with GCN-GRU, the MAE and RMSE decrease by approximately 34% and 30%, respectively, while R^2 increases by approximately 13%.

The performance differences among the baseline models can be explained by their feature representation mechanisms. GRU and LSTM mainly model temporal dependencies and do not explicitly describe the spatial response relationships among different monitoring points. Therefore, they are limited when the return-air corner concentration is affected by gas emission, return-airflow transport, corner accumulation, and drainage variation simultaneously. GCN-GRU and GAT-GRU introduce graph-based spatial modeling, but they still operate directly on the original gas concentration sequences. As a result, low-frequency trends and high-frequency disturbances are mixed in the same input space, which may weaken the stability of spatial feature extraction. In contrast, the proposed model first separates the raw sequence into low- and high-frequency components and then extracts their spatiotemporal features through different branches. This design reduces the interference between trend information and local disturbance information, leading to better prediction performance under different horizons.

Overall, Table 4 and Figure 6 show that the proposed model outperforms the four baseline models at $H = 1$, $H = 3$, and $H = 6$. The improvement is most pronounced at $H = 1$, indicating that the proposed VMD-based dual-branch fusion prediction method can achieve high short-term prediction accuracy while maintaining good medium-short-term forecasting performance.

To further evaluate the temporal stability of different models, a rolling-window analysis was conducted on the 1-step prediction results of the test set after inverse normalization. The trained models were kept fixed, and the test samples were divided into overlapping windows with a window length of 300 samples and a step size of 100 samples. MAE, RMSE, and R^2 were calculated within each window, and the mean and standard deviation values are summarized in Table 5.

Table 5. Rolling-window performance statistics under the 1-step prediction condition.

Model	MAE Mean $\pm$ Std.	RMSE Mean $\pm$ Std.	R ² Mean $\pm$ Std.
GRU	0.0686 $\pm$ 0.0152	0.0747 $\pm$ 0.0187	0.8529 $\pm$ 0.0612
LSTM	0.0541 $\pm$ 0.0118	0.0664 $\pm$ 0.0145	0.8813 $\pm$ 0.0487
GCN-GRU	0.0856 $\pm$ 0.0194	0.1043 $\pm$ 0.0231	0.7436 $\pm$ 0.0834
GAT-GRU	0.0519 $\pm$ 0.0093	0.0655 $\pm$ 0.0112	0.8921 $\pm$ 0.0356
Proposed	0.0347 $\pm$ 0.0053	0.0474 $\pm$ 0.0076	0.9417 $\pm$ 0.0249

Bold values indicate the best performance under each prediction horizon.

To further evaluate the robustness and statistical significance of the comparison with the basic baseline models, rolling-window analysis and Wilcoxon signed-rank tests were conducted. In the rolling-window analysis, the trained models were kept fixed, and the 1-step test samples were divided into overlapping windows with a window length of 300 samples and a step size of 100 samples. As shown in Table 5, the proposed model achieves the best rolling-window performance among GRU, LSTM, GCN-GRU, GAT-GRU, and the proposed model, with MAE, RMSE, and R² values of 0.0347 $\pm$ 0.0053, 0.0474 $\pm$ 0.0076, and 0.9417 $\pm$ 0.0249, respectively. The smaller standard deviations indicate stable prediction performance over different test periods. In addition, the Wilcoxon signed-rank test based on the absolute prediction errors shows that the proposed model achieves statistically significant differences from the basic baseline models in most prediction settings, further confirming the effectiveness of the proposed method.

To further evaluate the predictive performance of the proposed model against stronger baselines, three additional models, namely VMD-GRU, Informer, and STGCN, were introduced for comparison. VMD-GRU was used to examine the effect of single-branch recurrent prediction after frequency decomposition, Informer was used to evaluate the performance of a long-sequence time series forecasting model, and STGCN was used to assess spatiotemporal prediction capability under a fixed graph structure. All models used the same data split, input length, prediction horizons, and evaluation metrics as the proposed model. Recent studies have introduced Informer-based deep forecasting, optimized SVR, and feature selection-based machine learning models for coal mine gas prediction, providing useful references for selecting stronger comparison models and interpreting predictive differences [27–29]. The results are shown in Table 6.

Table 6. Comparison of strong baseline experimental results.

Models	MAE (%)			RMSE (%)			R ²		
	1-Step	3-Step	6-Step	1-Step	3-Step	6-Step	1-Step	3-Step	6-Step
VMD-GRU	0.0352	0.0423	0.0491	0.0486	0.0581	0.0684	0.9421	0.9183	0.8892
Informer	0.0364	0.0439	0.0505	0.0497	0.0593	0.0698	0.9378	0.9135	0.8846
STGCN	0.0381	0.0456	0.0522	0.0512	0.0609	0.0713	0.9315	0.9072	0.8785
Proposed	0.0338	0.0408	0.0478	0.0471	0.0566	0.0666	0.9499	0.9276	0.8997

Bold values indicate the best performance under each prediction horizon.

As shown in Table 6, the proposed model outperforms VMD-GRU, Informer, and STGCN under all three prediction horizons. Compared with VMD-GRU, the MAE of the proposed model decreases by 4.0%, 3.5%, and 2.6% at $H = 1$, $H = 3$, and $H = 6$, respectively, while the RMSE decreases by 3.1%, 2.6%, and 2.6%, respectively. This indicates that although VMD can improve prediction performance, a single-branch recurrent structure is still insufficient for separately representing low-frequency trends and high-frequency disturbances. Compared with Informer and STGCN at $H = 6$, the proposed model reduces MAE by 5.3% and 8.4%, reduces RMSE by 4.6% and 6.6%, and improves R² by 1.5% and

2.1%, respectively. These results show that the proposed model maintains lower prediction errors and higher fitting accuracy as the prediction horizon increases.

The comparison with the stronger baselines further demonstrates the necessity of combining frequency decoupling with differentiated spatiotemporal modeling. VMD-GRU benefits from frequency decomposition and therefore obtains closer results than models without decomposition. However, it still uses a single recurrent structure to process the reconstructed components, which makes it difficult to separately capture slow trend evolution and rapid local disturbances. Informer is effective for long-sequence temporal dependency modeling, but it does not explicitly incorporate the monitoring-point relationships formed by ventilation, drainage, and gas migration processes. STGCN introduces spatiotemporal graph modeling, but its fixed graph structure is less flexible for describing the dynamic relationships among monitoring points under changing production conditions. Compared with these models, the proposed method combines VMD-based frequency decoupling, self-adaptive graph learning, graph attention, and dual-branch temporal encoding, which provides a more suitable representation for nonstationary multi-point gas concentration prediction.

Overall, VMD-GRU, Informer, and STGCN all show certain predictive capability. However, the proposed model separately models low- and high-frequency components and further integrates self-adaptive graph learning with graph attention. It can therefore better extract dynamic correlations among multiple monitoring points and frequency-specific features, maintaining superior performance even against stronger baseline models.

4.4. Ablation Experiment Result Analysis

To verify the contribution of each component to model performance, four ablation models were constructed based on the complete model: removing the low-frequency branch, removing the high-frequency branch, removing the short-window skip connection, and removing the compensation module. The results are shown in Table 7. The proposed complete model achieves the highest R^2 at $H = 1$, $H = 3$, and $H = 6$, indicating the best overall fitting performance.

Table 7. Ablation experiment results.

Models	MAE (%)			RMSE (%)			R^2		
	1-Step	3-Step	6-Step	1-Step	3-Step	6-Step	1-Step	3-Step	6-Step
<i>w/o Low</i> *	0.0405	0.0490	0.0606	0.0574	0.0703	0.0874	0.8612	0.8352	0.7613
<i>w/o High</i> *	0.0395	0.0478	0.0581	0.0559	0.0679	0.0825	0.9041	0.8548	0.8085
<i>w/o Skip</i> *	0.0376	0.0448	0.0545	0.0529	0.0638	0.0769	0.9217	0.8913	0.8439
<i>w/o Comp</i> *	0.0369	0.0438	0.0530	0.0514	0.0620	0.0755	0.9357	0.9092	0.8633
Proposed	0.0338	0.0408	0.0478	0.0471	0.0566	0.0666	0.9499	0.9276	0.8997

* *w/o Low*, *w/o High*, *w/o Skip*, and *w/o Comp* denote the models without the low-frequency branch, high-frequency branch, short-window skip connection, and compensation module, respectively. Bold values indicate the best performance under each prediction horizon.

To further compare the effects of different modules under each prediction horizon, the relative scores of the evaluation metrics for the ablation models are presented in Figure 7.

The corresponding metric of the proposed model was used as the baseline and set to 100. A lower relative score indicates a more significant performance degradation after removing the corresponding module. For MAE and RMSE, a lower relative score indicates a larger prediction error, while for R^2 , it indicates weaker fitting ability. This representation provides a more intuitive comparison of the contribution of each module to prediction performance.

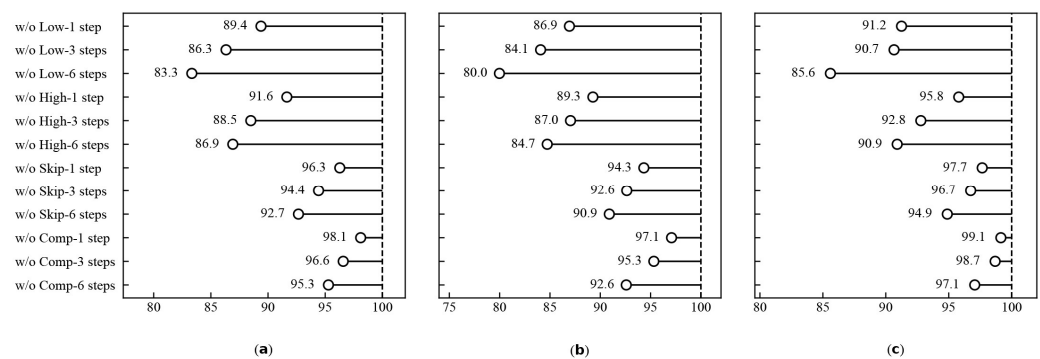


Figure 7. Relative scores of ablation models under different prediction horizons, with the proposed model set as 100. (a) Relative MAE score; (b) Relative RMSE score; (c) Relative R^2 score.

As shown in Figure 7, the relative scores of the ablation models generally decrease as the prediction horizon increases, indicating that the performance advantage of the complete model becomes more evident for longer horizons. Among all ablation models, removing the low-frequency branch causes the most significant degradation. At $H = 6$, the relative scores of MAE, RMSE, and R^2 decrease to 83.3, 80.0, and 85.6, respectively. Removing the high-frequency branch results in the second-largest degradation, with corresponding scores of 86.9, 84.7, and 90.9. In contrast, the relative scores after removing the short-window skip connection and the compensation module remain closer to 100, suggesting that these two components provide additional performance gains but contribute less than the dual-frequency branches.

To ensure consistency with the main experimental results, the performance of the complete proposed model in the ablation study was aligned with the corresponding results reported in the main comparison experiments.

Overall, the contribution of each module follows the order of low-frequency branch > high-frequency branch > short-window skip connection > compensation module. This indicates that the performance improvement of the proposed model mainly comes from the separate modeling of trend information and disturbance information by the low- and high-frequency branches, while the short-window skip connection and compensation module play auxiliary enhancement roles. These results demonstrate the rationality and effectiveness of the proposed dual-branch fusion prediction structure.

4.5. Generalization Ability Validation

To further verify the generalization ability of the proposed model, an additional experiment was conducted using the gas monitoring dataset from Ref. [22]. The sampling interval of this dataset was also 2 min, and the monitoring period ranged from 00:00 on 1 December 2017 to 13:40 on 20 December 2017. Compared with the main experiment, the monitoring points in the new dataset were changed to mixed return-air gas concentration, return-airway gas concentration, return-air corner gas concentration, and intake-airway gas concentration. The prediction target remained the gas concentration at the return-air corner, and the other experimental settings were kept unchanged.

For the cross-dataset validation, the model was retrained from scratch on the new dataset rather than using transfer learning. The same model architecture and hyperparameter settings were retained, while the normalization parameters were fitted only on the training subset of the new dataset. This setting was used to evaluate the adaptability of the proposed modeling framework under different mining conditions, rather than the direct transferability of pretrained weights.

The comparison results of different models on this dataset are shown in Table 8. The proposed model still achieves the best performance at $H = 1$, $H = 3$, and $H = 6$, indicating good stability and adaptability on the new dataset.

Table 8. Generalization experiment results.

Models	MAE (%)			RMSE (%)			R^2		
	1-Step	3-Step	6-Step	1-Step	3-Step	6-Step	1-Step	3-Step	6-Step
GRU	0.0422	0.0311	0.0402	0.0634	0.0503	0.0642	0.8566	0.9095	0.8526
LSTM	0.0314	0.0381	0.0467	0.0527	0.0607	0.0696	0.9008	0.8685	0.8271
GCN-GRU	0.0230	0.0290	0.0341	0.0346	0.0447	0.0551	0.9573	0.9287	0.8915
GAT-GRU	0.0192	0.0327	0.0388	0.0322	0.0480	0.0621	0.9631	0.9177	0.8622
Proposed	0.0161	0.0267	0.0257	0.0303	0.0442	0.0461	0.9672	0.9301	0.9240

Bold values indicate the best performance under each prediction horizon.

As shown in Table 8, under $H = 1$, the proposed model reduces MAE and RMSE by approximately 16.1% and 5.9%, respectively, compared with GAT-GRU, and by approximately 61.8% and 52.2%, respectively, compared with GRU. Under $H = 3$, the MAE decreases by approximately 7.9% compared with the better-performing GCN-GRU. Under $H = 6$, the advantage of the proposed model becomes more evident, with MAE and RMSE reduced by approximately 24.6% and 16.3%, respectively, compared with GCN-GRU. These results indicate that the proposed model can still maintain good prediction accuracy when the monitoring-point configuration changes and the prediction horizon increases.

Overall, the proposed model maintains good prediction performance under different datasets and monitoring-point combinations, demonstrating certain generalization ability and engineering application value. This advantage is mainly related to the hierarchical extraction of low- and high-frequency features and the dynamic modeling of relationships among monitoring points.

4.6. Short-Term Warning Strategy for Gas Exceedance

The practical purpose of predicting gas concentration at the return-air corner is to provide an early basis for identifying gas exceedance risk. To convert multi-step prediction results into field-applicable warning information, a lightweight short-term warning strategy is established in this study. This strategy does not introduce an additional classification model. Instead, it performs risk classification based only on the 1-step, 3-step, and 6-step prediction values, the warning threshold, and a continuous judgement rule. Therefore, the calculation process is simple and suitable for integration into mine safety monitoring systems.

Let the warning threshold of gas concentration at the return-air corner be C_{th} . This threshold is not involved in model training and can be set according to regulatory alarm limits or internal warning values used at the mine site. At the current time t , the future prediction values output by the model are denoted as $\hat{C}_{t+1}$, $\hat{C}_{t+3}$, and $\hat{C}_{t+6}$. To reduce false alarms caused by single-point fluctuations, a “multi-step prediction–trend confirmation–level judgement” strategy is adopted for warning classification. The specific rules are shown in Table 9.

As shown in Table 9, this strategy combines the multi-step prediction results with the gas exceedance threshold, thereby considering both short-term changes and future development trends. When the predicted values continuously approach or exceed the threshold, the warning level increases with the degree of risk. When an abnormal value appears at only one prediction step, the subsequent prediction horizons are used for trend confirmation, reducing misjudgement caused by local fluctuations.

Table 9. Short-term warning levels for gas exceedance risk.

Warning Level	Judgement Condition	Risk Implication
Normal	$\hat{C}_{t+1}$, $\hat{C}_{t+3}$, and $\hat{C}_{t+6}$ are all below the warning threshold.	No obvious exceedance risk in the short term.
Attention	$\hat{C}_{t+3}$ or $\hat{C}_{t+6}$ approaches the warning threshold, and the predicted sequence shows an increasing trend.	Gas concentration may continue to increase.
Warning	Any predicted value exceeds the warning threshold.	Short-term exceedance risk exists.
High risk	Multiple predicted values continuously exceed the threshold, or the 1-step predicted value exceeds the threshold and remains high.	The exceedance risk is high, and timely response measures are required.

The short-term alert strategy in Table 9 is designed as a lightweight post-processing rule based on the existing prediction outputs. It does not introduce an additional classification model, retraining process, or extra optimization procedure. The predicted gas concentration is only compared with the preset warning threshold to provide a rapid risk-estimation reference. Therefore, Table 9 is intended to illustrate the practical use of the prediction results for short-term warning support rather than to provide a complete event-level alarm evaluation.

Compared with threshold alarms based only on current monitoring values, this strategy can identify the rising trend of gas concentration at the return-air corner in advance. Compared with single-step prediction warning, the joint judgement based on multiple prediction horizons improves the stability of warning results. From an engineering application perspective, this strategy only requires the model outputs and the site-specific threshold, without any additional training process, and therefore has low computational cost. When the warning level reaches the attention level or above, the result can be further checked together with the on-site ventilation, gas drainage, and operating conditions. When the high-risk level is reached, measures such as strengthening ventilation, checking the drainage status, and restricting related operations should be taken in time.

5. Discussion

Although the proposed model shows good prediction performance, several issues should be further considered before field deployment. First, the current experiments were mainly conducted using data from limited mine scenarios. The cross-dataset validation was performed by retraining the same model architecture on the additional dataset, rather than by directly applying transfer learning or zero-shot cross-mine prediction. Therefore, the results mainly demonstrate the structural adaptability of the proposed model under different monitoring-point configurations, but they do not fully prove its direct transferability across mines. Future work should further evaluate single-mine training and cross-mine testing, few-shot fine-tuning, and transfer learning strategies under different ventilation systems, gas drainage conditions, and mining stages.

Second, the multi-step prediction tasks in this study used a direct multi-output forecasting strategy rather than a recursive strategy. For each prediction horizon, an independent model with the same architecture was trained, and only the output dimension was adjusted. This design avoids the explicit step-by-step error accumulation caused by feeding previous predicted values back into the model. However, prediction uncertainty still increases as the horizon becomes longer because future gas concentration may be affected by sudden mining disturbance, ventilation adjustment, drainage fluctuation, and abnormal gas emission. Therefore, the stability of 3-step and 6-step prediction under abrupt working-condition changes still needs further verification using more field data.

Third, the short-term warning strategy proposed in this study is a lightweight application-level judgement method based on prediction values and threshold rules. It does not introduce an additional classification model and is mainly intended to provide early reference information for site personnel. Due to the limited number of real gas exceedance events in the available data, quantitative warning indicators such as precision, recall, false alarm rate, missed alarm rate, and warning response time have not yet been systematically evaluated. In future studies, historical exceedance events, near-threshold events, alarm logs, manual inspection records, and field response results should be introduced to validate and calibrate the warning rules. This will help further transform the current lightweight warning strategy into a more complete closed-loop warning and decision-support method.

Finally, the robustness of the model under abnormal monitoring conditions remains to be improved. In practical mine monitoring systems, missing data, sensor drift, communication interruption, noise interference, and rapid production-state changes may lead to input distribution shifts. Future work should incorporate anomaly detection, sensor reliability weighting, online data correction, and model updating mechanisms to improve long-term operational stability. Field deployment should also be evaluated from the perspectives of edge-side inference, real-time data access, model update cycles, and linked visualization of warning information.

6. Conclusions and Future Studies

6.1. Conclusions

This study proposed a VMD-based dual-branch spatiotemporal graph model for short-term prediction of gas concentration at the return-air corner. The model establishes a prediction framework consisting of VMD, low-/high-frequency reconstruction, dual-branch graph spatiotemporal feature extraction, feature fusion, and lightweight warning judgement. The main conclusions are as follows:

- (1) VMD separates the original gas concentration sequence into low-frequency trend components and high-frequency disturbance components. This reduces the interference caused by mixed frequency information and provides clearer inputs for frequency-specific feature extraction.
- (2) The proposed dual-branch structure improves the representation of multi-point gas concentration relationships. The low-frequency branch captures stable trend-level spatiotemporal relationships through self-adaptive graph learning, graph convolution, and GRU encoding, while the high-frequency branch uses graph attention and GRU encoding to enhance the representation of local disturbance responses. The ablation results show that the low-frequency branch contributes most to the model performance, followed by the high-frequency branch, short-window skip connection, and compensation module.
- (3) The proposed model achieves MAE, RMSE, and R^2 values of 0.0338, 0.0471, and 0.9499 under the 1-step prediction condition, respectively. Under the 3-step and 6-step prediction conditions, the model also maintains better performance than GRU, LSTM, GCN-GRU, GAT-GRU, VMD-GRU, Informer, and STGCN. These results indicate that the proposed method can effectively improve the short-term and medium-short-term prediction accuracy of gas concentration at the return-air corner.
- (4) The cross-dataset validation shows that the proposed architecture maintains good prediction performance after retraining under different data sources and monitoring-point configurations. In addition, the lightweight warning strategy based on multi-step prediction results can convert continuous prediction values into graded risk information without adding a complex classification model, providing practical support for early identification of gas exceedance risk.

6.2. Future Studies

Future studies will further extend the proposed method from offline prediction to field-oriented application. More gas monitoring data from different mines, working faces, ventilation systems, gas drainage conditions, and mining stages will be collected to evaluate the model under more diverse operating conditions, especially considering that complex fluid-related engineering processes are often influenced by multiple physical and operational factors [31,32]. Cross-mine experiments, including direct testing, few-shot fine-tuning, and transfer learning, will also be conducted to examine the adaptability of the proposed graph learning structure to new monitoring layouts and production environments.

In addition, real alarm records, near-threshold events, manual inspection records, and field response information will be introduced to further validate the lightweight warning strategy. Quantitative warning indicators, such as hit rate, false alarm rate, missed alarm rate, and early warning lead time, will be evaluated when sufficient historical events are available. Future work will also consider sensor missing data, drift, noise interference, and communication interruption, and will explore online correction and model updating mechanisms to improve the long-term stability of the proposed method in practical mine safety monitoring systems.

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

Funding: This work was financially supported by the National Natural Science Foundation of China (52174181) and the Fundamental Research Funds for the Central Universities (3142015105).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

VMD	Variational mode decomposition
IMF	Intrinsic mode function
SAGL	Self-adaptive graph learning
GCN	Graph convolutional network
GAT	Graph attention network
GRU	Gated recurrent unit
LSTM	Long short-term memory
STGCN	Spatiotemporal graph convolutional network
VMD-GRU	Variational mode decomposition-gated recurrent unit
GCN-GRU	Graph convolutional network-gated recurrent unit
GAT-GRU	Graph attention network-gated recurrent unit
MLP	Multilayer perceptron
MAE	Mean absolute error
RMSE	Root mean square error
MSE	Mean squared error
R ²	Coefficient of determination

References

- Box, G.E.P.; Jenkins, G.M. *Time Series Analysis: Forecasting and Control*; Holden-Day: San Francisco, CA, USA, 1976.
- Li, C.; Fang, X.; Yan, Z.; Huang, Y.; Liang, M. Research on gas concentration prediction based on the ARIMA-LSTM combination model. *Processes* **2023**, *11*, 174. [\[CrossRef\]](#)
- Wang, X.; Xu, N.; Meng, X.; Chang, H. Prediction of gas concentration based on LSTM-LightGBM variable weight combination model. *Energies* **2022**, *15*, 827. [\[CrossRef\]](#)
- Yang, G.; Zhu, Q.; Wang, D.; Feng, Y.; Chen, X.; Li, Q. Method and validation of coal mine gas concentration prediction by integrating PSO algorithm and LSTM network. *Processes* **2024**, *12*, 898. [\[CrossRef\]](#)
- Tutak, M.; Krenicky, T.; Pirník, R.; Brodny, J.; Grebski, W.W. Predicting methane concentrations in underground coal mining using a multi-layer perceptron neural network based on mine gas monitoring data. *Sustainability* **2024**, *16*, 8388. [\[CrossRef\]](#)
- Liang, R.; Chang, X.; Jia, P.; Xu, C. Mine gas concentration forecasting model based on an optimized BiGRU network. *ACS Omega* **2020**, *5*, 28579–28586. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wen, H.; Yan, L.; Jin, Y.; Wang, Z.; Guo, J.; Deng, J. Coalbed methane concentration prediction and early-warning in fully mechanized mining face based on deep learning. *Energy* **2023**, *264*, 126208. [\[CrossRef\]](#)
- Brodny, J.; Felka, D.; Tutak, M. The use of the neuro-fuzzy model to predict the methane hazard during the underground coal mining production process. *J. Clean. Prod.* **2022**, *368*, 133258. [\[CrossRef\]](#)
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017*; NeurIPS: La Jolla, CA, USA, 2017; pp. 5998–6008.
- Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 11106–11115. [\[CrossRef\]](#)
- Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430.
- Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; Long, M. iTransformer: Inverted transformers are effective for time series forecasting. In *Proceedings of the International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024*.
- An, L.; Kong, S.; Li, K. Research on intelligent early warning and emergency response mechanism for tunneling face gas concentration based on an improved KAN-iTransformer. *Processes* **2025**, *13*, 3748. [\[CrossRef\]](#)
- Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017*.
- Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph attention networks. In *Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018*.
- Yu, B.; Yin, H.; Zhu, Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018*; AAAI Press: Palo Alto, CA, USA, 2018; pp. 3634–3640.
- Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018*.
- Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Zhang, C. Graph WaveNet for deep spatial-temporal graph modeling. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019*; AAAI Press: Palo Alto, CA, USA, 2019; pp. 1907–1913.
- Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [\[CrossRef\]](#)
- Liu, H.; Mi, X.; Li, Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. *Energy Convers. Manag.* **2018**, *159*, 54–64. [\[CrossRef\]](#)
- Zhang, J.; Cui, Y.; Yan, Z.; Huang, Y.; Zhang, C.; Zhang, J.; Guo, J.; Zhao, F. Time series prediction of gas emission in coal mining face based on optimized variational mode decomposition and SSA-LSTM. *Sensors* **2024**, *24*, 6454. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hu, Q.; Cheng, Y.; Zheng, S.; Chen, J.; Li, Y. Coal mine methane gas concentration prediction based on PSO-VMD-LSTM and error quadratic decomposition. *Min. Metall. Explor.* **2025**, *42*, 647–663. [\[CrossRef\]](#)
- Peng, Y.; Song, D.; Qiu, L.; Wang, H.; He, X. Combined prediction model of gas concentration based on deep learning and data decomposition. *Sensors* **2023**, *23*, 2883. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yang, H.; Wang, J.; Zhang, H. Research on gas multi-indicator warning method of coal mine working face based on MOA-Transformer. *ACS Omega* **2024**, *9*, 22136–22144. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cheng, Y.; Hu, Q.; Luo, Y.; Zhang, Y. Gas concentration prediction for mine environmental safety: A multi-step prediction framework incorporating decomposition prediction and cyclic error compensation. *J. Environ. Chem. Eng.* **2025**, *13*, 118318. [\[CrossRef\]](#)
- Xing, W.; Yang, Y.; Zhang, Y.; Yang, Y. Gas concentration prediction based on SSA algorithm with CNN-BiLSTM-attention. *Sci. Rep.* **2025**, *15*, 34214. [\[CrossRef\]](#) [\[PubMed\]](#)

27. Olca, K.D.; Yücel, Ö. Deep learning applications for biomass gasifier monitoring: A dual approach with direct and recursive long short-term memory forecasting. *Renew. Energy* **2026**, *268*, 125779. [[CrossRef](#)]
28. Li, Q.; Li, Q.; Wang, F.; Wu, J.; Wang, Y.; Jin, J. Effects of geological and fluid characteristics on the injection filtration of hydraulic fracturing fluid in the wellbores of shale reservoirs: Numerical analysis and mechanism determination. *Processes* **2025**, *13*, 1747. [[CrossRef](#)]
29. Song, W.; Han, X.; Qi, J. Prediction of gas emission in the working face based on LASSO-WOA-XGBoost. *Atmosphere* **2023**, *14*, 1628. [[CrossRef](#)]
30. Lin, H.; Li, W.; Li, S.; Wang, L.; Ge, J.; Tian, Y.; Zhou, J. Coal mine gas emission prediction based on multifactor time series method. *Reliab. Eng. Syst. Saf.* **2024**, *252*, 110443. [[CrossRef](#)]
31. Li, Q.; You, D.; Li, Q.; Wang, F.; Wang, Y.; Yang, Y. Analysis of sedimentation behavior and influencing factors of solid particles in CO₂ fracturing fluid. *Processes* **2025**, *13*, 4049. [[CrossRef](#)]
32. Li, Q.; Li, Q.; Wu, J.; He, K.; Xia, Y.; Liu, J.; Wang, F.; Cheng, Y. Wellhead stability during development process of hydrate reservoir in the Northern South China Sea: Sensitivity analysis. *Processes* **2025**, *13*, 1630. [[CrossRef](#)]

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