Information Fusion 113 (2025) 102607

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

MGSFformer: A Multi-Granularity Spatiotemporal Fusion Transformer for air quality prediction

Chengqing Yu [a,b], Fei Wang [a,b,*], Yilun Wang [c], Zezhi Shao [a,b], Tao Sun [a], Di Yao [a], Yongjun Xu [a,b,*]

[a] Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China
[b] University of Chinese Academy of Sciences, Beijing, 100049, China
[c] Institute of Navigation and Control Technology, China North Industries Group Corporation, Beijing, China

ARTICLE INFO

Keywords:
Air quality prediction
Multi-Granularity Spatiotemporal Fusion Transformer
Spatiotemporal correlation
Multi-source information fusion

ABSTRACT

Air quality spatiotemporal prediction can provide technical support for environmental governance and sustainable city development. As a classic multi-source spatiotemporal data, effective multi-source information fusion is key to achieving accurate air quality predictions. However, due to not fully fusing two pieces of information, classical deep learning models struggle to achieve satisfactory prediction results: (1) Multi-granularity: each air monitoring station collects air quality data at different sampling intervals, which show distinct time series patterns. (2) Spatiotemporal correlation: due to human activities and atmospheric diffusion, there exist correlations between air quality data from different air monitoring stations, necessitating the consideration of other air monitoring stations' influences when modeling each air quality time series. In this study, to achieve satisfactory prediction results, we propose the Multi-Granularity Spatiotemporal Fusion Transformer, comprised of the residual de-redundant block, spatiotemporal attention block, and dynamic fusion block. Specifically, the residual de-redundant block eliminates information redundancy between data with different granularities and prevents the model from being misled by redundant information. The spatiotemporal attention block captures the spatiotemporal correlation of air quality data and facilitates prediction modeling. The dynamic fusion block evaluates the importance of data with different granularities and integrates the prediction results. Experimental results demonstrate that the proposed model surpasses 11 baselines by 5% in performance on three real-world datasets.

1. Introduction

With the emission of industrial exhaust gasses, the concentration of pollutants in the atmosphere gradually increases, leading to respiratory-related diseases and hindering the development and construction of sustainable cities [1]. As one of the important means for sustainable cities development, accurate air quality monitoring and prediction technologies can provide important decision guidance for urban environmental governance [2]. Currently, countries around the world mainly use air monitoring stations to monitor air quality, aiming to analyze data in a timely manner and take governance measures. As a classic form of multi-source spatiotemporal data [3] and time series data [4], effectively achieving multi-source information fusion has gradually become an important technical support for air quality prediction tasks [5]. Due to the complexity of human activities and atmospheric diffusion, spatiotemporal air quality data presents complex patterns of variation, greatly increasing the difficulty of data fusion and prediction [6]. In recent years, based on the powerful nonlinear modeling and data mining capabilities of deep learning, there has been rapid development in the field of air quality prediction [7]. Classic deep learning frameworks such as Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) have been widely applied [8]. However, these methods often do not fully utilize and fuse two key pieces of information in air quality data (multi-granularity and spatiotemporal correlation), resulting in limited forecasting performance [9]. Below, we will introduce these two key pieces of information:

Multi-granularity: As shown in Fig. 1, air monitoring stations typically collect multiple air quality data with varying granularities based on different sampling intervals [10]. Coarse-grained data indicates air quality data with daily sampling intervals, primarily

* Corresponding authors at: No.6 Kexueyuan South Road Zhongguancun, Haidian District Beijing, China.
E-mail addresses: wangfei@ict.ac.cn (F. Wang), xyj@ict.ac.cn (Y. Xu).

https://doi.org/10.1016/j.inffus.2024.102607
Received 30 April 2024; Received in revised form 27 July 2024; Accepted 27 July 2024
Available online 31 July 2024

reflecting overall trends and seasonality [11]. Fine-grained data refers to air quality data collected at hourly intervals, highlighting local details [12]. Utilizing multi-granularity data can provide the model with a broader range of information, ultimately enhancing the accuracy of predictions [13].

Spatiotemporal correlation: Due to the impact of atmospheric diffusion, human activities, and other factors, there are often mutual influences between air quality data from various air monitoring stations [14]. This necessitates the consideration of Air quality data from other sites when modeling each air quality data [15]. Therefore, if the model can successfully capture the spatiotemporal correlation among air quality data from different sites, it can acquire crucial information and obtain more accurate prediction results [16].

If the two important information can be fully utilized and fused, the model can extract more valuable information and enhance its performance. However, there are still some technical challenges that must be addressed: (1) While multi-granularity features can offer additional information to the model, there is often redundancy between coarse-grained and fine-grained data. As shown in Fig. 1, coarse-grained data is a new form of data processed from fine-grained data and they contain similar global patterns. If redundancy is not eliminated, the model may overly focus on these global patterns, leading to a decrease in predictive performance. (2) The spatial correlation of Air quality data is highly dynamic at both local and global levels, influenced by factors like atmospheric diffusion and human activities. This complexity makes it challenging for simple methods based on correlation coefficients and distances to accurately capture spatial correlations between data collected from different sites. Therefore, it is crucial to develop a method that can deeply analyze the spatiotemporal correlation among different air quality datasets. (3) In multi-step prediction scenarios, the modeling effects of data with different granularities vary for each time step due to differences in information content. Combining multi-granularity data using weighted summation often limits the predictive power of the model. Therefore, it is essential for the model to analyze the impact of data with different granularities on multi-step prediction accurately and assign suitable weights accordingly.

To address the aforementioned three technical challenges, we propose three blocks: the Residual De-redundant (RD) block, the Spatiotemporal Attention (STA) block, and the Dynamic Fusion (DF) block. These blocks aim to resolve the current issues from the following perspectives:

RD block: Eliminating the redundant information between coarse-grained and fine-grained data can prevent the model from being misled by redundant information and impacting the results [17]. However, how can we effectively eliminate this redundant information? Given that time series data is a unique sequence with historical inertia, the correlation between coarse and fine granularity is most significant within the same time frame [18]. Therefore, we design the RD block, which utilizes piecewise sampling and the multilayer perceptron to establish the correspondence of redundant information across different granularities. Once the correspondence of redundant information is established, the RD block leverages residual connections to eliminate this redundancy.

STA block: Compared with CNN, RNN and other methods, the attention mechanism can effectively capture temporal information and spatial correlation from a global perspective [19]. Considering the highly dynamic spatiotemporal correlation of air quality data, this paper proposes the spatiotemporal attention block, designed to explore the air quality data in terms of its spatiotemporal aspects. Within this block, we propose spatial attention and temporal attention, individually mining spatial correlation and temporal information from the original data. Next, to fully integrate spatiotemporal information, we use the idea of the parallel modeling strategy and incorporate cross-attention to combine the tensors derived from spatial attention and temporal attention. Contrasted with the approach of stacking multiple layers, the parallel modeling strategy allows the model to extract information more efficiently [20]. Through these procedures, the STA block can effectively mine and integrate the spatial-temporal correlation of air quality data.

DF block: As coarse-grained data and fine-grained data contain different information, they have varying effects on different prediction time steps [21]. Taking this into account, we employ the concept of piecewise ensemble to achieve the dynamic fusion of prediction results. [22]. To implement the piecewise ensemble based on the varying granularities of data, this paper introduces the dynamic fusion block. This block utilizes the attention mechanism to assess the significance of different granularities across diverse prediction time steps. This allows the model to assign distinct weights at different time steps, ultimately enhancing prediction performance.

Based on the above three blocks, we propose the Multi-Granularity Spatiotemporal Fusion Transformer (MGSFformer), which leverages the two essential information (multi-granularity and spatiotemporal correlation) of air quality data to enhance high-precision air quality spatiotemporal prediction. In conclusion, the primary contributions of this paper can be outlined as follows:

- We propose the Multi-Granularity Spatiotemporal Fusion Transformer for air quality spatiotemporal prediction. MGSFformer can

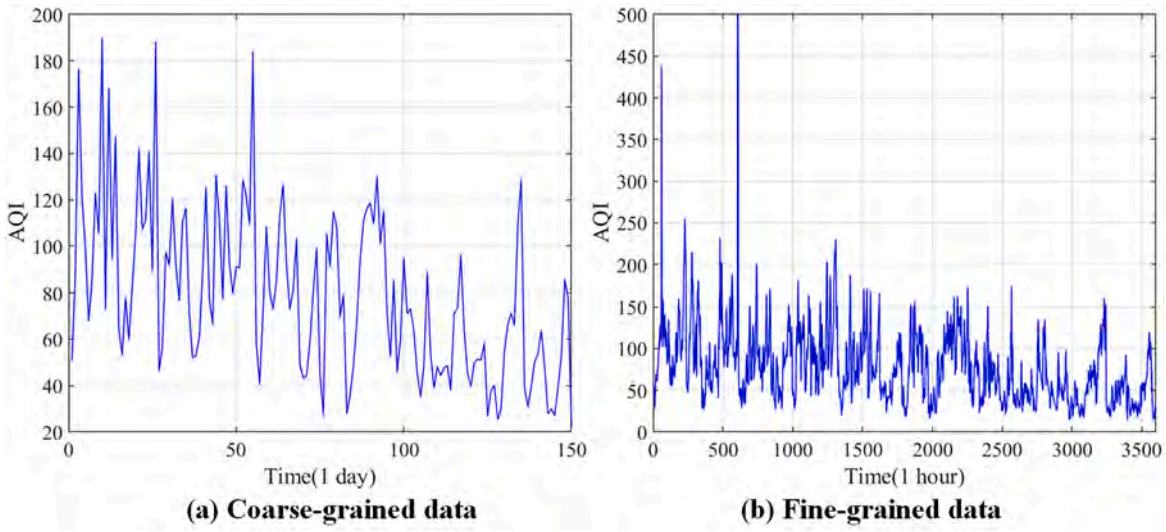

Fig. 1. The air quality index data under different granularity.

effectively utilize and fuse the multi-granularity and spatiotemporal correlation of air quality data.

- We design three blocks to address three challenges in modeling air quality spatiotemporal data. Firstly, the RD block is designed to eliminate information redundancy among multi-granularity features. Additionally, the STA block is proposed to extract spatiotemporal correlations among different air quality data for spatiotemporal prediction. Finally, the DF block is designed to analyze the impact of various granularities on multi-step prediction effectively and generate the final prediction results.
- To show the performance of MGSFformer, we conduct comparative experiments on three real-world datasets. All experimental results indicate that the proposed model can effectively outperform 8 different state-of-the-art (SOTA) baselines.

The remaining sections of this paper are organized as follows: Section 2 shows the existing related work. Section 3 presents the definition of the problem as well as the framework of the proposed MGSFformer. Section 4 reports the main experimental results. Section 5 concludes this paper and gives future works.

2. Related works

In this section, we discuss related works from three aspects: classic deep learning methods, multi-granularity modeling methods and spatiotemporal modeling methods.

2.1. Classic deep learning methods

In air quality spatiotemporal prediction, deep learning-based methods can enhance the model's feature analysis ability by stacking multiple hidden layers [23]. As one of the mainstream air quality prediction methods, RNN mainly adopts the sequential modeling concept to achieve accurate time series prediction. Compared to traditional methods, RNN can yield certain results [24]. The long short-term memory network (LSTM) [25] and the gated recurrent unit (GRU) [26], as the main variants of RNN, utilize the gated structure to enhance the model's training effectiveness and comprehensively capture time series dependencies [27]. Apart from RNN-based frameworks, CNN-based models have also demonstrated excellent performance. The temporal convolutional network (TCN) utilizes dilated convolution for sequence modeling and achieves similar results to RNN [28]. Although traditional deep learning models have achieved certain successes, they often fail to the deep information of multivariate time series in both temporal and among variables, which limits their effectiveness.

To fully exploit the deep information of multivariate time series in both temporal and among variables, the Transformer [29] and the Spatiotemporal Graph Neural Network (STGNN) [30] have been widely adopted in recent years. The multi-resolution interactive transformer [31] combines multi-resolution modeling with the attention mechanism, achieving impressive performance in time series forecasting. Furthermore, the spatiotemporal transformer [32] successfully predicts air quality by exploring the spatiotemporal dependencies between sites. Shao et al. [33] combined adaptive graph convolution and RNN to effectively improve the accuracy of spatiotemporal prediction models. Ji et al. [34] used the dynamic graph learning to optimize the performance of STGNN in the field of spatiotemporal prediction. Song et al. [35] designed a variety of different graph structures to improve the prediction of STGNN. Xu et al. [36] proposed the generic dynamic graph convolutional network to realize high-precision spatiotemporal information fusion and time series prediction. While above deep learning methods have shown promising results, they fail to effectively utilize and fuse all air quality data information (the multi-granularity and spatiotemporal correlation), limiting their modeling efficacy [37].

2.2. Multi-granularity modeling methods

At present, several scholars have proposed ensemble methods [38] and concatenation methods [39] to incorporate multi-granularity information. Liu et al. [40] utilized support vector machine (SVM) and long short-term memory network (LSTM) to model coarse-grained and fine-grained air quality data, respectively. They employed the non-dominated sorting genetic algorithm II (NSGA-II) to combine models with varying granularities, resulting in a 5% performance improvement. Teng et al. [41] applied LSTM and attention mechanisms to analyze air quality data with different granularities. Subsequently, they utilized a weighted summation approach to fuse the ensemble results. Ji et al. [42] utilized LSTM and Extreme learning machine (ELM) for coarse-grained and fine-grained modeling, respectively. The prediction outcomes with different granularities were integrated using the whale optimization algorithm (WOA). Yu et al. [43] used hierarchical structure to improve the ability of neural network to mine multi-granularity information of time series. Compared with single-grain model, the introduction of multi-grain information can effectively improve the prediction accuracy. Wang et al. [44] employed the concatenation method and deep learning to achieve the fusion of multi-granularity information and demonstrate the importance of multi-granularity through experimentation. Despite the effectiveness of these frameworks, the aforementioned methods often overlook two issues: (1) Information redundancy exists between air quality data with different granularities due to the nature of the data. (2) Data with different granularities contain distinct information, contributing differently at various prediction stages. In conclusion, existing methods do not fully leverage the multi-granularity information available, which limits their performance.

2.3. Spatiotemporal modeling methods

Currently, some researchers utilize other sites that are highly correlated with the target site as auxiliary inputs to achieve air quality spatiotemporal prediction [45]. Mi et al. [46] utilized the Pearson correlation coefficient to identify the most significant auxiliary sites for the target site. The experiments demonstrate that incorporating auxiliary sites can enhance the performance of the target site by 5%. Liu et al. [47] employed a clustering approach to categorize monitoring stations into multiple groups. Stations within the same group were then used as auxiliary inputs for the target stations. Apart from these techniques, spatiotemporal graph neural networks have also been employed for air quality spatiotemporal prediction [48]. Liu et al. [28] utilized the graph convolutional network (GCN) and reinforcement learning (RL) to construct a air quality spatiotemporal prediction model. Tan et al. [49] employed the graph attention network (GAT) to identify spatial correlations among air quality levels at different sites and developed a high-performing prediction model. Liu et al. [50] used Gaussian distribution and attention mechanisms to optimize the performance of graph convolution, achieving good results in spatiotemporal prediction. Ahmed et al. [51] realized high-precision multivariate time series prediction based on distance information and graph convolutional networks. Experimental results prove the effectiveness of spatiotemporal modeling. Generally, existing spatiotemporal forecasting methods primarily rely on correlation coefficients and distances to analyze spatiotemporal correlations [52]. Nevertheless, as the number and distances between air monitoring stations increase, these methods often encounter challenges in accurately capturing spatial correlations [53].

3. Methodology

3.1. Preliminaries

The air quality spatiotemporal prediction task can be defined as time series prediction and multi-granularity modeling. A brief introduction to

these two tasks is provided below:

Time series prediction [54]: The spatiotemporal prediction of air quality proposed in this paper can be defined as a time series prediction technique [55]. For N different air monitoring stations, given input features for $X = [x_1, x_2, \ldots, x_H]$, the model can get prediction results $Y = [Y_1, Y_2, \ldots, Y_L]$. And H represents the length of the looking back time window. L stands for the step size of the prediction. The dimensions of X and Y are $N*H$ and $N*L$ respectively.

Multi-granularity modeling [56]: Compared with traditional time series modeling, the main difference of multi-granularity modeling is the input features of the model. The input features of the model are transformed to $X = [X_1, X_2, \ldots, Xm]$. X_1 is the coarsest granularity, and its dimension is $N*H_1$. Xm is the finest granularity, and its dimension is $N*H_m$. The output of the model is the most coarse-grained Air quality for the next L time steps.

3.2. Framework of MGSFformer

Fig. 2 illustrates the main framework of the proposed air quality spatiotemporal prediction model, which consists of three main blocks: the residual de-redundant block, spatiotemporal attention block, and dynamic fusion block. The main modeling steps of the MGSFformer are outlined as follows:

Step 1: The data with different granularities are individually input into the corresponding RD blocks. The RD block comprises two main components: information embedding and residual connection. Information embedding establishes the correspondence of redundant information positions in data with different granularities and converts them into the same dimension. The residual connection eliminates redundant information. Section 3.2 elaborates on the specific modeling process of the RD block.

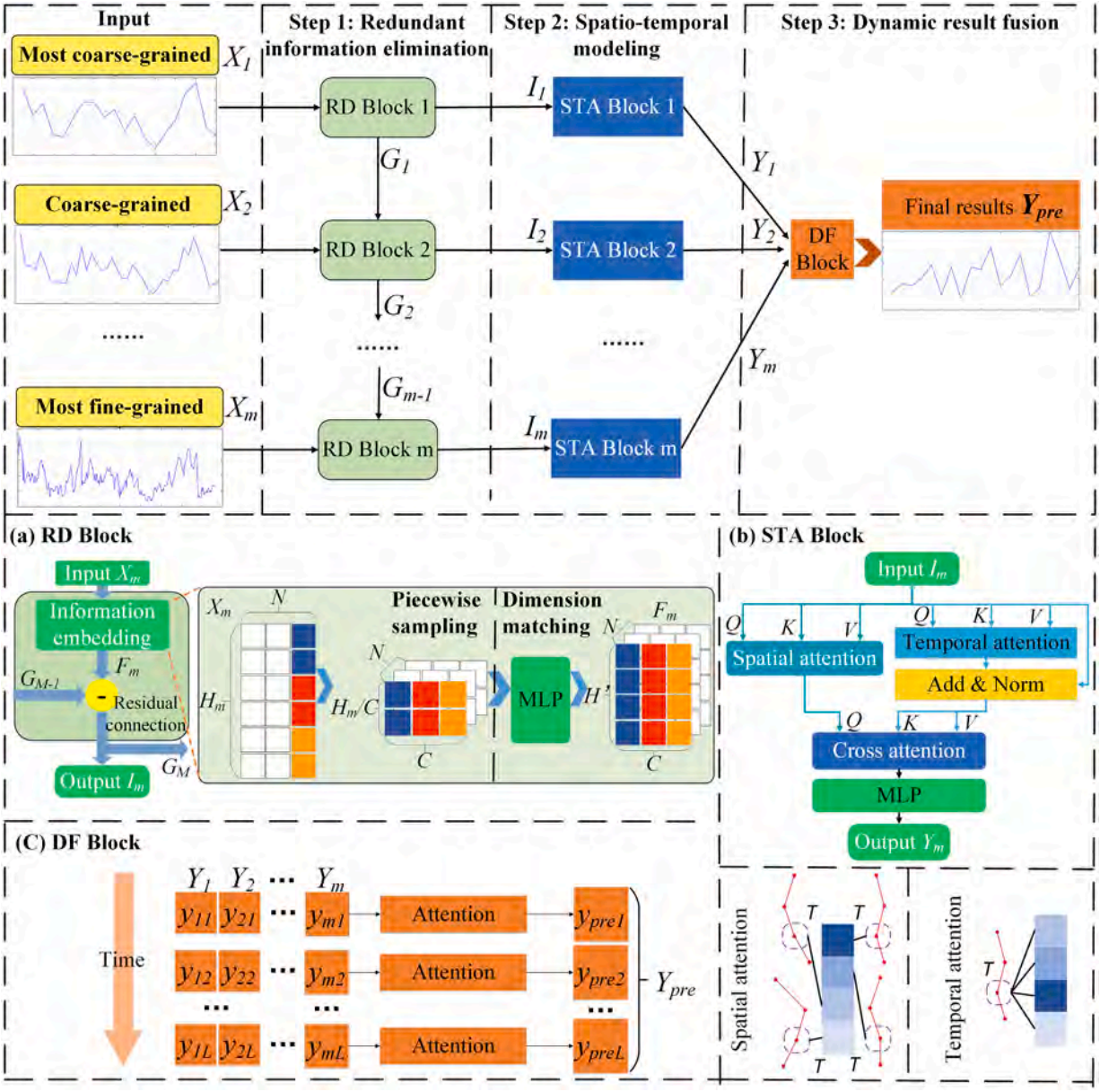

Fig. 2. The overall framework of the proposed model. The input of the model is air quality data with different granularities. The output of the model is the air quality data at the coarsest granularity. (a) The specific structure of the residual de-redundant block. (b) The specific structure of the spatiotemporal attention block. (c) The specific structure of the dynamic fusion block.

Step 2: The data processed by the RD block is then transmitted to the STA block. The STA block employs temporal attention, spatial attention, and cross-attention to conduct spatiotemporal modeling of air quality data across different granularities, resulting in prediction results. Section 3.3 provides a detailed explanation of the implementation process of the STA block.
Step 3: The prediction results of models with varying granularities are directed to the DF block. By assessing the significance of different granularities in multi-step prediction, this block effectively accomplishes dynamic result fusion to generate the final prediction results. The essential aspects of the DF block are introduced in Section 3.4.

3.3. Residual de-redundant block

The main function of the RD block is to embed information while eliminating redundant information between multi-granularity data. Compared to natural language data, the time series is a special sequence format, which makes the coarse-grained data and fine-grained data in the corresponding time period contain the strongest correlation of information [18]. The residual de-redundant block is proposed based on the above characteristics, which contains three important functions: corresponding to the redundant information positions between multiple granularities, matching the dimensions of features with different granularities, and eliminating of redundant information. The specific modeling steps are given as follows:

Step 1: The input feature of each RD block is$X_m \in R^{N*H_m}$. N represents the number of series and H_m represents the length of the look-back window. H_m is different for each granularity. The first modeling step is to realize the correspondence of redundant information locations between coarse-grained and fine-grained by the piecewise sampling method. For the coarsest granularity, the dimensions of $X_1 \in R^{N*H_1}$ are transformed as follows:

$$N * H_1 \rightarrow N * C * 1 \tag{1}$$

where, the size of H_1 and C are equal, that is, the dimension of X_1 is transformed from two to three dimensions.

Step 2: Then, based on the dimension of the coarsest granularity, all the fine granularities are transformed into 3D tensors by piecewise sampling. After transformation, the correspondence of redundant information positions between data with different granularities can be realized. The conversion of fine-grained data is shown as follows:

$$N * H_m \rightarrow N * C * \frac{H_m}{C} \tag{2}$$

where, the dimension C of all fine-grained X_m is equal to the dimension C of the coarsest grainedX_1. For the transformedX_m, the composition of the jth subsequence X_m^j obtained in the piecewise sampling is given as follows:

$$X_m^j = \left[x_{1+(j-1)*c}, x_{2+(j-1)*c}, x_{3+(j-1)*c}, \ldots, x_{j*c}\right] \tag{3}$$

Step 3: In the previous step, the fine-grained input feature X_m is transformed by the piecewise sampling method. Before residual modeling, it is necessary to ensure that the dimensions of different X_m are uniform. To unify the dimensions of these features, MLP is used to transform the dimension of X_m from $N * C * \frac{H_m}{C}$ to $N * C * H'$. The basic formula is given as follows:

$$F_m = \mathrm{FC}(ReLU(\mathrm{FC}(X_m))) \tag{4}$$

where, FC(·) is the fully connected layer. ReLU(·) stands for the activation function

Step 4: The above steps complete the information embedding. The feature tensor $F_m \in R^{N*C*H'}$ is obtained for each granularity. Next, redundant information of $F_m \in R^{N*C*H'}$ needs to be eliminated before spatiotemporal modeling.
Step 5: Piecewise sampling effectively matches the positions of redundant information across data of different granularities. To fully eliminate the redundant information between the fine-grained data and the coarse-grained data, this paper uses the residual connection. The basic formula for the residual connection is given below:

$$\begin{cases} I_m = F_m, m = 1 \\ I_m = F_m - G_{m-1}, m > 1 \\ G_m = COPY(I_m) \end{cases} \tag{5}$$

where, $I_m \in R^{N*C*H'}$ represents the input of the STA block. $F_m \in R^{N*C*H'}$ represents the input features after preliminary processing through the information embedding. $G_{m-1} \in R^{N*C*H'}$ represents the previous coarse-grained information. $COPY(.)$ means that we use the processed $I_m \in R^{N*C*H'}$ as $G_m \in R^{N*C*H'}$ and pass it to the $m+1$ RD blcok.

Step 6: After the above steps, we get $I_m \in R^{N*C*H'}$ and $G_m \in R^{N*C*H'}$. $I_m \in R^{N*C*H'}$ is transmitted to STA block to realize air quality spatiotemporal prediction. The $G_m \in R^{N*C*H'}$ is transmitted to the next RD block to eliminate redundant information of finer grained data.

3.4. Spatiotemporal attention block

The STA block is used to extract key information from the feature tensor $I_m \in R^{N*C*H'}$ obtained by the RD block, which contains the most important information at the current granularity, in order to achieve high-precision prediction. The main structure of the STA block includes temporal attention, spatial attention, cross attention and MLP. The temporal attention is used to mine the temporal correlation of the data at each site. The spatial attention is used to mine the spatial correlation between the air quality data from different sites. The cross attention is used to fuse temporal and spatial information. The spatiotemporal prediction results are obtained through the MLP layer. The main modeling steps of the STA block are given as follows:

Step 1: The temporal attention is used to mine the temporal information of the feature tensor $I_m \in R^{N*C*H'}$ and obtains a new feature tensor $I_m^{te} \in R^{N*C*H'}$. The specific formula is shown as follows:

$$\begin{cases} Q = W_Q^{te} I_m \\ K = W_K^{te} I_m \\ V = W_V^{te} I_m \end{cases} \tag{6}$$

$$I_m^{te} = F_{LN}\left(soft\max\left(Q * K^T\right)V + I_m\right) \tag{7}$$

where, W represents the weight coefficient. $F_{LN}(\cdot)$ represents the layer normalization. $K^T$is the transpose of K.

Step 2: The spatial attention is used to mine the spatial correlation between air quality data from N different monitoring stations in the feature tensor $I_m \in R^{N*C*H'}$. Then, a new feature tensor $I_m^s \in R^{N*C*H'}$ is obtained according to the following formula:

$$\begin{cases} Q = W_Q^s I_m \\ K = W_K^s I_m \\ V = W_V^s I_m \end{cases} \tag{8}$$

$$I_m^s = soft\max\left(Q * K^T\right) V \tag{9}$$

Temporal attention primarily models the dimension C of $I_m \in R^{N*C*H'}$, that is, the contextual association. Spatial attention mainly models the dimension N of $I_m \in R^{N*C*H'}$, that is, the association between different time series.

Step 3: After mining the temporal information and spatial correlation of Air quality data at different monitoring stations, this paper uses cross attention to fuse the $I_m^{te} \in R^{N*C*H'}$ and $I_m^s \in R^{N*C*H'}$, and realizes the fusion of spatiotemporal information. $I_m^{st} \in R^{N*C*H'}$ is obtained based on the following formulas:

$$\begin{cases} Q = W_Q^{ca} I_m^s \\ K = W_K^{ca} I_m^{te} \\ V = W_V^{ca} I_m^{te} \end{cases} \tag{10}$$

$$I_m^{st} = soft\max\left(Q * K^T\right) V \tag{11}$$

In cross-attention, we use the feature vector I_m^s obtained from spatial attention as the Query (Q), and the feature vector obtained from temporal attention I_m^{te} as the Key (K) and Value (V). Through this way, the feature vector I_m^{te} can query the most critical information from feature vector I_m^s and integrate it, avoiding information redundancy caused by direct adding.

Step 4: Before finally realizing the prediction through MLP, the three-dimensional tensor $I_m^{st} \in R^{N*C*H'}$ needs to be transformed into two-dimensional tensor $O_m \in R^{N*L'}$ by the following formula:

$$I_m^{st} \in R^{N*C*H'} \rightarrow O_m \in R^{N*L'} \tag{12}$$

$$L' = C * H' \tag{13}$$

$I_m^{st}$is transformed into $O_m$by reshaping.

Step 5: Finally, the Air quality spatiotemporal prediction is realized based on the MLP layer and the prediction results $Y_m \in R^{N*L}$ are obtained.

$$Y_m = \mathrm{ReLU}(\mathrm{FC}(O_m)) \tag{14}$$

where, $Y_m \in R^{N*L}$ represents the prediction of the model at the mth granularity. In this paper, each granularity of the model can obtain the corresponding prediction value $Y_m \in R^{N*L}$. After that, by effectively fusing these prediction values, the DF blocks can obtain the final prediction results $Y_{pre} \in R^{N*L}$.

3.5. Dynamic fusion block

Through the above modeling process, the corresponding prediction results $Y_m \in R^{N*L}$ of multiple STA block with different granularities are obtained. To fully fuse these prediction results and obtain the optimal multi-step prediction results $Y_{pre} \in R^{N*L}$, an attention-based dynamic fusion block is proposed. The modeling steps in this block are given as follows:

Step 1: Based on the STA block, the mth granularity model obtained the prediction result

$$Y_m = [y_{m1}, y_{m2}, \ldots, y_{mL}].$$

where, y_{mL} represents the prediction result of the mth granularity at the L step.

Step 2: For the first prediction time step, the prediction result for all granularities is $Y_{pre1} = [y_{11}, y_{21}, \ldots, y_{m1}]$. The attention distribution is defined by the following formula:

$$\alpha_1 = \mathrm{softmax}\left(Y_{pre1}\right) = [\alpha_{11}, \alpha_{21}, \ldots, \alpha_{m1}] \tag{15}$$

Step 3: The prediction result y_{pre1} is obtained by weighted summation of α and y based on the following formula:

$$y_{pre1} = \sum_{t=1}^{m} (\alpha_{t1} * y_{t1}) \tag{16}$$

Step 4: For each subsequent time step, the above method is used to fuse the results of all granularities and obtain the final air quality multi-step prediction results $Y_{pre} \in R^{N*L}$.

4. Experiments

4.1. Experimental design

Dataset: To fully compare and analyze the performance of different models, experiments are conducted using three real-world datasets from China. The basic statistics of these datasets are shown in Table 1. The spatial distribution of these data is shown in Fig. 3. As the most important indicator in these datasets, this paper mainly compared and analyzed the Air Quality Index (AQI). In addition, to further demonstrate the model's adaptability, we also compared and analyzed four important air pollutants (PM2.5, PM10, SO2, NO2) [57]. All datasets can be found at the following link: https://quotsoft.net/air/.

A brief description of these datasets is shown below:

- Beijing sites: This dataset is air quality data from 35 air monitoring stations in Beijing, China. The time range of the dataset is from 2015 to 2021. Granularity timestamps for data include 1 day, 12 h, 6 h, 3 h, and 1 hour.
- China cities: This dataset consists of air quality data collected from 350 cities in China. The time range of the dataset is from 2015 to

Table 1
Basic statistics of all datasets.

Datasets	Beijing sites	China cities	China sites
Number of time series	35	350	1200
Time steps	59,712	59,688	59,720
Initial granularity	1 h	1 h	1 h

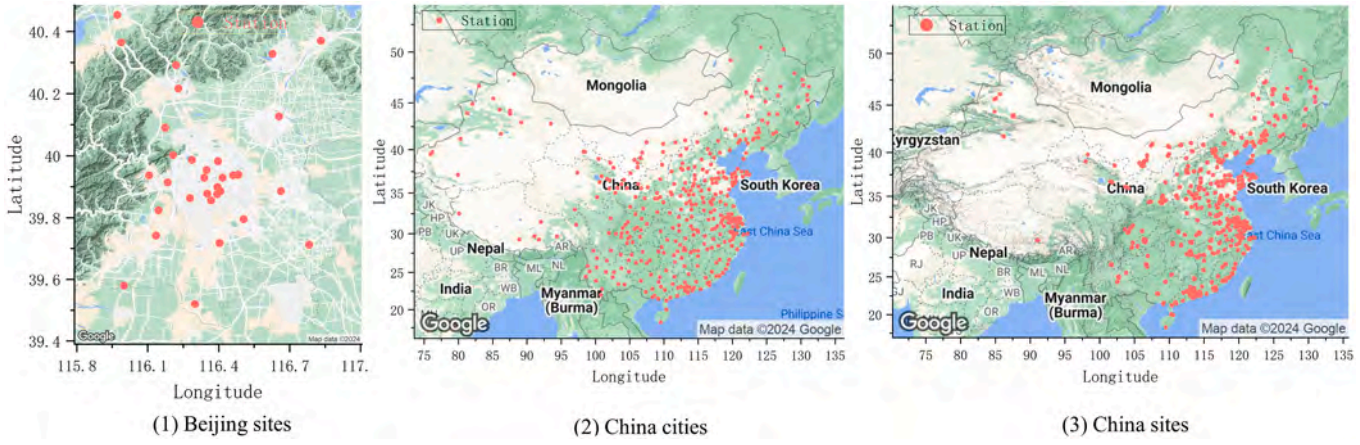

Fig. 3. Spatial distribution of different air monitoring stations in all datasets. The red dots represent the locations of the air monitoring stations.

2021. Granularity timestamps for data include one day, 12 h, 6 h, 3 h, and 1 hour.

- China sites: This dataset consists of air quality data collected from 1200 sites in China. The time range of the dataset is from 2015 to 2021. Granularity timestamps for data include one day, 12 h, 6 h, 3 h, and 1 hour.

Baselines. To demonstrate the performance of the proposed model, this paper selects baselines from the following six aspects: statistics model (VAR [58]), RNN-based networks (GRU [27] and LSTNet [59]), CNN-based model (TimesNet [60] and SciNet [61]), GNN-based model (MegaCRN [62] and AGCRN [63]), multi-granularity-based models (TimeMixer [44] and PatchFormer [64]) and Transformer-based models (DSformer [65] and Airformer [66]). A basic introduction to these baselines is shown below:

- VAR: Vector autoregression is a classical multivariate time series prediction model based on statistics.
- GRU: As a classical RNN variant, GRU can better model the context of air quality time series.
- LSTNet: It combines LSTM and attention mechanisms to effectively optimize the prediction performance.
- SciNet: The model uses hierarchical structure and sample convolution to improve the ability of the model to mine time series.
- TimesNet: It uses frequency domain conversion and convolutional network to mine the time information of data.
- AGCRN: It introduces adaptive graph learning to improve the ability of models to model spatiotemporal associations.
- MegaCRN: It combines memory bank and adaptive graph learning to improve the ability of the model to mine temporal and spatial correlations.
- PatchFormer: It employs multi-scale attention and integrates data from different patches to establish correlation among features with different scales.
- TimeMixer: It utilizes a fully MLP-based architecture with Past-Decomposable-Mixing (PDM) and Future-Multipredictor-Mixing (FMM) blocks, aiming to fully leverage the multi-scale information of time series.
- DSformer: It introduces the double sampling block and attention mechanism to effectively mine temporal and spatial information of data.
- Airformer: It proposes the Dartboard Spatial attention and Causal Temporal attention to mine the spatiotemporal correlation of air quality data.

Setting. The main hyperparameters of the proposed model are shown in Table 2. Besides, to ensure a comprehensive and unbiased comparison, we structured the experiment as follows: (1) The three datasets were segmented into training, validation, and test sets in a 6:2:2 ratio. (2) To align with the modeling configurations of SciNet, TimesNet, and other baselines, this paper employed the z-score standardized method [67] uniformly for preprocessing the raw air quality data. (3) In the proposed model, we utilized a 7-day window of historical observations as input features. The dimensions of features, varying in granularity, differ (the coarsest granularity has a history length of 7, while the finest granularity has a history length of 168). In addition, we further evaluated the influence of the length of historical observations on the experimental results in Section 4.7. The model's output comprises the air quality for all air monitoring stations for the subsequent 7 days. (4) For the other baseline models, we generated prediction results based on multi-granularity features and single-granularity features separately. The optimal results were then compared with those of the proposed model. (5) For fair comparison, all baselines uniformly use raster search [68] to optimize their hyperparameters. (6) The code can be found at the following link: https://github.com/ChengqingYu/MGSFformer.

Table 2
The main hyperparameters of the proposed model.

Config	Values
optimizer	Adam [69]
learning rate	0.0002
number of multi-head attention	2
dropout	0.3
dimensions H' after matching	32
learning rate schedule	MultiStepLR
milestone	[1,25,50,75]
gamme	0.5
batch size	32
epoch	100

Evaluation index. It is the core technology to evaluate the overall performance of all models adopted in this paper. Three classic time-series evaluation indexes, which include MAE (Mean Absolute Error) [70], MSE (Mean Square Error) [71], and CORR (R Squared index) [72], were used to analyze MGSFformer and other baselines. The core calculation formulas of these evaluation indexes are shown as follows:

$$MAE = \left(\sum_{T=1}^{n} \left|Y_{true}(T) - Y_{pre}(T)\right|\right) \Big/ n \tag{17}$$

$$MSE = \left(\sum_{T=1}^{n} \left[Y_{true}(T) - Y_{pre}(T)\right]^2\right) / n \tag{18}$$

$$CORR = 1 - \left(\sum_{T=1}^{n} \left| Y_{true}(T) - Y_{pre}(T) \right|^2 \right) \Big/ \left(\sum_{T=1}^{n} \left| Y_{true}(T) - Y_{ave} \right|^2 \right) \tag{19}$$

where $Y_{true}(T)$ represents actual air quality data. $Y_{pre}(T)$ represents the air quality data calculated by the proposed model. $Y_{ave}$is the average power of the power series. n represents the sample size.

4.2. Main results

To prove the advancement of the proposed model in air quality spatiotemporal prediction, this section compares the proposed model with 11 existing baselines. Table 3 shows the efficiency of different models. Table 4 shows the experimental results of all models on AQI datasets. The Horizon represents the prediction result of the corresponding time step, and the Average represents the average prediction result of the multi-step prediction. In addition to the air quality index, Table 5 further compares the predictive performance of MGSFformer and several baselines on four types of air pollution data (PM2.5, PM10, SO2, NO2).

As shown in Tables 4 and 5, the proposed model demonstrates superior performance across all datasets, providing strong evidence of its effectiveness. Based on the experimental results, the following conclusions can be drawn:

(1) The statistics model, RNN-based models, and CNN-based models exhibit subpar prediction performance due to their failure to fully leverage two crucial data patterns (multi-granularity and spatial correlation).
(2) GNN-based models excel in mining spatial correlation and temporal information, ensuring their performance. However, they do not fully capitalize on multi-granularity information, thereby limiting their efficacy.
(3) Multi-granularity-based models can achieve decent predictive performance. However, on one hand, they overlook the redundant information among multi-granularity data, and on the other hand, they fail to fully model spatial associations, which limits their prediction effectiveness.
(4) Transformer-based models effectively utilize both multi-granularity information and spatial correlation. Nevertheless, existing Transformer-based models struggle to eliminate redundant information and achieve seamless information fusion, ultimately impacting their performance.
(5) In comparison to other classical models and the state-of-the-art models, the proposed MGSFformer achieves the best results. The RD block eliminates redundant information from data with varying granularities. The STA block integrates temporal attention, spatial attention, and cross-attention to effectively facilitate air quality spatiotemporal prediction. Lastly, through the DF block, the model assigns weights to different granularities based on their contributions at distinct prediction steps. Consequently, the MGSFformer holds significant practical value.

Table 3
Efficiency of different models. The result is the average training time for each epoch.

Method	Datasets		
	Beijing sites	China cities	China sites
MGSFformer	24.67	74.16	189.34
Airformer [66]	31.84	89.84	192.56
DSformer [65]	20.75	64.31	174.83
MegaCRN [62]	34.86	93.17	231.49
AGCRN [63]	38.55	101.29	273.76

4.3. Performance evaluation of different components

To fully verify the importance of each block in the proposed network, component replacement experiments are taken in this section. This paper replaces the key components from the following three perspectives: (1) **Pro-GRU**: it stands for replacing the information embedding in the RD block with GRU. (2) **Pro-GWO** and **Pro-RL**: we use the gray wolf optimization algorithm and reinforcement learning to replace the DF block, respectively. (3) **Pro-MLP**: it stands for replacing the STA block with the MLP. (4) **Pro-SA:** it represents that STA blocks achieve spatiotemporal modeling by stacking temporal attention and spatial attention.

Table 6 presents the error evaluation indices for several methods and MGSFformer. Based on the results, the following conclusions can be drawn:

(1) The proposed information embedding method outperforms GRU-based information embedding in achieving better results. This improvement is attributed to GRU's limited capability in multi-step prediction and spatiotemporal modeling.
(2) The proposed attention-based DF block surpasses result fusion based on reinforcement learning and multi-objective optimization algorithms, delivering an average performance improvement of over 5%. This enhancement is due to the attention-based DF block's ability to evaluate the significance of different granularities at various prediction steps, thereby optimizing result fusion.
(3) The proposed STA block outperforms the MLP-based modeling component in establishing a more precise air quality prediction model. This superiority stems from the STA block's three attention mechanisms, which effectively extract and integrate temporal information and spatial correlation. Consequently, the proposed model can analyze spatiotemporal correlation more accurately than MLP, ensuring its superior performance.
(4) Compared with stacking temporal attention and spatial attention, the STA block based on the parallel modeling structure can achieve better experimental results. The main reason is that the parallel modeling structure enhances the ability of the model to mine key characteristics from the raw data, which guarantees the forecasting effect of the model.

4.4. Ablation experiments

To demonstrate the importance of several components, this paper also employs ablation experiments, which were performed from the following four aspects: (1) **w/o RE**: To demonstrate the importance of removing information redundancy, we remove the residual connection in the RD block. (2) **w/o DF**: To demonstrate the importance of the proposed DF block, we remove the result fusion block and use linear weighting as the alternative. (3) **w/o IE**: To prove the importance of our proposed information embedding method, we replace it with a linear layer. (4) **w/o RD**: We remove the RD block to prove the importance of input information processing.

Table 7 illustrates the results of the ablation experiments. The experimental findings indicate a significant degradation in the model's performance upon the removal of components. Based on these results, the following conclusions can be drawn:

(1) The results of w/o RE indicate that the residual connection effectively eliminates information redundancy between data with different granularities, enhancing the model's modeling capability.
(2) The results of w/o DF highlight the importance of dynamic result fusion. Attention, as opposed to linear weighting, accurately assesses the significance of different granularities at different prediction steps, thereby optimizing overall prediction accuracy.
(3) The experimental results of w/o IE reveal that direct dimension transformation and residual connection on original features fail

Table 4
Performance comparison results of all baselines and the proposed model on AQI datasets (The best results are shown in **bold**. An upward arrow represents accuracy, while a downward arrow represents error.).

Dataset	Model	Horizon 1			Horizon 4			Horizon 7			Average		
		MSE↓	MAE↓	CORR↑	MSE↓	MAE↓	CORR↑	MSE↓	MAE↓	CORR↑	MSE↓	MAE↓	CORR↑
Beijing sites	VAR [58]	0.2437	0.3845	0.814	0.7044	0.6795	0.504	0.8156	0.7025	0.454	0.6684	0.6252	0.528
	GRU [27]	0.2279	0.3618	0.836	0.6495	0.6154	0.553	0.7638	0.6817	0.497	0.5294	0.5630	0.606
	LSTNet [59]	0.2265	0.3566	0.847	0.5771	0.5987	0.566	0.7581	0.6823	0.482	0.5129	0.5559	0.618
	SciNet [61]	0.2169	0.3464	0.854	0.5028	0.5714	0.592	0.5755	0.6356	0.533	0.4640	0.5312	0.632
	TimesNet [60]	0.2151	0.3453	0.859	0.4917	0.5858	0.587	0.5651	0.6394	0.527	0.4429	0.5302	0.645
	AGCRN [63]	0.2099	0.3168	0.874	0.4622	0.5065	0.653	0.5022	0.5781	0.589	0.4037	0.4705	0.724
	MegaCRN [62]	0.2093	0.3165	0.878	0.4503	0.4947	0.679	0.5079	0.5774	0.591	0.4075	0.4675	0.729
	PatchFormer [64]	0.1956	0.3094	0.886	0.4356	0.4916	0.684	0.4915	0.5648	0.604	0.3987	0.4602	0.731
	TimeMixer [44]	0.1934	0.3052	0.897	0.4301	0.4873	0.709	0.4897	0.5622	0.602	0.3965	0.4591	0.735
	DSformer [65]	0.1894	0.3015	0.908	0.4138	0.4719	0.721	0.4812	0.5593	0.611	0.3826	0.4577	0.742
	Airformer [66]	0.1859	0.2913	0.916	0.4215	0.4826	0.715	0.4771	0.5519	0.616	0.3803	0.4529	0.753
	MGSFformer	**0.1767**	**0.2788**	**0.932**	**0.3975**	**0.4536**	**0.746**	**0.4514**	**0.5311**	**0.625**	**0.3542**	**0.4246**	**0.775**
China cities	VAR [58]	0.5136	0.4618	0.725	0.7543	0.6072	0.534	0.7796	0.6184	0.512	0.7135	0.5829	0.572
	GRU [27]	0.4912	0.4474	0.749	0.7241	0.5721	0.597	0.7425	0.5913	0.559	0.6658	0.5401	0.623
	LSTNet [59]	0.4817	0.4436	0.755	0.6978	0.5568	0.611	0.7432	0.5929	0.548	0.6541	0.5342	0.637
	SciNet [61]	0.4899	0.4415	0.764	0.6584	0.5541	0.619	0.7394	0.5896	0.574	0.6424	0.5315	0.645
	TimesNet [60]	0.4683	0.4406	0.773	0.6483	0.5495	0.628	0.7597	0.5787	0.589	0.6386	0.5260	0.661
	AGCRN [63]	0.4312	0.4156	0.819	0.6272	0.5322	0.641	0.6487	0.5658	0.603	0.5822	0.5010	0.694
	MegaCRN [62]	0.4337	0.4167	0.814	0.6213	0.5306	0.647	0.6412	0.5547	0.617	0.5786	0.5038	0.687
	PatchFormer [64]	0.4257	0.4149	0.829	0.6196	0.5294	0.653	0.6384	0.5497	0.625	0.5776	0.4956	0.699
	TimeMixer [44]	0.4209	0.4125	0.836	0.6184	0.5246	0.664	0.6337	0.5432	0.631	0.5718	0.4909	0.705
	DSformer [65]	0.4135	0.4028	0.863	0.6017	0.5139	0.675	0.6227	0.5281	0.657	0.5627	0.4816	0.721
	Airformer [66]	0.4196	0.4097	0.857	0.6132	0.5223	0.667	0.6298	0.5416	0.635	0.5697	0.4839	0.716
	MGSFformer	**0.3934**	**0.3840**	**0.894**	**0.5863**	**0.5046**	**0.689**	**0.6046**	**0.5187**	**0.673**	**0.5413**	**0.4694**	**0.728**
China sites	VAR [58]	0.4612	0.4833	0.718	0.8627	0.7319	0.477	1.0418	0.8016	0.388	0.8624	0.7307	0.486
	GRU [27]	0.4271	0.4517	0.743	0.7979	0.7108	0.495	0.9974	0.7703	0.407	0.8016	0.6734	0.518
	LSTNet [59]	0.4319	0.4673	0.735	0.7809	0.6921	0.518	0.9814	0.7614	0.425	0.8194	0.6674	0.539
	SciNet [61]	0.4093	0.4396	0.756	0.7394	0.6717	0.532	0.8968	0.7325	0.471	0.7684	0.6595	0.545
	TimesNet [60]	0.4046	0.4388	0.763	0.7216	0.6631	0.544	0.8923	0.7383	0.467	0.7769	0.6554	0.559
	AGCRN [63]	0.3915	0.4346	0.769	0.7161	0.6559	0.556	0.8609	0.7201	0.489	0.7391	0.6501	0.563
	MegaCRN [62]	0.3814	0.4374	0.775	0.7115	0.6431	0.577	0.8763	0.7188	0.493	0.7374	0.6471	0.574
	PatchFormer [64]	0.3791	0.4326	0.782	0.6774	0.6402	0.583	0.7843	0.6796	0.524	0.6679	0.6246	0.609
	TimeMixer [44]	0.3752	0.4318	0.796	0.6735	0.6378	0.598	0.7715	0.6742	0.526	0.6615	0.6213	0.617
	DSformer [65]	0.3718	0.4295	0.809	0.6617	0.6341	0.604	0.7418	0.6519	0.562	0.6408	0.6155	0.631
	Airformer [66]	0.3684	0.4251	0.814	0.6529	0.6227	0.615	0.7538	0.6682	0.537	0.6475	0.6179	0.629
	MGSFformer	**0.3657**	**0.4225**	**0.825**	**0.6373**	**0.6124**	**0.634**	**0.7064**	**0.6255**	**0.608**	**0.6261**	**0.5890**	**0.697**

Table 5
Performance comparison results of all baselines and the proposed model on other air pollution data (The best results are shown in **bold.** An upward arrow represents accuracy, while a downward arrow represents error.).

Dataset	Model	PM2.5			PM10			SO2			NO2		
		MSE↓	MAE↓	CORR↑	MSE↓	MAE↓	CORR↑	MSE↓	MAE↓	CORR↑	MSE↓	MAE↓	CORR↑
Beijing sites	PatchFormer [64]	0.3774	0.4437	0.741	0.3916	0.4615	0.719	0.4069	0.4634	0.715	0.4586	0.4962	0.688
	TimeMixer [44]	0.3752	0.4403	0.744	0.3928	0.4632	0.725	0.3954	0.4576	0.733	0.4435	0.4897	0.694
	DSformer [65]	0.3705	0.4384	0.759	0.3794	0.4521	0.743	0.3697	0.4398	0.761	0.4078	0.4756	0.702
	Airformer [66]	0.3657	0.4335	0.762	0.3897	0.4586	0.739	0.3742	0.4415	0.755	0.4029	0.4684	0.714
	MGSFformer	**0.3418**	**0.4196**	**0.784**	**0.3617**	**0.4362**	**0.765**	**0.3585**	**0.4249**	**0.771**	**0.3864**	**0.4533**	**0.729**
China cities	PatchFormer [64]	0.5737	0.4874	0.695	0.5845	0.5026	0.677	0.6045	0.5089	0.662	0.5432	0.4712	0.705
	TimeMixer [44]	0.5725	0.4852	0.702	0.5834	0.4983	0.682	0.6037	0.5076	0.667	0.5468	0.4753	0.717
	DSformer [65]	0.5584	0.4759	0.715	0.5782	0.4915	0.697	0.5821	0.4913	0.689	0.5413	0.4678	0.729
	Airformer [66]	0.5679	0.4804	0.723	0.5643	0.4879	0.706	0.5934	0.5024	0.674	0.5378	0.4625	0.731
	MGSFformer	**0.5374**	**0.4631**	**0.734**	**0.5519**	**0.4758**	**0.712**	**0.5617**	**0.4825**	**0.706**	**0.5126**	**0.4531**	**0.742**
China sites	PatchFormer [64]	0.6715	0.6276	0.594	0.6415	0.6043	0.621	0.6232	0.5804	0.639	0.6275	0.5892	0.638
	TimeMixer [44]	0.6764	0.6259	0.606	0.6394	0.6024	0.625	0.6298	0.5834	0.631	0.6198	0.5796	0.642
	DSformer [65]	0.6692	0.6238	0.615	0.6258	0.5871	0.644	0.6201	0.5779	0.649	0.6085	0.5687	0.663
	Airformer [66]	0.6561	0.6154	0.634	0.6314	0.5923	0.635	0.6184	0.5742	0.654	0.6104	0.5753	0.657
	MGSFformer	**0.6318**	**0.5944**	**0.685**	**0.6134**	**0.5719**	**0.708**	**0.6054**	**0.5627**	**0.715**	**0.5974**	**0.5584**	**0.714**

to effectively eliminate crucial redundant information, limiting the model's prediction ability.

(4) Complete removal of the RD block results in a model equivalent to a conventional multi-granularity ensemble learning model, leading to a substantial increase in prediction error. This outcome is attributed to the removal of essential components associated with redundant information elimination, impacting the model's analytical capability due to the presence of redundant information across multiple granularities.

4.5. Performance evaluation of different granularities

Considering that data with different granularities have different functions, this section evaluates the effect of fine-grained and coarse-grained data in the following ways: (1) The proposed model uses the

Table 6
Performance evaluation of different components.

Dataset	Model	Horizon 1		Horizon 4		Horizon 7		Average	
		MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓
Beijing sites	Pro-MLP	0.2245	0.3330	0.4396	0.5296	0.5429	0.5933	0.4146	0.4887
	Pro-GWO	0.2030	0.3256	0.4379	0.5142	0.5271	0.5829	0.4016	0.4776
	Pro-RL	0.2074	0.2968	0.4415	0.5065	0.4994	0.5657	0.3951	0.4597
	Pro-GRU	0.1916	0.2928	0.5205	0.5502	0.5036	0.5433	0.4175	0.4655
	Pro-SA	0.1825	0.2845	0.4038	0.4627	0.4693	0.5344	0.3679	0.4319
	Proposed	**0.1767**	**0.2788**	**0.3975**	**0.4536**	**0.4514**	**0.5311**	**0.3542**	**0.4246**
China cities	Pro-MLP	0.4558	0.4299	0.6517	0.5405	0.6711	0.5518	0.6052	0.5106
	Pro-GWO	0.4523	0.4233	0.6301	0.5335	0.6738	0.5481	0.5977	0.5048
	Pro-RL	0.4489	0.4203	0.6245	0.5312	0.6457	0.5499	0.5853	0.5037
	Pro-GRU	0.4287	0.4129	0.6426	0.5394	0.6964	0.5623	0.6015	0.5081
	Pro-SA	0.4015	0.3978	0.6078	0.5124	0.6235	0.5376	0.5649	0.4812
	Proposed	**0.3934**	**0.3840**	**0.5863**	**0.5046**	**0.6046**	**0.5187**	**0.5413**	**0.4694**
China sites	Pro-MLP	0.3907	0.4313	0.7151	0.6357	0.8826	0.7090	0.7217	0.6240
	Pro-GWO	0.3941	0.4348	0.7159	0.6423	0.8177	0.6906	0.6888	0.6215
	Pro-RL	0.3702	0.4338	0.7152	0.6285	0.7469	0.6484	0.6601	0.6049
	Pro-GRU	0.3755	0.4274	0.6818	0.6449	0.8435	0.7241	0.6760	0.6226
	Pro-SA	0.3704	0.4259	0.6532	0.6198	0.7235	0.6381	0.6449	0.5979
	Proposed	**0.3657**	**0.4225**	**0.6373**	**0.6124**	**0.7064**	**0.6255**	**0.6261**	**0.5890**

Table 7
The results of ablation experiments.

Model	Beijing sites		China cities		China sites	
	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓
Proposed	**0.3542**	**0.4246**	**0.5413**	**0.4694**	**0.6261**	**0.5890**
w/o RE	0.3879	0.4418	0.5947	0.5045	0.6941	0.6237
w/o DF	0.3816	0.4397	0.5923	0.5075	0.6831	0.6164
w/o IE	0.4013	0.4517	0.6137	0.5128	0.7694	0.6457
w/o RD	0.4109	0.4547	0.6314	0.5236	0.8012	0.6569

data with five granularities to achieve modeling. We appropriately remove some of these granularities to demonstrate the importance of multiple granularities. (2) We evaluate the prediction effect using single-granularity data and compare it with multi-granularity modeling. (3) In this paper, we employ an approach that aligns with the dimensions of the coarsest-grained data to help the RD block eliminate redundant information. To further assess the RD block, we also evaluated the impact of aligning dimensions with data of other granularities on the experimental results.

Table 8 displays the results of the proposed model utilizing features with different granularities (Beijing sites dataset). g1 to g5 denote five granularities from coarse to fine, respectively. Table 9 shows the experimental results under the alignment of dimensions of data with different granularities.

Based on the results, the following conclusions can be drawn:

(1) In comparison to single-granularity modeling, multi-granularity modeling can yield superior prediction outcomes, validating the effectiveness of the multi-granularity modeling concept. The primary reason for this improvement is that data with diverse granularities can provide additional information, enhancing the model's prediction accuracy.
(2) Fine-grained data significantly enhances the accuracy of short-term prediction, while coarse-grained data improves prediction accuracy for longer forecast intervals. Consequently, the effective utilization of multi-granularity data can boost the MGSFformer's performance in multi-step prediction tasks.
(3) As the granularity of the aligned data becomes finer, the prediction error of the model continuously increases. The main reason is that aligning coarse-grained data to fine-grained data requires initially mapping the coarse-grained data to a latent space using an MLP, which may cause misalignment of information redundancy positions, leading to decreased prediction performance.

Table 9
The experimental results under the alignment of dimensions of data with different granularities.

Aligned Granularity	Beijing sites		China cities		China sites	
	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓
g1	**0.3542**	**0.4246**	**0.5413**	**0.4694**	**0.6261**	**0.5890**
g2	0.3618	0.4305	0.5582	0.4739	0.6418	0.6014
g3	0.3785	0.4387	0.5729	0.4813	0.6872	0.6119
g4	0.3884	0.4454	0.5904	0.4952	0.7119	0.6253
g5	0.3978	0.4497	0.6045	0.5076	0.7429	0.6337

Table 8
Performance comparison of the proposed model using features with different granularities (Beijing site dataset). g1 to g5 represent five granularities from coarse to fine, respectively.

Granularity Combination	Horizon 1		Horizon 4		Horizon 7		Average	
	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓
g1	0.2237	0.3418	0.4789	0.5536	0.5416	0.6127	0.4228	0.5164
g2	0.2158	0.3392	0.4776	0.5524	0.5433	0.6137	0.4207	0.5138
g3	0.2119	0.3285	0.4718	0.5389	0.5341	0.6029	0.4153	0.5019
g4	0.2063	0.3207	0.4695	0.5236	0.5267	0.5978	0.4087	0.4829
g5	0.2018	0.3142	0.4625	0.5174	0.5152	0.5893	0.4029	0.4682
g1+g5	0.1953	0.3076	0.4349	0.4912	0.5024	0.5716	0.3916	0.4577
g1+g3+ g5	0.1916	0.2996	0.4186	0.4798	0.4873	0.5583	0.3827	0.4492
g2+g3+g4+g5	0.1834	0.2891	0.4029	0.4633	0.4716	0.5479	0.3728	0.4391
g1+g2+g3+g4	0.1897	0.2932	0.4058	0.4672	0.4685	0.5437	0.3794	0.4413
All	**0.1767**	**0.2788**	**0.3975**	**0.4536**	**0.4514**	**0.5311**	**0.3542**	**0.4246**

4.6. Visualization

To visually evaluate our model, we visualized the prediction results in temporal and spatial dimensions.

AQI time series visualization. Because these datasets contain air quality data from multiple sites, this paper shows the visualization of the first site. Fig. 4 shows the prediction results and the residual series of the proposed model. Based on the visual experimental results, it can be found that the proposed model is close to the real AQI time series. Specifically, the proposed model achieves satisfactory results whether predicting mutated values or smooth sequences.

Spatial distribution visualization. Fig. 5 shows the visualizations of spatial prediction results of several models on these three datasets. Based on the results, it can be found that AQI presents a distribution form with a high concentration in local areas and a low concentration in other areas. Compared with other models, the prediction results of the proposed model are closer to the real air quality data. Besides, the proposed model can predict the areas with high AQI, which provides a reference for environmental governance and personal protection.

4.7. Hyperparameter analysis

The setting of hyperparameters can significantly impact the performance of MGSFformer. Therefore, it is necessary to conduct a comprehensive analysis and evaluation of the impact of key hyperparameters. In this study, we evaluated four important hyperparameters: the number of multi-head attentions, dropout, dimensions H' after matching, and the length of input features. Fig. 6 presents the CORR values of MGSFformer under different hyperparameters. Based on the experimental results, the following conclusions are drawn:

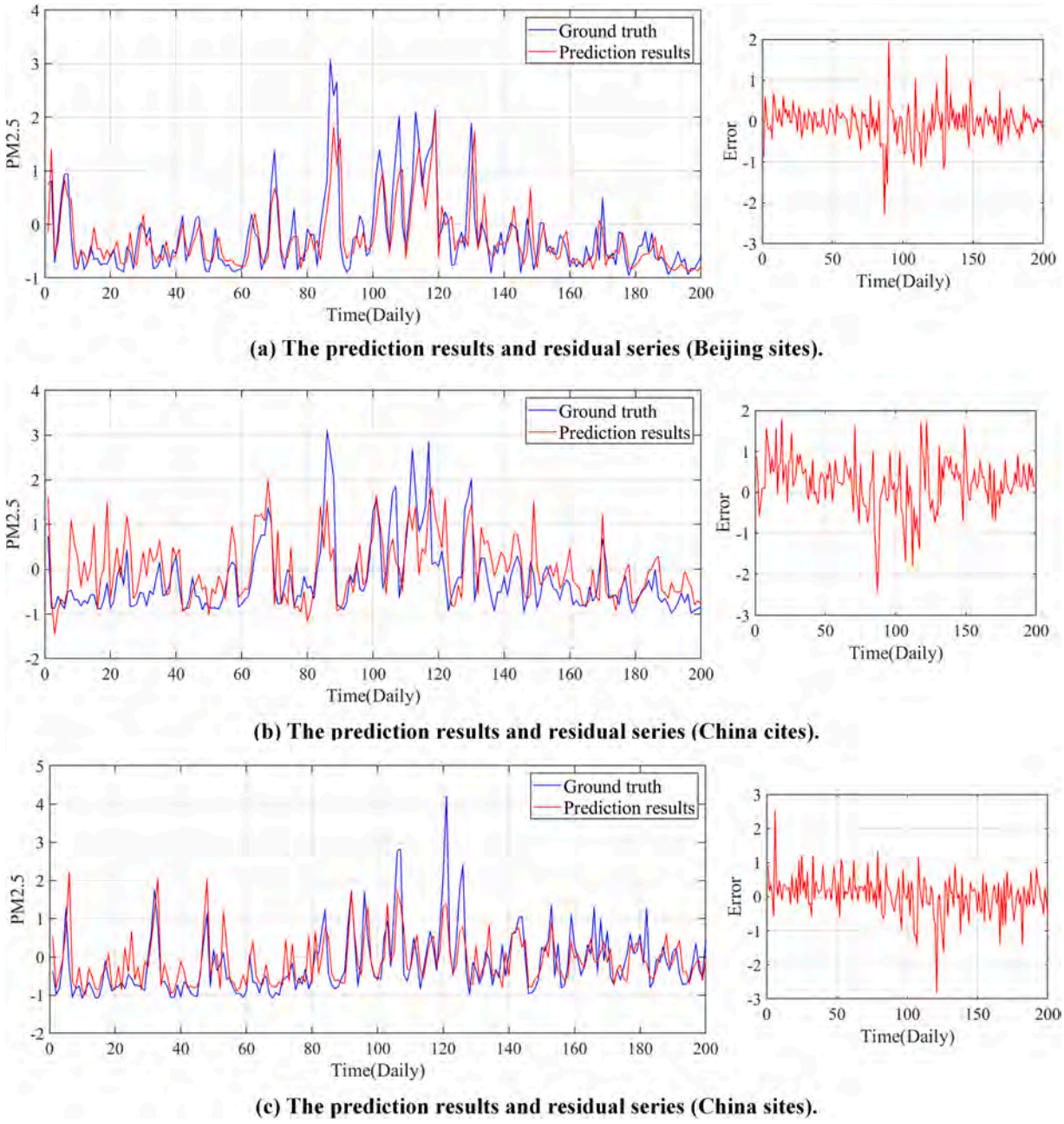

Fig. 4. The prediction results (left) and residual series (right) of the proposed model. The residual series represents the difference between the prediction results and the real value.

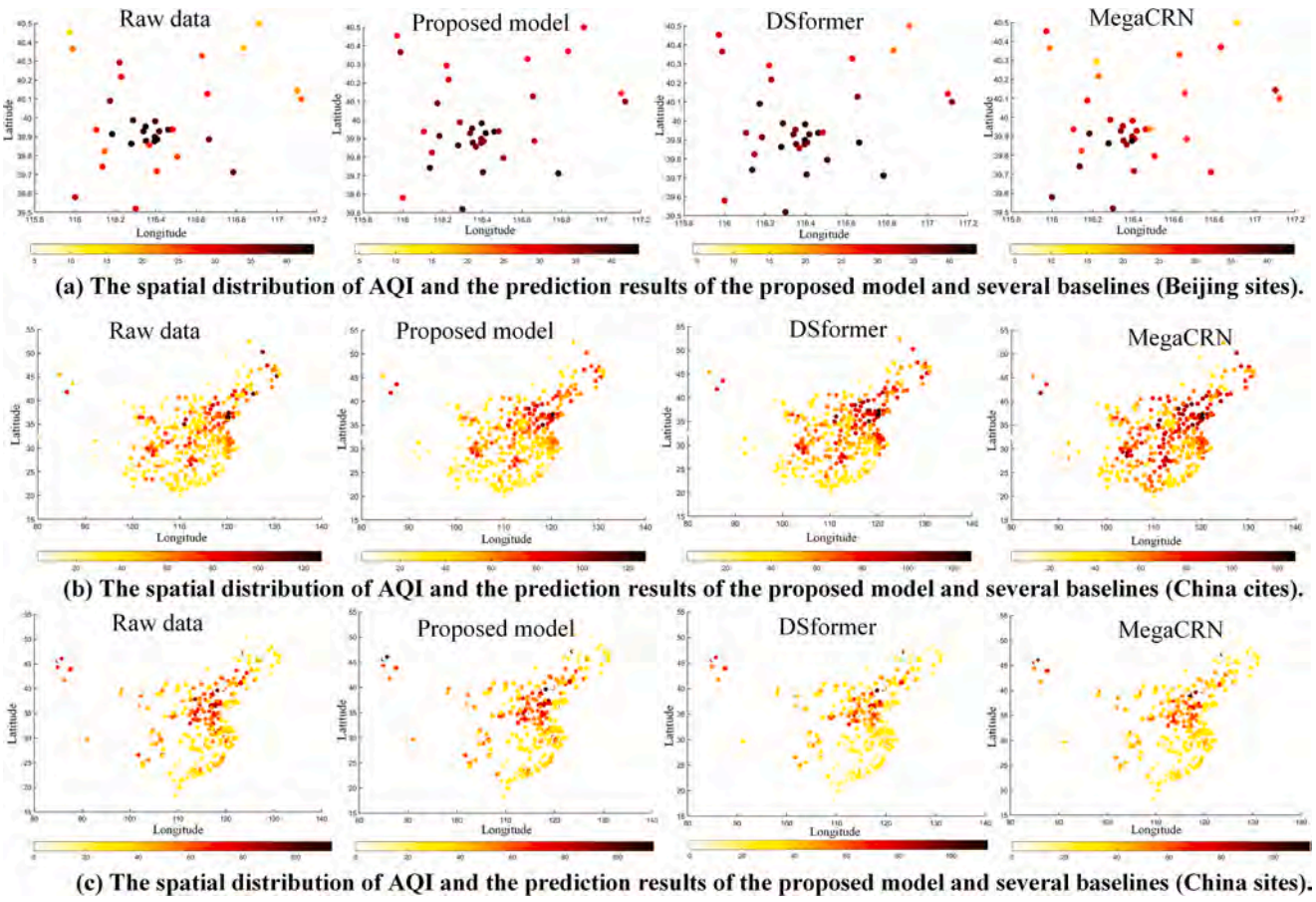

Fig. 5. The real spatial distribution of AQI and the prediction results of the proposed model and several baselines. Darker areas have higher concentrations of air quality.

(1) The number of multi-head attentions and dropout have a minor impact on the experimental results. Their main role is to affect the size of the model's parameters, but they do not lead to significant performance fluctuations.
(2) The length of input features significantly affects the predictive performance of MGSFformer. The primary reason is that the length of input features influences the amount of information MGSFformer can access. When the length of input features is large, MGSFformer tends to capture more noise, which can lead to overfitting. Conversely, when the length of input features is small, MGSFformer may not gather enough critical information, resulting in a decline in predictive performance.

4.8. Statistical test analysis

Statistical analysis is crucial for verifying the stability and reproducibility of a model. In this paper, we use the T-test to evaluate the performance of the MGSFformer compared to four baselines (Airformer, DSformer, MegaCRN, and AGCRN). Specifically, we employ the T-test to assess the residual sequences of the models, which represent the differences between the actual values and the predicted results. From this analysis, we obtained the mean, variance, and p-values. When the mean and variance of the prediction model are close to zero, it is more stable. A higher p-value indicates greater certainty in the prediction model. The results of the statistical tests are presented in Table 10. The experimental results show that the mean and variance of the residual series of MGSFformer are significantly lower than those of other baselines. In addition, the p-value of the residual series of MGSFformer is higher than that of other baselines. These results further demonstrate the stability and practicality of MGSFformer.

5. Conclusion and future work

The spatiotemporal prediction of Air quality can offer valuable references for environmental governance and personal protection. This study aims to leverage two crucial data patterns (multi-granularity and spatiotemporal correlation) of Air quality data and address existing technical challenges. To achieve this, we introduce the Multi-Granularity Spatiotemporal Fusion Transformer, comprising three intricately designed blocks: the residual de-redundant block, spatiotemporal attention block, and dynamic fusion block. This paper is summarized in the following three aspects:

Performance: The MGSFformer outperforms several state-of-the-art baselines on three real-world datasets. Additionally, ablation experiments and component replacement experiments underscore the significance of the designed blocks.

Application: The MGSFformer proficiently predicts the temporal trend and spatial distribution of Air quality, aiding governments and individuals in effective environmental management. Furthermore, experimental results indicate that fully recognizing and leveraging the two key data patterns (multi-granularity and spatiotemporal correlation) can significantly enhance prediction performance, offering novel insights into air quality spatiotemporal data pattern analysis.

Limits and future work: At present, we need to manually set the number of multi-granularity. In future research, we plan to optimize the quality of input features by appropriately filtering different granularities. We will explore decision-making techniques such as reinforcement learning [73,74] to achieve adaptive multi-granularity feature selection. In addition, we will consider researching optimization algorithms that can further achieve result fusion, such as the Linear Population Size Reduction Success-History based Adaptive Differential Evolution

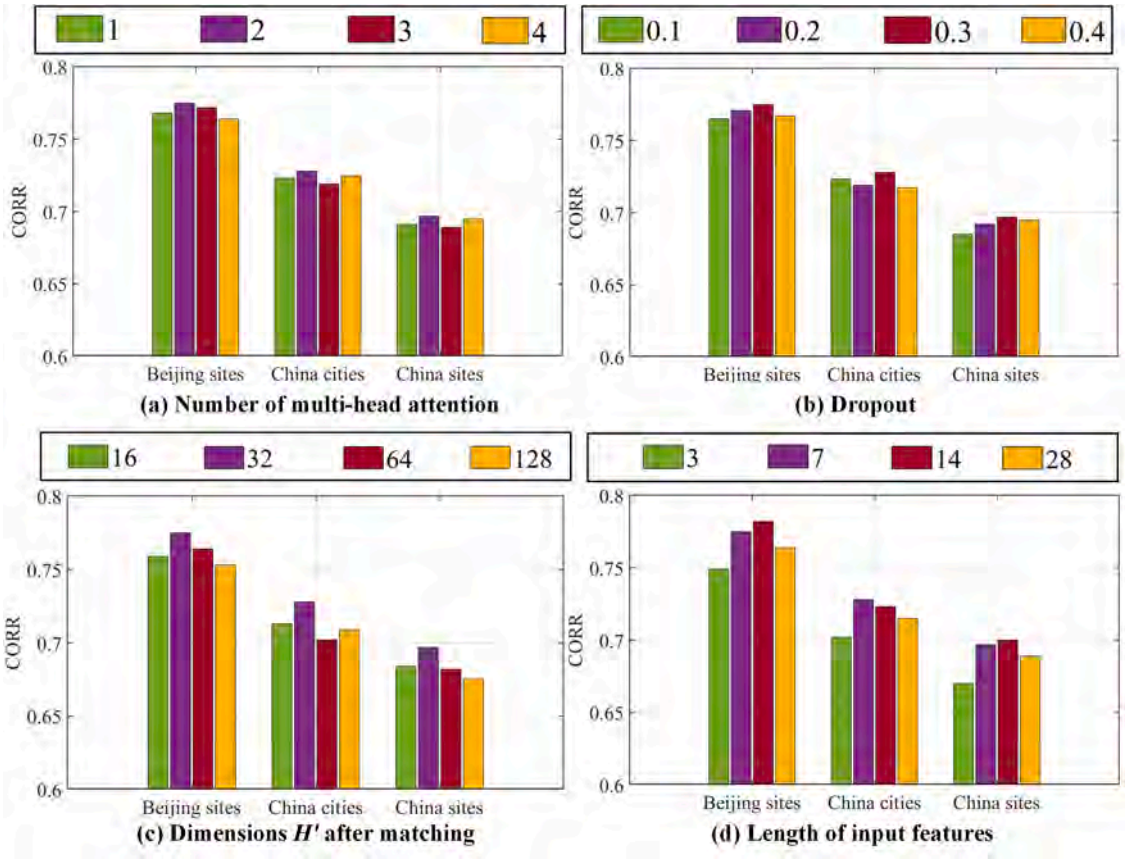

Fig. 6. The CORR values of MGSFformer under different hyperparameters.

Table 10
Statistical analysis of MGSFformer and other baselines.

Dataset	Models	Mean↓	Variance↓	P-value↑
Beijing sites	**MGSFformer**	**−0.0125**	**0.315**	**0.816**
	Airformer [66]	−0.0167	0.347	0.742
	DSformer [65]	−0.0179	0.358	0.736
	MegaCRN [62]	0.0204	0.411	0.682
	AGCRN [63]	0.0215	0.435	0.659
China cities	**MGSFformer**	**0.0157**	**0.356**	**0.775**
	Airformer [66]	−0.0194	0.407	0.717
	DSformer [65]	0.0188	0.398	0.724
	MegaCRN [62]	−0.0229	0.476	0.665
	AGCRN [63]	0.0244	0.497	0.646
China sites	**MGSFformer**	**0.0174**	**0.475**	**0.706**
	Airformer [66]	0.0216	0.519	0.649
	DSformer [65]	−0.0205	0.506	0.658
	MegaCRN [62]	0.0257	0.558	0.612
	AGCRN [63]	−0.0279	0.582	0.597

(L-SHADE) [75].

CRediT authorship contribution statement

Chengqing Yu: Writing – original draft, Validation, Methodology. **Fei Wang:** Writing – original draft, Funding acquisition, Conceptualization. **Yilun Wang:** Software, Formal analysis. **Zezhi Shao:** Writing – review & editing, Visualization. **Tao Sun:** Validation, Software. **Di Yao:** Visualization, Methodology. **Yongjun Xu:** Validation, Software, Conceptualization.

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.

Acknowledgments

This work is supported by NSFC No. 62372430, NSFC No. 62206266, and the Youth Innovation Promotion Association CAS No.2023112.

References

[1] M.A. Fadhel, A.M. Duhaim, A. Saihood, A. Sewify, M.N.A. Al-Hamadani, A. S. Albahri, L. Alzubaidi, A. Gupta, S. Mirjalili, Y. Gu, Comprehensive systematic review of information fusion methods in smart cities and urban environments, Inf. Fusion 107 (2024) 102317.
[2] W. Huang, T. Li, J. Liu, P. Xie, S. Du, F. Teng, An overview of air quality analysis by big data techniques: monitoring, forecasting, and traceability, Inf. Fusion 75 (2021) 28–40.
[3] H. Xing, G. Wang, C. Liu, M. Suo, PM2.5 concentration modeling and prediction by using temperature-based deep belief network, Neural Netw. 133 (2021) 157–165.
[4] Y. Xu, X. Liu, X. Cao, C. Huang, E. Liu, S. Qian, X. Liu, Y. Wu, F. Dong, C.-W. Qiu, Artificial intelligence: a powerful paradigm for scientific research, The Innovation 2 (2021).
[5] Z.L. Li, G.W. Zhang, J. Yu, L.Y. Xu, Dynamic graph structure learning for multivariate time series forecasting, Pattern Recognit. 138 (2023) 109423.
[6] C.-j. Li, Z. Qu, S.-y. Wang, L. Liu, A method of cross-layer fusion multi-object detection and recognition based on improved faster R-CNN model in complex traffic environment, Pattern Recognit. Lett. 145 (2021) 127–134.
[7] F. Wang, D. Yao, Y. Li, T. Sun, Z. Zhang, AI-enhanced spatial-temporal data-mining technology: new chance for next-generation urban computing, The Innovation 4 (2023).
[8] Y. Xu, F. Wang, Z. An, Q. Wang, Z. Zhang, Artificial intelligence for science—bridging data to wisdom, The Innovation 4 (2023) 100525.

[9] K. Trebing, T. Stańczyk, S. Mehrkanoon, SmaAt-UNet: precipitation nowcasting using a small attention-UNet architecture, Pattern Recognit. Lett. 145 (2021) 178–186.
[10] L. Miao, S. Tang, Y. Ren, M.-P. Kwan, K. Zhang, Estimation of daily ground-level PM2. 5 concentrations over the Pearl River Delta using 1 km resolution MODIS AOD based on multi-feature BiLSTM, Atmos. Environ. 290 (2022) 119362.
[11] Y. Li, L. Xue, Y. Tao, Y. Li, Y. Wu, Q. Liao, J. Wan, Y. Bai, Exploring the contributions of major emission sources to PM2. 5 and attributable health burdens in China, Environ. Pollut. 322 (2023) 121177.
[12] Y. Yu, H. Li, S. Sun, Y. Li, PM2. 5 concentration forecasting through a novel multi-scale ensemble learning approach considering intercity synergy, Sustain. Cities. Soc. 85 (2022) 104049.
[13] Q. Dai, J.-w. Liu, Y. Liu, Multi-granularity relabeled under-sampling algorithm for imbalanced data, Appl. Soft Comput. 124 (2022) 109083.
[14] Z. Xu, M. Bashir, W. Zhang, Y. Yang, X. Wang, C. Li, An intelligent fault diagnosis for machine maintenance using weighted soft-voting rule based multi-attention module with multi-scale information fusion, Inf. Fusion 86-87 (2022) 17–29.
[15] Q. Liu, H. Li, W.-l. Shang, K. Wang, Spatio-temporal distribution of Chinese cities' air quality and the impact of high-speed rail, Renew. Sustain. Energy Rev. 170 (2022) 112970.
[16] H. Hewamalage, C. Bergmeir, K. Bandara, Global models for time series forecasting: a Simulation study, Pattern Recognit. 124 (2022) 108441.
[17] H. Wang, H. Pan, K. Zhang, S. He, C. Chen, M2FNet: multi-granularity feature fusion network for medical visual question answering, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2022, pp. 141–154.
[18] T. Zhang, Y. Zhang, W. Cao, J. Bian, X. Yi, S. Zheng, J. Li, Less is more: fast multivariate time series forecasting with light sampling-oriented mlp structures, arXiv preprint arXiv:2207.01186, (2022).
[19] Z. Xia, X. Pan, S. Song, L.E. Li, G. Huang, Vision transformer with deformable attention, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 4794–4803.
[20] M. Panja, T. Chakraborty, U. Kumar, N. Liu, Epicasting: an Ensemble Wavelet Neural Network for forecasting epidemics, Neural Netw. 165 (2023) 185–212.
[21] X. Yan, B. Zhang, G. Liu, K. Fan, AHU sensor minor fault detection based on piecewise ensemble empirical mode decomposition and an improved combined neural network, Sci. Technol. Built. Environ. 28 (2022) 1184–1200.
[22] R. Gao, R. Li, M. Hu, P.N. Suganthan, K.F. Yuen, Online dynamic ensemble deep random vector functional link neural network for forecasting, Neural Netw. 166 (2023) 51–69.
[23] M. Su, H. Liu, C. Yu, Z. Duan, A novel AQI forecasting method based on fusing temporal correlation forecasting with spatial correlation forecasting, Atmos. Pollut. Res. 14 (2023) 101717.
[24] X. Guo, Y. Wang, S. Mei, C. Shi, Y. Liu, L. Pan, K. Li, B. Zhang, J. Wang, Z. Zhong, Monitoring and modelling of PM2. 5 concentration at subway station construction based on IoT and LSTM algorithm optimization, J. Clean. Prod. 360 (2022) 132179.
[25] T. Sun, F. Wang, Z. Zhang, L. Wu, Y. Xu, Human mobility identification by deep behavior relevant location representation, in: International Conference on Database Systems for Advanced Applications, Springer, 2022, pp. 439–454.
[26] T. Qian, Y. Chen, G. Cong, Y. Xu, F. Wang, AdapTraj: a multi-source domain generalization framework for multi-agent trajectory prediction, arXiv preprint arXiv:2312.14394, (2023).
[27] G. Zheng, H. Liu, C. Yu, Y. Li, Z. Cao, A new PM2. 5 forecasting model based on data preprocessing, reinforcement learning and gated recurrent unit network, Atmos. Pollut. Res. 13 (2022) 101475.
[28] X. Liu, M. Qin, Y. He, X. Mi, C. Yu, A new multi-data-driven spatiotemporal PM2. 5 forecasting model based on an ensemble graph reinforcement learning convolutional network, Atmos. Pollut. Res. 12 (2021) 101197.
[29] C. Yu, G. Yan, C. Yu, X. Mi, Attention mechanism is useful in spatio-temporal wind speed prediction: evidence from China, Appl. Soft Comput. 148 (2023) 110864.
[30] J. Deng, X. Chen, R. Jiang, D. Yin, Y. Yang, X. Song, I.W. Tsang, Disentangling structured components: towards adaptive, interpretable and scalable time series forecasting, in: IEEE Transactions on Knowledge and Data Engineering, 2024.
[31] C. Yu, G. Yan, C. Yu, X. Liu, X. Mi, MRIformer: a multi-resolution interactive transformer for wind speed multi-step prediction, Inf. Sci. (Ny) 661 (2024) 120150.
[32] M. Yu, A. Masrur, C. Blaszczak-Boxe, Predicting hourly PM2. 5 concentrations in wildfire-prone areas using a SpatioTemporal Transformer model, Sci. Total Environ. 860 (2023) 160446.
[33] Z. Shao, Z. Zhang, W. Wei, F. Wang, Y. Xu, X. Cao, C.S. Jensen, Decoupled dynamic spatial-temporal graph neural network for traffic forecasting, Proc. VLDB Endowment 15 (2022) 2733–2746.
[34] S. Ji, X. Lu, M. Liu, L. Sun, C. Liu, B. Du, H. Xiong, Community-based dynamic graph learning for popularity prediction, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, pp. 930–940.
[35] C. Song, Y. Lin, S. Guo, H. Wan, Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting, in: Proceedings of the AAAI conference on artificial intelligence, 2020, pp. 914–921.
[36] Y. Xu, L. Han, T. Zhu, L. Sun, B. Du, W. Lv, Generic dynamic graph convolutional network for traffic flow forecasting, Inf. Fusion 100 (2023) 101946.
[37] L. Sun, M. Liu, G. Liu, X. Chen, X. Yu, FD-TGCN: fast and dynamic temporal graph convolution network for traffic flow prediction, Inf. Fusion 106 (2024) 102291.
[38] X. Wu, J. Zhan, W. Ding, TWC-EL: a multivariate prediction model by the fusion of three-way clustering and ensemble learning, Inf. Fusion 100 (2023) 101966.
[39] M. Hou, C. Xu, Z. Li, Y. Liu, W. Liu, E. Chen, J. Bian, Multi-granularity residual learning with confidence estimation for time series prediction, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 112–121.
[40] H. Liu, Z. Duan, C. Chen, A hybrid multi-resolution multi-objective ensemble model and its application for forecasting of daily PM2. 5 concentrations, Inf. Sci. (Ny) 516 (2020) 266–292.
[41] X. Teng, X. Zhang, Z. Luo, Multi-scale local cues and hierarchical attention-based LSTM for stock price trend prediction, Neurocomputing 505 (2022) 92–100.
[42] C. Ji, C. Zhang, L. Hua, H. Ma, M.S. Nazir, T. Peng, A multi-scale evolutionary deep learning model based on CEEMDAN, improved whale optimization algorithm, regularized extreme learning machine and LSTM for AQI prediction, Environ. Res. 215 (2022) 114228.
[43] H. Yu, Z. Wang, Y. Xie, G. Wang, A multi-granularity hierarchical network for long- and short-term forecasting on multivariate time series data, Appl. Soft Comput. 157 (2024) 111537.
[44] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J.Y. Zhang, J. Zhou, Timemixer: decomposable multiscale mixing for time series forecasting, arXiv preprint arXiv:2405.14616, (2024).
[45] C. Li, X. Qi, Y. Yang, Z. Zeng, L. Zhang, J. Mao, FAST-CA: fusion-based Adaptive Spatial–Temporal Learning with Coupled Attention for airport network delay propagation prediction, Inf. Fusion 107 (2024) 102326.
[46] X. Mi, C. Yu, X. Liu, G. Yan, F. Yu, P. Shang, A dynamic ensemble deep deterministic policy gradient recursive network for spatiotemporal traffic speed forecasting in an urban road network, Digit. Signal Process. 129 (2022) 103643.
[47] H. Liu, Z. Long, Z. Duan, H. Shi, A new model using multiple feature clustering and neural networks for forecasting hourly PM2. 5 concentrations, and its applications in China, Engineering 6 (2020) 944–956.
[48] K. Liang, L. Meng, M. Liu, Y. Liu, W. Tu, S. Wang, S. Zhou, X. Liu, F. Sun, Reasoning over different types of knowledge graphs: static, temporal and multi-modal, arXiv preprint arXiv:2212.05767, (2022).
[49] J. Tan, H. Liu, Y. Li, S. Yin, C. Yu, A new ensemble spatio-temporal PM2. 5 prediction method based on graph attention recursive networks and reinforcement learning, Chaos, Soliton. Fractal. 162 (2022) 112405.
[50] Y. Liu, S. Rasouli, M. Wong, T. Feng, T. Huang, RT-GCN: gaussian-based spatiotemporal graph convolutional network for robust traffic prediction, Inf. Fusion 102 (2024) 102078.
[51] S.F. Ahmed, S.A. Kuldeep, S.J. Rafa, J. Fazal, M. Hoque, G. Liu, A.H. Gandomi, Enhancement of traffic forecasting through graph neural network-based information fusion techniques, Inf. Fusion 110 (2024) 102466.
[52] D.S.d.O. Santos Júnior, P.S.G. de Mattos Neto, J.F.L. de Oliveira, G.D.C. Cavalcanti, A hybrid system based on ensemble learning to model residuals for time series forecasting, Inf. Sci. (Ny) 649 (2023) 119614.
[53] Z. Shao, Z. Zhang, F. Wang, Y. Xu, Pre-training Enhanced Spatial-temporal graph neural network for multivariate time series forecasting, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 1567–1577.
[54] C. Yu, F. Wang, Z. Shao, T. Qian, Z. Zhang, W. Wei, Y. Xu, arXiv preprint, 2024.
[55] Z. Wu, S. Pan, G. Long, J. Jiang, X. Chang, C. Zhang, Connecting the dots: multivariate time series forecasting with graph neural networks, in: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, 2020, pp. 753–763.
[56] J. Chen, L. Xie, W. Lin, Y. Wu, H. Xu, Multi-granularity spatio-temporal correlation networks for stock trend prediction, IEEe Access. (2024).
[57] C. Yu, J. Tan, Y. Cheng, X. Mi, Data analysis and preprocessing techniques for air quality prediction: a survey, Stochastic Environ. Res. Risk Assess. 38 (2024) 2095–2117.
[58] J.H. Stock, M.W. Watson, Vector autoregressions, J. Econ. Perspect. 15 (2001) 101–115.
[59] G. Lai, W.-C. Chang, Y. Yang, H. Liu, Modeling long-and short-term temporal patterns with deep neural networks, in: The 41st international ACM SIGIR conference on research & development in information retrieval, 2018, pp. 95–104.
[60] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, M. Long, TimesNet: temporal 2D-variation modeling for general time series analysis, in: The Eleventh International Conference on Learning Representations, 2022.
[61] M. Liu, A. Zeng, M. Chen, Z. Xu, Q. Lai, L. Ma, Q. Xu, SCINet: time series modeling and forecasting with sample convolution and interaction, Adv. Neural Inf. Process. Syst. 35 (2022) 5816–5828.
[62] R. Jiang, Z. Wang, J. Yong, P. Jeph, Q. Chen, Y. Kobayashi, X. Song, S. Fukushima, T. Suzumura, Spatio-temporal meta-graph learning for traffic forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2023, pp. 8078–8086.
[63] L. Bai, L. Yao, C. Li, X. Wang, C. Wang, Adaptive graph convolutional recurrent network for traffic forecasting, Adv. Neural Inf. Process. Syst. 33 (2020) 17804–17815.
[64] P. Chen, Y. Zhang, Y. Cheng, Y. Shu, Y. Wang, Q. Wen, B. Yang, C. Guo, Pathformer: multi-scale transformers with adaptive pathways for time series forecasting, arXiv preprint arXiv:2402.05956, (2024).
[65] C. Yu, F. Wang, Z. Shao, T. Sun, L. Wu, Y. Xu, Dsformer: a double sampling transformer for multivariate time series long-term prediction, in: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, 2023, pp. 3062–3072.
[66] Y. Liang, Y. Xia, S. Ke, Y. Wang, Q. Wen, J. Zhang, Y. Zheng, R. Zimmermann, Airformer: predicting nationwide air quality in china with transformers, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2023, pp. 14329–14337.
[67] Y. Wang, Z. Shao, T. Sun, C. Yu, Y. Xu, F. Wang, Clustering-property matters: a cluster-aware network for large scale multivariate time series forecasting, in: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, 2023, pp. 4340–4344.

[68] P. Shang, X. Liu, C. Yu, G. Yan, Q. Xiang, X. Mi, A new ensemble deep graph reinforcement learning network for spatio-temporal traffic volume forecasting in a freeway network, Digit. Signal Process. 123 (2022) 103419.
[69] Z. Shao, Z. Zhang, F. Wang, W. Wei, Y. Xu, Spatial-temporal identity: a simple yet effective baseline for multivariate time series forecasting, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 4454–4458.
[70] Z. Shao, F. Wang, Y. Xu, W. Wei, C. Yu, Z. Zhang, D. Yao, G. Jin, X. Cao, G. Cong, Exploring progress in multivariate time series forecasting: comprehensive benchmarking and heterogeneity analysis, arXiv preprint arXiv:2310.06119, (2023).
[71] C. Yu, J. Qiao, C. Chen, C. Yu, X. Mi, TFEformer: a new temporal frequency ensemble transformer for day-ahead photovoltaic power prediction, J. Clean. Prod. 448 (2024) 141690.
[72] L. Shen, Y. Wei, Y. Wang, GBT: two-stage transformer framework for non-stationary time series forecasting, Neural Netw. 165 (2023) 953–970.
[73] Y. Zhou, Z. Yang, Q. Sun, C. Yu, C. Yu, An artificial intelligence model based on multi-step feature engineering and deep attention network for optical network performance monitoring, Optik (Stuttg) 273 (2023) 170443.
[74] M. Fan, Y. Wu, T. Liao, Z. Cao, H. Guo, G. Sartoretti, G. Wu, Deep reinforcement learning for uav routing in the presence of multiple charging stations, IEEe Trans. Veh. Technol. 72 (2022) 5732–5746.
[75] Z. Ma, G. Wu, P.N. Suganthan, A. Song, Q. Luo, Performance assessment and exhaustive listing of 500+ nature-inspired metaheuristic algorithms, Swarm. Evol. Comput. 77 (2023) 101248.