IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 June 2023, accepted 9 July 2023, date of publication 13 July 2023, date of current version 20 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3294983

RESEARCH ARTICLE

A Graph Convolutional Encoder-Decoder Model for Methane Concentration Forecasting in Coal Mines

YIFEI GAO[1], XIAOHANG ZHANG[1], (Member, IEEE), TIANBAO ZHANG[2], AND ZHENGREN LI[3]

[1]School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Department of Mathematics, University of Washington, Seattle, WA 98195, USA
[3]School of Modern Post, Beijing University of Posts and Telecommunications, Beijing 100876, China

Corresponding author: Xiaohang Zhang (zhangxiaohang@bupt.edu.cn)

This work was supported by the Beijing University of Posts and Telecommunications (BUPT) Excellent Ph.D. Students Foundation under Grant CX2021132.

ABSTRACT Methane is one of the most dangerous gases produced in the process of coal mining. Because of its flammable and explosive characteristics, it has seriously threatened the life and property safety of coal miners. As a result, accurate and real-time gas concentration forecasting is becoming a crucial but challenging issue for reducing methane risks and accidents. To further improve the efficiency and accuracy of methane concentration forecasting, this paper proposes a graph convolutional encoder-decoder (GCN-ED) network, which can train and infer all the sensors of a coal face as a unified entity. The proposed GCN-ED is composed of the GCN module and the ED module with a parallel structure. The GCN module constructs a priori graph structure through the adjacency relation between sensors in reality and uses a learnable self-adaptive dependency matrix to precisely capture the hidden spatial dependency in the data. The ED module is used to learn complex temporal features with LSTM cells and generate multi-step results of the gas concentration prediction. Experiments are conducted on real coal mine datasets, whose results demonstrate that the GCN-ED achieves the better performance than various state-of-the-art solutions and largely improves the efficiencies of training processes.

INDEX TERMS Graph convolution, encoder-decoder model, spatio-temporal data, methane concentration forecasting, coal mine safety.

I. INTRODUCTION

Throughout the world, there are too many deep underground coal mines with naturally complex and frequently varying environmental conditions. Coal mine safety needs to be given more attention to address different hazards. Among those underground accidents, coal mine gas hazard often accounts for the highest proportion, and causes the most casualties. Coal mine gas accident and other accidents caused by it, like mine fire and explosion [1], have posed a great threat to the safety of miners [2]. As shown in Table 1, methane and coal dust explosions have caused major damage in different countries [3], [4]. As demand for coal resources is still going up every year, the mining conditions of mines is more severe and the gas capacity of coal seams increases gradually. Then, gas accident prevention becomes more difficult and challenging [5]. Fortunately, before gas accidents occur, the change in gas concentration can be observed [6]. Nevertheless, if operators make adjustments until the occurrence of exceed-limit alarms, the operable window periods will inevitably shorten and the probability of hazards occurring will increase. Therefore, it is necessary to monitor and predict precisely the real-time gas concentration in the areas of ongoing mining exploitation [7].

Under the above background, timely and accurate prediction of gas concentration is of great importance for coal mine safety production. At the same time, the scientific efforts of methane concentration forecasting have been a research

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

TABLE 1. Major underground coal mine accidents due to methane explosion in different countries.

Year	Country	Name of mine	Cause of accident	Casualties
2005	China	Sunjiawan Colliery	Methane explosion	214
2006	Poland	Halemba Coal Mine	Methane and coal dust explosion	23
2006	Mexico	Pastade Conchos	Methane explosion	65
2007	China	Ruizhiyuan Mine	Methane explosion	105
2009	China	Xinxing Coal Mine	Methane explosion	108
2010	New Zealand	Pike River Mine	Methane explosion	29
2014	Poland	Myslowice-Wesola Mine	Methane explosion	5
2014	Turkey	Soma Coal Mine	Methane and coal dust explosion	301
2015	China	Xiangyangqu Mine	Methane explosion	19
2017	China	Xingyu Mine	Outburst of methane	12
2019	China	Lijiagou Mine	Methane and coal dust explosion	21
2020	China	Liaoyuan Mine	Methane explosion	7
2021	Russia	Listvyazhnaya Mine	Methane explosion	51

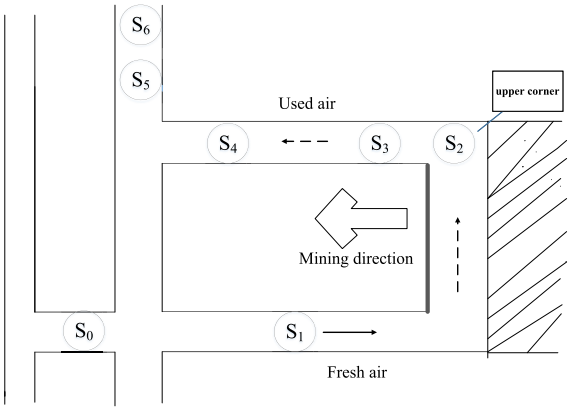

FIGURE 1. The distribution of methane sensors of a coal face.

focus. Two mainstreams of methane concentration prediction methods are the physical methods and data-driven methods. *Physical methods* (e.g., [8], [9], [10]), which include analytical methods and numerical simulations methods, mainly predict the gas concentration by establishing complex equations or using numerical simulations. However, the necessary data of these methods are always difficult to accurately obtain in practice. Moreover, some empirical assumptions beneath the physical methods are too idealistic to react to the real working conditions of coal mines. Although the physical methods can, to some extent, reveal the general rules of gas migration, they lack the ability to make real-time predictions.

Given the widespread deployment of sensor platforms, *data-driven models*(e.g., [11], [12], [13]), which mainly include time-series analysis methods, traditional machine learning methods, and deep learning methods, have gained increasing popularity. As shown in Fig. 1, numerous sensors have been deployed in different locations of coal mines to continuously and cooperatively collect amounts of data. Apparently, the collected data has spatial and temporal characteristics. Thus, the data-driven approaches need to solve two challenging tasks.

(1) Capture the complex spatio-temporal correlations of methane concentration data. Regarding this, many researchers have already made great efforts. First, time series analysis models are employed for gas concentration prediction problems. Yet, they are usually used to predict a single time series, ignoring the potential correlation between multiple time series in relatively complex mining conditions. Second, traditional machine learning methods, such as random forest and support vector machine are developed to model gas concentration data, but it is difficult for them to handle the spatio-temporal data with shallow learning in practice. Third, deep neural networks (DNNs) have been widely applied in many fields in recent years and redefine state-of-the-art performance [14]. Deep learning methods have powerful feature representation and function approximation capabilities by stacking nonlinear layers [15], which makes deep learning methods the most popular solution for current gas concentration prediction. For example, recurrent neural networks (RNN) and their variants, such as Long Short-term Memory (LSTM), are widely adopted to extract the temporal features of data. However, RNN-based models only exploit the temporal features in single sensor data and ignore the underlying spatial dependencies in sensor networks. In other areas, some studies commonly employ a convolutional neural network (CNN) to capture the spatial dependencies [16], [17]. However, CNN-based models are more suitable for large-scale datasets rather than several sensors on the same coal face. As a result, many studies simply use a full connection (FC) layer to learn spatial relationships.

(2) Reduce the training cost and the complexity of models. When predicting the gas concentration of a certain sensor, all data of the other sensors are extracted and fed into the FC layer as spatial topological features. This method, while easy to implement, uses a lot of invalid features for forecasting because of the ignorance of topological structure in sensor networks, such as adjacency. Hence, it neither achieves accurate prediction nor reduces the efficiency of the model. In addition, existing methods predict the gas concentration only for a single target sensor; thus, we have to train a different model for each sensor, which lead to a huge waste of resources.

To tackle the above challenges, we propose a novel deep learning model, consisting of a graph convolutional network module (GCN) and an encoder-decoder module (ED), to collectively predict methane concentration on the coal face.

In contrast to CNN and FC, the GCN module provides a more efficient and feasible way to capture spatial correlation. It can naturally utilize the inherent graph structure of sensor networks to model the dependencies between sensors and achieve the final task through the transfer, aggregation and update of node information [18]. In the GCN module, we directly adapt the adjacency within a sensor network to build graphs based on prior knowledge. However, this graph structure based on prior knowledge does not necessarily represent the real dependencies between sensors. So, in our model, a learnable self-adaptive dependency matrix is simultaneously used to capture the hidden spatial dependency among sensors. The ED module is used to learn complex temporal features with LSTM cells and generate multi-steps predictions for gas concentrations. This model can predict the methane concentration directly on the original sensor data and effectively capture the complex spatio-temporal patterns. The main contributions of this paper are summarized as follows:

- We present an effective and efficient model composed of the GCN and ED modules, which considers the spatial and temporal dependencies, respectively. The proposed model can simultaneously predict the gas concentration for all the sensors on a coal face instead of a certain sensor.
- A self-adaptive adjacency matrix is designed and applied into the GCN module to improve the ability of capturing real spatial correlations.
- The ED module can capture long-term temporal dependency, including temporal closeness and daily period, and can generate a multi-step prediction.
- We evaluate the proposed model on real-world coal mine data collected from coalface sensors and achieve state-of-the-art results with low computation costs.

II. RELATED WORK

A. METHANE CONCENTRATION FORECASTING IN COAL MINES

After years of continuous research and practice, many achievements have been made in the field of methane concentration forecasting. In general, the methane concentration forecasting methods fall into five major categories: analytical methods, numerical simulations methods, time-series analysis methods, traditional machine learning methods, and deep learning methods.

1) ANALYTICAL METHODS

Through the continuous study of gas concentration laws, the constitutive equations for the gas concentration from different gas sources are established [8]. These equations provide meaningful insights and aid the gas extraction and ventilation design [19]. However, the methane concentration is influenced by geological conditions, the gas content of coal and rock, the depth of coal, the way of mining, and so on [20]. There are dynamic nonlinear relationships among these factors [21]. However, in practice these factors are difficult to be accurately measured and obtained in the mine, so it is difficult for analytical methods to predict gas concentration in advance and give early warning.

2) NUMERICAL SIMULATIONS METHODS

Numerical simulations have been a key player in recent progress in coal mine methane forecasting [22]. The flow of methane within coal satisfies the gas diffusion law. When methane is produced in the coal mining process, it together with the air forms a mixed gas and diffuses in the roadway network along with the wind [23]. To reveal the movement law of gas specifically and directly in the coal face, a lot of numerical simulation models have been proposed [9], [10]. However, the numerical simulation of the gas migration is based on some ideal assumptions. For example, the state of mining equipment, the movement of workers, and the change in environmental factors are not easy to simulate in reality [24]. So, it is difficult to apply numerical simulation models in forecasting methane concentration in practice.

3) TIME-SERIES ANALYSIS METHODS

Compared with the above methods, data-driven methods provide more flexible alternatives for gas concentration forecasting. So, many researchers have turned their attention to gas time series prediction. Many methods have been proposed in the gas forecasting field, including chaos time series [11], [25], [26], gray theory [27], [28], fuzzy mathematics [29], [30], [31]. Although these methods have yielded good results in some situations, there is still a desire for improvement regarding the effectiveness and efficiency of the methods. For example, ARIMA and its variants are commonly adopted for linear predictions of stationary single time series. However, owing to the non-stationary fluctuations and chaotic properties of the gas concentration time series, the models only utilize the original raw data from a single sensor and often cannot provide satisfying forecast results.

4) TRADITIONAL MACHINE LEARNING METHODS

Traditional machine learning methods are important approaches for gas concentration forecasting. Many effective methods have been used, e.g., traditional neural networks [32], [33], support vector machine [12], Gaussian process regression [34]. To incorporate useful features into the forecasting model, some feature extraction methods have been proposed, such as WELM (wavelet-based ELM) [35], EMD-MFOA-ELM [36] and CAPSO-ENN [37]. However, these methods have one common limitation: they cannot process spatio-temporal data efficiently and often rely on prior knowledge.

5) DEEP LEARNING METHODS

Recently, deep learning methods have gained intensive growth in capabilities and wide usages in many domains, such as marketing, finance, e-commerce, and the health field. Owing to its powerful feature extraction ability, deep

learning models hardly rely on the manual work of feature engineering, which offers new ideas for methane concentration forecasting [3], [38]. As a popular model in sequential learning tasks, the RNN network has been used for methane concentration forecasting. However, most practices still utilize the data from a single sensor, lacking the fusion of data from multiple sensors. To fusion multi-sensor information and achieve multi-step prediction, [13] propose to use an LSTM-based encoder-decoder model for short-term predictions, and obtain good experimental results. However, the model may ignore the hidden and dynamic spatial dependencies from the data. Encouraged by this, [39] proposed a gas concentration prediction method based on attention mechanism. However, most of the existing methods are adopted to predict the gas concentration of a single target sensor by taking all measured data as inputs. As a result, there is still an improvement in the accuracy and efficiency.

B. GRAPH CONVOLUTION NETWORKS

The application of traditional convolution on standard grid data such as images has achieved great success due to the automatic extraction and learning of local patterns of the data. In order to generalize this efficient feature learning method to data with graph structure, graph convolution theory and methods have developed rapidly in recent years. Existing graph convolution approaches mainly fall into two categories: the spatial methods and the spectral methods.

The key idea of the spatial method is to aggregate information on the nodes of the graph and its neighboring nodes. Therefore, the neighborhood selection of nodes becomes the main challenge of such methods. Reference [40] proposed a neighborhood heuristic linear method for determining each central node in a social network. Reference [41] introduced graph convolutions into human action recognition tasks. Several partitioning strategies were proposed here to divide the neighborhood of each node into different subsets and to ensure the numbers of each node's subsets are equal.

Spectral methods learn local patterns of graph-structured data through spectral analysis. Reference [42] proposed a graph convolution framework based on graph Laplacian, which was further optimized by using Chebyshev polynomial approximation to achieve eigenvalue decomposition [43]. In these methods, the adjacency matrix is considered as prior knowledge, which is fixed throughout the training process. To address this problem, [44] proposed a model that learns the weights of node neighbors through a Gaussian kernel. Furthermore, a method to update the weights of node neighbors through an attention mechanism was proposed by [45]. Although these methods set each neighbor's contribution to the central node as a learnable parameter, they are still based on a predefined Laplacian matrix. In addition, GCN-based hybrid models are increasingly popular by combining temporal features. For instance, [46] solve the temporal-spatio prediction of traffic flow by proposing a diffusion convolutional recurrent neural network (DCRNN).

Motivated by the aforementioned studies, we employ the graph convolutions with an adaptive dependency matrix and a LSTM-based ED network to model the complex spatio-temporal patterns of the methane concentration data collected from the coal mine sensor network.

III. METHODOLOGY

In this section, we discuss the motivation and detailed steps for the construction of the proposed model. At the very beginning, we briefly introduced the related symbols in this paper and the mathematical definition of the methane concentration forecasting task. Then, we elaborate on the two constituent modules of the framework, the graph convolution module (GCN) for spatial dependency capture and the encoder-decoder module (ED) for temporal feature learning. Finally, we describe the overall framework of the graph convolutional encoder-decoder model (GCN-ED) model, as shown in Fig. 2

A. PROBLEM DEFINITION

Considering the spatial correlation of sensors and the temporal characteristics of gas diffusion, gas concentration prediction is a typical spatio-temporal forecasting problem. We intend to predict the methane concentration in the future by leveraging historical methane concentration data and topological structure of sensor network. The sensor network can be denoted as $G = (V, E)$, where $V = \{v_1, \cdots, v_N\}$ is a set of N methane sensor nodes and E is a set of edges connecting these nodes. The graph can be represented by a symmetric 0-1 adjacency matrix $A \in R^{N\times N}$, where $A[i, j] = 1$ denotes the existence of a link between nodes v_i and v_j, and 0 otherwise. The existence of a link is determined by the geographical proximity between the two sensors.

We regard the methane concentration as the attributes of the nodes in the network, denoted as $X_t \in R^N$ (t=1, ...T), where T represents the length of the historical time series and $x_t^i \in R$ denotes the value of the node i at time t. Assume that the current time is t_0 and the size of prediction interval is τ. Given a graph G and its historical T step graph features, our problem is to learn a function f that can forecast its next τ step graph features. The mapping relation is represented as follows:

$$\left[X_{t_0+1}, \ldots, X_{t_0+\tau}\right] = f\left(G; \left(X_{t_0-T+1}, \ldots, X_{t_0-1}, X_{t_0}\right)\right) \tag{1}$$

B. HISTORICAL TIME SERIES SAMPLING

The methane concentration in a location is a typical time series, which is affected by historical data, both near and far. To make better use of historical information and capture high-levels of temporal features for methane concentration data, we categorize the historical time series data into two parts: the recent samples and daily-periodic samples. Suppose the sampling frequency is q times per day. Assume that the current time is t_0 and the size of predicting window is τ. As shown in Fig. 3, we intercept two time series segments of

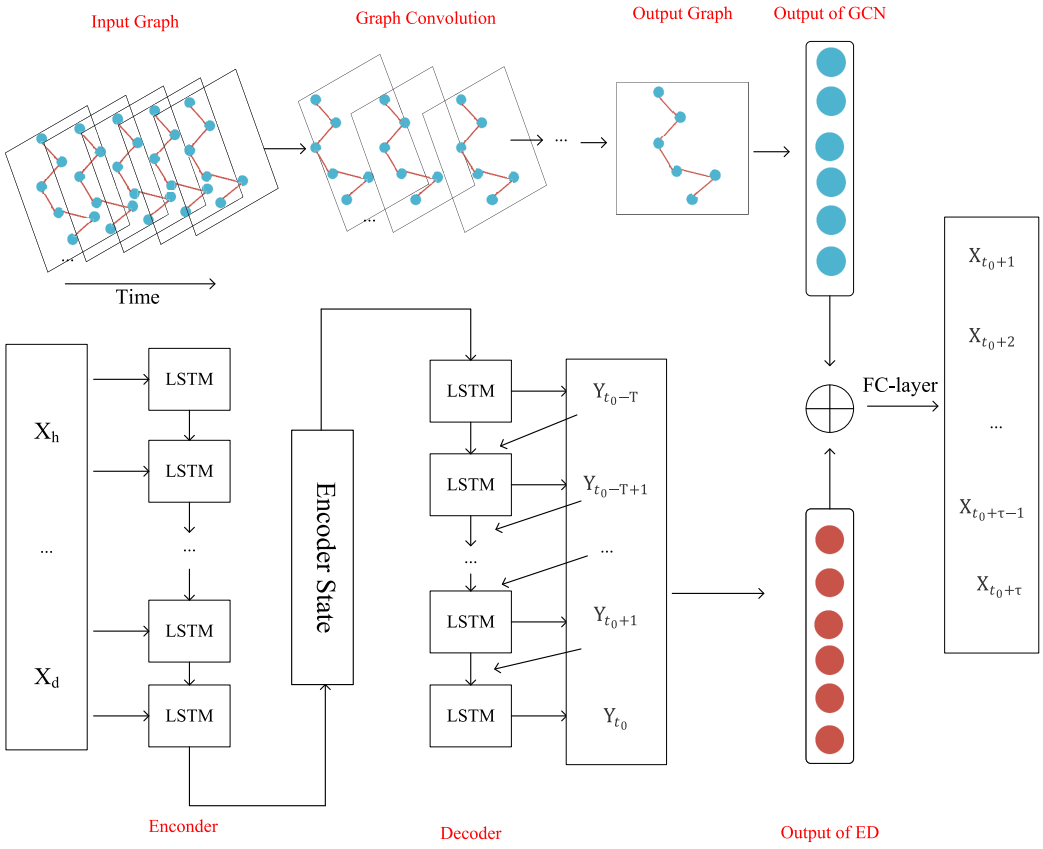

FIGURE 2. The framework of the GCN-ED model.

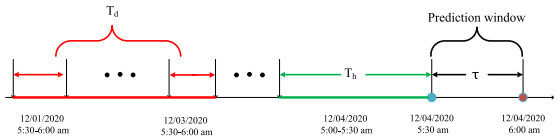

FIGURE 3. An example of historical time series sampling (Suppose that the prediction interval is 30 minutes.).

length T_h and T_d along the time axis as the input of the recent and daily-period component respectively, where T_h and T_d are all integer multiples of τ. Details about the two time series segments are as follows:

(i) The recent segment: $X_h = \left(X_{t-T_h+1}, X_{t-T_h+2}, \cdots, X_t\right) \in R^{N \times T_h}$, T_h is a segment adjacent to the predicting period, as shown by the green part of Fig. 3. Intuitively, the adjacent period has a greater impact on predictions than the far periods.

(ii) The daily-periodic segment: $X_d = (X_{t-(T_d/\tau)*q+1}, X_{t-(T_d/\tau)*q+2}, \ldots, X_{t-(T_d/\tau)*q+\tau}, X_{t-(T_d/\tau-1)*q+1}, X_{t-(T_d/\tau-1)*q+2}, \ldots, X_{t-(T_d/\tau-1)*q+\tau}, \ldots, X_{t-q+1}, X_{t-q+2}, \ldots, X_{t-q+\tau}) \in R^{N \times T_d}$ consists of the segments on the past few days at the same period as the predicting period, as shown by the red part of Fig. 3. Due to the regular daily routine of coal miners, such as the daily lunch break, methane concentrations may show similar patterns from day to day. Hence, we use the daily-period segment to capture the daily periodicity of methane concentration, which does not occur in most existing methods.

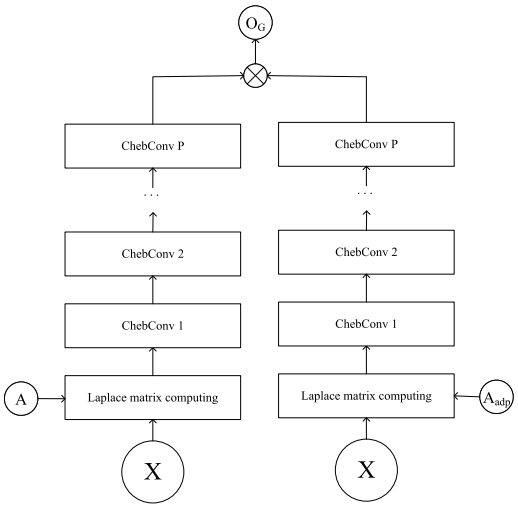

FIGURE 4. The architecture of graph convolution module.

C. GRAPH CONVOLUTION MODULE

In our model, a novel graph convolution module is proposed to model the complex spatial dependencies of sensor

networks. In this section, we introduced the structure and operation of graph convolution module. Due to using different graph matrices to generate the Laplace matrix, graph convolution module is composed of two parts, as shown in Fig. 4. In the left part, we use an adjacency matrix based on prior knowledge to build the initial graph structure, while in the right part, a self-adaptive adjacency matrix is used to capture hidden spatial relations. After obtaining the Laplace matrix, both parts learn the final spatial features through the stacked Chebyshev convolution layers.

1) CHEBYSHEV CONVOLUTION LAYER

In this study, we regard the methane sensor network as a graph, and the features of each node can be regarded as the signals on the graph [47]. The spectral method transforms a graph into an algebraic form to analyze the topological attributes of a graph, such as the connectivity in the graph structure. In spectral graph analysis, a graph is represented by its corresponding Laplacian matrix. The properties of the graph structure can be obtained by analyzing the Laplacian matrix and its eigenvalues. The Laplacian matrix of a graph is defined as $L = D - A$, and its normalized form is $L = I_N - D^{-1/2}AD^{-1/2} \in R^{N\times N}$, where A is the adjacent matrix, I_N is a unit matrix, and the degree matrix $D \in R^{N\times N}$ is a diagonal matrix, consisting of node degrees, $D_{ii} = \sum_j A_{ij}$. The eigenvalue decomposition of the Laplacian matrix is $L = U\Lambda U^T$, where $\Lambda = diag([\lambda_0, \cdots, \lambda_{(N-1)}])$ is a diagonal matrix, and U is the Fourier basis. Taking the methane concentration at time t as input, the signal all over the graph is $X_t \in R^N$, and the graph Fourier transform of the signal is defined as $\hat{X}_t = U^T X_t$. According to the properties of the Laplacian matrix, U is an orthogonal matrix, so the corresponding inverse Fourier transform is $X_t = U\hat{X}_t$. The graph convolution is a convolution operation implemented by using linear operators that diagonalize in the Fourier domain to replace the traditional convolution operator [48]. Based on this, the signal X_t on the graph G is filtered by a kernel:

$$g_\theta *_G X_t = g_\theta(L)X_t = g_\theta\left(U\Lambda U^T\right)X_t = Ug_\theta(\Lambda)U^T X_t \tag{2}$$

where $*_G$ denotes a graph convolution operation. Since the convolution operation of the graph signal is equal to the product of these signals which have been transformed into the spectral domain by graph Fourier transform, the above formula can be understood as Fourier transforming g_θ and X_t respectively into the spectral domain, then multiplying their transformed results, and doing the inverse Fourier transform to get the final result of the convolution operation. However, due to the computational complexity, when the eigenvalue decomposition of the Laplacian matrix of a large-scale graph is directly performed, the computational resource requirements are very demanding. Therefore, Chebyshev polynomials are adopted in this paper to achieve the approximate result efficiently [49]:

$$g_\theta *_G X_t = g_\theta(\mathbf{L})X_t = \sum_{k=0}^{K-1}\theta_k T_k(\tilde{L})X_t \tag{3}$$

where the parameter $\theta \in R^K$ is a vector of polynomial coefficients. $\tilde{L} = \frac{2}{\lambda_{max}}L - I_N$, λ_{max} is the maximum eigenvalue of the Laplacian matrix. The recursive definition of the Chebyshev polynomial is $T_k(X_t) = 2X_t T_{k-1}(X_t) - T_{k-2}(X_t)$, where $T_0(X_t) = 1$, $T_1(X_t) = X_t$. Solving this formula using an approximate expansion of the Chebyshev polynomial is equivalent to aggregating 0 to $(K-1)^{th}$ order neighbor information for each node in the graph through a convolution kernel g_θ. The graph convolution module uses the Rectified Linear Unit (ReLU) as the final activation function, i.e., $ReLU(g_\theta *_G X_t)$.

2) SELF-ADAPTIVE ADJACENCY MATRIX

To dynamically adjust the correlations between nodes and precisely capture the hidden spatial dependency in the data, we propose a self-adaptive adjacency matrix A_{adp}. A_{adp} is a learnable matrix that trains along with the model. We achieve this by randomly initializing two transformation matrices with learnable parameters $M_1, M_2 \in R^{N\times N}$. Specifically, the self-adaptive adjacency matrix is defined as follows:

$$A_{adp} = M_1 E M_2 \tag{4}$$

where E is a matrix of all ones. We use the E to represent all possible connections in the graph, and derive the actual spatial dependency weights through the matrix transformation. In the same way, we can get $\tilde{L}_{adp}$ by taking A_{adp} as the matrix of the graph and define the graph convolution based on the $\tilde{L}_{adp}$ as:

$$g_\theta *_G X_t = \sum_{k=0}^{K}\theta'_k T_k(\tilde{L}_{adp})X_t \tag{5}$$

In our GCN-ED model (see Figure. 4.), we stack P Chebyshev convolution layers as follows,

$$z_{p+1} = \mathcal{G}(z_p; \theta_p) \tag{6}$$

where z_{p+1} and z_p are the input and output of the p^{th} Chebyshev convolution layer, respectively; $\mathcal{G}$ is the graph convolution function (i.e., two combinations of "ReLU + Chebyshev convolution"), and θ_p includes all learnable parameters in the p^{th} Chebyshev convolution layer. And, the impacting weights of the two matrixes for each node are different at each time slice, and they should be learned from the data. By combining spatial dependencies of predefined graph structures and hidden spatial dependencies learned from data, the final result of the GCN module after the fusion is:

$$O_G = W_a \odot z_P^a + W_{adp} \odot z_P^{adp} \tag{7}$$

where $\odot$ is the Hadamard product. W_a and W_{adp} are learning parameters, reflecting the influence degrees of the two matrixes on the forecasting target. z_P^a and z_P^{adp} are results after P-layer graph convolution.

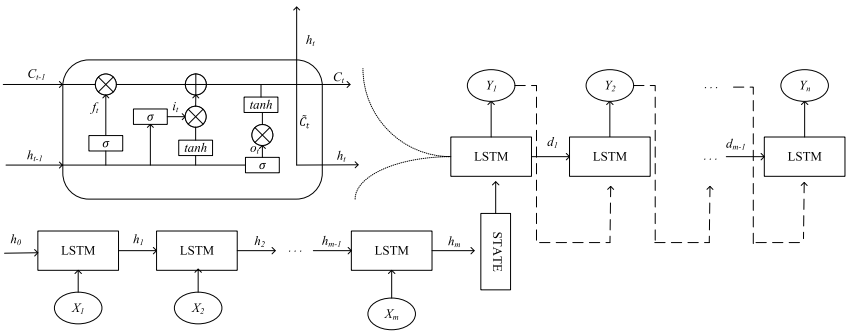

FIGURE 5. The architecture of an LSTM-based encoder-decoder network and an LSTM memory cell.

D. LSTM-BASED ENCODER-DECODER MODULE

The data generated by methane sensors are typically time-series data with periodicity and proximity characteristics. There exist complex correlations between the gas concentration in different time slices, and the correlations are also varying under different situations. To capture the temporal dependency, we adopt the LSTM-based encoder-decoder model to process the sequence information. LSTM is an improved model of RNN to solve the gradient vanishing problem. The major innovation of LSTM is its memory cell c, which essentially is used to store temporal states. As shown in Fig. 5, each memory cell is accessed, updated, and cleared by forget gate f, input gate i, and output gate o to control the flow of sequential information. As a result, LSTM can effectively extract and transfer important historical data features, which is crucial in modeling time dependence. The LSTM maps the input sequence to an output sequence by calculating various unit activations using the following equations:

$$\begin{aligned} f_t &= \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \\ i_t &= \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \\ \tilde{c}_t &= \tanh\left(W_c\left[h_{t-1}, x_t\right] + b_c\right) \\ c_t &= f_t * c_{t-1} + i_t * \tilde{c}_t \\ o_t &= \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \\ h_t &= o_t * \tanh\left(c_t\right) \end{aligned} \tag{8}$$

where the matrix W (e.g., W_f) and the bias vector b (e.g., b_f) are learnable parameters and σ represents the standard sigmoid function. The h_t denotes the hidden state at time t and x_t is the input at time t, and $\tilde{c}_t$ is the temporary cell state.

The encoder-decoder is a common framework in the deep learning model and is firstly used for machine translation. It has been proven to be stable and powerful for extracting long-range dependencies of historical data in various previous studies. The LSTM-based encoder-decoder model takes in a sequence of data and outputs a new sequence of data, which is a way to solve the sequence-to-sequence problem [50]. As shown in Fig. 5, the model consists of two parts - an encoder and a decoder. The encoder is mainly used to learn the important information from the historical time series data. We feed the X_t as the input to the encoder and update the hidden state at time t by using $h_t = f_e(h_{t-1}, X_t)$, where f_e is an LSTM unit. In the decoder, once we get the hidden state $d_{\tilde{t}-1}$ at a future time step $\tilde{t}$, we combine it with the last output of decoder $Y_{\tilde{t}-1}$ to update the decoder hidden state with $d_{\tilde{t}} = f_d(d_{\tilde{t}-1}, Y_{\tilde{t}-1})$, where f_d is an LSTM unit used in the decoder. Finally, we use a linear transformation to generate the final output of LSTM based encoder-decoder module as follows:

$$O_T = W_y Y + b_y \tag{9}$$

where the matrix W_y and the vector b_y are learning parameters.

E. COMBINATION OF THE GCN AND ED MODULES

Most of the existing studies commonly model the spatial dependencies first and then the temporal dependencies. For example, in [51], the methane concentration data are sequentially fed into the FC layer and the LSTM network. However, the original temporal patterns may be distorted after complex spatial convolution so that the outputs generated by the neural network layers no longer fully represent the patterns of the raw methane concentration data. Therefore, in order to better capture the spatio-temporal dependencies in the data separately, a novel framework is proposed in this paper. The framework maximizes the effectiveness of the spatial and temporal modules by running the GCN and ED modules in parallel and then combining their respective results. As shown in Fig. 2, the methane concentration data is first organized into two forms, sensor graphs G and time series X; then these two forms of data are used as the inputs of the GCN and ED modules to capture the spatial and temporal dependencies, respectively, and their outputs are represented by OG and OT, respectively; Finally, the results of the two modules are concatenated and a fully connected layer is used to generate predictions for future time, which can be expressed by the following formula:

$$\hat{X} = W_{st}(O_G \oplus O_T) + b_{st} \tag{10}$$

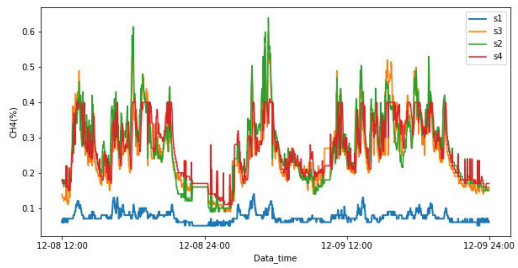

FIGURE 6. Plots of the methane concentration values of SXMine.

TABLE 2. Statistics of the two datasets used in experiments.

Dataset	Sensors	Sampling interval	Training set	Test set
SXMine	6	1 min	17280	4320
CQMine	3	2 min	4320	720

where W_{st} and b_{st} are the model parameter used to obtain the prediction results $\hat{X}$. $\oplus$ is used to represent the concatenation operation.

Our GCN-ED model is trained by minimizing the loss function, which is defined as the mean squared error (MSE) between the predicted value and the ground truth data.

$$L(\theta) = \frac{1}{NT} \sum_{i=1}^{i=N} \sum_{t=1}^{j=T} \left| \hat{x}_t^i - x_t^i \right|^2 \tag{11}$$

where θ are all learnable parameters in the GCN-ED model.

Unlike previous works such as [51], our GCN-ED model outputs $\hat{X} = [X_{t+1}, \ldots, X_{t+\tau}]$ as a whole rather than generating $\hat{X}_t$ recursively through T steps, which solves the problem of multi-step prediction for methane concertation. More importantly, our GCN-ED model can train and inferences all sensors on the same coal face as a whole. Compared with existing approaches that target a single sensor, the GCN-ED model undoubtedly greatly improves the efficiency in practical applications.

IV. EXPERIMENTS

A. DATA DESCRIPTION

A real sensor dataset called SXMine from a coal mine monitoring system in Shanxi, China, was used to validate the effectiveness of the proposed GCN-ED model. To accurately monitor the gas concentration of coal face, the administrative department of coal mine safety has specified the number and locations of gas sensors. This dataset contains six methane sensors installed at different locations on the same coal face, as shown in Fig. 1. The sampling interval of the sensors is 1min, so each sensor can generate 1,440 samples per day. The missing values are filled by linear interpolation. For effective training and testing, we used the methane concentration data of the first 12 days to construct the training set and the remaining data to construct the test set. The specific statistical information of the dataset can be referred to in Table 2. The time series values of gas concentration at some measuring points are shown in Fig. 6.

B. EXPERIMENT DESIGN

Different from the previous point prediction task, the multi-time step prediction task on SXMine is a typically sequence-to-sequence problem. As a result, classic coal mine methane prediction methods, including ARIMA [52], LSTM [53], and FC-LSTM [51] are not suitable for comparison. LSTM-based encoder-decoder [13] is the state-of-the-art model for predictions of multi-time step gas concentration. To better illustrate the effectiveness of the GCN-ED model, multi-time step prediction tasks were conducted by taking an LSTM-based encoder-decoder as the control method. Specifically, we design two tasks on SXMine to verify the performance of the GCN-ED model:

- Task 1 (30 to 15) is to generate the next 15 sample prediction results given 30 historical samples, which include 15 periodic samples and 15 recent samples.
- Task 2 (60 to 30) to obtain the next 30 sample prediction results using 30 periodic samples and 30 recent samples.

To verify the impact of the ED module, we also design a degraded version of the GCN-ED model, named GCN, which gets rid of the LSTM-based ED module. The settings of GCN are the same as those of GCN-ED, except for removing the ED module. In addition, to evaluate the impact of different graph structures, we build a simplified version of the GCN-ED model that uses only the adjacency matrix without the self-adaptive matrix, denoted by GCN-ED(a priori), and a model that uses only the self-adaptive adjacency matrix without the a priori adjacency matrix denoted by GCN-ED(adaptive).

C. EXPERIMENTAL SETUP

We compiled and tested the model on a desktop with an AMD processor Ryzen 7 4800HS and an NVIDIA RTX 2060 running Windows 10. In addition, we implemented each prediction model based on the Pytorch framework. During the training phase, the batch size is 64 and the learning rate is 0.001. We built our model with two GCN layers, which proved to be the most effective in many studies. According to [49], we test the number of the terms of Chebyshev polynomial $K \in \{1, 2, 3\}$. And we directly build the prior adjacency matrix according to whether the sensors are adjacent in the coal face. Additionally, we use three commonly used metrics to evaluate the performance of either model, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE), which are expressed as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i}^{n} \left| \hat{X}_i - X_i \right|$$

$$\text{MAPE} = \frac{1}{n} \sum_{i}^{n} \frac{\left| \hat{X}_i - X_i \right|}{X_i}$$

TABLE 3. Experimental results of different models on task 1.

Model	MAE	RMSE	MAPE	TIME(s/epoch)
ED	0.0435	0.0626	0.0812	$4*$ N
FC-ED	0.0395	0.0568	0.0750	$6*$ N
GCN	0.0369	0.0551	0.0824	**2**
GCN-ED(a priori)	0.0343	0.0550	0.0725	5
GCN-ED(adaptive)	0.0361	0.0542	0.0781	5
GCN-ED	**0.0329**	**0.0513**	**0.0698**	5

TABLE 4. Experimental results of different models on task 2.

Model	MAE	RMSE	MAPE	TIME(s/epoch)
ED	0.0769	0.1077	0.1102	$7*$ N
FC-ED	0.0564	0.0771	0.1041	$9*$ N
GCN	0.0633	0.0919	0.1137	3
GCN-ED(a priori)	0.0526	0.0716	0.1105	8
GCN-ED(adaptive)	0.0476	0.0725	0.1084	8
GCN-ED	**0.0469**	**0.0660**	**0.1057**	8

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i}^{n}\left(\hat{X}_i - X_i\right)^2} \tag{12}$$

As an early warning model, the proposed model should be sensitive to outliers in methane concentration data. Since MAPE and MAE are relatively more robust to abnormal value, we use RMSE, as the main evaluation metric for our experimental comparison.

D. COMPARISON AND RESULT ANALYSIS

Tables 3 and 4 show the performance of different models on Tasks 1 and 2, respectively. It can be seen that our GCN-ED model obtains superior results on both tasks in terms of the evaluation metrics, MAE, RMSE, and MAPE. Among all tested models, ED obtains the worst accuracies with RMSE values of 0.0626 and 0.1077 for Tasks 1 and 2, which suggests the limited effectiveness of learning only temporal features for methane concentration prediction. By using both FC to capture spatial and LSTM to capture temporal dependencies, FC-ED can achieve better performance than the ED model, with RMSE values reduced to 0.0568 and 0.07711 in Task 1 and Task 2, respectively. However, the model that used a FC layer to capture the spatial dependencies is easy to ignore the relationship from the realistic topology and adopts a lot of invalid features for forecasting. Hence, it reduces not only the accuracy but also the efficiency of the model. The GCN model significantly improves the accuracy with an RMSE value lower than 0.0551 in Task 1. But in the long-term prediction task 2, it performs significantly worse than the FC-ED model. This indicates that the graph convolution operation can achieve information fusion in a short term and the RNN-based models have a stronger ability to capture long-term time dependence. Because the proposed GCN-ED model combines the GCN module and ED module to capture complex spatio-temporal dependencies, it achieves the lowest RMSE values of 0.0513 for Task 1 and 0.0660 for Task 2.

The results of the ablation experiments further emphasize the effectiveness of each adjacency matrix of graph in GCN-ED: (1) compared with the results of the GCN-ED, the GCN-ED(a priori) achieved worse performance, i.e., RMSE of 0.0550 and 0.0716, which illustrates the importance of incorporating adaptive matrices into the GCN module; (2) the accuracy of the GCN-ED (adaptive) further decreased, whose RMSE is 0.0542 and 0.0725, indicating the positive effect of a priori adjacency matrix in GCN module. Although the prediction accuracy of these ablation models decreased to varying degrees. However, compared with other test models, it still shows obvious superiority, which verifies the validity of the overall structure of the model.

In order to test the effect of different K values on model performance, we set K ranging from 1 to 3 in our experiments, and the RMSE values of the corresponding models are shown in Fig. 7. It can be seen that RMSE reach its minimum at $K = 2$ for task 1 and 2, and then increases as K becomes larger. Larger K means that more spatial dependence information is used in the prediction model. The reasonable dependence information is helpful for improving prediction accuracy; however, overmuch information may result in overfitting. So, considering the forecasting performance of the model, we set $K = 2$ in the experiment.

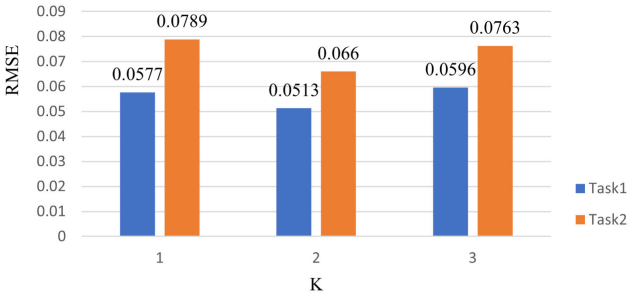

FIGURE 7. Results of GCN-ED with different values of K.

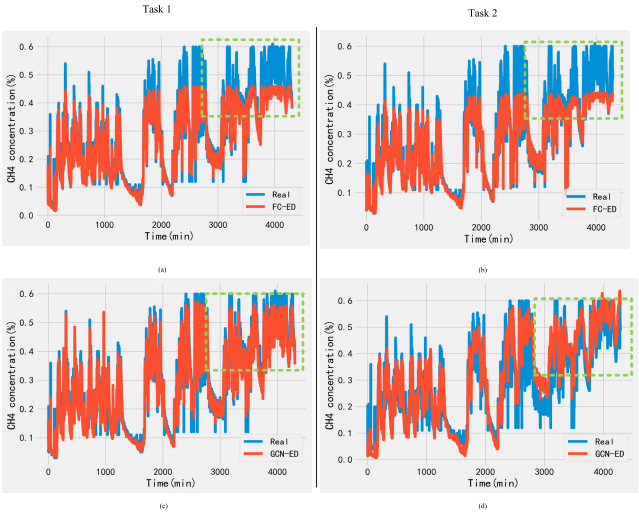

FIGURE 8. Comparison of ground-truth and prediction results.

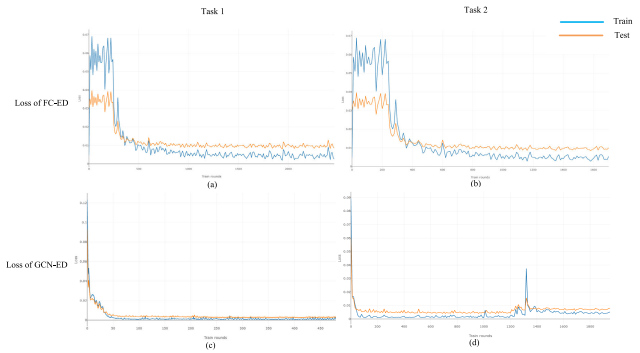

FIGURE 9. Loss of FC-ED and GCN-ED model.

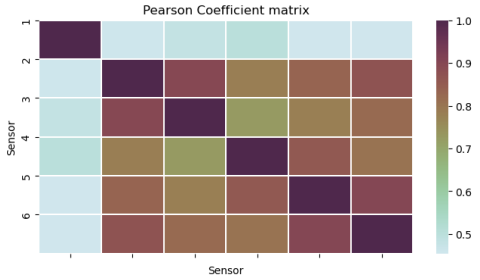

FIGURE 10. Visualization of the Pearson correlation coefficient matrix.

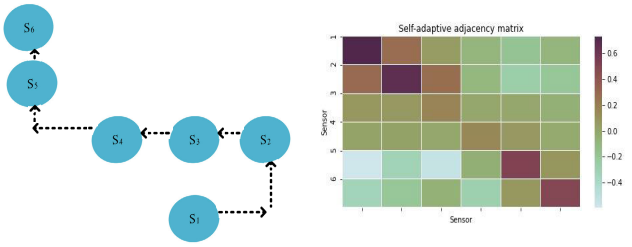

FIGURE 11. Visualization of self-adaptive adjacency matrix.

1) VISUALIZATION OF PREDICTION RESULTS

We plot predicted values vs. real values of the GCN-ED model and FC-ED model on a snapshot of the test data in Fig. 8. It shows that the GCN-ED model generates more accurate predictions than the FC-ED model. In particular, when methane concentrations fluctuate wildly at high levels, our model is closer to the real value. Since gas accidents are usually caused by the increase and accumulation of gas concentration, the prediction models should be sensitive to the fluctuation of data in the danger range. In summary, the proposed GCN-ED achieves better predictions both in the whole data and in the large-value intervals.

2) ANALYSIS OF THE COMPUTATIONAL EFFICIENCY

To apply the model in practice, the efficiency of the model is also crucial. Tables 3 and 4 summarize the training time per epoch for all tested models. We can see that GCN is the most efficient model among all the tested models, and the GCN-ED proposed in this paper is the second most efficient model with requirements of 5s and 8s for task 1 and task 2, respectively. Although the average training time of our model is slightly higher than that of GCN, it is much lower than the other models, which shows the competitiveness of our model in terms of efficiency. In contrast, the FC-ED model with the second-best prediction accuracy on both tasks is the least efficient model. Because it is modeled for a single sensor, the training time of the model is proportional to the number of sensors. Specifically, FC-ED takes (6*N)s and (9*N)s during the training process of each epoch of tasks 1 and 2, respectively, which is more than N times longer than the proposed GCN-ED. The reason of this huge difference in efficiency is that the FC-ED is modeled for a single sensor, whereas our model is built for all sensors across the entire working face. Therefore, our model can realize the simultaneous training and inference of all sensors, which saves a lot of manpower and computing resources. Fig. 9 outlines the training and validation accuracy and loss curves respectively. Compared with task 1, the convergence speed of task 2 is relatively slow, which indicates that longer sequence data needs longer training. Due to the superiority of the model, compared with the FC-ED model, the proposed GCN-ED model not only converges faster but also achieves better results after convergence.

3) ANALYSIS OF MODEL INTERPRETABILITY

In order to provide better explanations for the model, we analyze model performance for three different adjacency matrix configurations. The last three rows Table 3 and Table 4 show the average score of MAE, RMSE, and MAPE on task1 and task 2. We find that the performance of the GCN-ED(adaptive) model is similar to that of GCN-ED(a priori) on RMSE. When the graph structure is uncertain, GCN-ED (adaptive) can still obtain good prediction accuracy. And the GCN-ED model significantly outperforms other models on all evaluation metrics. It indicates that the combination of the graph structural information and the self-adaptive adjacency matrix could introduce new and useful information to the model.

As shown in Fig. 6, there is a clear similarity in the trend of gas concentration values among different monitoring points within the same workface. The Pearson correlation coefficient is a widely used measure for quantifying the degree of correlation between two sequences. Therefore, in this paper, the correlation coefficient is used to describe this dependency effect. Its calculation is given by Equation (13).

$$\mathrm{COR}(X, Y) = \frac{\sum_1^n \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_1^n \left(X_i - \bar{X}\right)^2 \sum_1^n \left(Y_i - \bar{Y}\right)^2}} \tag{13}$$

In this equation, X_i represents the gas concentration at point X at time i, Y_i represents the gas concentration at point Y at time i, $\bar{X}$ represents the mean gas concentration at point X, $\bar{Y}$ represents the mean gas concentration at point Y. The visualization of the correlation coefficients between the gas

concentration data of all sensors and the gas concentration data of the target sensor is shown in Fig. 10.

Similarly, to emphasize that our model can learn the dependency patterns in the data, we performed an experiment: picking out the coal face with 6 methane sensors from the SXMine and showing learned self-adaptive adjacency matrix under the configuration of the GCN-ED model trained on task 2. The self-adaptive adjacency matrix is shown on the right side of Fig. 11, where the i-th row represents the impact strength from all sensors to the i-th sensor. By comparing two heatmaps, we can clearly observe that our model can learn the correlations among the gas concentration data from different sensors. Furthermore, based on the two components of Fig. 11 and the actual mining environment, we have provided explanations for the data correlations. Some columns have more high-value points than others, such as columns 1 and 2. It illustrates that some nodes have more influence than others in the sensor network. This may be because these influential sensors are located at the upstream points of wind flow, and therefore control the trend of gas concentration change. So, the self-adaptive adjacency matrix can help capture the direction of methane diffusion in the prediction tasks. In addition, distance is also an important factor affecting spatial dependence. Combined with the left side of Fig. 11, we can know adjacent sensors has a great influence on the target sensor. In particular, s_5 and s_6, because they are in a separate tunnel, have a significantly stronger influence between them. Through the above analysis, we can find that the hidden dependencies learned by the matrices demonstrate an interpretability advantage of the GCN-ED model.

4) ANALYSIS OF MODEL GENERALIZATION

To validate the effectiveness of our proposed model on different datasets, we conducted experiments using real data from another coal mine. The dataset was collected from April 21 to April 27, 2020 based on sensors provided by the coal mine monitoring system in Chongqing, China, hence, we call it CQMine. The sensors have a sampling interval of two minutes, resulting in the generation of 720 data points per day. We divided the dataset into a training set and a test set, using the first six days of data as the training set and the data from the last day as the test set. The specific statistical information of the dataset can be referred to in Table 2. The safety management department has specific regulations regarding the number and placement of methane sensors in the coal mine working face. As a result, the quantity and locations of methane sensors may vary among different working faces in the coal mine. However, every working face has an upper corner monitoring point (s_2), a working face monitoring point (s_3) and a return air monitoring point (s_4), as shown in Fig. 1. Therefore, we used the data from these three sensors in the CQMine dataset as our prediction targets to validate the generalization and effectiveness of our model. Due to its superior performance and its similarity to the proposed model, we chose the FC-ED model as the comparative method in our generalization experiments. The detailed experimental results are presented in Table 5.

The experimental results in Table 5 demonstrate that the proposed model outperforms the baseline model across all evaluation metrics. This indicates the robustness and generalization capability of our model, further showcasing its potential application in the production operations of different coal mines.

TABLE 5. Experimental results of different models on CQMine dataset.

Interval	Model	MAE	RMSE	MAPE	TIME(s/epoch)
40 min	FC-ED	0.0344	0.0454	0.0620	$1*\mathrm{N}$
	GCN-ED	**0.0241**	**0.0361**	**0.0397**	**1**
60 min	FC-ED	0.0415	0.0531	0.0636	$2*\mathrm{N}$
	GCN-ED	**0.0256**	**0.0390**	**0.0412**	**3**

5) ANALYSIS OF MODEL UNCERTAINTY

In the field of coal mine methane prediction, prior works have mostly focused on point estimates without quantifying the uncertainty of the predictions. However, in this high-risk domain, being able to generate probabilistic forecasts with confidence intervals is critical to risk assessment and decision making. Therefore, building upon relevant works, this study conducted uncertainty quantification of the proposed model and the baseline model using three different methods. We utilized the Mean Interval Score (MIS) metric provided by [54] to quantify the uncertainties in our paper. It rewards narrower confidence or credible intervals and encourages intervals that include the observations (coverage). Therefore, MIS is defined to estimate the upper and lower confidence bounds for the predictions. For a one-dimensional random variable $Z \sim \mathbb{P}_Z$, if the estimated upper and lower confidence bounds are u and l, where u and l are the $(1-\frac{\rho}{2})$and $\frac{\rho}{2}$ quantiles for the $(1-\rho)$ confidence interval, MIS is defined using samples $z_i \sim \mathbb{P}_Z$ and $\mathbb{1}\{\cdot\}$ is an indicator function:

$$\mathrm{MIS}_N(u,l;\rho) = \frac{1}{N}\sum_{i=1}^{N}\{(u-l) + \frac{2}{\rho}(z_i-u)\,\mathbb{1}\{z_i > u\} + \frac{2}{\rho}(l-z_i)\,\mathbb{1}\{z_i < l\}\}. \quad (14)$$

We adapt FC-ED and GCN-ED to implement various UQ methods with the following settings.

- **Quantile Regression** We can use the one-sided quantile loss function to generate predictions for a fixed confidence level ρ. Given an input x, and the output $f(x)$ of a neural network, parameterized by θ, quantile loss is defined as follows:

$$L_{\text{Quantile}}(y, f(x); \theta, \rho) = \min_{\theta}\left\{\mathbb{E}_{(x,y)\sim\mathcal{D}}[(y-f(x))(\rho - \mathbb{1}\{y < f(x)\}]\right\} \quad (15)$$

  In the experiments, we set the corresponding quantile to (0.025, 0.5, 0.975) and kept the rest of the settings the same as point estimation.

- **MIS Regression** For a fixed confidence level ρ, we can directly minimize MIS to obtain estimates of the confidence intervals. Specifically, we use MIS as a loss function for deep neural networks, we use a multi-headed model to jointly output the upper bound $u(x)$, lower bound $l(x)$, and the prediction $f(x)$ for a give input x, and minimize the neural network parameter θ:

$$L_{\text{MIS}}(y, u(x), l(x), f(x); \theta, \rho) = \min_{\theta} \left\{ \mathbb{E}_{(x,y)\sim\mathcal{D}}[(u(x) - l(x)) + \frac{2}{\rho}(y - u(x)) \times \mathbb{1}\{y > u(x)\} + \frac{2}{\rho}(l(x) - y)\mathbb{1}\{y < l(x)\} + |y - f(x)|] \right\} \tag{16}$$

- **MC Dropout** Monte Carlo (MC) drop-out method serves as a simple alternative to variational Bayes methods which involves performing multiple forward passes with dropout to obtain a distribution of predictions, allowing for the quantification of uncertainty. We implement the algorithm provided by [55] and simplify the model by only considering the model uncertainty. We apply random dropout through the testing process with 5% drop rate and use the average value of the 10 trails as the final comparison result.

Table 6 presents the results of uncertainty quantification for the GCN-ED model and FC-ED model using three different methods. It is evident that the proposed model exhibits better performance in terms of both prediction accuracy and reliability compared to the baseline model. Furthermore, in terms of different uncertainty quantification methods, quantile regression has demonstrated better performance in both MAE and MIS. In summary, by quantifying the uncertainty of the model, we have further revealed the credibility of our model, which can provide reliable support for safety decision-making in coal mines.

TABLE 6. Performance comparison of different models for uncertainty quantification using three evaluation methods.

Interval	Model	MAE	MIS
MIS			
30 min	FC-ED	0.0479	0.3977
	GCN-ED	0.0441	0.3698
60 min	FC-ED	0.0591	0.5651
	GCN-ED	0.0575	0.4158
Quantile			
30 min	FC-ED	0.0395	0.1893
	GCN-ED	0.0342	0.1875
60 min	FC-ED	0.0497	0.2618
	GCN-ED	0.0510	0.2287
MC Dropout			
30 min	FC-ED	0.0488	0.4236
	GCN-ED	0.0420	0.3975
60 min	FC-ED	0.0602	0.4918
	GCN-ED	0.0556	0.4782

V. CONCLUSION AND IN FUTURE

Methane concentration prediction is one of the crucial and challenging tasks for coal mine safety production. Considering the complex spatio-temporal dependence of the sensor data in a coal face, deep learning models show strong superiority, owing to their revolutionary feature representation capability. However, due to the negligence of the real topology of sensor networks or the deficiency in capturing representative spatio-temporal patterns, existing deep learning models utilizing the recurrent neural network can hardly provide accurate prediction. In addition, the current methods are based on a single target sensor model, which results in inefficiency and waste of resources in practical applications. Focusing on improving the accuracy and efficiency of methane concentration prediction, a new methane concentration prediction model GCN-ED is proposed in this paper, which consists of GCN and ED modules with a parallel structure. In the GCN module, a fixed graph structure through the adjacency relation between sensors in reality and a novel adaptive dependency matrix are used to precisely capture the hidden spatial dependency in the data. The ED module can learn complex temporal features with LSTM cells and generate multi-step results of the gas concentration prediction. In addition, the parallel structure of the overall framework maintains the independence between the spatial dependency learning module and the temporal dependency learning module; thus, it can avoid the uncertainty caused by mutual interference. In experiments based on real datasets, the GCN-ED model achieves the best prediction accuracy compared with various state-of-the-art solutions and largely enhances the efficiencies of training processes.

Certainly, our model also has some limitations and constraints. Firstly, our approach is based on a unified model constructed using data from all sensors in the working face. Although this approach can greatly improve the efficiency and effectiveness of the model, it relies on the segmentation of sensors based on the working face using information from the raw data, which is currently not feasible in some coal mines. Additionally, our model is a standard data-driven end-to-end model. Therefore, like typical deep learning models, there is still room for improvement in terms of interpretability in the proposed model. Hense, there are still some meaningful works to be explored in the future. First, we can consider more environmental factors and human factors that may affect the diffusion of gas concentration to improve the prediction accuracy of the model. Second, we will try the application of GCN-ED model on large-scale datasets and explore methods for learning the spatial correlation of dynamic graph-structured data, which allows our model to solve more practical problems.

REFERENCES

[1] S. Mahdevari, K. Shahriar, and A. Esfahanipour, "Human health and safety risks management in underground coal mines using fuzzy TOPSIS," *Sci. Total Environ.*, vols. 488–489, pp. 85–99, Aug. 2014.
[2] C. Ö. Karacan, F. A. Ruiz, M. Cotè, and S. Phipps, "Coal mine methane: A review of capture and utilization practices with benefits to mining safety and to greenhouse gas reduction," *Int. J. Coal Geol.*, vol. 86, nos. 2–3, pp. 121–156, May 2011.
[3] M. Tutak and J. Brodny, "Predicting methane concentration in longwall regions using artificial neural networks," *Int. J. Environ. Res. Public Health*, vol. 16, no. 8, p. 1406, Apr. 2019.

[4] P. Dey, S. K. Chaulya, and S. Kumar, "Hybrid CNN-LSTM and IoT-based coal mine hazards monitoring and prediction system," *Process Saf. Environ. Protection*, vol. 152, pp. 249–263, Aug. 2021.

[5] T. Xia, F. Zhou, X. Wang, Y. Zhang, Y. Li, J. Kang, and J. Liu, "Controlling factors of symbiotic disaster between coal gas and spontaneous combustion in longwall mining gobs," *Fuel*, vol. 182, pp. 886–896, Oct. 2016.

[6] Y. Song, S. Yang, X. Hu, W. Song, N. Sang, J. Cai, and Q. Xu, "Prediction of gas and coal spontaneous combustion coexisting disaster through the chaotic characteristic analysis of gas indexes in goaf gas extraction," *Process Saf. Environ. Protection*, vol. 129, pp. 8–16, Sep. 2019.

[7] W. Yang, B.-Q. Lin, C. Zhai, X.-Z. Li, and S. An, "How in situ stresses and the driving cycle footage affect the gas outburst risk of driving coal mine roadway," *Tunnelling Underground Space Technol.*, vol. 31, pp. 139–148, Sep. 2012.

[8] B. Jensen, S. Gillies, N. Jones, and J. M. Anderson, "Review of methane emission and prediction research in longwall coal mines," in *Proc. Australas. Inst. Min. Metall. (AusIMM)*, vol. 1, 1992, pp. 11–17.

[9] T. Xia, F. Zhou, X. Wang, J. Kang, and Z. Pan, "Safety evaluation of combustion-prone longwall mining gobs induced by gas extraction: A simulation study," *Process Saf. Environ. Protection*, vol. 109, pp. 677–687, Jul. 2017.

[10] J. Cao and W. Li, "Numerical simulation of gas migration into mining-induced fracture network in the goaf," *Int. J. Mining Sci. Technol.*, vol. 27, no. 4, pp. 681–685, Jul. 2017.

[11] X. Ni, "Forecasting of gas emissions based on chaotic time series," *Sci. Technol. Inf.*, vol. 31, p. 3420, Nov. 2008.

[12] M. Y. Qiao, X. P. Ma, J. Y. Lan, and Y. Wang, "Time series short-term gas prediction based on weighted LS-SVM," *Caikuang yu Anquan Gongcheng Xuebao/J. Mining Saf. Eng.*, vol. 28, no. 2, pp. 310–314, 2011.

[13] P. Lyu, N. Chen, S. Mao, and M. Li, "LSTM based encoder–decoder for short-term predictions of gas concentration using multi-sensor fusion," *Process Saf. Environ. Protection*, vol. 137, pp. 93–105, May 2020.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[15] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *J. Manuf. Syst.*, vol. 48, pp. 144–156, Jul. 2018.

[16] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, "DNN-based prediction model for spatio-temporal data," in *Proc. 24th ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, 2016, pp. 1–4.

[17] Z. Yuan, Z. Yang, Y. Ling, C. Wu, and C. Li, "Spatiotemporal attention mechanism-based deep network for critical parameters prediction in chemical process," *Process Saf. Environ. Protection*, vol. 155, pp. 401–414, Nov. 2021.

[18] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-GCN: A temporal graph convolutional network for traffic prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[19] K. Noack, "Control of gas emissions in underground coal mines," *Int. J. Coal Geol.*, vol. 35, nos. 1–4, pp. 57–82, Feb. 1998.

[20] K.-X. Wang, X.-H. Fu, Y.-A. Zhou, Y. He, and H. Wu, "Dynamic development characteristics of amounts of gas and levels of pressure in the Pan-1 coal mine of Huainan," *Mining Sci. Technol.*, vol. 19, no. 6, pp. 740–744, Nov. 2009.

[21] L. Wang, Y.-P. Cheng, L. Wang, P.-K. Guo, and W. Li, "Safety line method for the prediction of deep coal-seam gas pressure and its application in coal mines," *Saf. Sci.*, vol. 50, no. 3, pp. 523–529, Mar. 2012.

[22] J. C. Kurnia, A. P. Sasmito, and A. S. Mujumdar, "CFD simulation of methane dispersion and innovative methane management in underground mining faces," *Appl. Math. Model.*, vol. 38, no. 14, pp. 3467–3484, Jul. 2014.

[23] Q. Yu, W. Kai, and S. Yang, "Study on pattern and control of gas emission at coal face in China," *J. China Univ. Mining Technol.*, vol. 29, no. 1, pp. 9–14, 2000.

[24] J. Wu, S. Yuan, C. Zhang, and X. Zhang, "Numerical estimation of gas release and dispersion in coal mine using ensemble Kalman filter," *J. Loss Prevention Process Industries*, vol. 56, pp. 57–67, Nov. 2018.

[25] J. Cheng, J. Y. Bai, J. S. Qian, and S. Y. Li, "Short-term forecasting method of coalmine gas concentration based on chaotic time series," *Zhongguo Kuangye Daxue Xuebao*, vol. 37, pp. 231–235, Mar. 2008.

[26] J. Diaz, Z. Agioutantis, D. T. Hristopulos, S. Schafrik, and K. Luxbacher, "Time series modeling of methane gas in underground mines," *Mining, Metall. Explor.*, vol. 39, no. 5, pp. 1961–1982, Oct. 2022.

[27] P. Lu, Y. G. Ma, and X. Q. Zhou, "Research and application on dynamic forecasting model of gas consistence in top corner," *J. China Coal Soc.*, vol. 31, no. 4, pp. 461–465, 2006.

[28] J. Pan, Z. P. Meng, and Y. C. Liu, "Grey smoothing model for predicting mine gas emission," *J. China Univ. Mining Technol.*, vol. 13, no. 1, pp. 76–78, 2003.

[29] J. Y. Zhang, J. Cheng, Y. H. Hou, J. Y. Bai, and X. F. Pei, "Forecasting coalmine gas concentration based on adaptive neuro-fuzzy inference system," *J. China Univ. Mining Technol.*, vol. 36, no. 4, pp. 494–498, 2007.

[30] X. Sun, "Coalmine gas prediction based on fuzzy neural network," *J. Anhui Univ. Technol. Natural Sci.*, vol. 3, pp. 229–232, Jul. 2012.

[31] J. Brodny, D. Felka, and M. Tutak, "The use of the neuro-fuzzy model to predict the methane hazard during the underground coal mining production process," *J. Cleaner Prod.*, vol. 368, Sep. 2022, Art. no. 133258.

[32] Z.-J. Qian-Xu, W. Xuehui, and Z. Jinhui, "Short-term prediction of coalmine gas concentration based on chaotic series and wavelet neural network," in *Proc. Int. Conf. Artif. Intell. Comput. Intell.*, vol. 3, Oct. 2010, pp. 240–244.

[33] L. Gao, Y. J. Hu, and H. Z. Yu, "Prediction of gas emission time series based on W-RBF," *Mei T'an Hsueh Pao, J. China Coal Soc.*, vol. 33, pp. 60–70, Jan. 2008.

[34] D. W. Dong, S. G. Li, X. T. Chang, and H. F. Lin, "Prediction model of gas concentration around working face using multivariate time series," *J. Mining Saf. Eng.*, vol. 29, no. 1, pp. 135–139, 2012.

[35] W. Xiang, Q. Jian-sheng, H. Cheng-hua, and Z. Li, "Short-term coalmine gas concentration prediction based on wavelet transform and extreme learning machine," *Math. Problems Eng.*, vol. 2014, Jul. 2014, Art. no. 858260.

[36] G. Lu, X. Li, B. Zu, and J. Dong, "Research on time-varying series forecasting of gas emission quantity based on EMD-MFOA-ELM," *J. Saf. Sci. Technol.*, vol. 13, no. 6, pp. 109–114, 2017.

[37] H. Fu and Y. H. Y. Y. Liu Li Xu Wang, "Short term forecasting model of gas concentration in coal mine using the CAPSO-ENN," *Chin. J. Sensors Actuators*, vol. 28, no. 5, pp. 717–722, 2015.

[38] X. Wang, N. Xu, X. Meng, and H. Chang, "Prediction of gas concentration based on LSTM-LightGBM variable weight combination model," *Energies*, vol. 15, no. 3, p. 827, Jan. 2022.

[39] Y. Gao, X. Zhang, Z. Li, and Y. Du, "An attention-based spatio-temporal model for methane concentration forecasting in coal mine," *Neural Process. Lett.*, pp. 1–22, Oct. 2022.

[40] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 2014–2023.

[41] C. Li, Z. Cui, W. Zheng, C. Xu, and J. Yang, "Spatio-temporal graph convolution for skeleton based action recognition," in *Proc. 32nd AAAI Conf. Artif. Intell.*, Apr. 2018, pp. 3482–3489.

[42] J. Bruna et al., "Spectral networks and locally connected networks on graphs," 2013, *arXiv:1312.6203*.

[43] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[44] F. Monti, D. Boscaini, J. Masci, E. Rodola, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5115–5124.

[45] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.

[46] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*.

[47] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 83–98, May 2013.

[48] M. Henaff, J. Bruna, and Y. LeCun, "Deep convolutional networks on graph-structured data," 2015, *arXiv:1506.05163*.

[49] M. Simonovsky and N. Komodakis, "Dynamic edge-conditioned filters in convolutional neural networks on graphs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 29–38.

[50] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3104–3112.

[51] Z. Cheng, L. Ma, and Y. Zhang, "Prediction of spatiotemporal distribution of gas concentration based on LSTM-FC model," *Comput. Eng. Appl.*, vol. 56, no. 16, pp. 258–264, 2020.
[52] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.
[53] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
[54] D. Wu, L. Gao, M. Chinazzi, X. Xiong, A. Vespignani, Y.-A. Ma, and R. Yu, "Quantifying uncertainty in deep spatiotemporal forecasting," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2021, pp. 1841–1851.
[55] L. Zhu and N. Laptev, "Deep and confident prediction for time series at Uber," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 103–110.

YIFEI GAO received the B.Admin. degree in information management and information systems from the Shandong University of Finance and Economics, in 2019. He is currently pursuing the Ph.D. degree in management science and engineering with the Beijing University of Posts and Telecommunications. His current research interest includes spatio-temporal data mining and analysis.

XIAOHANG ZHANG (Member, IEEE) received the Ph.D. degree in management science and engineering from the Beijing University of Posts and Telecommunications, in 2003. From 2015 to 2017, he was a Research Fellow with the Department of Statistics, University of Michigan, Ann Arbor. He is currently a Professor of information systems with the Beijing University of Posts and Telecommunications. His current research interests include machine learning, business intelligence, and risk modeling.

TIANBAO ZHANG is currently a Senior Student with the Department of Mathematics, University of Washington. His current research interests include machine learning and deep learning.

ZHENGREN LI received the Ph.D. degree in management science and engineering from the Beijing University of Posts and Telecommunications (BUPT), in 2014. Currently, he is an Associate Professor with BUPT. His current research interests include user behavior analysis, data mining, and business intelligence.