Temporal Difference-based Graph Transformer Networks for Air Quality PM2.5 Prediction: A Case Study in China

Zhen Zhang  
Taizhou University

Shiqing Zhang  
Taizhou University  https://orcid.org/0000-0001-8184-5088

Xiaoming Zhao  (✉ tzxyzxm@163.com)  
Taizhou University

Linjian Chen  
Yuhuan branch of Taizhou Ecological Environment Bureau

Jun Yao  
Taizhou University

Research Article

Keywords: Air quality prediction, deep learning, temporal difference, graph attention, transformer, long-term dependency

Posted Date: January 11th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1168251/v1

License: ☕️  This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Temporal difference-based graph transformer networks for air quality PM2.5 prediction: a case study in China

Zhen Zhang¹, Shiqing Zhang², Xiaoming Zhao*, Linjian Chen³, and Jun Yao¹

¹Zhejiang Provincial Key Laboratory of Evolutionary Ecology and Conservation, Taizhou University, Taizhou 318000, Zhejiang, China
²Institute of Intelligent Information Processing, Taizhou University, Taizhou 318000, Zhejiang, China
³Yuhuan branch of Taizhou Ecological Environment Bureau, Yuhuan 317600, Zhejiang, China

Abstract: The acceleration of industrialization and urbanization has recently brought about serious air pollution problems, which threaten human health and lives, the environmental safety, and sustainable social development. Air quality prediction is an effective approach for providing early warning of air pollution and supporting cleaner industrial production. However, existing approaches have suffered from a weak ability to capture long-term dependencies and complex relationships from time series PM2.5 data. To address this problem, this paper proposes a new deep learning model called temporal difference-based graph transformer networks (TDGTN) to learn long-term temporal dependencies and complex relationships from time series PM2.5 data for air quality PM2.5 prediction. The proposed TDGTN comprises of encoder and decoder layers associated with the developed graph attention mechanism. In particular, considering the similarity of different time moments and the importance of temporal difference between two adjacent moments for air quality prediction, we first construct graph-structured data from original time series PM2.5 data at different moments without explicit graph structure. Then, based on the constructed graph, we improve the self-attention mechanism with the temporal difference information, and develop a new graph attention mechanism. Finally, the developed graph attention mechanism is embedded into the encoder and decoder layers of the proposed TDGTN to learn

*Corresponding author. Tel: +86 57688661937.
E-mail:tzxyzxm@163.com
Address: Institute of Intelligent Information Processing, Taizhou University, Taizhou 318000, P.R. China
long-term temporal dependencies and complex relationships from a graph prospective on air
quality PM2.5 prediction tasks. To verify the effectiveness of the proposed method, we conduct
air quality prediction experiments on two real-world datasets in China, such as Beijing PM2.5
dataset ranging from 01/01/2010 to 12/31/2014 and Taizhou PM2.5 dataset ranging from
01/01/2017 to 12/31/2019. Compared with other air quality forecasting methods, such as
autoregressive moving average (ARMA), support vector regression (SVR), convolutional neural
network (CNN), long short-term memory (LSTM), the original Transformer, our experiment
results indicate that the proposed method achieves more accurate results on both short-term (1
hour) and long-term (6, 12, 24, 48 hours) air quality prediction tasks.

*Key words*: Air quality prediction; deep learning; temporal difference; graph attention; transformer;
long-term dependency
1. Introduction

With the rapid development of economy, industrialization, and urbanization, a large number of urban cities throughout the world are undergoing increasingly serious air pollution problems, thereby threatening human health and lives, the environment, and sustainable social development (Darçın 2014, Ke et al. 2021). In particular, long exposure to polluted air leads to a variety of cardiovascular and respiratory sicknesses like lung cancer, bronchial asthma, atherosclerosis, chronic obstructive pulmonary diseases, etc (Chang et al. 2020a, Schwartz 1993, Yan et al. 2020). Wherein, PM2.5 (fine particulate with an aerodynamic diameter of 2.5 μm or smaller) has become the primary factor of air pollution, and the increasing PM2.5 concentration directly threatens to human health (Zhang et al. 2020a). As a result, real-time, accurate and long-term PM2.5 concentration predictions in advance play a significant role in preventing and curbing air pollution, government decision-making, as well as protecting human health, and so on.

So far, a large number of studies have explored the performance of various methods for air quality PM2.5 prediction (Liao et al. 2020, Liu et al. 2021). Prior methods for air quality prediction can be mainly grouped into two categories, namely physical prediction methods and statistical prediction methods. The physical prediction methods are a numerical simulation model on the basis of aerodynamics, atmospheric physics, and chemical reactions for studying pollutant diffusion mechanism (Geng et al. 2015). The well-known physical prediction models include chemical transport models (CTMs) (Mihailovic et al. 2009, Ponomarev et al. 2020), community multiscale air quality (CMAQ) (Zhang et al. 2014), weather research and forecasting (WRF) (Powers et al. 2017), the GEOS-Chem model (Lee et al. 2017), and so on. Nevertheless, owing to the complicated pollutant diffusion mechanism, leveraging these models leads to several limitations such as expensive computation, the complexity of processing, uncertainty of
parameters, and low prediction accuracy (Wang et al. 2019a). Statistical prediction methods employ a statistical modeling strategy to forecast future air quality on the basis of the observed historical time series PM2.5 data. In comparison with physical prediction methods, statistical prediction methods have low computation since they can avoid the complicated pollutant diffusion mechanism. In this case, they can still obtain competitive performance to physical prediction methods on air quality PM2.5 prediction tasks (Suleiman et al. 2019). Owing to these advantages, leverage statistical prediction methods for air quality PM2.5 prediction is more extensive. There are two kinds of statistical prediction models: linear and nonlinear. The commonly-used linear statistical prediction models, which is based on the supposed linearity of real-world observed data, are autoregressive moving average (ARMA) (Graupe et al. 1975), and autoregressive integrated moving average (ARIMA) (Cekim 2020, Jian et al. 2012). Considering the nonlinearity of real-world observed data, the conventional nonlinear statistical prediction methods are machine learning (ML) models. At present, Various common machine learning (ML) algorithms, including multiple linear regression (MLR) (Donnelly et al. 2015), artificial neural network (ANN) (Agarwal et al. 2020, Arhami et al. 2013), support vector regression (SVR) (Chu et al. 2021, Yang et al. 2018), random forest (RF) (Gariazzo et al. 2020), as well as ensemble learning of multiple ML models (Xiao et al. 2018), have been employed for air quality PM2.5 prediction. Among these models, ANN has become the most popular one. The widely-used ANN models contain multilayer perceptron (MLP) (Feng et al. 2015), back propagation neural network (BPNN) (Wang et al. 2019b), and general regression neural network (GRNN) models (Zhou et al. 2014). These ML methods have distinct mathematical logic in which the correlation between input and output data is relatively definite. Additionally, they have relatively shallow network structure, resulting in the
limited ability of modelling dependency on time series PM2.5 data.

To address the above-mentioned issue, the recently-emerged deep learning (Hinton &Salakhutdinov 2006, LeCun et al. 2015) methods may provide a possible clue. With the aid of multi-layer network architecture, deep learning algorithms are able to automatically extract multiple levels of abstract feature representations from input data. Due to such powerful feature learning capability, deep learning methods have made great breakthroughs (LeCun et al. 2015, Pouyanfar et al. 2018) in object detection and image classification, natural language processing (NLP), speech signal processing, and so on.

In recent years, various deep learning models have also been successfully employed for air quality PM2.5 prediction (Aggarwal &Toshniwal 2021, Liao et al. 2020, Zaini et al. 2021). In particular, Ragab et al., presented a method of air pollution index (AQI) prediction by means of using one-dimensional convolutional neural network (1D-CNN) and exponential adaptive gradients optimization for Klang city, in Malaysia (Ragab et al. 2020). In addition, recurrent neural network (RNN) (Elman 1990), and its variants such as long short term memory (LSTM) (Hochreiter &Schmidhuber 1997) and gated recurrent unit (GRU) (Chung et al. 2014), have become popular techniques for forecasting time series PM2.5 data. This is attributed to the fact these RNN-based models have excellent capability of capturing temporal dependency from input time series PM2.5 data. A bidirectional LSTM (BiLSTM) consisting of both forward and backward LSTM units was provided for univariate air quality PM2.5 prediction (De Melo et al. 2019). In this study, they also adopted transfer learning techniques to further improve air quality prediction performance at wider daily and weekly temporal intervals. Jin et al., proposed a new model integrating multiple nested long short term memory networks (MN-LSTMs) for accurate
AQI forecasting enlightened with the federated learning (Jin et al. 2021).

At present, several hybrid deep learning framework (Aggarwal & Toshniwal 2021, Chang et al. 2020b, Du et al. 2021, Zhang et al. 2021) have attracted extensive attention for air quality PM2.5 forecasting. Specially, a hybrid deep learning model, based on one-dimensional CNNs (1D-CNN) and bidirectional LSTMs for spatial-temporal feature learning, was developed for air quality prediction (Du et al. 2021). This hybrid deep learning framework focused on learning the spatial-temporal correlation features and interdependence of multivariate air quality data. A spatio-temporal CNN and LSTM (CNN-LSTM) model (Pak et al. 2020) was provided to forecast the next day’s daily average PM2.5 concentration in Beijing City. In this CNN-LSTM model, the mutual information (MI) was used for the spatio-temporal correlation analysis, which took into account both the linear and nonlinear correlation between target and observed parameter values. A new spatial-temporal deep learning method with bidirectional gated recurrent unit integrated with attention mechanism (BiAGRU) (Zhang et al. 2020b), was proposed for accurate air quality forecasting. A hierarchical deep learning framework comprising of three components like the encoder, STAA-LSTM, and the decoder was presented for forecasting the real-world air quality data of Delhi (Abirami & Chitra 2021). In their work, the encoder was used to encode all spatial relations from input data. The STAA-LSTM, as a variant of LSTM, aimed to forecast future spatiotemporal relations in the latent space. The decoder was leveraged to decode these relations for the actual forecasting.

In addition, graph neural networks (GNN) (Scarselli et al. 2008) have become an emerging active research subject in machine learning, and obtained great success in processing graph-structured data owing to their powerful graph-based feature learning ability. The
representative GNN method is graph convolutional neural network (GCN) (Kipf & Welling 2016). GCN is a generalization of conventional CNNs to deal with homogeneous graph, in which the graph nodes and edge types should be identical. Considering the fact that different air quality monitoring stations may have different topological structure in space, GCNs can be intuitively used to capture the spatial dependencies among multiple air quality monitoring stations. Specially, Xu et al. proposed a hierarchical GCN Method called HighAir (Xu et al. 2021) for air quality prediction, in which a city graph and station graphs were constructed to take into account the city-level and station-level patterns of air quality, respectively. Chen et al. presented the group-aware graph neural network (GAGNN) (Chen et al. 2021) for nationwide city air quality prediction. GAGNN aimed to build up a city graph and a city group graph to learn the spatial and latent dependencies between cities, respectively. In addition, combining GNN with LSTM has recently become a popular method to model spatio-temporal dependencies for air quality PM2.5 forecasting. Specially, a graph-based LSTM (GLSTM) model (Gao & Li 2021) was developed to forecast PM2.5 concentration in Gansu Province of Northwest China. They regarded all air quality monitoring stations as a graph, and yielded a parameterized adjacency matrix based on the adjacency matrix of the graph. Then, integrating the parameterized adjacency matrix with LSTMs was employed to learn spatio-temporal dependencies for air quality PM2.5 prediction. These existing graph-based works aim to capture the spatial dependencies among multiple air quality monitoring stations rather than the single air quality monitoring station.

More recently, the attention mechanism (Niu et al. 2021) has become an important direction in the field of deep learning. In particular, the temporal attention mechanism is capable of adaptively assigning greater weights to input data at different times from a sequence with higher
correlations for target prediction tasks. Moreover, it can be also calculated in parallel, thereby improving the computational efficiency. Among attention-based deep learning methods, the recently-developed Transformer (Vaswani et al. 2017) technique achieving great success for machine translation tasks in NLP, has become fashionable at present. The original Transformer model does not contain any recurrent structures and convolutions and aims to model temporal dependencies in machine translation tasks with the aid of the powerful self-attention mechanism. So far, the Transformer models have exhibit better performance than RNN and LSTM in capable of learning long-range dependencies in a number of areas ranging from NLP (Neishi & Yoshinaga 2019, Vaswani et al. 2017), object detection and classification (Bazi et al. 2021, Duke et al. 2021, Lanchantin et al. 2021), to electricity consuming load analysis (Yue et al. 2020, Zhou et al. 2021). Although the recently-emerged Transformer techniques have achieved promising performance in various domains, few studies focus on the applications of Transformer techniques to air quality PM2.5 prediction. Additionally, how to integrate the advantages of GNNs and Transformer techniques for air quality PM2.5 prediction is a challenging problem, which is under-exploited in existing works.

Inspired by the recent great success of GNNs and Transformer, this paper combines the advantages of GNNs processing graph-structured data and Transformer modeling temporal dependencies, and proposes a novel graph attention-based deep learning model called temporal difference-based graph transformer networks (TDGTN) for air quality PM2.5 prediction. The main contributions of this paper are summarized as follows:

(1) Considering the similarity of different time moments and the importance of temporal difference between two adjacent moments for air quality prediction, for the single air monitoring
station we aim to construct graph-structured data from the obtained time series PM2.5 data at
different moments without explicit graph structure. To the best of our knowledge, this is the first
attempt to exploit graph-based air quality PM2.5 prediction for the single air monitoring station
from a graph-based perspective.
(2) This paper combines the advantages of both GNNs and Transformer, and proposes a new
deep learning model called TDGTN to learn long-term temporal dependencies and complex
relationships from time series PM2.5 data for air quality PM2.5 prediction. In the proposed
TDGTN model, we improve the self-attention mechanism with the temporal difference
information and develop a new graph attention mechanism.
(3) This paper evaluates the performance of the proposed TDGTN on two real-world datasets,
in China, including Beijing and Taizhou PM2.5 datasets and compares it with the state-of-the-art
models such as ARMA, SVR, CNN, LSTM, and the original Transformer. Experimental results
demonstrate that TDGTN outperforms existing models both short-term (1 hour) and long-term (6,
12, 24, 48 hours) air quality prediction tasks.
2. Data and methods
2.1 Study area and data collection
To verify the effectiveness of the proposed method on air quality prediction tasks, we adopt two
real-world hourly air quality PM2.5 datasets to perform air quality prediction experiments. They
are Beijing PM2.5 dataset (Liang et al. 2015), and Taizhou PM2.5 dataset. Fig.1 shows the
location of Beijing and Taizhou air quality monitoring stations in China. In particular, as the
Capital of China, Beijing city is located at 116°66 east longitude and 40°13 north latitude.
Taizhou is located at 121°42′ east longitude and 28°65′ north latitude. Taizhou city lies in the southeast of Zhejiang Province, China.

Fig. 1 The location of Beijing and Taizhou air quality monitoring stations in China

The used Beijing PM2.5 dataset contains around 43,800 samples, each of which was recorded with an hourly interval ranging from 01/01/2010 to 12/31/2014. In this dataset, the PM2.5 data (http://www.mee.gov.cn/) was collected from the US Embassy in Beijing, and the corresponding meteorological data (http://tianqi.2345.com/) was collected from Beijing Capital International Airport. This dataset comprises of eight feature items, including PM2.5 concentration (ug/m$^3$), dew point, temperature, pressure, combined wind direction, cumulated wind speed (m/s), cumulated hours of snow, cumulated hours of rain. Fig. 2 presents an illustration of hourly PM2.5 values from 5/01/2014 to 5/31/2014 on Beijing PM2.5 dataset.

The used Taizhou PM2.5 dataset contains about 26,000 hourly records ranging from 01/01/2017 to 12/31/2019. They were collected by our teams from the single Hongjia monitoring station,
which is located in Jiaojiang urban district from Taizhou city in the southeast of Zhejiang Province.

This dataset also include eight feature items, such as PM2.5 concentration (ug/m$^3$), dew point, temperature, pressure, combined wind direction, cumulated wind speed (m/s), cumulated hours of rain, cumulated hours of relative humidity. Fig. 3 provides an illustration of hourly PM2.5 values from 10/01/2019-10/31/2019 on Taizhou PM2.5 dataset.

Fig. 2 An illustration of hourly PM2.5 values (ug/m$^3$) from 5/01/2014 to 5/31/2014 on Beijing PM2.5 dataset (Each observation point in the horizontal axis denotes a timescale (hour) related to the collected PM2.5 value, as described in the vertical axis in this figure)

Fig. 3 An illustration of hourly PM2.5 values (ug/m$^3$) from 10/01/2019-10/31/2019 on Taizhou PM2.5 dataset (Each observation point in the horizontal axis denotes a timescale (hour) related to the collected PM2.5 value, as described in the vertical axis in this figure)

2.2 Method

Fig. 4 presents an overview of our proposed temporal difference-based graph transformer
networks (TDGTN) for air quality PM2.5 prediction. Like the original Transformer (Vaswani et al. 2017), our proposed TDGTN model comprises of encoder and decoder layers associated with the graph attention mechanism, as depicted in Fig.4. Compared with the original Transformer (Vaswani et al. 2017), TDGTN has two distinct properties. One is that we embed the graph attention into the encoder and decoder instead of the common multi-head attention in the original Transformer except for the used masked multi-head attention. The other is that based on time series PM2.5 data, the first-order backward difference information between two adjacent moments is embedded into the constructed graph so as to learn long-term dependency and complex relationships from a graph perspective. In the following, we will describe the relevant details of TDGTN.

Fig.4 An overview of our proposed temporal difference-based graph transformer networks (TDGTN) for air quality PM2.5 prediction

2.2.1 Problem description

Given a length $L_x$ of input time series data $X = \{x_1, x_2, \ldots, x_{L_x}\}$ ($x_i \in \mathbb{R}^{d_x}$) with feature
dimension $d_y$, the used air quality prediction methods aim to forecast the corresponding time
series PM2.5 data $Y = \{y_1, y_2, \ldots, y_L\}$ (where $y_i \in \mathbb{R}^{d_y}$) with a length $L_y$ and feature dimension $d_y$. The encoder aims to learn hidden continuous feature representations $Z = \{z_1, z_2, \ldots, z_L\}$ with a length $L_z$ from input time series data $X$. Then, the decoder produces an output of $Y = \{y_1, y_2, \ldots, y_L\}$ from the obtained hidden continuous feature representations in the encoder.

This inference is performed by means of an step-by-step implementation, in which the decoder computes a new hidden feature representation $z_{k+1}$ from the previous feature representation $z_k$ and other outputs in $k$-th step, and then predict $(k+1)$-th time series data $y_{k+1}$.

### 2.2.2 Graph construction

Graphs, as a special form of data, aim to characterize the relationships between different entities. GNNs endow each node in a graph with an ability of learning its neighborhood context by means of propagating information through graph-based structures. In this case, air quality PM2.5 prediction for the single air monitoring station can be intuitively regarded as a problem of graph-based multivariate time series forecasting from a view point of graphs. Considering the similarity of different time moments and the importance of temporal difference between two adjacent moments for air quality prediction, for the single air monitoring station we first construct graph-structured data from the obtained time series PM2.5 data at different moments without explicit graph structure, as described below. Based on the constructed graph-structured data, modeling time series PM2.5 data from a graph prospective may be a good way to maintain their temporal trajectory while exploring the temporal dependencies among time series PM2.5 data.

A graph is defined as $G = (\Lambda, \Gamma)$ where $\Lambda$ represents its nodes, and $\Gamma$ denotes its edges. The number of nodes in a graph is denoted by $n$. The graph adjacency matrix $U \in \mathbb{R}^{n \times n}$
is used to characterize the relationships among nodes.

Fig.5 (a) an illustration of constructing graph-structured data from time series PM2.5 data at four different moments, (b) the graph adjacency matrix is computed by using a Hadamard product $A \circ E$ between the attention scores $A$ and first-order backward difference matrix $E$.

As shown in Fig.5, the moment $t_i$ ($i = 1, 2, \ldots, n$) from time series PM2.5 data can be regarded as the $i$-th node in a graph, and they are interconnected by using their hidden dependency relationships. Therefore, all the nodes in a graph can be defined as

$$\Lambda = (t_1, t_2, \ldots, t_n)$$

(1)

The graph adjacency matrix is computed by a Hadamard product

$$U = A \circ E$$

(2)

where $A \in R^{n \times n}$ denotes the initial multiplicative attention scores calculated by $A = XX'$, and $E \in R^{n \times n}$ represents the first-order backward difference matrix, which is obtained by

$$E = E' \cdot W$$

(3)

$$E' = (\nabla_{10}, \nabla_{21}, \nabla_{32}, \ldots, \nabla_{n,n-1})$$

(4)

$$\nabla_{i,i-1} = x_i - x_{i-1}, i = 1, 2, \ldots, n$$

(5)

where $W$ is the linear transformation of the original first-order backward difference matrix $E' \in R^{n \times 1}$, and $\nabla_{i,i-1}$ is the first-order backward difference between two adjacent nodes in a graph representing the meteorological dynamical changes between two adjacent moments. In this
way, the edge values in $\Gamma$ from a graph correspond to the element values in the produced graph adjacency matrix $U$, as depicted in Fig.5 (a).

It’s worth pointing that there are two distinct properties about our constructed graph-structured data from original time series PM2.5 data without explicit graph structures. First, the attention score $A$ is used to weigh the similarity of different nodes (i.e., time moments). The higher the similarity of different nodes is, the larger the attention score values are. Moreover, the dynamical changing information in time series PM2.5 data between two adjacent moments is very important for air quality prediction. Therefore, the Hadamard product $A \odot E$ between the attention score $A$ and first-order backward difference matrix $E$ is designed to weigh the similarity of different nodes, and the dynamical changing in time series PM2.5 data simultaneously. Second, it is known that in long-term time series data the obtained information with close nodes (such as two neighboring nodes) is more important for air quality prediction than the obtained information with far nodes. Multiplying the attention scores of far nodes with the first-order backward difference of neighboring nodes, thus makes the graph adjacency matrix $U$ not only focuses on the information of far nodes, but also pays more attention to the neighboring nodes.

After constructing graph-structured data, the self-attention mechanism in Transformer (Vaswani et al. 2017) can be operated on these graph-structured data. Then, we improve the self-attention mechanism with the temporal difference information. This yields our graph attention mechanism used in the proposed TDGTN model for learning long-term temporal dependencies from a graph prospective on air quality PM2.5 prediction tasks.

2.2.3 The details of TDGTN model
Similar to the original Transformer (Vaswani et al. 2017), TDGTN contains encoder and decoder blocks with the developed graph attention mechanism, as described in Fig.4.

**Encoder:** The encoder is composed of a graph attention layer and a fully connected feed-forward network layer. Around each of two sub-layers, the residual connection (He et al. 2016) is adopted, each of them is followed by an addition and layer-normalization layer (Add&Norm). Given input time series data $X$, the encoder aims to learn the interrelationship of PM2.5 related data in time series data from a graph perspective.

**Decoder:** The decoder comprises of a masked multi-head attention layer, a graph attention layer and a fully connected feed-forward network layer, and each of them is followed by an addition and layer-normalization layer (Add&Norm). Similar to the encoder, Similar to the encoder, the residual connection is employed around each of two sub-layers. The decoder accepts the input time series data $X_{de} = \{X_{token}, X_0\}$, in which $X_{token}$ denotes the started tokens, and $X_0$ represents the placeholder for target time series data. The decoder aims to produce the output of predicted PM2.5 concentration data in a generative manner based on the obtained hidden continuous feature representations in the encoder.

**Graph attention:** Based on the constructed graph, we improve the self-attention mechanism with the temporal difference information and embed it into the produced graph-structured data to calculate the hidden representations of each node in the graph. As shown in (Vaswani et al. 2017), the canonical self-attention consists of three parts: query, key and value, and is computed by performing scaling dot product calculation:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V$$  \hspace{1cm} (6)

where $Q \in R^{L \times d}$ is the query matrix, $K \in R^{L \times d}$ is the key matrix, $V \in R^{L \times d}$ is the value
matrix, $L$ is the length of input data, and $d$ is the feature dimension of input data.

As mentioned above, the first-order backward difference between two adjacent nodes in a graph can be used to represent the meteorological dynamical changes between two adjacent moments. This difference information is useful for air quality prediction, and can be embedded into the canonical self-attention. In order to simultaneously capture the interrelation and dynamical changes among different nodes, we modify the attention calculation in Eq. (6) by multiplying the first-order backward difference matrix $E$ as follows:

$$\text{Attention} = \text{softmax}((Q \cdot K^T) \cdot E)V$$

(7)

For graph-based time series PM2.5 data prediction, we employed fixed position encoding with the nonlinear sine and cosine functions (Vaswani et al. 2017) to provide the temporal information of time series data for graph attention calculation.

3 Evaluation criteria

To verify the performance of air quality PM2.5 prediction methods, two representative evaluation metrics, including mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE), were employed for experiments. MAE, RMSE, and MAPE are defined as:

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

(8)

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

(9)

$$\text{MAPE}(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

(10)

where $y$ and $\hat{y}$ separately denotes the ground-truth and predicted PM2.5 value, and $m$
represents the whole number of testing data. The smaller the values of MAE, RMSE, and MAPE are, the higher the final prediction results are.

4 Experiments

4.1 Implementation details

We implement all the experiments on a PC server with a GPU NVIDIA Quadro P6000 with 24G memory. The open source Pytorch tools are leveraged to conduct all machine learning models for air quality prediction. For deep learning models, the open source Tensorflow library is installed and configured. The Adam optimizer is adopted, and the initial learning rate is set to le-4. The batch size is set to 32, the maximum epoch number is 200, and the mean squared error loss function is employed. Normalization is conducted to be [0, 1] for air quality time series data. The lookup size (window size), which is used to represent historical observations as input size of all machine learning models, is 24 for its promising performance. We evaluate the performance of our method in comparison with other representative methods, such as the traditional ARMA and SVR, as well as the recently-emerged deep models like CNNs, LSTMs, original Transformer methods, as describe below.

ARMA is a traditional linear statistical method for time series data prediction. Here, ARMA is just used for single-step air quality prediction since it is limited multi-step air quality prediction strategy. SVR is a kernel method on the basis of non-linear statistical machine learning theories. We adopt the linear kernel for SVR on air quality PM2.5 prediction tasks.

CNNs are a well-known deep model originally processing two-dimension (2D) image data. Due to the used 1D time-series PM2.5 data, 1D-CNN is adopted in this work. The network configuration for 1D-CNN is that it consists of 256 convolution kernels with a kernel width of 5 and a stride of 1. Then, a batch normalization layer, max-pooling layer, rectified linear units (RLU)
layer, a dropout (0.3) layer, and a fully-connected (FC) layer are used after the convolution layers.

LSTMs are a typical kind of recurrent architecture modeling long-range dependencies of time series data. Bidirectional LSTM (BiLSTM) is employed for air quality prediction. BiLSTM contains a forward LSTM and a backward LSTM. We leveraged a two-layer BiLSTM for air quality forecasting in this work. Each layer of BiLSTM contains 256 hidden neurons, followed by a dropout (0.05) layer. For the original Transformer model (Vaswani et al. 2017) and our proposed model, we leverage three encoders and two decoders for their promising performance on air quality PM2.5 prediction.

In this work, we adopt a year-independent strategy for air quality forecasting experiments which is definitely close to the real-world sceneries. More specially, the training, and testing sets are selected from different years. In detail, on the used Beijing PM2.5 dataset, the first four-year data (01/01/2010 to 12/31/2013) is selected as the training net, and the last year data (01/01/2014-12/31/2014) is adopted for testing. On the used Taizhou PM2.5 dataset, the first two-year data (01/01/2017 to 12/31/2018) is employed for training, and the last year data (01/01/2019 to 12/31/2019) is adopted for testing. During the training of deep models, we randomly select 10% of the entire training set as the validation set for model validation.

4.2 Results and analysis

To verify the performance of different air quality PM2.5 prediction methods, we presented two types of experimental results: single-step prediction for the next 1 hour, and multi-step prediction for the next multiple hours.

4.2.1 Single-step prediction results

Fig.6 provides performance comparisons of different air quality prediction methods on Beijing
and Taizhou PM2.5 datasets for single-step PM2.5 prediction tasks when the forward-step prediction size is 1 for the next 1 hour (h1). These comparing methods contain ARMA, the linear SVR, CNN, LSTM, the original Transformer (abbreviated as Transformer), as well as our method. As shown in Fig. 6, it can be seen that our method outperforms other used methods on Beijing and Taizhou PM2.5 datasets for single-step PM2.5 prediction tasks. In detail, our method obtains the lowest RSME, MAE, and MAPE on these two datasets. More specially, our method is able to reduce RMSE to 18.51 (ug/m$^3$), MAE to 11.06 (ug/m$^3$), and MAPE to 22.91 (%) on Beijing PM2.5 dataset, whereas on Taizhou PM2.5 dataset our method can reduce RMSE to 5.70 (ug/m$^3$), MAE to 3.66 (ug/m3), and MAPE to 20.23 (%). This indicates the effectiveness of our proposed method for air quality PM2.5 prediction from a graph perspective. In comparison with other methods like ARMA, SVR, CNN, LSTM, and Transformer, our method has stronger capability of capturing long-term dependency and complex relationships from time series PM2.5 data for air quality prediction. In addition, our method yields better performance than Transformer, showing the advantages of our method on the basis of graph attention.

Besides, compared with traditional shallow learning methods like ARMA and SVR, deep learning methods, including LSTM, Transformer and our method, produce better performance for air quality prediction. This demonstrates the superiority of deep learning techniques over traditional shallow learning techniques on air quality prediction tasks. However, the used 1D-CNN obtains slight lower performance than SVR on single-step PM2.5 prediction tasks. This indicates that CNN may not very effective to learn long-term dependency and complex relationships from 1D time series PM2.5 data.
Fig. 6 Comparisons of different methods for single-step PM2.5 prediction tasks for the next 1 hour

4.2.2 Multi-step prediction results

For multi-step prediction results, we provided performance comparisons of different air quality prediction methods for the next multiple hours (6, 12, 24, 48). For the next 6 hours, the average prediction results in the next forward 6 hours were reported as the testing error of different methods. For more than the next 6 hours, we divided them into a number of adjacent intervals and trained individual models corresponding to every interval. Then, we figured out the average prediction results for every interval. In particular, for the next 12 hours prediction, we split it into three intervals: 0-3h, 3-6h, and 6-12h. For the next 24 hours prediction, we split it into four intervals: 0-3h, 3-6h, 6-12h, and 12-24h. For the next 48 hours prediction, we split it into four intervals: 0-6h, 6-12h, 12-24h and 24-48h.

Fig. 7 presents the obtained results (RMSE, MAE and MAPE) of different methods for the next 6 hours on Beijing and Taizhou PM2.5 datasets. It can be seen from the results in Fig. 7, compared with other methods, our method achieves the smaller RSME, MAE and MAPE on Beijing and Taizhou PM2.5 datasets. This indicates the superiority of the proposed method on long-term air quality prediction tasks. More specially, our method reduces RMSE to 36.27 (ug/m³), MAE to
22.61 (ug/m$^3$), MAPE to 51.88 (%) on Beijing PM2.5 dataset, and RMSE to 11.19 (ug/m$^3$), MAE to 7.40 (ug/m$^3$), and MAPE to 44.37 (%) on Taizhou PM2.5 dataset, respectively. The ranking order for other methods is Transformer, LSTM, CNN, and SVR. Note that CNN provides slightly smaller RMSE, MAE, and MAPE than SVR on multi-step PM2.5 prediction tasks for the next 6 hours. This is opposite to single-step PM2.5 prediction tasks for the next 1 hour, as shown in Fig.6. This shows that CNN is capable of promoting the prediction performance with the increasing forward-step prediction size from the next 1 hour to the next 6 hours. This finding of CNN will be verified further in the next 12, 24 and 48 hours.

Fig.7 Comparisons of different methods for multi-step prediction results for the next 6 hours

Fig.8 and 9 separately show the prediction results (RMSE, MAE and MAPE) of different methods for the next 12 hours (three intervals) on Beijing and Taizhou PM2.5 datasets. Fig.10 and 11 individually depict the prediction results (RMSE, MAE and MAPE) of different methods for the next 24 hours (four intervals) on Beijing and Taizhou PM2.5 datasets. Fig.12 and 13 independently present the prediction results (RMSE, MAE and MAPE) of different methods for the next 48 hours (four intervals) on Beijing and Taizhou PM2.5 datasets. From the results in Fig.8-13, we can observe that when the forward prediction size increases from 12 hours to 48
hours, the multi-step PM2.5 prediction accuracies of all used methods clearly drop down. This may be attributed to the fact that the larger the forward prediction size is, the more difficult and challenging the accurate air quality prediction task is. In addition, Fig.8-13 show that our method still presents the lowest prediction error (RMSE, MAE, MAPE) among all used methods when the forward prediction size changes from 12 hours to 48 hours. Besides, CNN performs better than SVR again for the next 12 to 48 hours, and outperforms LSTM for the next 48 hours. This shows that CNN is more appropriate to implement long-term air quality prediction compared with short-term air quality prediction tasks.

Fig.8 Comparisons of different methods for multi-step prediction results for the next 12 hours on Beijing PM2.5 dataset

Fig.9 Comparisons of different methods for multi-step prediction results for the next 12 hours on Taizhou PM2.5 dataset
Fig. 10 Comparisons of different methods for multi-step prediction results for the next 24 hours on Beijing PM2.5 dataset

Fig. 11 Comparisons of different methods for multi-step prediction results for the next 24 hours on Taizhou PM2.5 dataset

Fig. 12 Comparisons of different methods for multi-step prediction results for the next 48 hours on Beijing PM2.5 dataset
Fig. 13: Comparisons of different methods for multi-step prediction results for the next 48 hours on Taizhou PM2.5 dataset.

To intuitively exhibit the superiority of our method over the original Transformer method, Fig. 14 and 15 separately provide the visualization of their single-step ground truth and predicted PM2.5 values for the next 1 hour, and multi-step ground truth and predicted PM2.5 values for the next 48 hours on Beijing and Taizhou PM2.5 datasets. The forward prediction size is 4/01/2014-4/30/2014 on Beijing PM2.5 dataset, and 4/01/2019-4/30/2019 on Taizhou PM2.5 dataset. Here, an illustration of their difference is labeled with a red circle in Fig. 14 and 15. As shown in Fig. 14 and 15, we can observe that both of them obtain promising performance on single-step prediction tasks for the next 1 hour. Nevertheless, our method slightly outperforms Transformer on subtle changes in the time period of wave valley and the wave peak of air quality PM2.5 testing data from these two datasets. Moreover, such superiority of our method over Transformer is more obvious for multi-step prediction results for the next 48 hours. The visualization in Fig. 14 and 15 show the advantages of our method over Transformer on short-term and long-term air quality PM2.5 prediction tasks, again.
Fig. 14 Comparisons of our method and Transformer on single-step (h1) and multi-step (h48) ground truth and air quality prediction tasks during one month (4/01/2014-4/30/2014) on Beijing PM2.5 dataset.

Fig. 15 Comparisons of our method and Transformer on single-step (h1) and multi-step (h48) ground truth and air quality prediction tasks during one month (4/01/2019-4/30/2019) on Taizhou PM2.5 dataset.
5. Conclusions and future work

In this work, a new deep learning model called TDGTN is proposed to learn long-term temporal dependencies and complex relationships from time series PM2.5 data for air quality PM2.5 prediction. Based on the constructed graph-structured data, we implement air quality PM2.5 prediction tasks for the single air monitoring station from a view point of graphs. In the proposed TDGTN model, the conventional self-attention mechanism in the original Transformer model is improved by means of integrating the temporal difference information, which gives rise to a new graph attention mechanism. Experiment results on Beijing and Taizhou PM2.5 datasets demonstrate the validity of the proposed method on both short-term and long-term air quality prediction tasks. In future, it is an interesting subject to extend our method for more cities throughout the world. Additionally, it is also challenging to exploit the applications of our method as well as more advanced deep models on real-time air quality PM2.5 prediction tasks.

Funding information:

This work was supported by Zhejiang Provincial National Science Foundation of China under Grant No. LY20E080013, LZ20F020002, and National Science Foundation of China (NSFC) under Grant No. 61976149.

Author contribution

Zhen Zhang Conceptualization, Methodology, Software, Writing- Original draft. Shiqing Zhang: Conceptualization, Software. Xiaoming Zhao: Project administration, Writing- Reviewing and Editing. Linjian Chen, and Jun Yao: Data collection, Pre-processing.

Availability of data and materials Not applicable.

Declarations

Ethics approval and consent to participate Not applicable.
Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References:

Abirami S, Chitra P (2021): Regional air quality forecasting using spatiotemporal deep learning. Journal of Cleaner Production 283, 125341

Agarwal S, Sharma S, R S, Rahman MH, Vranckx S, Maiheu B, Blyth L, Janssens S, Gargava P, Shukla VK, Batra S (2020): Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions. Science of The Total Environment 735, 139454

Aggarwal A, Toshniwal D (2021): A hybrid deep learning framework for urban air quality forecasting. Journal of Cleaner Production, 129660

Arhami M, Kamali N, Rajabi MM (2013): Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. Environmental Science and Pollution Research 20, 4777-4789

Bazi Y, Bashmal L, Rahhal MMA, Dayil RA, Ajlan NA (2021): Vision transformers for remote sensing image classification. Remote Sensing 13, 516

Cekim HO (2020): Forecasting PM 10 concentrations using time series models: a case of the most polluted cities in Turkey. Environmental Science and Pollution Research 27, 25612-25624

Chang Q, Zhang H, Zhao Y (2020a): Ambient air pollution and daily hospital admissions for respiratory system–related diseases in a heavy polluted city in Northeast China. Environmental Science and Pollution Research, 10.1007/s11356-021-01321-3

Chang Y-S, Abimannan S, Chiao H-T, Lin C-Y, Huang Y-P (2020b): An ensemble learning based hybrid model and framework for air pollution forecasting. Environmental Science and Pollution Research 27, 38155-38168

Chen L, Xu J, Wu B, Qian Y, Du Z, Li Y, Zhang Y (2021): Group-Aware Graph Neural Network for Nationwide City Air Quality Forecasting. arXiv preprint arXiv:2108.12238

Chu J, Dong Y, Han X, Xie J, Xu X, Xie G (2021): Short-term prediction of urban PM 2.5 based on a hybrid modified variational mode decomposition and support vector regression model. Environmental Science and Pollution Research 28, 56-72

Chung J, Gulcehre C, Cho K, Bengio Y (2014): Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555

Darçın M (2014): Association between air quality and quality of life. Environmental Science and Pollution Research 21, 1954-1959, 10.1007/s11356-013-2101-3

De Melo WC, Granger E, Hadid A (2019): Depression detection based on deep distribution learning, 2019 IEEE International Conference on Image Processing (ICIP). IEEE, Taipei, Taiwan, pp. 4544-4548

Donnelly A, Misstear B, Broderick B (2015): Real time air quality forecasting using integrated parametric and non-parametric regression techniques. Atmospheric Environment 103, 53-65

Du S, Li T, Yang Y, Horng SJ (2021): Deep Air Quality Forecasting Using Hybrid Deep Learning Framework. IEEE Transactions on Knowledge and Data Engineering 33, 2412-2424, 10.1109/TKDE.2019.2954510
Duke B, Ahmed A, Wolf C, Aarabi P, Taylor GW (2021): Sstvos: Sparse spatiotemporal transformers for video object segmentation, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5912-5921.

Elman JL (1990): Finding structure in time. Cognitive science 14, 179-211.

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 107, 118-128.

Gao X, Li W (2021): A graph-based LSTM model for PM2.5 forecasting. Atmospheric Pollution Research 12, 101150.

Gariazzo C, Carlino G, Silibello C, Renzi M, Finardi S, Pepe N, Radice P, Forastiere F, Michelozzi P, Viegi G, Stafoggia M (2020): A multi-city air pollution population exposure study: Combined use of chemical-transport and random-Forest models with dynamic population data. Science of The Total Environment 724, 138102.

Geng G, Zhang Q, Martin RV, van Donkelaar A, Huo H, Che H, Lin J, He K (2015): Estimating long-term PM2.5 concentrations in China using satellite-based aerosol optical depth and a chemical transport model. Remote sensing of Environment 166, 262-270.

Graupe D, Krause D, Moore J (1975): Identification of autoregressive moving-average parameters of time series. IEEE Transactions on Automatic Control 20, 104-107, 10.1109/TAC.1975.1100855.

He K, Zhang X, Ren S, Sun J (2016): Deep residual learning for image recognition, Proceedings of the IEEE conference on computer vision and pattern recognition, Nevada, USA, pp. 770-778.

Hinton GE, Salakhutdinov RR (2006): Reducing the dimensionality of data with neural networks. Science 313, 504-507.

Hochreiter S, Schmidhuber JJNe (1997): Long short-term memory. 9, 1735-1780.

Jian L, Zhao Y, Zhu Y-P, Zhang M-B, Bertolatti D (2012): An application of ARIMA model to predict submicron particle concentrations from meteorological factors at a busy roadside in Hangzhou, China. Science of The Total Environment 426, 336-345.

Jin N, Zeng Y, Yan K, Ji Z (2021): Multivariate Air Quality Forecasting With Nested Long Short Term Memory Neural Network. IEEE Transactions on Industrial Informatics 17, 8514-8522, 10.1109/TII.2021.3065425.

Ke J, Zhang J, Tang M (2021): Does city air pollution affect the attitudes of working residents on work, government, and the city? An examination of a multi-level model with subjective well-being as a mediator. Journal of Cleaner Production 295, 126250.

Kipf TN, Welling M (2016): Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907.

Lanchantin J, Wang T, Ordonez V, Qi Y (2021): General Multi-label Image Classification with Transformers, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16478-16488.

LeCun Y, Bengio Y, Hinton G (2015): Deep learning. Nature 521, 436-444.

Lee H-M, Park RJ, Henze DK, Lee S, Shim C, Shin H-J, Moon K-J, Woo J-H (2017): PM2.5 source attribution for Seoul in May from 2009 to 2013 using GEOS-Chem and its adjoint model. Environmental Pollution 221, 377-384.

Liang X, Zou T, Guo B, Li S, Zhang H, Zhang S, Huang H, Chen SX (2015): Assessing Beijing's PM2.5 pollution: severity, weather impact, APEC and winter heating. Proceedings of the Royal...
Liao Q, Zhu M, Wu L, Pan X, Tang X, Wang Z (2020): Deep Learning for Air Quality Forecasts: a Review. Current Pollution Reports 6, 399-409, 10.1007/s40726-020-00159-z

Liu H, Yan G, Duan Z, Chen C (2021): Intelligent modeling strategies for forecasting air quality time series: A review. Applied Soft Computing, 106957

Mihailovic DT, Alapaty K, Podrascanin Z (2009): Chemical transport models. Environmental Science and Pollution Research 16, 144-151

Neishi M, Yoshinaga N (2019): On the relation between position information and sentence length in neural machine translation. Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL), Hong Kong, China, pp. 328-338

Niu Z, Zhong G, Yu H (2021): A review on the attention mechanism of deep learning. Neurocomputing 452, 48-62

Pak U, Ma J, Ryu U, Ryok K, Juhyok U, Pak K, Pak C (2020): Deep learning-based PM2.5 prediction considering the spatiotemporal correlations: A case study of Beijing, China. Science of The Total Environment 699, 133561

Ponomarev N, Elansky N, Kirsanov A, Postylyakov O, Borovski A, Verevkin YM (2020): Application of atmospheric chemical transport models to validation of pollutant emissions in Moscow. Atmospheric and Oceanic Optics 33, 362-371

Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Shyu M-L, Chen S-C, Iyengar SS (2018): A survey on deep learning: Algorithms, techniques, and applications. ACM Computing Surveys (CSUR) 51, 1-36

Powers JG, Klemp JB, Skamarock WC, Davis CA, Dudhia J, Gill DO, Coen JL, Gochis DJ, Ahmadov R, Peckham SE (2017): The weather research and forecasting model: Overview, system efforts, and future directions. Bulletin of the American Meteorological Society 98, 1717-1737

Ragab MG, Abdulkadir SJ, Aziz N, Al-Tashi Q, Alyousifi Y, Alhussian H, Alqashaibi A (2020): A Novel One-Dimensional CNN with Exponential Adaptive Gradients for Air Pollution Index Prediction. Sustainability 12, 10090

Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G (2008): The graph neural network model. IEEE transactions on neural networks 20, 61-80

Schwartz J (1993): Particulate air pollution and chronic respiratory disease. Environmental research 62, 7-13

Suleiman A, Tight MR, Quinn AD (2019): Applying machine learning methods in managing urban concentrations of traffic-related particulate matter (PM10 and PM2.5). Atmospheric Pollution Research 10, 134-144

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017): Attention is all you need, Advances in neural information processing systems, Long Beach, CA, USA, pp. 5998-6008

Wang J, Bai L, Wang S, Wang C (2019a): Research and application of the hybrid forecasting model based on secondary denoising and multi-objective optimization for air pollution early warning system. Journal of Cleaner Production 234, 54-70

Wang W, Zhao S, Jiao L, Taylor M, Zhang B, Xu G, Hou H (2019b): Estimation of PM2.5 concentrations in China using a spatial back propagation neural network. Scientific reports 9, 1-10

Xiao Q, Chang HH, Geng G, Liu Y (2018): An ensemble machine-learning model to predict historical
PM2.5 concentrations in China from satellite data. Environmental science & technology 52, 13260-13269

Xu J, Chen L, Lv M, Zhan C, Chen S, Chang J (2021): HighAir: A Hierarchical Graph Neural Network-Based Air Quality Forecasting Method. arXiv preprint arXiv:2101.04264

Yan X, Zang Z, Luo N, Jiang Y, Li Z (2020): New interpretable deep learning model to monitor real-time PM2.5 concentrations from satellite data. Environment International 144, 106060

Yang W, Deng M, Xu F, Wang H (2018): Prediction of hourly PM2.5 using a space-time support vector regression model. Atmospheric Environment 181, 12-19

Yue Z, Witzig CR, Jorde D, Jacobsen H-A (2020): BERT4NILM: A Bidirectional Transformer Model for Non-Intrusive Load Monitoring, Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring, New York, USA, pp. 89-93

Zaini Na, Ean LW, Ahmed AN, Malek MA (2021): A systematic literature review of deep learning neural network for time series air quality forecasting. Environmental Science and Pollution Research, DOI: 10.1007/s11356-021-17442-1, 10.1007/s11356-021-17442-1

Zhang B, Wu B, Liu J (2020a): PM2.5 pollution-related health effects and willingness to pay for improved air quality: Evidence from China’s prefecture-level cities. Journal of Cleaner Production 273, 122876

Zhang H, Chen G, Hu J, Chen S-H, Wiedinmyer C, Kleeman M, Ying Q (2014): Evaluation of a seven-year air quality simulation using the Weather Research and Forecasting (WRF)/Community Multiscale Air Quality (CMAQ) models in the eastern United States. Science of the Total Environment 473, 275-285

Zhang K, Thé J, Xie G, Yu H (2020b): 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 277, 123231

Zhang Z, Zeng Y, Yan K (2021): A hybrid deep learning technology for PM2.5 air quality forecasting. Environmental Science and Pollution Research 28, 39409-39422

Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, Zhang W (2021): Informer: Beyond efficient transformer for long sequence time-series forecasting, Proceedings of AAAI, pp. 11106-11115

Zhou Q, Jiang H, Wang J, Zhou J (2014): A hybrid model for PM2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network. Science of the Total Environment 496, 264-274