

Original paper

Distance adaptive graph convolutional gated network-based smart air quality monitoring and health risk prediction in sensor-devoid urban areas

Shahzeb Tariq ^{a,+}, Shahroz Tariq ^{c,+}, SangYoun Kim ^a, Simon S. Woo ^b, ChangKyoo Yoo ^{a,*}^a Integrated Engineering, Dept. of Environmental Science and Engineering, College of Engineering, Kyung Hee University, 1732 Deogyeong-daero, Giheung-gu, Yongin-si, Gyeonggi-do, 17104, Republic of Korea^b Department of Computer Science and Engineering, College of Computing and Informatics, Sungkyunkwan University, 2066 Seobu-Ro, Jangan-Gu, Suwon-si, Gyeonggi-do, 16419, Republic of Korea^c Data61, CSIRO, Sydney, New South Wales, Australia

ARTICLE INFO

Keywords:

Sustainable development
Remote sensing
Early warning system
Distance adaptive graph
Graph convolutional gated network

ABSTRACT

Robust air quality early warning systems for sustainable development of cities have garnered significant attention. These systems enable health risk warnings and regulatory plans when harmful pollutant levels are forecasted. However, with current frameworks, forecasting is only possible at locations with sufficient sensor data and cannot be predicted at new areas. Furthermore, long-term sensor failure also limits their real-time monitoring and forecasting applications. We propose a distance adaptive graph convolutional gated network that provides simultaneous forecasts of primary air pollutants at multiple temporal horizons and locations of a megacity via spatio-temporal sensor fusion. The framework also solves critical problems of early warning systems related to long-term sensor failure and prediction at a new location of city. The results suggest that for 12 h ahead PM_{2.5} forecast, the proposed model reduces prediction errors by 43.89% and 52.59% compared to the timeseries-Transformer and convolutional long-short term memory network, respectively. For long-term sensor failure imputation, the mean absolute error for CO, NO₂, O₃, PM₁₀, and PM_{2.5} ranged between 0.081–0.160, 0.004–0.007, 0.005–0.0155, 9.41–13.5, and 5.75–8.12, respectively. Whereas the remotely forecasted concentrations at sensor-less locations showed a close similarity to the actual air quality distribution in the city area.

1. Introduction

Rapid industrialization, societal change, and energy consumption have led to substantial air pollution problems over the last two decades. Numerous studies and media outlets have repeatedly emphasized the detrimental effects of air pollution on ecosystems, human health, and life expectancy (Faganeli Pucer and Štrumbelj, 2018; Shi and Wu, 2020). Consequently, the European Union (EU) has enacted legislation compelling member states to limit their industrial and transportation emissions of hazardous pollutants (European Commission 2016). According to the World Air Quality Report, small improvements have occurred in major cities around the globe; however, approximately 78% of the monitored countries exceeded the World Health Organization (WHO) pollutant standards in 2020 (IQAir 2020). Thus, it is necessary to develop reliable systems for air quality monitoring, prediction, and traceability for sustainable and healthy development of cities. In this

context, accurate identification of pollutant dynamics based on the spatio-temporal realization of air quality patterns by sensor fusion might aid governing authorities in adopting effective air pollution management strategies that minimize exposure effects.

Air quality is primarily measured by considering the distribution of primary air pollutants, including particulate matter (PM₁₀), fine particulate matter (PM_{2.5}), carbon monoxide (CO), nitrogen dioxide (NO₂), and ground-level ozone (O₃) levels (Zhang, Thé, Xie, and Yu, 2020). Constant exposure to primary air pollutants may have both short-and long-term detrimental effects on respiratory and cardiovascular health. For example, short-term exposure to NO_x may result in respiratory morbidity and impaired lung function (Shaw et al., 2020). Li et al. found that an increase in ambient CO content by 1 mg/m³ increased hospital admissions for cardiovascular and coronary heart diseases by 2.8% and 3%, respectively (Li et al., 2018). While exposure to elevated surface O₃ levels is strongly associated with respiratory disorders (Chen

^{*} Corresponding AuthorE-mail address: ckyoo@knu.ac.kr (C. Yoo).⁺ These authors have contributed equally to this research.

et al., 2021). Additionally, the Global Burden of Disease 2019 study determined that ambient PM_{2.5} was responsible for nearly 4,140,971 all-cause deaths and 307,680 lung cancer deaths worldwide (Vos, Lim, and Abbaftati, 2020, Yun and Zhao, 2022). Therefore, simultaneous monitoring of primary air pollutants is required for true realization of air quality. However, measuring concentrations alone does not adequately inform the general public or policymakers about the state of air quality.

To simplify the public's understanding of the health risks caused by ambient air, the US Environmental Protection Agency (EPA) developed an air quality index (AQI) (Benchrif et al., 2021). The index categorizes air quality into six distinct zones, ranging from "good" to "hazardous," which indicate levels of health risk concern. With regard to air quality monitoring and risk assessments, researchers have suggested implementing early warning systems to alert citizens to possible health hazards as well as intelligent urban pollution control by government agencies (Li and Zhu, 2018, Wang, Li, and Lu, 2018). Nonetheless, developing early warning systems for future health risks remains a challenging task because air quality is influenced by a variety of variables, including meteorological conditions, traffic pollution, and spatial dependence (Seng et al., 2021).

Current air quality prediction models are divided into two main categories: numerical simulation-based models and machine-learning (ML) models (Wang, Li, Yang, and Wang, 2021). Simulation-based models such as CMAQ by the US EPA (Djalalova, Delle Monache, and Wilczak, 2015), and WRF/Chem (Grell et al., 2005), explain environmental conditions by identifying the underlying physiochemical processes and dynamic mutability of pollutants. In contrast, various researchers have applied traditional ML models for air quality predictions, including multiple linear regression (MLR) (Vlahogianni et al., 2011), least square support vector machines (LSSVM) (Li, Wang, Li, and Lu, 2019), gradient boosting algorithm (Yang, Shi, Tang, and Yang, 2022), and random forest regression (RFR) (Wang et al., 2020). In general, conventional ML models can realize the temporal relationships among pollutants and have shown superior performance compared to deterministic models (Feng et al., 2015). However, these frameworks require substantial feature engineering and cannot correctly capture the non-stationary dynamism of timeseries correlations.

In recent years, deep neural networks (DNNs), with the assistance of enormous air quality data, remote sensing (Su et al., 2021), and sophisticated algorithms, have overcome the limitations of traditional ML architectures, especially with respect to time series forecasting (Liu, Yan, Duan, and Chen, 2021). Sequential data processing and hidden memory states reveal that recurrent neural networks (RNNs), and their variants are appropriate to capture time-varying dynamics of air quality relationships (Wu and Lin, 2019). Bai et al. proposed an empirical mode decomposition and ensemble long short-term memory (LSTM) model to predict hourly PM_{2.5} concentrations (Bai, Zeng, Li, and Zhang, 2019). The ensemble LSTM demonstrated superior performance in comparison to the regular LSTM and feed-forward network by effectively mapping each of the decomposed modes. Moreover, Ma et al. proposed the integration of transfer learning and bidirectional LSTM (Bi-LSTM) to improve model performance where insufficient data are available for univariate predictions (Ma et al., 2019). Although these studies provide efficient temporal learning frameworks, the spatial relationships between pollutants are still ignored.

To further improve air quality forecasts, several researchers have proposed hybrid spatio-temporal prediction models that provide early estimations for future pollutant levels (Zhao, Deng, Cai, and Chen, 2019). Yin et al. combined data augmentation and a residual network structure to estimate surface PM_{2.5} concentrations by remote sensing (Yin, Li, Cheng, and Wu, 2022). The findings demonstrated an accurate estimation of high PM_{2.5} concentrations, supported by R² of 0.82, which is beneficial for PM exposure and epidemiological analysis. Zhang et al. combined a self-attention mechanism with a gated recurrent unit (GRU) network for the long-term prediction of citywide AQI distributions (Zhang, Thé, Xie, and Yu, 2020). The AQI distribution of pollutants

provides a clear view of health risks and potentially enhances decision-making capacities for environmental management. To further facilitate spatiotemporal dependency, researchers have incorporated spatial pollutant data into a graph structure utilizing graph convolutional networks (GCNs) (Zhou, Zhang, Du, and Liu, 2021).

The GCN utilizes an adjacency matrix to integrate spatial correlations from neighboring nodes to effectively improve forecasting at a target node (Holzinger, Malle, Saranti, and Pfeifer, 2021). GCN have been utilized to expand the learning capability of DNNs for a variety of fascinating tasks including chest water distribution systems (Zanfei et al., 2022), geolocation identification (Zhou, Wang, Zhong, and Trajcevski, 2022), among others (Yao et al., 2022, Li et al., 2022, Chae, Lee, Han, and Seong, 2022). On the other hand, the GCNs have also been applied to facilitate spatio-temporal learning for timeseries prediction tasks (Choudhury, Middya, and Roy, 2022). Chen et al. proposed an attention-aware GCN framework for multi-scale temporal feature extraction and multi-variate forecasting for financial and traffic data (Chen, Ding, and Zhai, 2022). Qi et al. proposed a hybrid graph convolutional LSTM (GC-LSTM) framework to predict PM_{2.5} at multiple temporal horizons (Qi, Li, Karimian, and Liu, 2019). The GC-LSTM examines the predictors as a unified graph; therefore, the combined spatial information improves the prediction performance compared to other models. However, the current graph representation approaches, as represented by the variants of GCN (Kipf and Welling, 2017), GAT (Veličković et al., 2018), and GraphSAGE (Hamilton, Ying, and Leskovec, 2017) are unweighted graph representations that do not include the realization of node locations. Thus, these methods overlook location and distance-based dependencies while identifying the spatial-temporal relations among different nodes.

Despite the awareness of spatial correlations, current early warning frameworks have several intrinsic drawbacks, including unique model development for each monitoring location or considering a single target pollutant at the expense of others. Several reasons hinder their real-time application, notably individual model updates, hardware costs, and incomprehensive air quality information. Missing data and sensor failure are another major issue that can be ascribed to uncontrolled factors such as electrical malfunctions, hostile monitoring environments, and long usage (Loy-Benitez, Heo, and Yoo, 2020, Loy-Benitez, Li, Nam, and Yoo, 2020). These events can further compromise the reliability and forecasting capabilities of traditional air quality early warning frameworks, which rely on continuous air quality monitoring. In addition, these models only provide future pollutant information at already established monitoring locations and do not offer remote sensing capability (Zhang, Thé, Xie, and Yu, 2020). Therefore, frameworks exhibiting remote sensing and multi-horizon forecasts at a specified location are limited. Ground station monitoring is the most direct and precise method for air quality monitoring. However, station distribution is sparse, and only a few monitoring stations are usually available in an urban environment (Huang et al., 2021). This problem is further incremented by the discrepancy between developed and underdeveloped regions.

In this study, we aim to overcome the aforementioned limitations of air quality early warning systems using a distance adaptive graph convolutional gated recurrent network (DAGCGN) incorporating sensor fusion. The study was conducted in three main phases. First, air quality monitoring data were collected from 23 monitoring stations located in Gyeonggi Province, South Korea, and pre-treated for outliers and missing observations. Then, statistical analysis was conducted to identify the underlying spatial correlations in the study area. Second, the DAGCGN framework was developed to simultaneously forecast CO, NO₂, O₃, PM_{2.5}, and PM₁₀ at multiple temporal horizons in all target monitoring locations. The performance of the proposed framework was compared with other spatiotemporal models, including LSTM, GRU, bidirectional-LSTM, convolutional-LSTM, and the timeseries transformer model. Third, the application of the DAGCGN was further analyzed in two scenarios. First, pollutant forecasting was conducted in

the event of a sensor failure at a monitoring location. Finally, remote sensing of pollutants and future forecasts at unknown locations is explored to estimate the health risk distribution given by the CAI.

2. Background

This study leveraged the synergy of graph convolutional operations and node sampling to detail the dynamics of numerous air pollutants in a spatiotemporal domain using sensor fusion to facilitate air quality monitoring and remote sensing. Therefore, this section is dedicated to the introduction of graph convolutional frameworks.

2.1. Graph convolutional network

To anticipate air pollution concentrations at a wide scale, we must analyze the geospatial relationship between ground monitoring stations and train them collectively. The distribution of stations and their associated data is not regular in nature, as in sequences or grids, but rather has an irregular topological structure similar to that of non-Euclidean data (i.e., a graph-like structure). Numerous studies have successfully used GCN to utilize the non-Euclidean topological information of weather stations (Wang et al., 2020), traffic flows (Yu, Lee, and Sohn, 2020), and PM_{2.5} concentration relationships (Liu et al., 2021) to obtain enticing forecasting results. This study also examines data with similar irregular structures and hence employs the GCN framework. Owing to their spatial localization capability and mathematical simplicity, GCNs represented by spectral filters have attracted substantial interest (Sandryhaila and Moura, 2013). The normalized Laplacian transform on graph G, based on the graph Fourier transform, is given by Eq. (1). The spectral graph transformation of a signal 'x' with the filter $f_\theta = \text{diag}(\theta)i.e., \theta\mathcal{CR}^n$ is represented by Eq. (2):

$$L = I_n - D^{-1/2}AD^{-1/2} = U\Lambda U^T \quad (1)$$

$$x * f_\theta = Uf_\theta U^T x \quad (2)$$

Here, I_n , A , and D represent the identity, adjacency, and the degree matrices, respectively. \mathcal{UCR}^n and $\Lambda\mathcal{CR}^n$ comprise matrices containing the eigenvectors and diagonal matrix of the eigenvalues of L , respectively. Moreover, $x * f_\theta$ represents the convolution operation and $U^T x$ represents the Fourier transform of x . Although Eq. (2) can be used to identify the spectral graph convolutions, the high computation of eigenvalue decomposition resulted in lower applicability and led to different approximations. Kipf et al. proposed that the graph convolutional process can be approximated well using a first-order Chebyshev polynomial expansion and can be extended to high-dimensional GCNs as follows (Kipf and Welling, 2017):

$$Z = (I_n + D^{-1/2}AD^{-1/2})X\Theta + b \quad (3)$$

Computationally, $X\mathcal{CR}^n$ and $Z\mathcal{CR}^n$ represent the inputs and the outputs of the GCN layer, respectively, whereas $\Theta\mathcal{CR}^n$ and $b\mathcal{CR}^n$ denote the learnable weight and bias parameters for the model. To further enhance the learning capability of the GCN framework, Bai et al. proposed data adaptive graph generation (DAGG) for automatically suggesting hidden interdependencies from data (Bai et al., 2020). In this study, learnable node dictionaries $C_A\mathcal{CR}^n$ are initialized for all nodes, and C_A represents the node embedding. The spatial dependencies between each pair of nodes are inferred by a product of C_A and C_A^T , whereas a softmax is utilized to normalize the adaptive matrix. The network automatically updates the C_A to detail the spatial relationships among different monitoring stations and builds an adaptive graph topology depending on the data. The DAGG module is represented by Eq. (4) and based on Eq. (3), the DAGG enhanced GCN can be formulated as:

$$D^{-1/2}AD^{-1/2} = \text{softmax}(\text{ReLU}(C_A C_A^T)) \quad (4)$$

$$Z = (I_n + \text{softmax}(\text{ReLU}(C_A C_A^T)))X\Theta + b \quad (5)$$

2.2. Distance adaptive graph convolutional network

The adaptive graph convolutional network efficiently discovers hidden relationships in the path of automated graph construction from a network of spatially distributed nodes. However, it has certain disadvantages. As the number of nodes increases, the cost of convolutions for graph formation becomes prohibitively expensive. Moreover, in larger graphs, it is not ideal to utilize all neighborhood information for each target node (Zhou et al., 2020). As a result, graph partitioning and node sampling techniques should be explored to effectively optimize network training. Chiang et al. suggested sampling several subgraphs and limiting the neighboring nodes within each subgraph (Chiang et al., 2019). Nevertheless, owing to the non-uniform distribution of monitoring stations, the preceding strategy may result in an unstable number of nodes, resulting in either incomprehensible spatial information or neighbor explosion, depending on the grid size chosen.

In comparison, Yang et al. suggested an importance-based node sampling technique that incorporates initiating random walks from target nodes and selecting the top N nodes based on their highest normalized visit count (Ying et al., 2018). Given the topology of air quality monitoring stations, a distance matrix S , and a set number of nearby nodes K , the latter technique may result in the automated identification of the most highly correlated nodes for inference and adaptive graph creation; more specifically, the distance adaptive graph convolutional (DAGC) module. A graphical representation of the DAGC is presented in Fig. 1. Additionally, the module enables the neural model to have a flexible memory footprint while simultaneously benefiting from the spatial dependency of the significant top-K nodes.

3. Materials and Methods

3.1. Data description and statistical analysis

Gyeonggi-do province, which accounts for 25% of South Korea's population, was selected as the study area. In recent years, this region has experienced rapid urbanization with associated increases in industrial outputs and energy consumption, leading to frequent and severe air pollution scenarios. Fig. 2 shows the geographical map of southern Gyeonggi province and the locations of 23 selected monitoring stations spread across different cities. This study is primarily focused on the southern side of the province, which neighbors the capital of the country.

The Korea Environment Corporation, supervised by the Ministry of Environment, maintains a database of pollutants recorded at 446 stations for nationwide air quality monitoring (Korea Environment Corporation August 20, 2021). These publicly accessible and quality-assured datasets have been utilized in prior studies (Ju, Oh, and Choi, 2021, Han et al., 2020). This study compiled data on five key pollutants, PM_{2.5}, PM₁₀, NO₂, CO, and O₃, from 23 monitoring sites between January 1 and December 31, 2019. The temporal distribution of pollutants at station 1 is illustrated in Fig. 3.

The missing pollutant observations in the raw data were filled using the shape-preserving piecewise cubic spline interpolation method. There were few outliers in the present data; however, obvious outliers were removed by employing and supplementing the observations with a cubic Hermitian polynomial transformation interpolation. All operations were performed using the matrix laboratory (MATLAB) provided by MathWorks. For intuitive insight, the basic statistics of the data are presented in Table 1. The statistics reveal that the majority of air pollutants were highly skewed, demonstrating a nonlinear distribution. However, the NO₂ dynamics were moderately skewed.

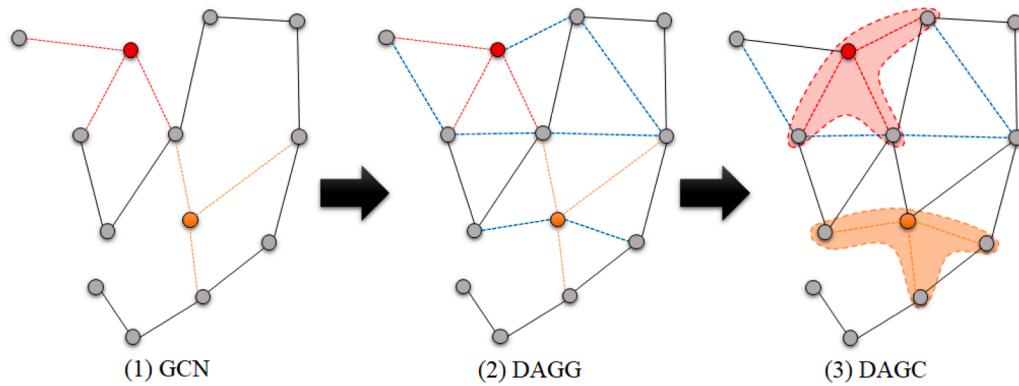


Fig. 1. Graphical representation of DAGC neighborhood node embeddings.

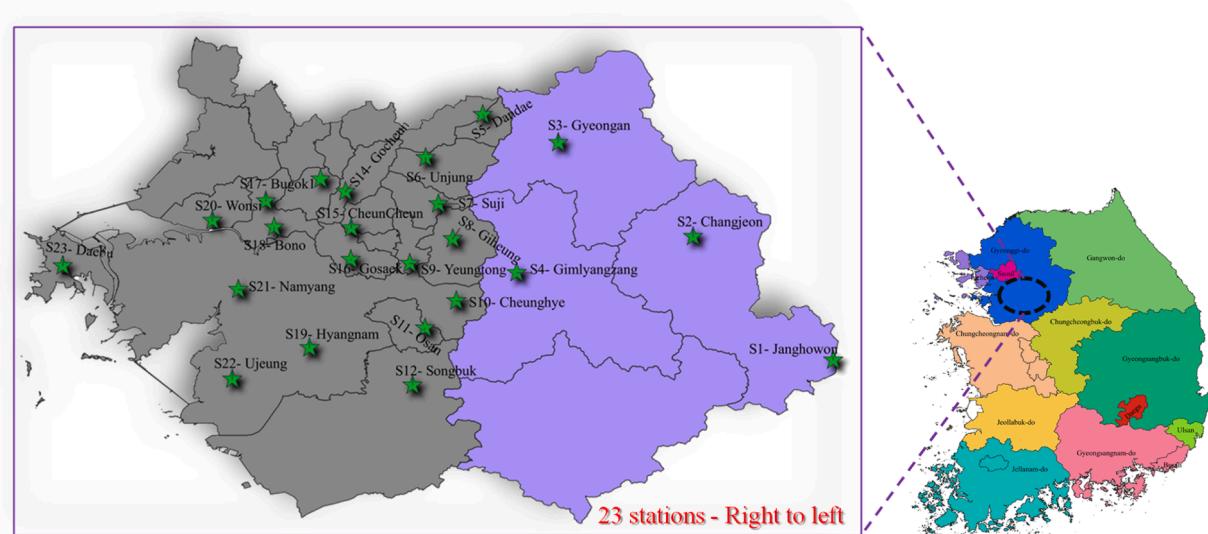


Fig. 2. Geographical location of monitoring stations.

3.2. Spatio temporal dependency

In air quality-related tasks, GCNs are mainly applied to facilitate temporal learning by capturing hidden spatiotemporal dependencies. However, it is necessary to first evaluate whether a spatial correlation exists among the data. To address this, the spatial correlation between monitoring stations was evaluated using Pearson's correlation coefficient (R). As presented in Fig. 4, a strong spatial correlation exists between the air quality distribution of the study area, with R scores for $\text{PM}_{2.5}$ -based relationships ranging between 0.7316 and 0.9610. Therefore, graph-like neural networks that efficiently extract spatial dependencies can help alleviate temporal learning.

Further investigation reveals that the R -score has an inverse relationship with the distance between monitoring stations, creating highly correlated smaller sub-regions. In this scenario, the prediction outputs from nearby nodes must be carefully analyzed while reducing noise from distant nodes. Moreover, as presented in Fig. 5, the spatiotemporal correlation can also be verified using a box plot of the PM variation. Whereas the dots represent the distribution of samples between the upper and lower whiskers of the boxplot. Despite having slightly different concentration profile, a similar trend was observed throughout the year. Owing to space limitations limitation, a box plot of 1st nine monitoring stations is shown, whereas a box plot of the remaining monitoring stations is presented in Fig. S1 in Supplementary Information (SI).

3.3. Comprehensive air quality index and distribution characteristics

Air quality monitoring and pollutant analyses are essential in identifying the temporal characteristics of air quality and their spatial dependence. However, monitoring alone does not convey the true status of health risks to the public or policymakers, which is needed to establish a mechanism for warning and pollution control. To overcome this issue, air quality standards have been implemented in many countries, as they are easily computed and effectively convey health risk concerns (Janarthanan, Partheeban, Somasundaram, and Elamparithi, 2021). The Korean Environment Cooperation introduced a comprehensive air quality index (CAI) that considers the concentration of several environmental pollutants and characterizes the health risk based on the pollutant with the highest AQI score (Korea Environment Corporation August 20, 2021). The CAI for each pollutant was calculated using Eqs. (6 and 7):

$$AQI_p = \frac{I_{HI} - I_{LO}}{BP_{HI} - BP_{LO}} (C_p - BP_{LO}) + I_{LO} \quad (6)$$

$$CAI = \max [AQI_{p_1}, AQI_{p_2}, \dots, AQI_{p_n}] \quad (7)$$

AQI_p represent the individual air quality index for each pollutant and C_p is the current round of air pollution. BP_{HI} and BP_{LO} are concentration breakpoints $\geq CP$ and $\leq Cp$, respectively. In addition, I_{HI} and I_{LO} are the index values corresponding to BP_{HI} and BP_{LO} , respectively. Further

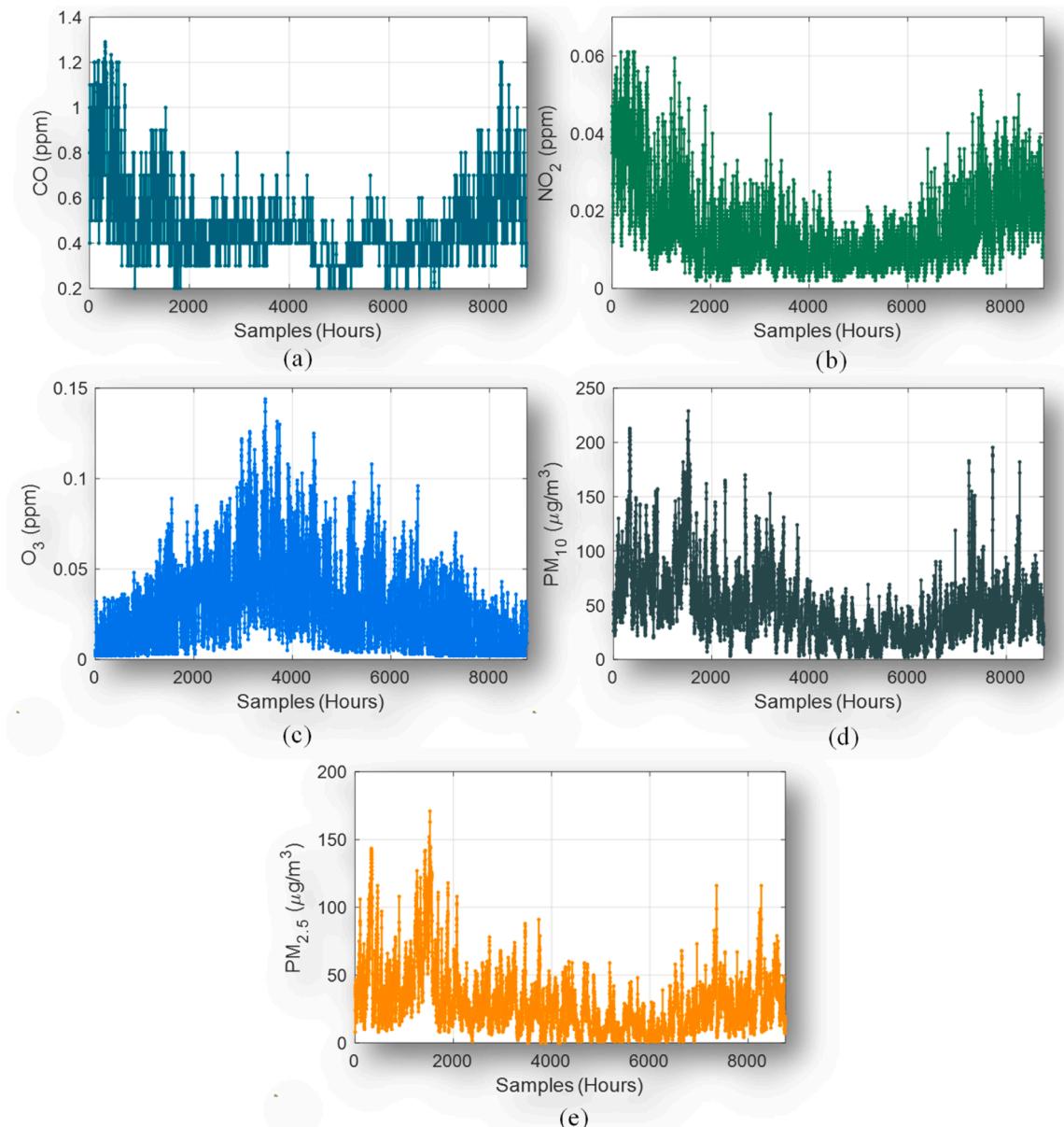


Fig. 3. Timeseries variation of air quality pollutants at monitoring station S1 (a) CO, (b) NO₂, (c) O₃, (d) PM₁₀, and (e) PM_{2.5}.

Table 1

Basic statistics of collected variables. Unit, range, mean, standard deviation and Skewness from 1 January 2019 to 1 December 2019.

Variable	Unit	Range	Mean	St. dev.	Skewness
CO	ppm	[0.1, 1.738]	0.51	0.19	1.34
NO ₂	ppm	[0.001, 0.127]	0.02	0.01	0.92
O ₃	ppm	[0.001, 0.144]	0.03	0.02	1.02
PM ₁₀	$\mu\text{g}/\text{m}^3$	[1, 283]	45.30	31.19	1.88
PM _{2.5}	$\mu\text{g}/\text{m}^3$	[1, 214]	25.41	22.26	2.24

details concerning health risk levels are presented in Table S1.

Notably, until recently, air quality evaluations mainly focused on the analysis and forecasting of individual pollutants, especially PM (Qi, Li, Karimian, and Liu, 2019, Liu et al., 2021). However, as described by CAI, individual pollutants cannot denote the true status of overall air quality. Fig. 6 shows a comparison of the health risks spread over the study area based on AQI_p and CAI, and to remove daily bias, the AQI and CAI were averaged for one month. The AQI only provides higher risk levels when

the breakpoints exceed a specific pollutant, thereby neglecting the effects of other pollutants. By contrast, CAI exclusively considers the most serious pollutants at each time step to portray the air quality status. The spatial distribution of CAI reiterates the importance of future information on all monitored pollutants in developing an accurate health risk warning system.

3.4. Distance adaptive graph convolutional gated recurrent network

The graph convolutional network (GCN) (Kipf and Welling, 2017, Defferrard, Bresson, and Vandergheynst, 2016) is a variant of the convolutional neural network (CNN) that is optimized for graph-structured data and is widely used in node classification, link prediction, and graph classification (Wu et al., 2021). Recent research has focused on developing complex graph neural network architectures for capturing shared patterns using predefined graphs. Additionally, learning node-specific patterns are critical in spatiotemporal outdoor air quality predictions. We propose to accomplish this task by utilizing a distance-adaptive graph convolutional gated recurrent network (DAGCGN). The

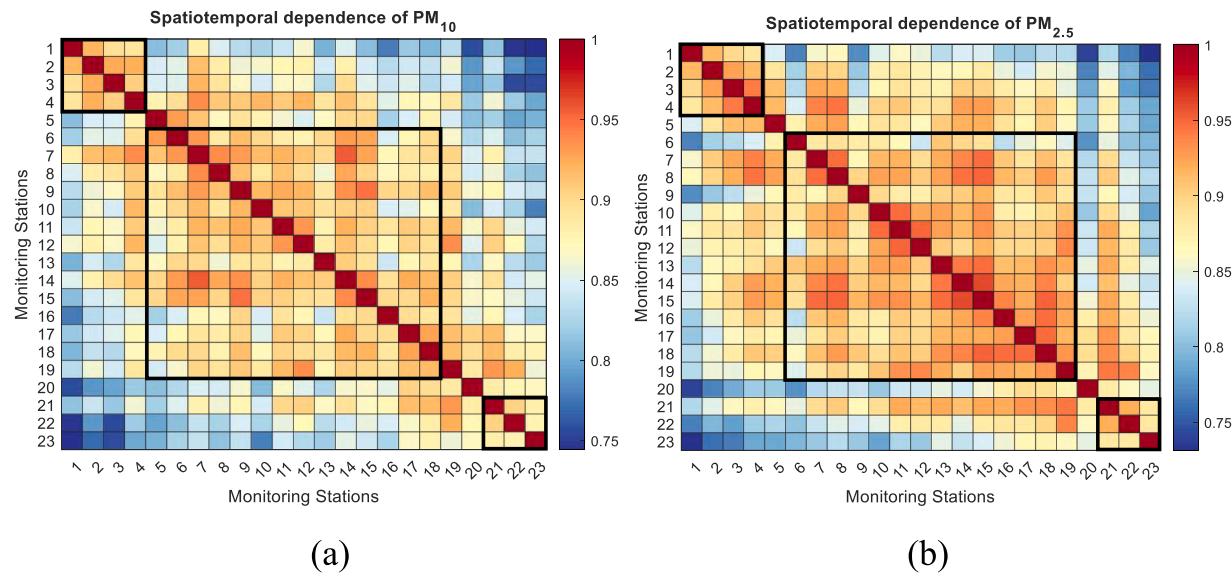


Fig. 4. Spatiotemporal dependence of PM between monitoring stations, (a) PM₁₀, (b) PM_{2.5}

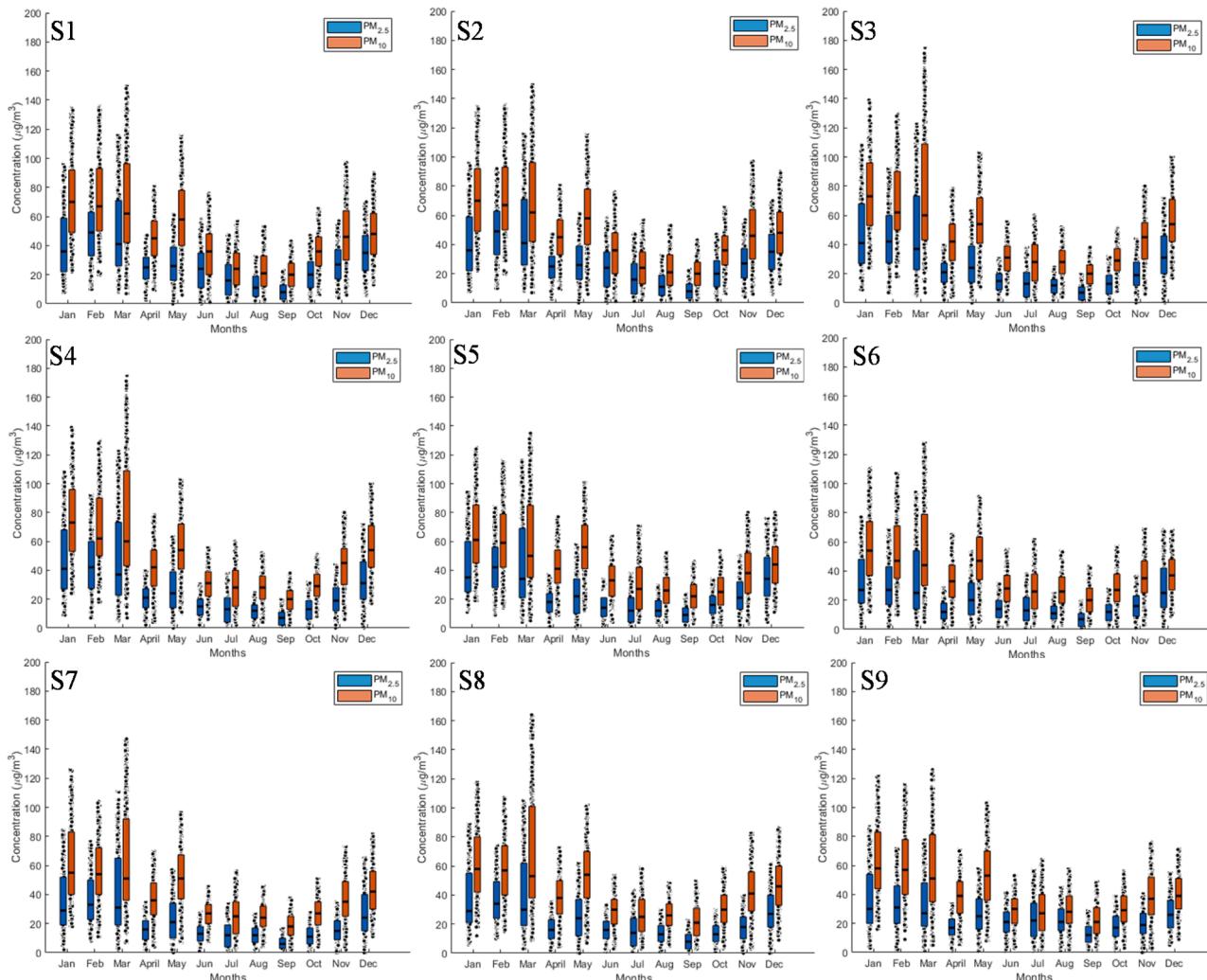


Fig. 5. Boxplot comparing monthly PM₁₀ and PM_{2.5} concentrations at S1-S9 monitoring stations.

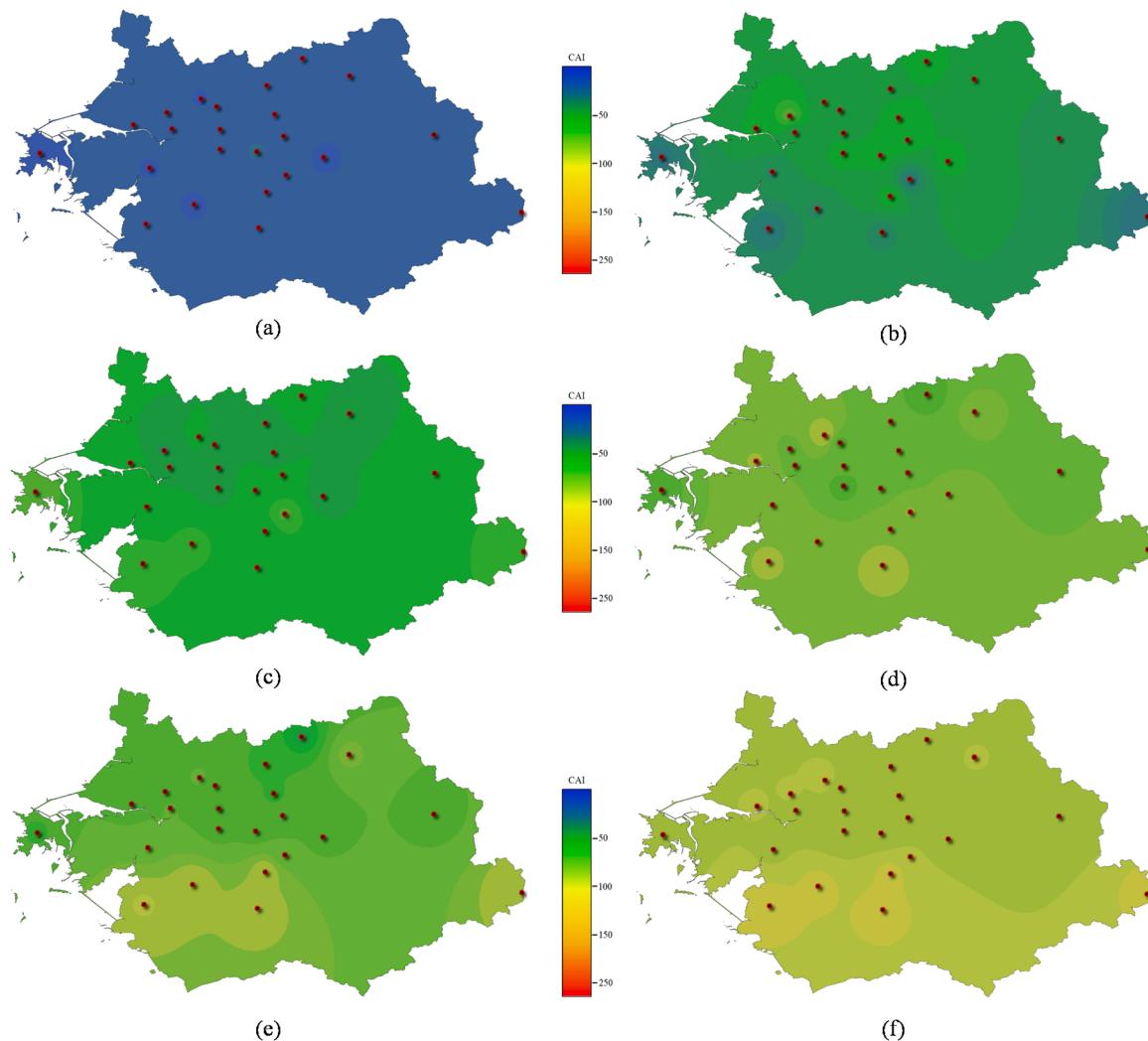


Fig. 6. Spatial distribution of AQI_p and CAI for the entire study area (01-04-2019 to 30-04-2019), (a) AQI_{CO} , (b) AQI_{NO_2} , (c) AQI_{O_3} , (d) $\text{AQI}_{\text{PM}10}$, (e) $\text{AQI}_{\text{PM}2.5}$ and (f) CAI.

following sections describe our methodology in detail.

3.4.1. Problem formulation

This section is focused on the multi-step and multi-site problem of forecasting outdoor air quality to provide early estimation of health risk levels by pollutant concentrations. The dataset is defined as $D = \{(S_i, Y_i)|i = 1, \dots, N\}$ containing N data samples from m sources as $S_i^t = \{s_{i,1}^t, \dots, s_{i,m}^t\} \in \mathbb{R}^{m \times p}$ and $Y_i^t = \{y_{i,1}^t, \dots, y_{i,m}^t\} \in \mathbb{R}^{m \times p}$, at time step t , respectively. Where p is the number of air quality metrics, and each data sample is composed of an $m \times p$ -dimensional multivariate input S_i^t and output Y_i^t timeseries. The problem is formulated as finding a function $\mathcal{F}_\theta(S_i^t) \in f : \mathbb{R}^N \rightarrow \hat{Y}_i^t$ capable of forecasting the future \hat{Y}_i^t based on the historical data S_i^t using a loss function $\mathcal{L}(Y_i^t, \hat{Y}_i^t)$ (e.g., mean squared error or mean absolute error), where θ denotes all learnable parameters in the model.

To obtain accurate spatial correlations between different outdoor station series, the problem is further formulated as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where \mathcal{V} comprises a set of nodes representing the air quality sensor sources with $|\mathcal{V}| = m$, \mathcal{E} is a set of edges, and $A \in \mathbb{R}^{m \times m}$ represents the adjacent matrix of the graph representing the proximity of the air quality sensors. As a result, the problem can be expressed as $\mathcal{F}_\theta(S_i^t; \mathcal{G}) \in f : \mathbb{R}^N \rightarrow \hat{Y}_i^t$, where an edge between two nodes is defined if and only if it is in a close neighborhood. Additionally, by adding distance to target location together with the air quality infor-

mation of selected spatial station as inputs, we force our model to learn the air quality distribution at remote locations and develop the underlying graph.

3.4.2. Time series transformation

The sequential raw sensor data of the outdoor air quality dataset is unsuitable for training a neural network-based method in its natural state. Therefore, preprocessing was first conducted by defining the size of the historical data (lag) l and multi-step futures (horizon) h . Using this information, the dataset (input S_i^t and output Y_i^t time series) was transformed using the sliding window technique (Norwawi, 2021). At each step, the method creates a sample by iterating over a sequence of features and target variables. The inputs for t_0 are created using the values between t_0 and Δt , and y_0 represents the target variables. Here Δt represents the window size, configured by the number of past input samples and the future temporal horizon of the target. For example, for an hourly time-resolution, if the lag and future horizon parameters are set to 5 and 2, respectively. At every time-step, the data is transformed into windows, containing concentrations of five primary pollutants from the previous five time-steps from other stations as inputs. Given these inputs, the model attempts to provide a two-hour ahead forecast of 5 primary pollutants as an output.

3.4.3. Network architecture

For effective outdoor air quality forecasting, it is critical to capture both the node-specific spatial and temporal correlations. Bai et al. proposed the first adaptive graph convolutional recurrent network for a univariate time series, that combines node adaptive parameter learning, data adaptive graph generation, and recurrent units to capture node-specific spatial and temporal correlations (Bai et al., 2020). We extend this concept to a multivariate time series and propose the following distance adaptive graph convolutional gated recurrent network-based encoder decoder framework for the simultaneous multi-step forecasting of several air quality pollutants:

$$\tilde{A} = \text{softmax}(ReLU(CC^T)) \quad (8)$$

$$z_t = \sigma(\tilde{A}[S_i^t, \mathcal{H}_{t-1}]CW_z + Cb_z) \quad (9)$$

$$r_t = \sigma(\tilde{A}[S_i^t, \mathcal{H}_{t-1}]CW_r + Cb_r) \quad (10)$$

$$\widehat{\mathcal{H}}_t = \tanh(\tilde{A}[S_i^t, r \odot \mathcal{H}_{t-1}]CW_{\widehat{\mathcal{H}}} + Cb_{\widehat{\mathcal{H}}}) \quad (11)$$

$$\mathcal{H}_t = z \odot \mathcal{H}_{t-1} + (1-z) \odot \widehat{\mathcal{H}}_t \quad (12)$$

where S_i^t and \mathcal{H}_t denote the input and output at time step t , respectively, $[.]$ denotes the concatenation operation, and z and r denote the reset and update gates, respectively. C , W_z , W_r , $W_{\widehat{\mathcal{H}}}$, b_z , b_r , and $b_{\widehat{\mathcal{H}}}$ are learnable parameters. For notation simplicity, we used GCN (S_i^t, A) to denote the output of a GCN model with inputs S_i^t and adjacency matrix A . The number of layers and choice of activation functions were suppressed in the notation. The task of the decoder is to reconstruct the latent space to a future time step Y_i^t using historical data from the input S_i^t and distance information from the adjacency matrix A . Specifically, the encoder and decoder can be summarized as follows:

$$\text{Encoder : } Z = \text{GCN}(S_i^t, A) \quad (13)$$

$$\text{Decoder : } Y_i^t = \text{BatchNorm}(\text{Conv2D}(Z)) \quad (14)$$

Fig. 7 illustrates the designed encoder-decoder architecture that includes two adaptive graph convolutional gated network (AGCGN) cells in the encoder. The basic structure of the proposed DAGCGN model is developed by taking inspiration from several state-of-the-art GCN frameworks such as spatio-temporal graph convolutional network

(STGCN) (Yu, Yin, and Zhu, 2018), attention-based spatio-temporal graph convolutional network (ASTGCN) (Guo et al., 2019), spatial-Temporal Synchronous Graph Convolutional Network (STGCN) (Song, Lin, Guo, and Wan, 2020), and adaptive graph convolutional recurrent network (AGCRN) (Bai et al., 2020). Additionally, findings in the previous research suggest that stacking sufficient GCN based layers can improve the model's ability to capture node-specific spatial and temporal dynamics. The DAGCGN network structure also contains a batch normalization layer to standardize the network's input data that aids in regularization and speeds up the learning process. In addition, a dropout layer is introduced to minimize overfitting and offer further regularization.

For model outputs, using only one convolution layer followed by a batch normalization layer, we kept the decoder extremely simple in our architecture. Therefore, we indirectly forced the encoder to create a latent space embedding that was both informative and sufficiently simple for the decoder to construct the future time step of multiple pollutants. This enabled our model to acquire a more accurate representation of the latent space of the input data and thus forecast outdoor air quality pollutants with greater accuracy. This study employs pytorch's ray tune library with random search method to select the best hyperparameters for the model. The hyperparameters to optimize include 1) number of hidden layers, 2) GRU units in AGCGN cells, 3) dropout rate, 4) node embedding dimension, and 5) activation functions. Table S2 of the SI presents the results of different model structures measuring the loss. The model is trained for 100 epochs with Adam as an optimizer and mean squared error (MSE) as loss function. Whereas the selected hyperparameters that yielded the lowest loss are given as follows:

- Number of hidden layers: 2
- Number of GRU units in AGCGN cells: 64, 64
- Dropout rate: 0.25
- Embedding dimension: 10
- Activation functions: Sigmoid, Tanh, ReLU, Softmax
- Learning rate: 0.003

The pseudo code for the data preprocessing and the training procedure for DAGCGN framework is described by [Algorithm 1](#) as follows:

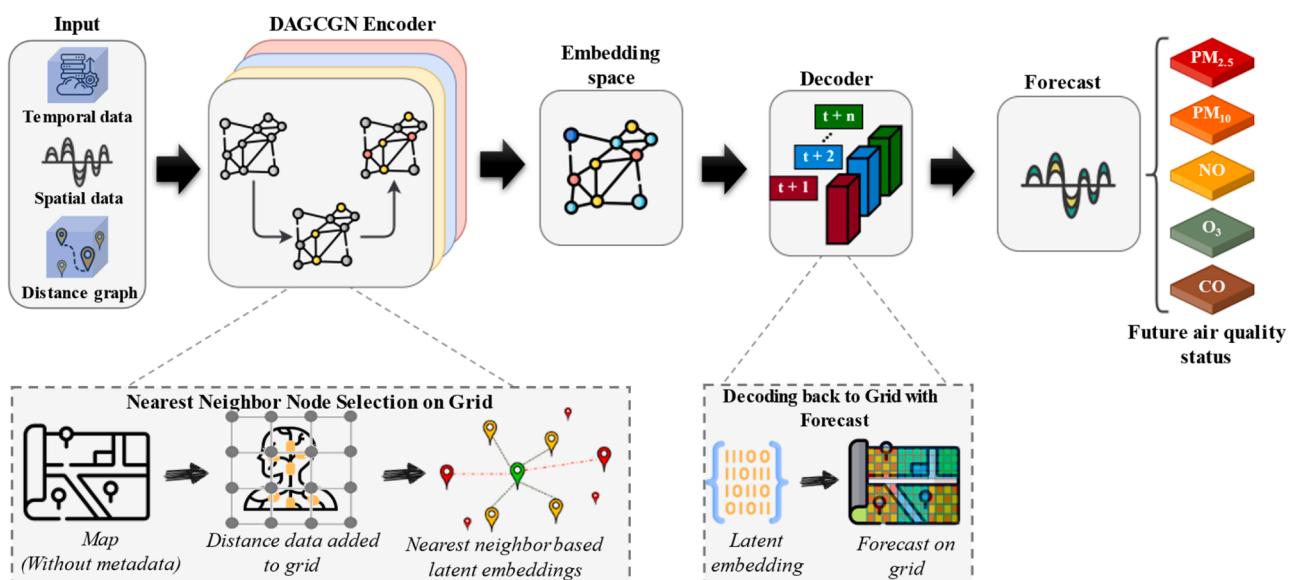


Fig. 7. Structure of distance adaptive encoder-decoder graph convolutional framework for multi-pollutant forecast.

Algorithm 1

Data preprocessing and training process of DAGCGN

```

Input:  $D = \{(S_i, Y_i) | i = 1, \dots, N\}$ ,  $M = [m][m]$ ,  $K = [k]$ ,  $P \leftarrow$  number of nearest neighbors i.e., proximity,  $H \leftarrow$  number of historical timesteps used for training i.e., historical window,  $F \leftarrow$  number of future timesteps for forecasts i.e., forecast horizon

1: for  $i$  in  $K$  do /* Data Preprocessing step */
2:   for  $j$  in  $K$  do
3:      $x, y = transform(D, i, j, M[i][j], P, H, F)$  /* transforms the dataset  $D$  into an input time-series  $S_i = \{s_{i,1}, \dots, s_{i,m}\} \in \mathbb{R}^{m \times p}$ , (i.e., Sensor values from  $m$  sources for  $H$  timesteps) and output time-series  $Y_i = \{y_{i,1}, \dots, y_{i,m}\} \in \mathbb{R}^{m \times p}$ , (i.e., Prediction for sensors  $S_i$  for  $F$  timesteps) */
4:      $S_i.append(x)$ ,  $Y_i.append(y)$ 
5:   end
6: end
7: while  $i < N$  do /* Training step */
8:    $\mathcal{F}_\theta(S_i) \in f: \mathbb{R}^N \rightarrow \hat{Y}_i$ 
9:    $\mathcal{L}(Y_i, \hat{Y}_i)$ , /* where the loss is calculated based on the difference between forecast and ground truth i.e.,  $\mathcal{L}(Y_i, \hat{Y}_i)$ , using rmse, mae or mase */
10:  update weights using backpropagation based on the loss
11:   $i = i + 1$ 
12: end
Output:  $\mathcal{F}_\theta(\cdot)$  /* Trained Model */

```

3.5. DAGCGN experimental settings and application

Sensors are extensively employed for collecting data and signals, especially for environmental monitoring. However, in the physical world, missing data from air quality monitoring equipment or sensor unavailability is a frequent issue and can be attributed to uncontrolled variables such as machine failure, frequent maintenance, unstable transmission, or expensive monitoring equipment. These events can lead to catastrophic consequences for health risk monitoring, which relies on continuous air quality data. Therefore, we considered two real-life validation scenarios to assess the performance of the proposed DAGCGN framework, as presented in Fig. 8. The first scenario involves long-term sensor failure at a monitoring station, which disables air quality realization at a monitoring station. The second scenario, by contrast, employs air quality remote sensing and generates forecasts at a new location that was not utilized in the training phase.

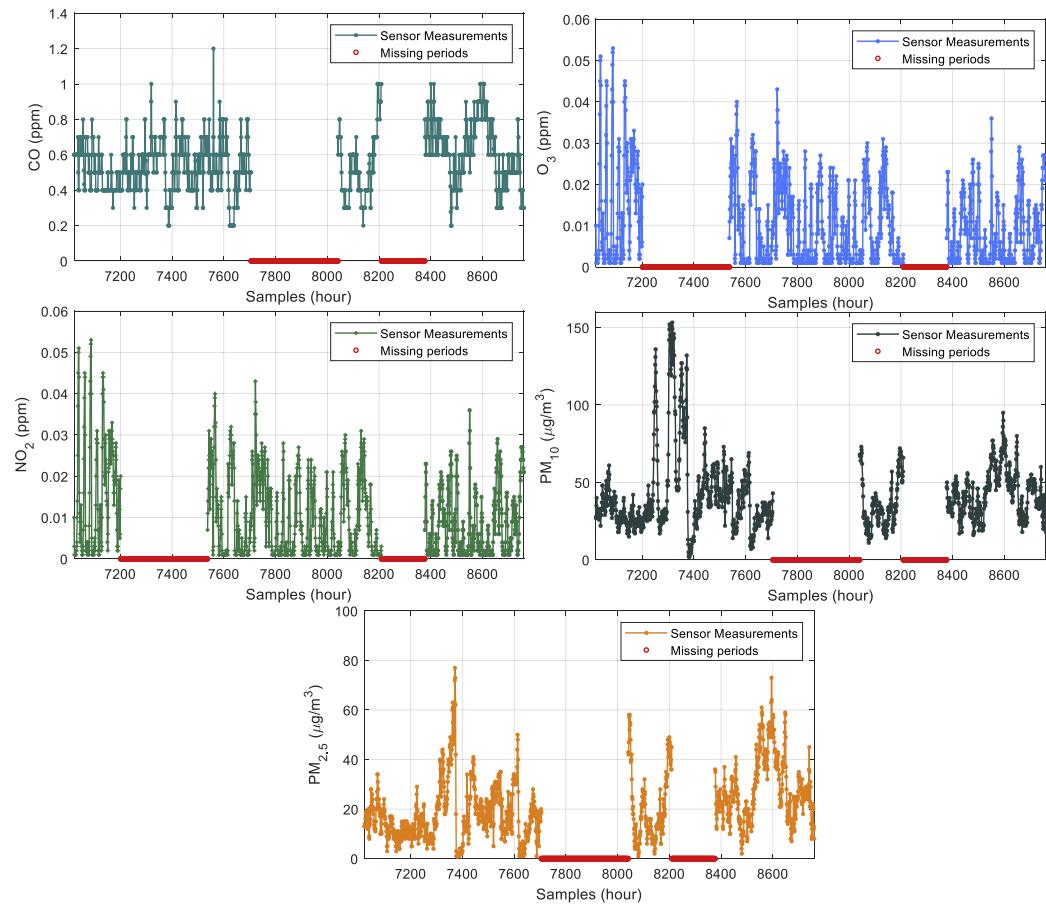
Scenario 1: The first validation scenario entailed purposely introducing two missing data intervals for each sensor measurement of CO, NO₂, O₃, PM₁₀, and PM_{2.5}, with different downtime periods (one week and two weeks) at the S7, S11, and S17 monitoring stations. Table 2 details the sensor failure incidence periods along with the data points where the missing sensor measurements were introduced. In the case of null outputs from a monitoring station, the network predicted future pollutants at that site using selected neighboring nodes. Moreover, Fig. 8 (a) illustrates the time series measurements of the air quality sensors and the missing data intervals at the S-7 monitoring station. Comparable

intervals were added for the other two stations.

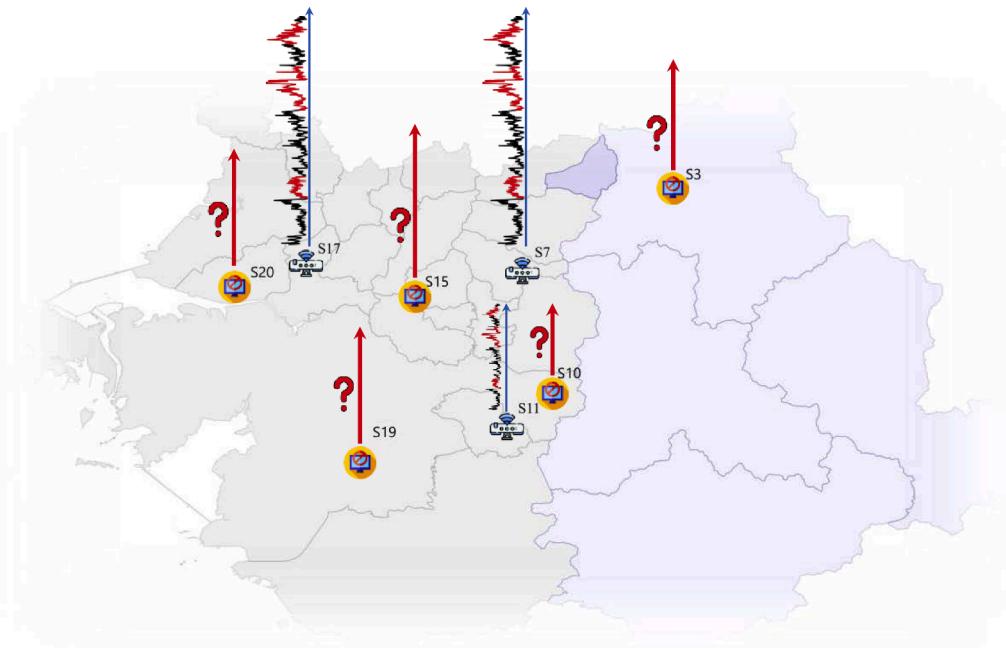
Scenario 2: The second scenario involves remote sensing and predicting the complete air quality status, specifically all five air pollutants, at an unknown location. For the demonstrations, we assumed that no sensor data were available for the five monitoring stations and labeled them as unknown locations. These monitoring stations were separated from the original dataset and were not utilized in any part of the network training. The DAGCGN framework constructed a secondary grid based on the nearest neighbor node sampling method and simultaneously forecasted NO₂, CO, O₃, PM₁₀, and PM_{2.5} based on the already developed dynamic graph. Furthermore, the locations of the target stations with missing data intervals and unknown stations for remote sensing (S3, S10, S15, S19, and S20) are presented in Fig. 8 (b). The pseudo code for these applications by the DAGCGN is described by Algorithm 2 as follows:

3.6. Performance evaluation

The constructed graph convolutional recurrent network was thoroughly tested for its ability to capture future air quality statuses using three distinct criteria often utilized in timeseries forecasting applications (Nam et al., 2020). These include the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute scaled error (MASE). The governing equations of the metrics are as follows:



(a)



(b)

Fig. 8. Timeseries distribution of measured and missing sensor data with different missing intervals of (a) CO, NO₂, O₃, PM₁₀, PM_{2.5} (Top Left to bottom) and, (b) location of monitoring stations with missing data intervals and no-sensor.

Table 2

Partial and complete sensor failure intervals for application scenario 1.

Monitoring station	Air quality sensors					Missing interval	
	CO	O ₃	NO ₂	PM ₁₀	PM _{2.5}	Start	End
S7	✓	❖	❖	✓	✓	7201	7537
	❖	✓	✓	❖	❖	7707	8043
	❖	❖	❖	❖	❖	8209	8377
S11	❖	❖	❖	❖	❖	7201	7369
	✓	❖	❖	✓	✓	7707	8043
	❖	✓	✓	❖	❖	8209	8545
S17	✓	❖	❖	✓	✓	7201	7537
	❖	❖	❖	❖	❖	7707	7873
	❖	✓	✓	❖	❖	8209	8545

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (15)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

$$\text{MASE} = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{\frac{1}{n-1} \sum_{i=2}^n |y_i - \hat{y}_{i-1}|} \right) \quad (17)$$

Here y and \hat{y} denote the observed and estimated values, respectively, at time step i .

3.7. Proposed Methodology

The suggested development of an air quality early warning framework is divided into four major parts. Fig. 9 illustrates the entire methodology. The first stage entailed collecting data on critical air quality indicators from 23 monitoring sites situated across Gyeonggi-do Province. Additionally, a dataset inspection was performed to determine missing values using piecewise cubic spline interpolation, and a cubic Hermitian polynomial transformation was used to remove various outliers that could bias the results. Second, a correlation analysis was used to determine whether the air pollutants had a spatial relationship in the

Algorithm 2

Forecast for sensor failure and remote sensing using trained DAGCGN

```

Input:  $D = \{(S_i, Y_i) | i = 1, \dots, N\}$ , /* dataset containing N data samples */
        $M = [m][m]$ , /* adjacency matrix containing distance between each sensor m */
        $U = [u]$ , /* list of u unknown stations */
        $F = [f]$ , /* list of f faulty sensors */
        $P \leftarrow$  number of nearest neighbors i.e., proximity
        $H \leftarrow$  number of historical timesteps used for training i.e., historical window
        $F \leftarrow$  number of future timesteps for forecasts i.e., forecast horizon
        $\mathcal{F}_\theta(\cdot)$  /* Trained Model */

-----
```

```

1: for  $i$  in  $K$  do /* Data Preprocessing step */
2:   for  $u$  in  $U$  do
3:      $x, y = \text{transform}(D, i, j, M[i][u], P, H, F)$ 
4:      $S_k^u.append(x)$ ,  $Y_u.append(y)$  /* here  $Y_u$  is the ground truth which can be
5:     used to validate the forecast  $\hat{Y}_u$  */
6:   end
7:   for  $f$  in  $F$  do
8:      $x, y = \text{transform}(D, i, j, M[i][f], P, H, F)$ 
9:      $S_k^f.append(x)$ ,  $Y_f.append(y)$  /* here  $Y_f$  is the ground truth which can be
10:    used to validate the forecast  $\hat{Y}_f$  */
11:   end
12: end
13:  $\hat{Y}_u = \mathcal{F}_\theta(S_k^u)$  /* Forecast for unknown locations */
14:  $\hat{Y}_f = \mathcal{F}_\theta(S_k^f)$  /* Forecast for faulty sensors */
-----
```

Output: \hat{Y}_u , \hat{Y}_f

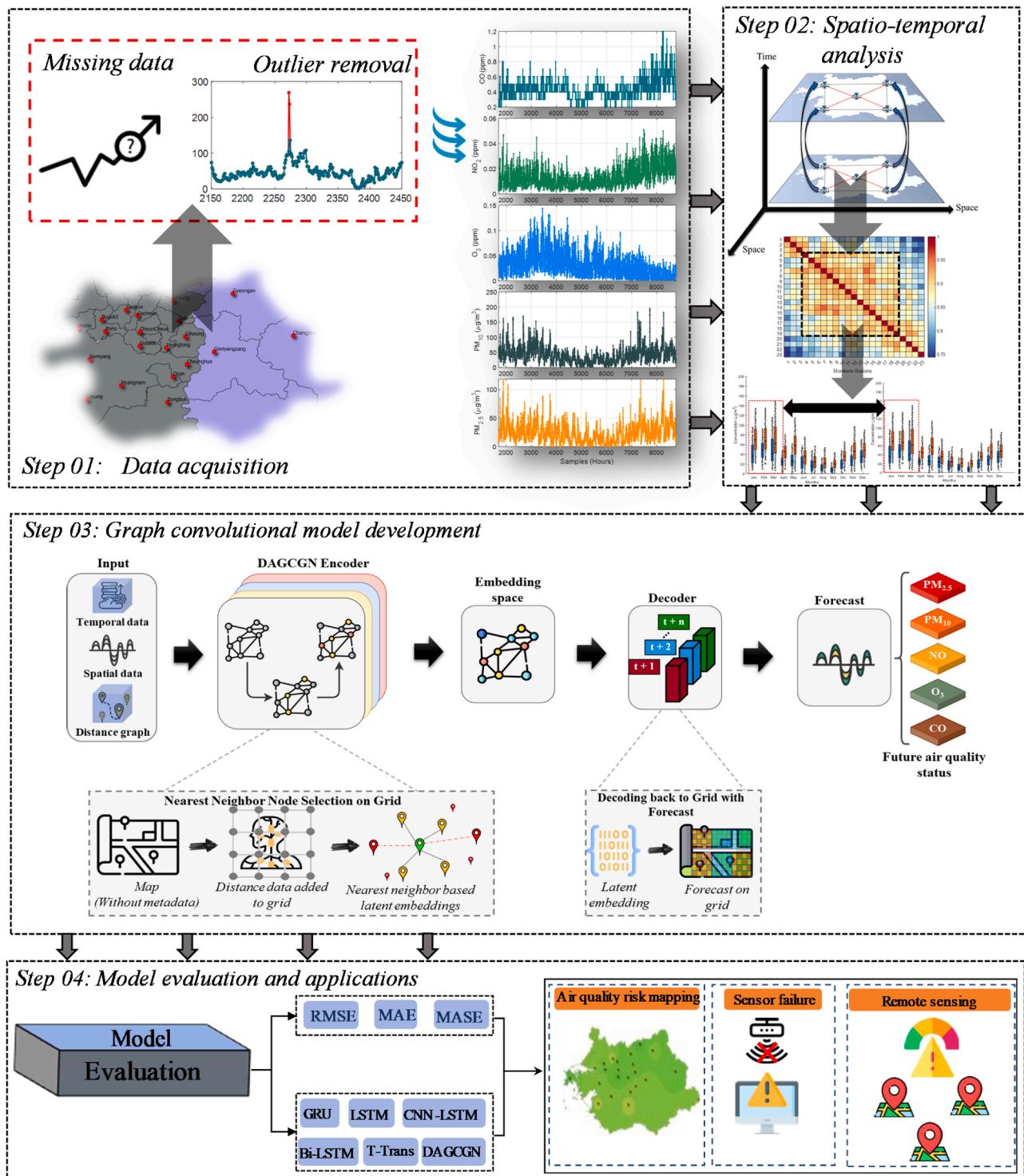


Fig. 9. Graphical representation of overall forecasting and application methodology of distance adaptive graph convolutional gated network (DAGCGN).

study region, and a CAI map was created to visualize the health risk distribution. At the conclusion of this step, the data were normalized and partitioned into training and test sets at an 80/20 ratio. Five monitoring stations (S3, S10, S15, S19, and S20) were separated from the original dataset to evaluate the model's performance for remote sensing application at unknown locations. This resulted in 18 stations labelled "known locations" for training and five stations labeled "unknown locations" to demonstrate the DAGCGN application.

In the third stage, to provide multi-step air quality forecasting, we utilized a sliding window to transform our data into input and output

windows. The proposed DAGCGN network was then constructed and incorporated into the sliding window architecture. The network was first trained on known locations to detail spatiotemporal dynamics and build the underlying graph. Various combinations of spatiotemporal forecasting models, including stacked spatial models and encoder-decoder frameworks were constructed for comparison. Notably, the DAGCGN network forecasted all five contaminants concurrently at each monitoring station. However, owing to the absence of a graph-based framework, unique and location-based models were developed for the stacked spatio-temporal baseline category.

Following that, the model evaluation, and applications of the proposed DAGCGN framework were performed. First, the performance was compared with the baselines at known stations with available sensor data to assess the early detection of pollutant levels across a range of temporal horizons, namely T_{+1} , T_{+3} , T_{+6} , and T_{+12} . Second, the proposed GNN was utilized to solve a critical problem in air quality monitoring involving sensor failures with variable lengths of downtime. Finally, the performance of the DAGCGN was evaluated for remote sensing at a new site without accessible pollution sensors for air quality monitoring. In addition, a CAI map was generated to demonstrate the efficacy of the suggested methodology in instances of limited monitoring for early identification of health risk distribution.

4. Results and Discussion

4.1. Air quality forecast and health risk estimation at known locations

In the first phase of the experiments, five different machine learning frameworks were employed as baselines to capture the spatiotemporal dynamics of environmental pollutants in the study area. For the model inputs, the time series was transformed into past sequences and future time-steps. The lookback period was set at 12 h with several temporal horizons, including T_{+1} , T_{+3} , T_{+6} , and T_{+12} , to provide an early estimation of the pollutant concentrations. The performance metrics quantified the error yielded by each neural network model, evaluated at the known monitoring locations, as characterized in the previous section. The most accurate model was identified using the lowest RMSE, MASE, and MAE values. The baselines consisted of a mix of spatiotemporal air quality forecasting architectures that were characterized into two categories.

Stacked spatial model (Zhao, Deng, Cai, and Chen, 2019, Vaswani et al., 2017): these models include implementing RNN variants (LSTM, GRU) and an attention-based timeseries transformer (T-Trans) framework. The LSTM and GRU models consist of a stacked architecture. The timeseries transformer also contains a GRU layer in the encoder. The stacked structures and attention mechanism enable effective learning of short- and long-term events by considering the air quality data from the spatial nodes located in the study area and generating predictions exclusively at the local node. This reduces the framework complexity at the output; however, unique multi-hour forecasting models are developed for each monitoring station. Each stacked spatial network has a dropout layer to prevent overfitting concerns. Additionally, the dataset for the multivariate stacked spatial models is prepared using the sliding window. Here at each station, the past 12 hours of sensor data from all stations located in the study area are used as inputs, whereas the output consists of 05 primary pollutants at different temporal horizons, i.e., T_{+1} , T_{+3} , T_{+6} , and T_{+12} .

Encoder-decoder spatial model (Zhang, Zhang, Zhao, and Lian, 2020, Yan et al., 2021): encoder-decoder frameworks were also considered for comparison between the baselines and the proposed GCN. In this study, the encoder mapped the hidden representation for each sensor based on the pollutant data available from all nodes in the study area. The decoder was designed to perform simultaneous multi-site predictions of all air quality pollutants. Therefore, a single encoder-decoder framework was developed for multi-site forecasts of pollutant sensors in the study region at different temporal horizons. However, the model complexity increased compared to stacked spatial models. To ensure an effective spatio-temporal representation, two network architectures were utilized. The first model employed a convolutional neural network (CNN) layer to extract spatial information from nearby nodes, whereas the second model employed a bidirectional LSTM layer, namely CNN-LSTM and Bi-LSTM, respectively. Additionally, a dropout layer is incorporated in the network structure to avoid overfitting issues. The inputs for stacked spatial models include 12 hours of past sensor data from all stations located in the study area, whereas the decoder generates simultaneous forecasts at all sensor locations

Table 3
RMSE of forecasting models at different temporal horizons on training set.

Time step	Model	CO (ppm)	O ₃ (ppm)	NO ₂ (ppm)	PM ₁₀ ($\mu\text{g}/\text{m}^3$)	PM _{2.5} ($\mu\text{g}/\text{m}^3$)
T+1	GRU	0.1207	0.0063	0.0082	12.9408	7.7823
	LSTM	0.1264	0.0066	0.0080	12.8433	8.5104
	T-Trans	0.1171	0.0059	0.0075	11.6309	7.3758
	Bi-LSTM	0.1038	0.0058	0.0067	10.4041	6.8155
	CNN-	0.1062	0.0058	0.0071	11.0244	7.0047
	LSTM					
	DAGCGN	0.0895	0.0058	0.0052	9.7251	5.6128
	GRU	0.1380	0.0095	0.0103	15.7429	9.0288
	LSTM	0.1382	0.0096	0.0100	15.7253	9.3886
	T-Trans	0.1316	0.0089	0.0100	14.5408	8.5553
T+3	Bi-LSTM	0.1343	0.0107	0.0102	14.9342	9.5729
	CNN-	0.1327	0.0099	0.0101	14.0942	8.4906
	LSTM					
	DAGCGN	0.1084	0.0065	0.0059	10.6319	6.1727
	GRU	0.1623	0.0134	0.0123	19.2531	11.3557
	LSTM	0.1592	0.0142	0.0124	18.9854	11.5454
	T-Trans	0.1582	0.0129	0.0125	18.6336	10.8019
	Bi-LSTM	0.1598	0.0147	0.0123	20.6623	12.9867
	CNN-	0.1585	0.0156	0.0128	19.4076	11.9800
	LSTM					
T+6	DAGCGN	0.1214	0.0065	0.0057	12.4452	6.7033
	GRU	0.1658	0.0126	0.0125	22.1422	13.3464
	LSTM	0.1665	0.0129	0.0122	21.9042	13.1019
	T-Trans	0.1660	0.0123	0.0125	21.6421	12.7739
	Bi-LSTM	0.1672	0.0144	0.0127	24.7832	15.2683
	CNN-	0.1699	0.0154	0.0130	22.9342	14.4250
	LSTM					
	DAGCGN	0.1119	0.0084	0.0067	13.4816	7.3994

Table 4
RMSE of forecasting models at different temporal horizons on test set.

Time step	Model	CO (ppm)	O ₃ (ppm)	NO ₂ (ppm)	PM ₁₀ ($\mu\text{g}/\text{m}^3$)	PM _{2.5} ($\mu\text{g}/\text{m}^3$)
T+1	GRU	0.1325	0.0069	0.0090	13.8834	8.2870
	LSTM	0.1365	0.0072	0.0087	13.9166	9.1804
	T-Trans	0.1259	0.0065	0.0083	12.4826	8.0007
	Bi-LSTM	0.1131	0.0062	0.0073	11.1668	7.2429
	CNN-	0.1165	0.0064	0.0078	11.8501	7.5375
	LSTM					
	DAGCGN	0.0939	0.0061	0.0055	10.1198	5.8345
	GRU	0.1557	0.0107	0.0115	17.2531	9.8081
	LSTM	0.1542	0.0106	0.0111	17.1980	10.1694
	T-Trans	0.1480	0.0100	0.0111	16.0728	9.4485
T+3	Bi-LSTM	0.1489	0.0119	0.0115	16.3197	10.3854
	CNN-	0.1476	0.0109	0.0112	15.4730	9.2479
	LSTM					
	DAGCGN	0.1157	0.0069	0.0065	11.3710	6.5597
	GRU	0.1832	0.0150	0.0139	21.7205	12.6986
	LSTM	0.1785	0.0158	0.0138	21.3737	12.8541
	T-Trans	0.1778	0.0145	0.0141	20.9687	11.9931
	Bi-LSTM	0.1814	0.0165	0.0139	22.9957	14.3716
	CNN-	0.1805	0.0178	0.0146	21.5322	13.2225
	LSTM					
T+6	DAGCGN	0.1327	0.0071	0.0063	13.5273	7.2783
	GRU	0.1895	0.0144	0.0142	25.2917	15.1690
	LSTM	0.1910	0.0149	0.0140	24.9522	14.9485
	T-Trans	0.1902	0.0141	0.0143	24.7803	14.6876
	Bi-LSTM	0.1908	0.0164	0.0144	28.1421	17.3821
	CNN-	0.1938	0.0175	0.0148	25.9419	16.3861
	LSTM					
	DAGCGN	0.1258	0.0095	0.0076	15.0464	8.2399

depending on the future temporal horizon, i.e., T+1, T+3, T+6, and T+12.

Table 3 and **Table 4** shows the air quality forecasting performance at 18 known monitoring stations quantified by the average RMSE metric for each pollutant sensor at 1, 3, 6, and 12 h temporal horizons at training and test sets, respectively. The results demonstrate a moderate variation in the RMSE scores obtained by forecasting frameworks on the

training and test sets. For shorter temporal horizons of 1 and 3 h, the baseline models and DAGCGN showed performance differences of up to 12% and 8%, respectively. Furthermore, the baselines had a slightly higher RMSE difference of 13–15% for the 12 h ahead temporal horizon. Whereas the DAGCGN exhibited a modest performance change of 12%. Since there is a difference between the outdoor air quality conditions for the training and test sets, the moderate variation at the 12 h forecasting horizon is reasonable. The forecasting performance comparison between the training and test sets indicates well-trained models with low overfitting concerns.

According to results for the test set, based on shorter temporal horizons at 1 and 3 h, the encoder-decoder framework outperformed the stacked spatial models despite having a higher complexity. This may be related to the efficacy with which the hidden spatiotemporal representations of the research region are learned. Considering PM_{2.5} for one and three hour-ahead temporal horizons, the CNN-LSTM presented a performance improvement of 9.05–5.71%, 17.89%–9.06%, and 5.78%–2.12% compared to the GRU, LSTM, and T-Trans frameworks, respectively. The CNN-LSTM model utilizes convolutional feature maps to extract spatial interactions from multiple neighboring nodes, and when paired with the LSTM layer, the model may learn temporal dependencies. Therefore, CNN-LSTM shows a superior performance at higher temporal horizons in comparison to the Bi-LSTM framework.

The forecasting efficiency of the encoder-decoder framework decreases when the prediction horizon is large. Although the CNN-LSTM and Bi-LSTM frameworks allow simultaneous prediction at all known stations, the model loses temporal dependence for greater than 3 h ahead forecasting tasks. Among the baselines, the T-Trans outperformed the other baselines at 6 and 12 h prediction horizons. The T-Trans employs a multi-head attention mechanism that jointly examines input information from multiple subspace representations to extract more comprehensive information. Moreover, when combined with the GRU layers, in addition to short-term events, the model shows superior performance for long-term dependencies. Based on these findings, the spatial models outperformed baselines over longer prediction horizons. However, these models entail training, updating, and maintaining them at each site to provide an early estimation of pollution concentrations.

Table 5
MASE of forecasting models at different temporal horizons on the test set.

Time step	Model	CO (ppm)	O ₃ (ppm)	NO ₂ (ppm)	PM ₁₀ ($\mu\text{g}/\text{m}^3$)	PM _{2.5} ($\mu\text{g}/\text{m}^3$)
T+1	GRU	5.0435	1.3695	3.5794	3.8341	3.4855
	LSTM	4.4820	1.3723	3.2065	3.1617	3.2490
	T-Trans	3.9986	1.1695	2.6531	2.7263	2.6632
	Bi-LSTM	3.7938	1.1357	2.4361	2.7117	2.7791
	CNN-LSTM	3.7626	1.1691	2.5654	2.6805	2.7532
	DAGCGN	1.8332	1.3774	0.8317	1.2500	1.2228
T+3	GRU	6.2253	2.1765	4.8476	4.5790	3.9713
	LSTM	5.3417	2.0319	4.3588	3.7492	3.4497
	T-Trans	4.7977	1.8260	3.9077	3.3305	2.9639
	Bi-LSTM	5.5327	2.2901	4.0742	4.0855	4.1536
	CNN-LSTM	5.2899	2.0130	4.1601	3.3323	3.2427
	DAGCGN	2.3522	1.4959	1.0302	1.4221	1.3469
T+6	GRU	8.2195	3.4637	7.1043	6.2023	5.3309
	LSTM	6.9995	3.4775	6.7133	5.2155	4.7077
	T-Trans	6.6562	3.0547	6.4254	4.6901	4.0059
	Bi-LSTM	6.9620	3.4087	5.2787	5.9812	5.6788
	CNN-LSTM	7.1100	3.6676	6.4663	4.9816	4.8383
	DAGCGN	2.6811	1.5596	1.0104	1.6043	1.4826
T+12	GRU	8.5225	3.5832	7.3970	7.5829	6.5461
	LSTM	7.8678	3.6801	6.8246	6.9339	5.8296
	T-Trans	7.4327	3.2954	6.7071	6.1767	5.2733
	Bi-LSTM	7.2707	3.4354	5.2928	7.1403	6.7850
	CNN-LSTM	7.6742	4.0352	6.5207	6.6361	6.2113
	DAGCGN	2.5418	2.0800	1.2425	1.7320	1.6681

The proposed DAGCGN framework, composed of a node sampling technique, efficiently minimized noise from distant nodes and learned spatio-temporal patterns for a variety of air quality pollutants across a range of temporal horizons. For PM_{2.5} at a 12 h temporal horizon, the framework offered RMSE performance improvements of 45.67%, 44.87%, 43.89%, 52.59%, and 49.71% compared to GRU, LSTM, T-Trans, CNN-LSTM, and Bi-LSTM, respectively. The forecasting performance improvements for CO, O₃, NO₂, and PM₁₀ ranged between 33–35%, 32–45%, 45–48%, and 39–46%, respectively. A MASE-based performance evaluation is presented in Table 5. According to the results, a variation in the learning capability of baseline models was observed for different pollutant sensors. For a one-hour forecast, the MASE scores range from 1.13 up to 5.04 for different baselines showing unstable learning capacity. In contrast, DAGCGN detected the temporal dynamics of all air quality contaminants with modest fluctuations in MASE scores, resulting in a reasonably robust early warning system. Moreover, a comparable evaluation, similar to RMSE can be done by the comparison among baselines.

To visually evaluate the performance of the baseline and the proposed model, the ability to capture the dynamic behavior of primary pollutant concentrations was first compared. The temporal predictions were compared, for example, with monitoring station S12 at the 1 h ahead horizon, as presented in Fig. 10. According to the results, all forecasting frameworks intend to capture the temporal trend of PM_{2.5}. However, a certain deviation between the measured and estimated concentrations can be observed at specific time intervals based on the forecasts of the baseline models. This may have resulted from incorrect recognition of the spatiotemporal information. However, DAGCGN captured the dynamic trend in PM_{2.5} and exhibited considerable stability. Furthermore, Fig. S2 presents the comparison of temporal predictions at monitoring station-S4, S7, 16, and 18 situated across the study region.

For an unbiased evaluation, the MAE scores of the neural network frameworks are compared at each sensor of the known monitoring stations across a range of temporal horizons. As presented in Fig. 11, the left axis depicts the MAE scores, whereas the right axis shows the mean pollutant concentration at the node location using a bar plot. According to Fig. 11 (a), the MAE is considerably low in comparison to the mean pollutant concentration at each node. For example, for PM₁₀ and PM_{2.5}, the mean concentrations ranged between 38.13–52.9 $\mu\text{g}/\text{m}^3$ and 20.93–31.72 $\mu\text{g}/\text{m}^3$. The MAE scores of the baselines considering PM₁₀ and PM_{2.5} ranged between 9–15 $\mu\text{g}/\text{m}^3$ and 5–10 $\mu\text{g}/\text{m}^3$ respectively, indicating that the forecasting models were properly trained. Moreover, the performance of the baseline changed with the concentration profiles at each location.

In comparison to the baselines, the DAGCGN model showed a steady performance with minimal node-specific fluctuations at each monitoring location for most of the air quality pollutants. However, certain fluctuations were observed for the CO sensor at monitoring locations S9 and S18. The improved performance of DAGCGN can be attributed to node-specific learning, where the model adopts node sampling to avoid noise coming from distant stations and reduce prediction variance. A significant performance boost from the neighborhood information provided by the distance adaptive graph generation can be observed specifically at nodes 12, 13, 17, and 18, where the baselines showed the lowest performance for different primary pollutants.

Finally, for an intuitive and simultaneous comparison of forecasting methods across multiple temporal horizons, the collective probability density distribution of the predictive MAEs was investigated. Fig. 12 illustrates the probability density curves of predictive MAE scores for the baseline and the suggested method. The vertical axis represents the probability of the overall predictive MAE from the 1 to 12 h temporal horizons. The MAE probability distributions for CO and NO₂ for all baseline methods have similar curves and do not present a clear advantage of one method over the other. According to the remainder of the density distribution results, the bulk of the GRU and T-Trans

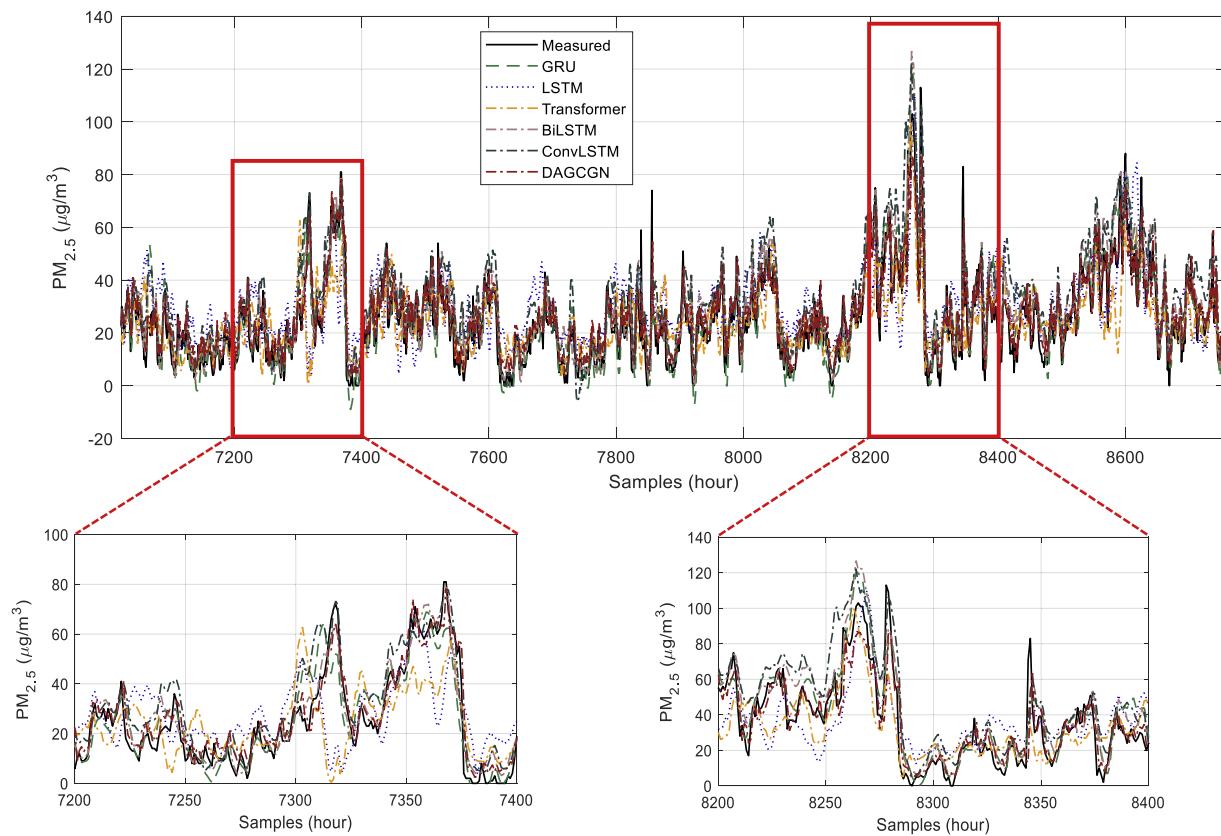


Fig. 10. Comparison of forecasted $\text{PM}_{2.5}$ concentrations at S12 monitoring station.

prediction errors were located on the left side in comparison to the LSTM, indicating smaller errors across various time spans.

Additionally, T-Trans appears to have a higher performance than the encoder-decoder frameworks represented by a narrow curve with a high probability of smaller errors. Bi-LSTM and CNN-LSTM have a lower probability of smaller errors and a relatively high dispersion. The larger errors generated by the encoder-decoder frameworks originate mainly at higher temporal horizons, where the prediction span increases, as demonstrated in Table 4. The low errors caused by DAGCGN across multiple temporal horizons shift the probability density distribution to the left with a relatively high probability. The MAE distribution for CO showed a comparable probability ratio with the baselines but with a lower error value. For the remainder of the pollutants, the results indicate that DAGCGN can generate smaller forecasting errors across multiple temporal horizons with a narrow curve, showing stable performance and high forecasting precision.

Finally, in order to evaluate the computational cost of neural network frameworks, the model parameters and training times of the baselines and DAGCGN are compared in Table S3 of the SI. For the category of stacked spatial models, a unique multivariate model is constructed for each station; as a result, the training time is calculated for each station and then summed to obtain the total training time per epoch. According to results, the encoder-decoder CNN-LSTM and Bi-LSTM frameworks have the lowest training times, between 12s and 14s per epoch, respectively. However, these frameworks also generate the highest forecasting errors. Additionally, as a trade-off for learning node-specific patterns, the DAGCGN contains twice as many model parameters as the baseline T-Trans model. Compared to T-Trans, which is the best performing baseline model, DAGCGN's training time is moderately increased by 17s. Nevertheless, the model demonstrates a considerable improvement in forecasting ability, with RMSE reductions of up to 52%. Based on an extensive performance evaluation and computational cost analysis, the results indicate that, in comparison to

state-of-the-art air quality early warning models, the proposed DAGCGN framework provides high forecasting performance at short and long time-intervals.

4.2. Applications of DAGCGN for early warning systems

Conventional early warning frameworks offer pollutant forecasts across a range of time periods based on historical sensor data. However, in real-world circumstances, sensor reliability is significantly compromised by a variety of variables, such as weather conditions, maintenance, and hostile monitoring environments. Therefore, the practicality of traditional frameworks is severely compromised because they are applicable only when all sensor data are in normal condition. Moreover, forecasts using conventional frameworks are only available at locations where sufficient monitoring equipment is available.

This section describes the application of graph-based frameworks in two scenarios: a long-term sensor failure and remote early health risk warnings at an unknown or new location. As described in Section 3.5, monitoring stations S7, S11, and S17 were selected for Scenario 1. At each station, the sensor data were purposely erased in three intervals: two partial failures, when two to three sensors fail concurrently for two weeks, and one week of complete failure, which reflects an entire sensor data failure for the monitoring station. If a sensor showed NaN values, the proposed graph-based framework considers it a sensor failure and stops receiving input from the identified node for local and spatial positions. Moreover, it generates a distance-adaptive grid for the identified station and forecasts the pollutant concentration until the failure is resolved.

The long-term sensor data imputation performance was first evaluated based on the discrepancy between actual and forecasted values. Table 6 lists the overall MAE scores obtained by partial and complete sensor failures at monitoring stations S7, S11, and S17. The time-step depicts the forecast generated by the DAGCGN utilizing neighboring

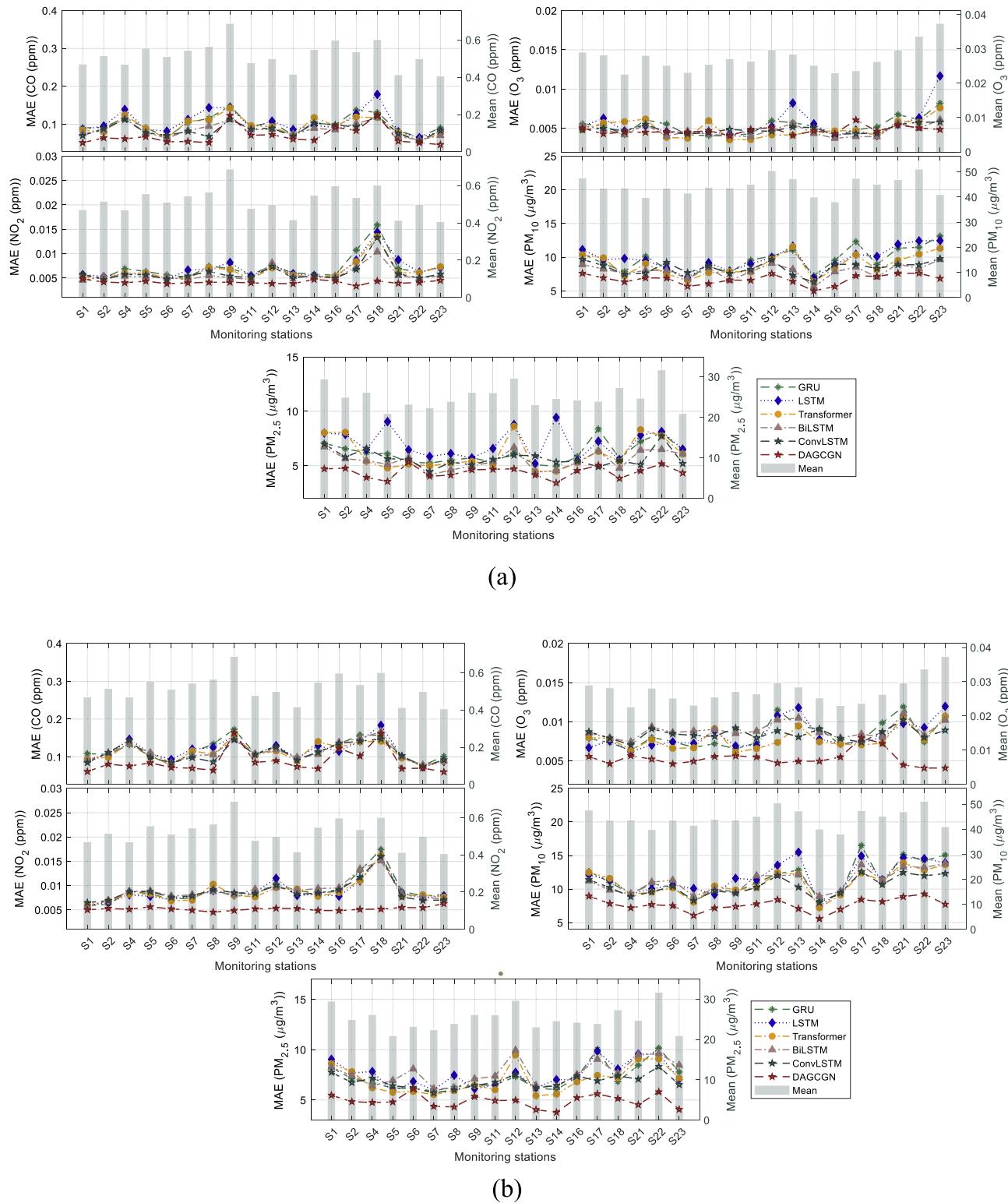


Fig. 11. Distribution of average primary air pollutant concentrations and the predictive MAE scores at each known station in study area under temporal horizons, (a) 1-hr, (b) 3-hr, (c) 6-hr, and (d) 12-hr.

nodes at various prediction spans. According to the results, the generated forecast for long-term sensor failure without the local node information shows a close similarity, as validated by the low MAE scores. The MAE values for CO, NO₂, O₃, PM₁₀, and PM_{2.5} ranged between 0.081-0.160, 0.004-0.007, 0.005-0.0155, 9.41-13.5, and 5.75-8.12,

respectively. The MAE for the pollutants differed across the stations, which was mainly associated with the spatiotemporal relations with the neighboring locations and the mean pollutant concentrations during the failure intervals. For instance, S11 showed the lowest concentration of PM_{2.5} (23.4 µg/m³) and obtained the lowest average MAE of 5.91 µg/m³

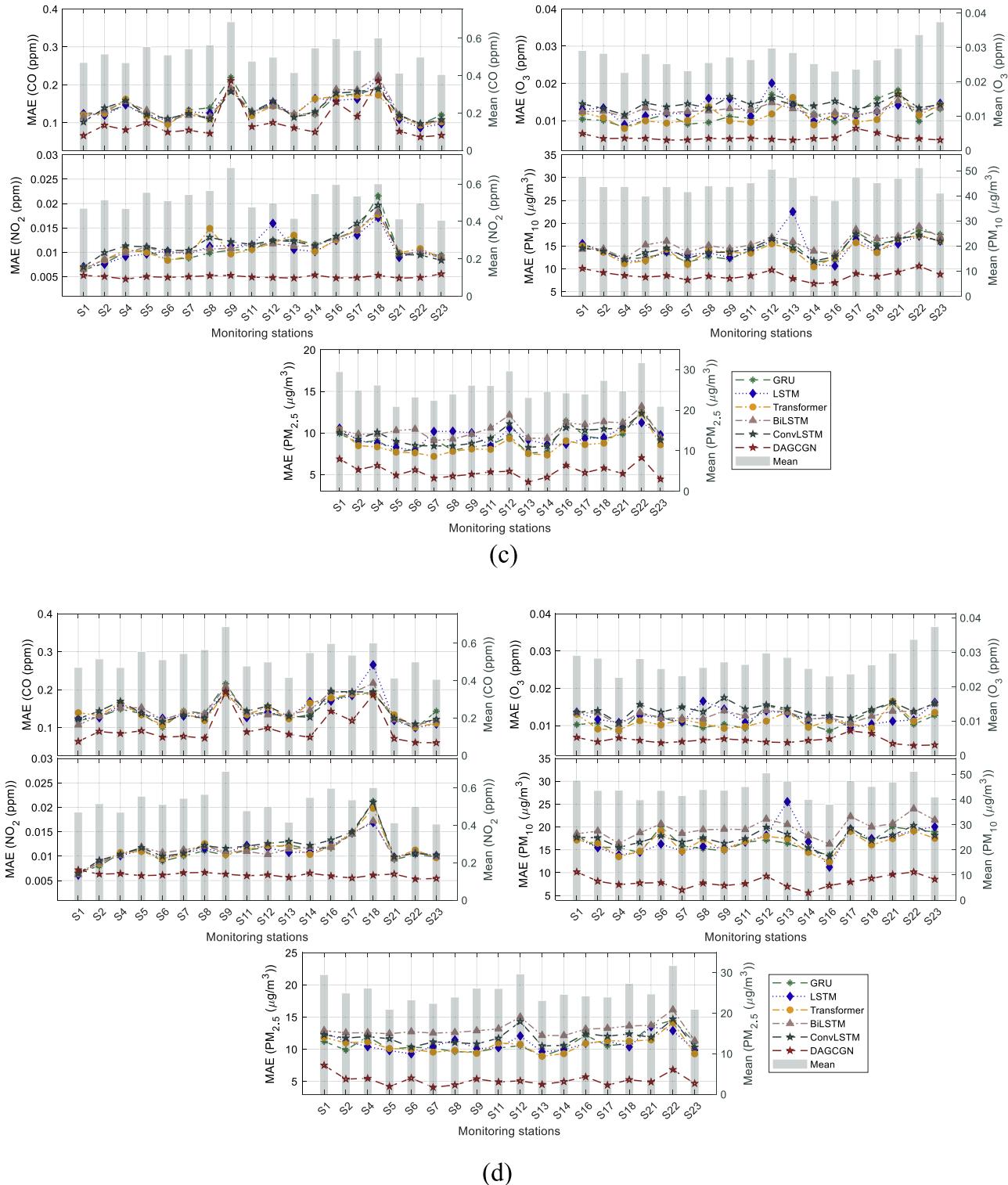


Fig. 11. (continued).

m^3 . Additionally, a low MAE score in comparison to the mean pollutant concentration reflects a high imputation performance.

Further evaluation by a visual comparison of original measurements treated as missing values and the 6h ahead forecasted concentrations at S-7 are presented in Fig. 13. A visual assessment demonstrates that the DAGCGN identified the majority of the dynamic trends of primary air pollutants for the long-term sensor failures. Whereas a few deviations were observed for PM₁₀ and PM_{2.5}, these discrepancies are associated

with the highly nonlinear behavior of PM in complex urban environments and originate from rare events. The investigation demonstrates the overall efficacy of the proposed methodology in resolving real-world sensor failure problems for real-time monitoring. These imputation techniques are important for generating accurate health risk warnings originating from elevated concentrations of various outdoor pollutants. Additionally, the results for stations S-11 and S-17 are presented in Fig. S3 and Fig. S4, respectively.

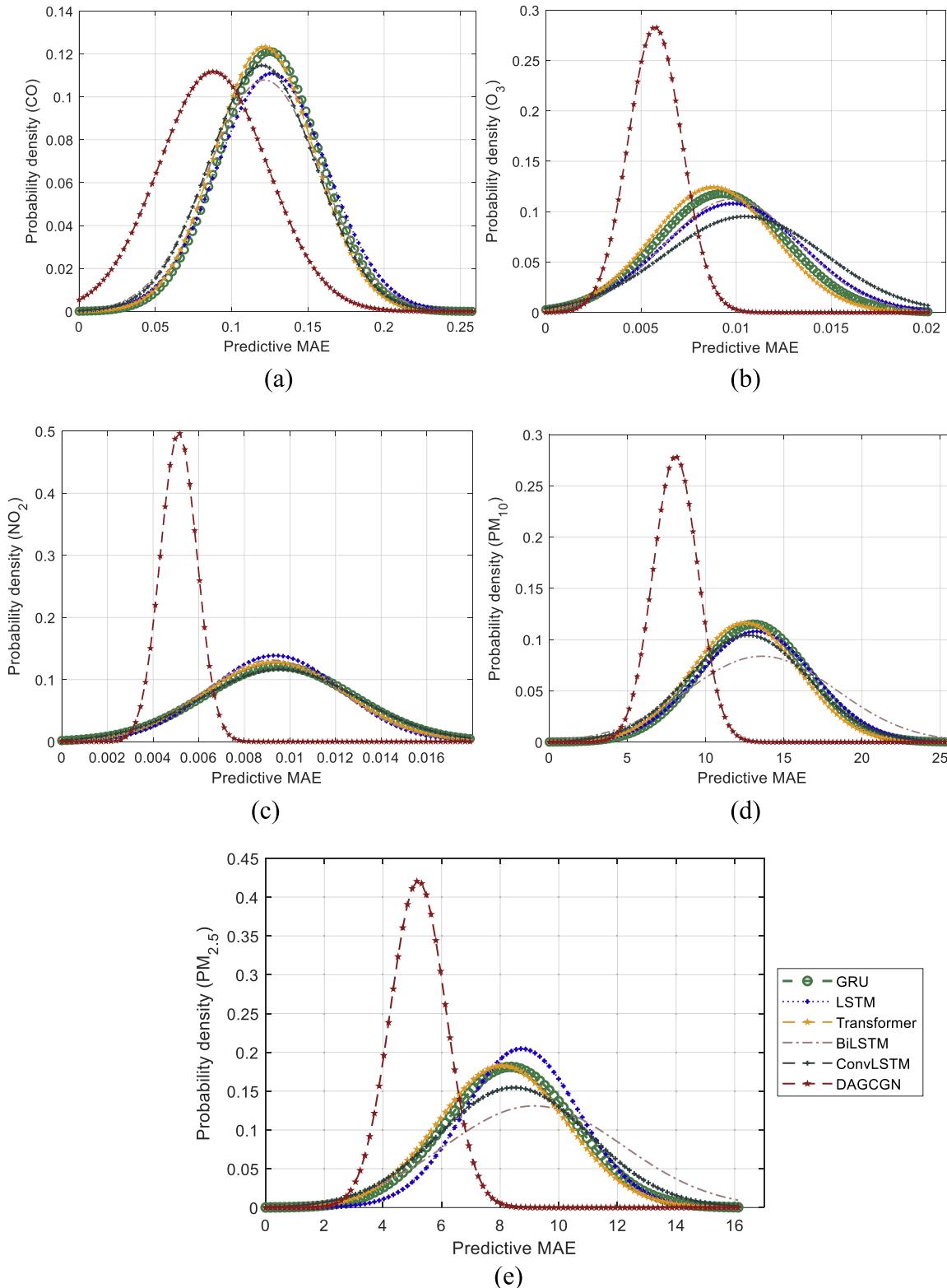


Fig. 12. Predictive MAE comparison at multiple temporal horizons across the study region (a) CO, (b) O₃, (c) NO₂, (d) PM₁₀, and (e) PM_{2.5}

Early warning systems are critical for notifying people in advance and initiating regulatory actions to prevent pollution. However, cities often have a limited number of air quality monitoring stations that are irregularly placed and do not cover the entire area (Huang et al., 2021). Additionally, there is a disparity in the monitoring capabilities of developed and undeveloped cities, thus restricting the ability of the general population to receive accurate health risk warnings. To resolve

this real-world problem at new or unknown locations, the remote sensing and prediction performance of the DAGCGN framework at five unknown stations was investigated. The monitoring data from these stations was not utilized in any part of the training procedure. The DAGCGN utilizes the already learned hidden representation by the dynamic graph, and prediction is performed by assigning the distance to the unknown locations with the air quality information available at

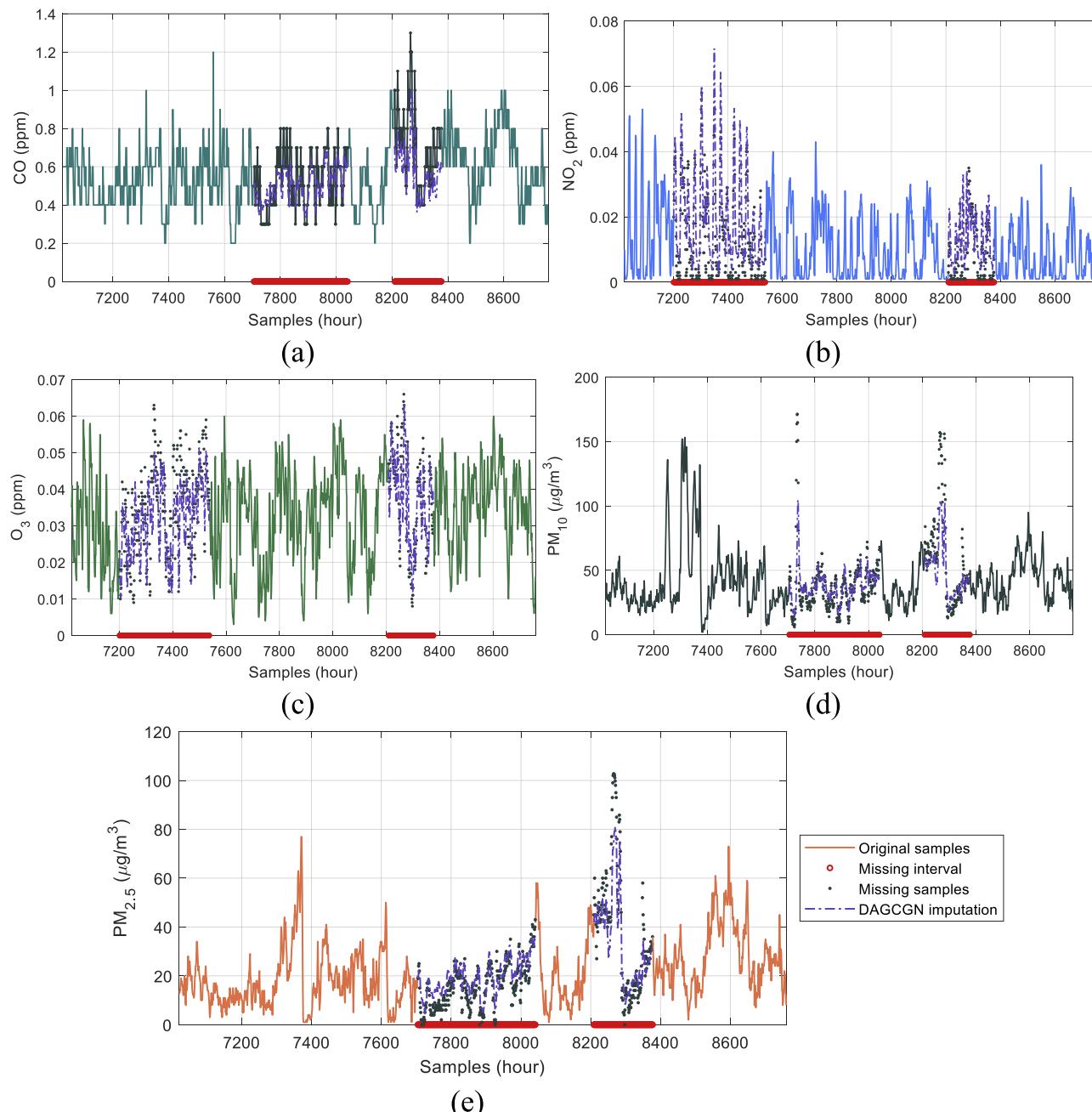
Table 6

MAE based imputation performance for partial and complete sensor failure.

Time step	Station	CO	O ₃	NO ₂	PM ₁₀	PM _{2.5}
T+1	S7	0.0811	0.0064	0.0063	9.53	5.75
	S11	0.1146	0.0041	0.0054	11.16	5.84
	S17	0.1476	0.006	0.0114	11.70	6.68
T+3	S7	0.0867	0.0045	0.0066	9.41	5.84
	S11	0.1198	0.0056	0.0064	12.29	5.93
	S17	0.1373	0.0055	0.0135	11.12	7.20
T+6	S7	0.095	0.0065	0.0062	10.50	5.93
	S11	0.1107	0.0052	0.0067	12.73	6.06
	S17	0.1609	0.0059	0.0129	13.18	7.66
T+12	S7	0.0892	0.0075	0.0078	10.42	6.11
	S11	0.1248	0.0059	0.0091	13.47	6.33
	S17	0.1515	0.0067	0.0156	13.19	8.12

known locations.

To demonstrate the effect of a sensor devoid unknown locations on health risk distribution, a comparison was made in terms of the overall CAI surface distribution, as presented in Fig. 14. The air quality distributions reveal that the southern section of the study region is predisposed to have a moderate to unhealthy CAI, resulting in high PM₁₀ and PM_{2.5} concentrations. Fig. 14 (a) shows the CAI distribution in the study region where the data at all monitoring stations are available and for when the monitoring data are not available at unknown locations, specifically, S3, S10, S15, S19, and S20. Visual analysis confirms that sensor unavailability due to a lack of monitoring can lead to an incorrect depiction of actual health risks at outdoor locations. Differences in the CAI surface distribution can be observed specifically at monitoring locations S3, S10, S15, 19, and S20. Sensor unavailability showed elevated CAI values at stations S3 and S10, whereas the actual health risk was in the moderate range. Additionally, a contrary effect can be observed at

**Fig. 13.** Imputation performance of DAGCGN for partial and complete sensor failure at S-7 station, (a) CO, (b) O₃, (c) NO₂, (d) PM₁₀, and (e) PM_{2.5}.

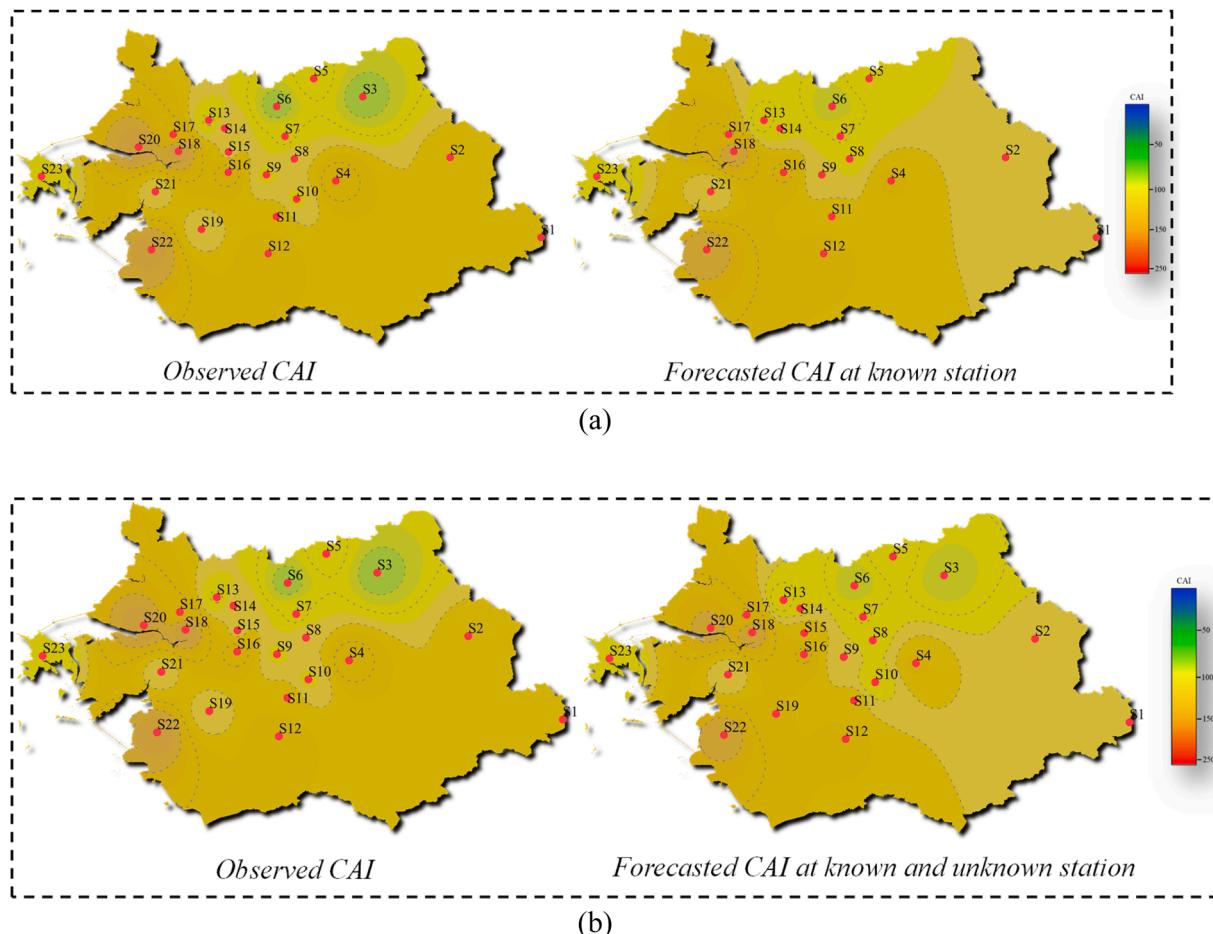


Fig. 14. Comparison of air quality associated health risk distribution in study area (24-12-2019 to 28-12-2019), (a) observed CAI and forecasted CAI at known stations, (b) observed CAI and forecasted CAI at known and unknown stations.

the S15 and S20 monitoring locations, where an improper estimation can result in an underestimation of unhealthy air quality conditions.

The proposed DAGCGN framework facilitated remote sensing and future forecasts of air quality pollutants at unknown, user-defined locations. Fig. 14 (b) shows the CAI distribution of the study area at a 3 h ahead temporal horizon that provides an early health risk warning. In comparison to the observed concentrations, the CAI distribution shows that the DAGCGN framework can adequately describe a spatiotemporal relationship in the study region and forecast at a new location. The CAI distribution is validated by observing a close similarity between the actual and forecasted CAI values, primarily at S3, S10, S15, and S20 monitoring stations. Notably, the performance of graph-based frameworks is sensitive to the spatial reliance of nearby nodes, where a large distance between the monitoring stations may increase the mismatch between the observed and predicted values. Based on extensive evaluation and visual comparisons, the DAGCGN simultaneously outperforms conventional early warning systems in terms of pollutant forecasting capability, allows for data reconstruction in the event of long-term sensor failure, and facilitates remote health risk estimations at an unknown location. The proposed DAGCGN framework can potentially fill up the gap in urban air quality monitoring and contribute to sustainable and healthy development of cities. Furthermore, to improve the effectiveness of DAGCGN, the present framework can be expanded to entire provinces for accurate air quality associated health risk estimations, which is particularly beneficial for regions with inadequate monitoring capacity.

5. Conclusions

In this work, a graph-based spatio-temporal air quality forecasting framework employing sensor fusion is proposed for early health risk warning of primary pollutants in an urban environment. The primary contribution of this study is the design of a distance adaptive graph convolutional gated network (DAGCGN) capable of identifying complicated spatio-temporal interactions between neighboring monitoring points to provide satisfactory prediction at multiple temporal horizons. Additionally, we use a distance adaptive grid to address challenges associated with traditional early warning frameworks, such as long-term sensor failure and remote sensing with health risk warning at a new site. The performance of the proposed model is compared with several spatio-temporal and encoder-decoder based air quality forecasting frameworks. The rest of the findings of our study are given below:

- DAGCGN shows robust air quality forecasting performance in comparison to conventional baselines. For $\text{PM}_{2.5}$ at a 12 h temporal horizon, the proposed model outperforms GRU, LSTM, T-Trans, CNN-LSTM, and Bi-LSTM with an RMSE reduction of 45.67%, 44.87%, 43.89%, 52.59%, and 49.71%, respectively.
- The probability density of prediction errors reveals that DAGCGN generates smaller forecasting errors across multiple temporal horizons, indicating steady performance and a high forecasting accuracy.
- DAGCGN demonstrates effective imputation for long-term sensor failure, with MAE scores for PM_{10} and $\text{PM}_{2.5}$ ranging between

- 9.41–13.5 and 5.75–8.12, respectively, five times lower than average pollutant concentrations.
- Remote sensing and air quality forecasting at an unknown location show close similarity with the health risk distribution shown by the air quality index.

The proposed methodology demonstrates the efficacy of distance adaptive graph-based methods for early warning systems and enables the assessment of air quality-related health risks in areas without adequate monitoring. Despite the fact that DAGCGN solves some of the critical problems of traditional forecasting frameworks, certain factors can restrict its prediction performance. Since DAGCGN generates predictions based on the spatial and temporal reliance of nearby nodes, a considerable distance between the monitoring stations can increase the discrepancy between the observed and predicted values. In addition, neighboring nodes may have different monitoring time scales, resulting in low-quality inputs that might reduce the accuracy of the model's predictions. This presents a new obstacle to the application of graph-based networks for air quality prediction. We consider this challenge to be the future direction of the research and application of graph neural networks for air quality warning systems in urban areas.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (No. 2021R1A2C2007838), from the Subway Fine Dust Reduction Technology Project of the Ministry of Land Infrastructure and Transport from the Republic of Korea (21QPPW-B152306-03), Institute for Information & Communication Technology Planning & Evaluation (IITP) grants funded by the Korean MSIT: (No. 2022-0-00688, and AI Platform to Fully Adapt and Reflect Privacy-Policy Changes), and (No. 2021-0-02068, Artificial Intelligence Innovation Hub).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.scs.2023.104445](https://doi.org/10.1016/j.scs.2023.104445).

References

- Bai, L., Yao, L., Li, C., Wang, X., & Wang, C. (2020). Adaptive graph convolutional recurrent network for traffic forecasting. *Advances in Neural Information Processing Systems*.
- Bai, Y., Zeng, B., Li, C., & Zhang, J. (2019). An ensemble long short-term memory neural network for hourly PM2.5 concentration forecasting. *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2019.01.121>
- Benchrif, A., Wheida, A., Tahri, M., Shubbar, R. M., & Biswas, B. (2021). Air quality during three covid-19 lockdown phases: AQI, PM2.5 and NO2 assessment in cities with more than 1 million inhabitants,. *Sustainable Cities and Society*. <https://doi.org/10.1016/j.scs.2021.103170>
- Chae, Y. H., Lee, C., Han, S. M., & Seong, P. H. (2022). Graph neural network based multiple accident diagnosis in nuclear power plants: Data optimization to represent the system configuration. *Nuclear Engineering and Technology*. <https://doi.org/10.1016/j.net.2022.02.024>
- Chen, G., Chen, J., hui Dong, G., yi Yang, B., Liu, Y., Lu, T., Yu, P., Guo, Y., & Li, S. (2021). Improving satellite-based estimation of surface ozone across China during 2008–2019 using iterative random forest model and high-resolution grid meteorological data,. *Sustainable Cities and Society*. <https://doi.org/10.1016/j.scs.2021.102807>
- Chen, Y., Ding, F., & Zhai, L. (2022). Multi-scale temporal features extraction based graph convolutional network with attention for multivariate time series prediction. *Expert Syst. Appl.*, 200, Article 117011. <https://doi.org/10.1016/j.eswa.2022.117011>
- Chiang, W. L., Li, Y., Liu, X., Bengio, S., Si, S., & Hsieh, C. J. (2019). Cluster-GCN: An efficient algorithm for training deep and large graph convolutional networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. <https://doi.org/10.1145/3292500.330925>
- Choudhury, A., Middya, A. I., & Roy, S. (2022). Attention enhanced hybrid model for spatiotemporal short-term forecasting of particulate matter concentrations. *Sustainable Cities and Society*, 86, Article 104112. <https://doi.org/10.1016/j.scs.2022.104112>
- Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in Neural Information Processing Systems*.
- Djalalova, I., Delle Monache, L., & Wilczak, J. (2015). PM2.5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. *Atmospheric Environment*. <https://doi.org/10.1016/j.atmosenv.2015.02.021>
- European Commission. (2016). DIRECTIVE (EU) 2016/2284 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the reduction of national emissions of certain atmospheric pollutants, amending Directive 2003/35/EC and repealing Directive 2001/81/EC. *Official Journal of the European Union*.
- Faganelli Pucer, J., & Štrumbelj, E. (2018). Impact of changes in climate on air pollution in Slovenia between 2002 and 2017. *Environmental Pollution*. <https://doi.org/10.1016/j.envpol.2018.06.084>
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., & Wang, J. (2015). Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. *Atmospheric Environment*. <https://doi.org/10.1016/j.atmosenv.2015.02.030>
- Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G., Skamarock, W. C., & Eder, B. (2005). Fully coupled “online” chemistry within the WRF model. *Atmos. Environ.* <https://doi.org/10.1016/j.atmosenv.2005.04.027>
- Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In *33rd AAAI Conference on Artificial Intelligence AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*. <https://doi.org/10.1609/aaai.v33i01.3301922>
- Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*.
- Han, B. S., Park, K., Kwak, K. H., Park, S. B., Jin, H. G., Moon, S., Kim, J. W., & Baik, J. J. (2020). Air quality change in Seoul, South Korea under covid-19 social distancing: Focusing on pm2.5. *International Journal of Environmental Research and Public Health*. <https://doi.org/10.3390/ijerph17176208>
- Holzinger, A., Malle, B., Sarantí, A., & Pfeifer, B. (2021). Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI. *Information Fusion*. <https://doi.org/10.1016/j.inffus.2021.01.008>
- Huang, W., Li, T., Liu, J., Xie, P., Du, S., & Teng, F. (2021). An overview of air quality analysis by big data techniques: Monitoring, forecasting, and traceability. *Information Fusion*. <https://doi.org/10.1016/j.inffus.2021.03.010>
- IQAir. (2020). *World air quality report*.
- Janarthanan, R., Partheeban, P., Somasundaram, K., & Elamparithi, P. Navin (2021). A deep learning approach for prediction of air quality index in a metropolitan city. *Sustainable Cities and Society*. <https://doi.org/10.1016/j.scs.2021.102720>
- Ju, M. J., Oh, J., & Choi, Y. H. (2021). Changes in air pollution levels after COVID-19 outbreak in Korea. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2020.141521>
- Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *5th Int. Conf. Learn. Represent. ICLR 2017 - Conf. Track Proc.*
- Korea Environment Corporation. (2021). *Air Korea*, (n.d.). <https://www.airkorea.or.kr/eng/> accessed.
- Li, C., & Zhu, Z. (2018). Research and application of a novel hybrid air quality early-warning system: A case study in China. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2018.01.195>
- Li, H., Wang, J., Li, R., & Lu, H. (2019). Novel analysis-forecast system based on multi-objective optimization for air quality index. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2018.10.129>
- Li, H., Wu, J., Wang, A., Li, X., Chen, S., Wang, T., Amsalu, E., Gao, Q., Luo, Y., Yang, X., Wang, W., Guo, J., Guo, Y., & Guo, X. (2018). Effects of ambient carbon monoxide on daily hospitalizations for cardiovascular disease: A time-stratified case-crossover study of 460,938 cases in Beijing, China from 2013 to 2017. *Environmental Health: A Global Access Science Source*. <https://doi.org/10.1186/s12940-018-0429-3>
- Li, Y., Liu, L., Wang, G., Du, Y., & Chen, P. (2022). EGNN: Constructing explainable graph neural networks via knowledge distillation. *Knowledge-Based System*, 241, Article 108345. <https://doi.org/10.1016/j.knosys.2022.108345>
- Liu, H., Yan, G., Duan, Z., & Chen, C. (2021). Intelligent modeling strategies for forecasting air quality time series: A review. *Applied Soft Computing*, 102. <https://doi.org/10.1016/j.asoc.2020.106957>
- Liu, X., Qin, M., He, Y., Mi, X., & Yu, C. (2021). A new multi-data-driven spatiotemporal PM2.5 forecasting model based on an ensemble graph reinforcement learning convolutional network. *Atmospheric Pollution Research*, 12, Article 101197. <https://doi.org/10.1016/j.apr.2021.101197>
- Loy-Benitez, J., Heo, S., & Yoo, C. (2020). Soft sensor validation for monitoring and resilient control of sequential subway indoor air quality through memory-gated

- recurrent neural networks-based autoencoders. *Control Engineering Practice*, 97. <https://doi.org/10.1016/J.CONENGPRAC.2020.104330>
- Loy-Benitez, J., Li, Q., Nam, K. J., & Yoo, C. K. (2020). Sustainable subway indoor air quality monitoring and fault-tolerant ventilation control using a sparse autoencoder-driven sensor self-validation. *Sustainable Cities and Society*, 52, Article 101847. <https://doi.org/10.1016/j.scs.2019.101847>
- Ma, J., Cheng, J. C. P., Lin, C., Tan, Y., & Zhang, J. (2019). Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. *Atmospheric Environment*, 214, Article 116885. <https://doi.org/10.1016/j.atmosenv.2019.116885>
- Nam, K. J., Heo, S. K., Li, Q., Loy-Benitez, J., Kim, M. J., Park, D. S., & Yoo, C. K. (2020). A proactive energy-efficient optimal ventilation system using artificial intelligent techniques under outdoor air quality conditions. *Applied Energy*, 266, Article 114893. <https://doi.org/10.1016/j.apenergy.2020.114893>
- Norwawi, N. M. (2021). *Sliding window time series forecasting with multilayer perceptron and multiregression of COVID-19 outbreak in Malaysia*. Elsevier Inc.. <https://doi.org/10.1016/B978-0-12-824536-1.00025-3>
- Qi, Y., Li, Q., Karimian, H., & Liu, D. (2019). A hybrid model for spatiotemporal forecasting of PM 2.5 based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, 664, 1–10. <https://doi.org/10.1016/j.scitotenv.2019.01.333>
- Sandryhaila, A., & Moura, J. M. F. (2013). Discrete signal processing on graphs. *IEEE Transactions on Signal Processing*. <https://doi.org/10.1109/TSP.2013.2238935>
- Seng, D., Zhang, Q., Zhang, X., Chen, G., & Chen, X. (2021). Spatiotemporal prediction of air quality based on LSTM neural network. *Alexandria Engineering Journal*. <https://doi.org/10.1016/j.aej.2020.12.009>
- Shaw, C., Boulle, M., Longley, I., Mitchell, T., Pierse, N., & Howden-Chapman, P. (2020). The association between indoor and outdoor NO₂ levels: A case study in 50 residences in an urban neighbourhood in New Zealand. *Sustainable Cities and Society*. <https://doi.org/10.1016/j.scs.2020.102093>
- Shi, K., & Wu, L. (2020). Forecasting air quality considering the socio-economic development in Xingtai. *Sustainable Cities and Society*. <https://doi.org/10.1016/j.scs.2020.102337>
- Song, C., Lin, Y., Guo, S., & Wan, H. (2020). Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting. In *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*. <https://doi.org/10.1609/aaai.v34i01.5438>
- Su, H., Zhang, T., Lin, M., Lu, W., & Yan, X. H. (2021). Predicting subsurface thermohaline structure from remote sensing data based on long short-term memory neural networks. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2021.112465>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- Veličković, P., Casanova, A., Liò, P., Cucurull, G., Romero, A., & Bengio, Y. (2018). Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings* (pp. 1–12).
- Vlachogianni, A., Kassomenos, P., Karppinen, A., Karakitsios, S., & Kukkonen, J. (2011). Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM10 in Athens and Helsinki. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2010.12.040>
- Vos, Theo, Lim, S. S., & Abbafati, Cristiana (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)
- Wang, H. W., Peng, Z. R., Wang, D., Meng, Y., Wu, T., Sun, W., & Lu, Q. C. (2020). Evaluation and prediction of transportation resilience under extreme weather events: A diffusion graph convolutional approach. *Transportation Research Part C: Emerging Technologies*. <https://doi.org/10.1016/j.trc.2020.102619>
- Wang, J., Li, H., & Lu, H. (2018). Application of a novel early warning system based on fuzzy time series in urban air quality forecasting in China. *Applied Soft Computing*. <https://doi.org/10.1016/j.asoc.2018.07.030>
- Wang, J., Li, H., Yang, H., & Wang, Y. (2021). Intelligent multivariable air-quality forecasting system based on feature selection and modified evolving interval type-2 quantum fuzzy neural network. *Environmental Pollution*. <https://doi.org/10.1016/j.envpol.2021.116429>
- Wang, L., Bi, J., Meng, X., Geng, G., Huang, K., Li, J., Tang, L., & Liu, Y. (2020). Satellite-based assessment of the long-term efficacy of PM2.5 pollution control policies across the Taiwan Strait. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2020.112067>
- Wu, Q., & Lin, H. (2019). Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network. *Sustainable Cities and Society*. <https://doi.org/10.1016/j.scs.2019.101657>
- Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*. <https://doi.org/10.1109/TNNLS.2020.2978386>
- Yan, R., Liao, J., Yang, J., Sun, W., Nong, M., & Li, F. (2021). Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2020.114513>
- Yang, N., Shi, H., Tang, H., & Yang, X. (2022). Geographical and temporal encoding for improving the estimation of PM2.5 concentrations in China using end-to-end gradient boosting. *Remote Sensing of Environment*. <https://doi.org/10.1016/j.rse.2021.112828>
- Yao, D., Zhi-li, Z., Xiao-feng, Z., Wei, C., Fang, H., Yao-ming, C., & Cai, W. W. (2022). Deep hybrid: Multi-graph neural network collaboration for hyperspectral image classification. *Defence Technology*. <https://doi.org/10.1016/j.dt.2022.02.007>
- Yin, S., Li, T., Cheng, X., & Wu, J. (2022). Remote sensing estimation of surface PM 2.5 concentrations using a deep learning model improved by data augmentation and a particle size constraint. *Atmospheric Environment*, 287, Article 119282. <https://doi.org/10.1016/j.atmosenv.2022.119282>
- Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L., & Leskovec, J. (2018). Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. <https://doi.org/10.1145/3219819.3219890>
- Yu, B., Lee, Y., & Sohn, K. (2020). Forecasting road traffic speeds by considering area-wide spatio-temporal dependencies based on a graph convolutional neural network (GCN). *Transportation Research Part C: Emerging Technologies*. <https://doi.org/10.1016/j.trc.2020.02.013>
- Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *International Joint Conferences on Artificial Intelligence*. <https://doi.org/10.24963/ijcai.2018.505>
- Yun, G., & Zhao, S. (2022). The imprint of urbanization on PM2.5 concentrations in China: The urban-rural gradient study. *Sustainable Cities and Society*, 86, Article 104103. <https://doi.org/10.1016/j.scs.2022.104103>
- Zanfei, A., Menapace, A., Brentan, B. M., Righetti, M., & Herrera, M. (2022). Novel approach for burst detection in water distribution systems based on graph neural networks. *Sustainable Cities and Society*, 86, Article 104090. <https://doi.org/10.1016/j.scs.2022.104090>
- Zhang, B., Zhang, H., Zhao, G., & Lian, J. (2020). Constructing a PM2.5 concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. *Environmental Modelling & Software*. <https://doi.org/10.1016/j.envsoft.2019.104600>
- Zhang, K., Thé, J., Xie, G., & Yu, H. (2020). Multi-step ahead forecasting of regional air quality using spatial-temporal deep neural networks: A case study of Huaihai Economic Zone. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2020.123231>
- Zhao, J., Deng, F., Cai, Y., & Chen, J. (2019). Long short-term memory - Fully connected (LSTM-FC) neural network for PM2.5 concentration prediction. *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2018.12.128>
- Zhou, F., Wang, T., Zhong, T., & Trajcevski, G. (2022). Identifying user geolocation with Hierarchical Graph Neural Networks and explainable fusion. *Information Fusion*, 81, 1–13. <https://doi.org/10.1016/j.inffus.2021.11.004>
- Zhou, H., Zhang, F., Du, Z., & Liu, R. (2021). Forecasting PM2.5 using hybrid graph convolution-based model considering dynamic wind-field to offer the benefit of spatial interpretability. *Environmental Pollution*. <https://doi.org/10.1016/j.envpol.2021.116473>
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open..* <https://doi.org/10.1016/j.aiopen.2021.01.001>