

RESEARCH ARTICLE

Spatial and temporal characteristics analysis and prediction model of PM_{2.5} concentration based on SpatioTemporal-Informer model

Zhanfei Ma^{1,2}, Wenli Luo^{2*}, Jing Jiang², Bisheng Wang², Ziyuan Ma³, Jixiang Lin², Dongxiang Liu¹

1 School of Information Science and Technology, Baotou Teachers' College, Baotou, Inner Mongolia, China, **2** School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia, China, **3** College of Mechanical and Electrical Engineering, Sichuan Agricultural University, Ya'an, Sichuan, China

* wenli131583@163.com



OPEN ACCESS

Citation: Ma Z, Luo W, Jiang J, Wang B, Ma Z, Lin J, et al. (2023) Spatial and temporal characteristics analysis and prediction model of PM_{2.5} concentration based on SpatioTemporal-Informer model. PLoS ONE 18(6): e0287423. <https://doi.org/10.1371/journal.pone.0287423>

Editor: Mahmud Iwan Solihin, UCSI University Kuala Lumpur Campus: UCSI University, MALAYSIA

Received: April 10, 2023

Accepted: June 6, 2023

Published: June 23, 2023

Copyright: © 2023 Ma et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data underlying the results presented in the study are available from The U.S. Embassy in Beijing (<http://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data>).

Funding: This study is fully supported by National Natural Science Foundation of China (Grant No. 61762071, Grant No.61163025). There was no additional external funding received for this study.

Abstract

The primary cause of hazy weather is PM_{2.5}, and forecasting PM_{2.5} concentrations can aid in managing and preventing hazy weather. This paper proposes a novel spatiotemporal prediction model called SpatioTemporal-Informer (ST-Informer) in response to the shortcomings of spatiotemporal prediction models commonly used in studies for long-input series prediction. The ST-Informer model implements parallel computation of long correlations and adds an independent spatiotemporal embedding layer to the original Informer model. The spatiotemporal embedding layer captures the complex dynamic spatiotemporal correlations among the input information. In addition, the ProbSparse Self-Attention mechanism in this model can focus on extracting important contextual information of spatiotemporal data. The ST-Informer model uses weather and air pollutant concentration data from numerous stations as its input data. The outcomes of the trials indicate that (1) The ST-Informer model can sharply capture the peaks and sudden changes in PM_{2.5} concentrations. (2) Compared to the current models, the ST-Informer model shows better prediction performance while maintaining high-efficiency prediction ($MAE \approx 7.50 \mu\text{g}/\text{m}^3$, $RMSE \approx 4.31 \mu\text{g}/\text{m}^3$, $R^2 \approx 0.88$). (3) The ST-Informer model has universal applicability, and the model was applied to the concentration of other pollutants prediction with good results.

Introduction

China's rapid industrialization and urbanization are worsening urban air pollution, which has detrimental effects on individuals' lives, health, and economic activity [1–3]. To reduce air pollution, relevant departments have proposed important initiatives such as haze control and improvement of environmental quality [4]. As a significant component of atmospheric pollutants, PM_{2.5} is a major factor in causing hazy weather, reducing visibility, and affecting traffic safety. Long-term exposure to high PM_{2.5} concentrations can result in cancer of the lungs, cardiovascular disease, respiratory disorders, and other health significantly impacts youngsters,

Competing interests: The authors have declared that no competing interests exist.

the elderly, and those with heart or lung disease. In addition, The development of aquatic organisms and plants can also be impacted by high PM_{2.5} concentrations that land in soil and water bodies. Therefore, accurate prediction of PM_{2.5} concentration is of great practical importance for national air pollution control and risk avoidance. However, PM_{2.5} concentrations are often influenced by various external factors and have complex time and space dependence [5]. Therefore, the development of highly precise and effective PM_{2.5} prediction systems is still an urgent scientific issue.

With the advancement of science and technology, monitoring tools, including weather observation stations, air quality monitoring stations, and meteorological satellites, have started to be extensively dispersed in diverse locations. Because the massive amount of air pollutant concentration data and meteorological data collected by this equipment provides a large amount of data support for air quality prediction and meteorological research, PM_{2.5} concentration prediction models built on data-driven models are receiving increasing attention from research scholars. Previous studies have used statistical models to predict PM_{2.5} concentrations, such as Auto-regressive Moving Average (ARMA) model [6, 7] and Auto-regressive Integrated Moving Average (ARIMA) model [8, 9], which are traditionally spartan in structure, less computationally intensive, and can effectively describe the variability of time series; however, these models tend to oversimplify the complexities between PM_{2.5} and other air pollutant concentrations non-linear relationships, which are prone to bias in the face of non-linear relationships. The development of pollutant concentration prediction models based on machine learning algorithms has been rapid in recent years [10], and the nonlinear regression prediction performance of these models is better than that of traditional statistical methods, and commonly used models include Support Vector Regression (SVR) [11], random forest models [12], and Artificial Neural Networks (ANN) [13–15]. In addition, artificial neural networks have strong adaptive and robust properties and show excellent performance in solving prediction problems, especially Recurrent Neural Network (RNN); RNN and their variants Long Short Term Memory Network (LSTM) in processing and are commonly used for pollutant concentration prediction in different time spans [16, 17]. Since various factors influence the concentration of pollutants, the data's long-term time dependence and dynamic spatial correlation need to be considered when constructing prediction models. There are studies to construct spatiotemporal prediction models by fusing different neural network structures, such as Convolutional Neural Network and Long Short Term Memory Network (CNN-LSTM) model; CNN-LSTM model by CNN layer Convolutional operations are performed on the input data, and the spatial feature information extracted from it is passed to the LSTM layer to capture the dynamic changes and long-term dependencies in the spatiotemporal sequence data [18, 19]. In addition, there are Graph Convolutional networks and Long Short Term Memory Networks (GC-LSTM) model [20], Long Short Term Memory-Fully Connected Neural Network (LSTM-FC) model [21], and other combined spatiotemporal prediction models that are more comprehensive in dealing with spatiotemporal data with multidimensional features. However, due to the limitations of convolutional filters and the shortcomings of LSTM, which cannot be computed in parallel, these models cannot adequately extract the long-term and complex correlations of spatiotemporal data when the input data is extended. Some studies have compensated for these shortcomings in the above models by introducing the Attention mechanism [22, 23], the addition of the Attention mechanism can be used to focus on extracting important information in the context, but their inherent characteristics limit their predictive power for long spatiotemporal sequences.

Vaswani et al. [24] proposed the Transformer model, which is a model with a new architecture built entirely based on the Attention mechanism and achieves higher parallelism, which has achieved excellent results in Natural Language Processing (NLP) [25], Computer Vision

(CV), and other fields. Considering that the model allows parallel processing and temporal feature extraction for time-series data, many studies have applied the Transformer model to time-series data prediction tasks [26]. For example, Li S et al. [27] proposed an Enhanced Local Transformer (ELM) model that enhances the Transformer's local and context-aware capabilities in processing time series data by introducing adaptive location coding and spatially aware attention mechanisms, and the ELM model achieves better performance on various time series prediction tasks. In addition, Zhou et al. [28] developed an Informer model, which improves the traditional Transformer model by introducing new structures and mechanisms to learn non-smooth and long-term time dependence and has higher efficiency and accuracy when dealing with long series time series data [29]. However, the Informer model focuses only on the "temporal attention" between learning time steps and ignores the complex spatial relationships between variables. Therefore, to predict PM_{2.5} concentrations with spatiotemporal characteristics more accurately and effectively, this paper proposes a spatiotemporal prediction model: the Spatio-Temporal-Informer (ST-Informer) model. The ST-Informer model has an independent spatiotemporal embedding layer, which enables a comprehensive analysis of the spatiotemporal correlation of the input data. Secondly, considering that meteorological characteristics and spatial characteristics of data can affect the prediction of pollutant concentrations [30], auxiliary data such as meteorological data and atmospheric pollutant concentration data from multiple stations were integrated into predicting PM_{2.5} concentrations to improve the performance of the model. The results show that the ST-Informer model has spatiotemporal prediction capability and high prediction accuracy. The main contributions of this paper are as follows.

1. In this paper, the influence of multi-source data is considered comprehensively, and adjacent stations' meteorological factors and pollutant concentration factors affect the PM_{2.5} concentration prediction at the central station. Therefore, the study selects meteorological data and pollutant concentration data from multiple stations as input data.
2. Due to anomalies and missing data collected, the ST-Informer model is relatively sensitive to outliers and noise in the input sequence, and the presence of many outliers or noise will affect the prediction results of this model. Therefore, this paper performs complete pre-processing work on the data. Then, Pearson correlation coefficients were used to analyze the strength of the data's temporal and spatial correlations to ensure the input data's strong correlation as the ST-Informer prediction model in PM_{2.5} prediction and enhance the model's prediction performance.
3. The independent temporal embedding layer and spatial embedding layer in the ST-Informer model proposed in this paper are used to process the input data and capture the complex spatiotemporal characteristics of the data. As well as the ProbSparse Self-Attention mechanism, which is unique to this model, concentrates on extracting information of data across time, space, and multiple dimensions. The model can analyze the spatiotemporal correlation of data more comprehensively. Its prediction effect is better than LSTM, Transformer, and Informer models through the comparative analysis of the results of three evaluation indexes. And the ST-Informer model is applied to the prediction of other pollutant concentrations, which proves its universality.

Materials

Description of the data

Because of the rapid economic growth, dense population, and industrialization of the Beijing-Tianjin-Hebei area, in this paper, we take the forecast of PM_{2.5} concentration in Beijing as an

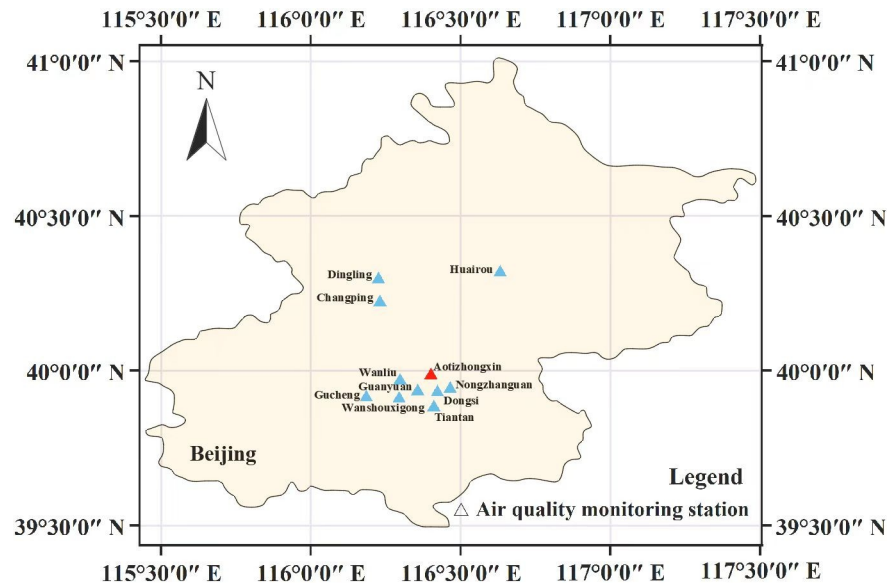


Fig 1. Coordinates of air quality monitoring stations. Location map of each air quality monitoring station in Beijing. <https://doi.org/10.1371/journal.pone.0287423.g001>

example. The U.S. Embassy in Beijing, China provided the hourly meteorological and pollutant concentration data from 11 air quality monitoring stations in Beijing that were collected between March 1, 2013, at 00:00 and February 28, 2017, at 23:00 (<http://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data>). Fig 1 depicts the locations of each station, and Table 1 lists the latitude and longitude of each station. The Aotizhongxin Center's air quality monitoring station, which is close to all the other sites, was chosen as the central site. The 10 other sites, including Changping, that had strong correlations with the central site were then identified by looking at the spatial and temporal correlation of the data. As shown in Table 2, the pollutant concentration factors include the following 6 variables: PM_{2.5}, PM₁₀, SO₂, NO₂, CO, and O₃. The weather variables primarily consist of 6 variables: temperature, atmospheric pressure, dew-point temperature, rainfall capacity, wind direction, and wind speed.

Data preprocessing

The factors predicted in this study include pollutant concentrations (PM_{2.5}, PM₁₀, SO₂, NO₂, CO, O₃) and meteorological factors (temperature, atmospheric pressure, dew-point temperature, rainfall capacity, wind direction, wind speed). Before using this dataset as input data for PM_{2.5} concentration prediction, data pre-processing is performed. The data preprocessing mainly includes outlier testing, missing value filling, and data normalization.

1. Test for outliers

In this paper, outliers are examined using the principle of box plot, which is a statistical graph used to display the dispersion of a set of data [31], and its structure is shown in Fig 2. The box plot provides a criterion for identifying outliers, i.e., a value greater than or less than the upper and lower bounds set by the box plot is considered an outlier.

The historical data collected from each air quality monitoring station were first examined for outliers. Taking the Aotizhongxin station as an example, the image of the meteorological data and air pollutant concentration data collected at this station after processing by the box-

Table 1. Name of air quality monitoring station.

Station Name	longitude	latitude
Aotizhongxin	116° 24' 3.028"	39° 59' 6.562"
Changping	116° 13' 51.722"	40° 13' 15.427"
Dingling	116° 13' 32.732"	40° 17' 41.435"
Dongsi	116° 24' 55.862"	39° 55' 46.423"
Guanyuan	116° 21' 24.548"	39° 55' 57.986"
Gucheng	116° 11' 6.130"	39° 54' 47.041"
Huairou	116° 37' 55.106"	40° 19' 1.211"
Nongzhanguan	116° 27' 54.468"	39° 56' 25.872"
Tiantan	116° 24' 38.984"	39° 52' 54.887"
Wanliu	116° 17' 53.063"	39° 58' 1.758"
Wanshouxigong	116° 17' 40.420"	39° 54' 34.171"
Aotizhongxin	116° 24' 3.028"	39° 59' 6.562"
Changping	116° 13' 51.722"	40° 13' 15.427"
Dingling	116° 13' 32.732"	40° 17' 41.435"
Dongsi	116° 24' 55.862"	39° 55' 46.423"
Guanyuan	116° 21' 24.548"	39° 55' 57.986"
Gucheng	116° 11' 6.130"	39° 54' 47.041"
Huairou	116° 37' 55.106"	40° 19' 1.211"
Nongzhanguan	116° 27' 54.468"	39° 56' 25.872"
Tiantan	116° 24' 38.984"	39° 52' 54.887"
Wanliu	116° 17' 53.063"	39° 58' 1.758"
Wanshouxigong	116° 17' 40.420"	39° 54' 34.171"

<https://doi.org/10.1371/journal.pone.0287423.t001>

line diagram principle is shown in Fig 3. The examined outliers are then marked as missing values.

2. Missing value padding

The method for filling in the missing values in this paper is the K Nearest Neighbor Classification (KNN) algorithm [32], which estimates the missing data by identifying K sample points adjacent to the missing values and using the average of these K sample points. As shown in Fig 4, a comparison graph of unprocessed and processed historical data from the air quality

Table 2. Attributes of the correlation factors input when the model is making predictions.

data name	data type	unit
PM _{2.5}	PM _{2.5} Treatment serum concentration	μm/m ³
PM ₁₀	PM ₁₀ Treatment serum concentration	μm/m ³
SO ₂	SO ₂ Treatment serum concentration	μm/m ³
NO ₂	NO ₂ Treatment serum concentration	μm/m ³
CO	CO Treatment serum concentration	μm/m ³
O ₃	O ₃ Treatment serum concentration	μm/m ³
TEMP	temperature	°C
PRES	atmospheric pressure	pa
DEWP	dew-point temperature	°C
RAIN	rainfall capacity	mm
WD	wind direction	orientation
WSPM	wind speed	m/s

<https://doi.org/10.1371/journal.pone.0287423.t002>

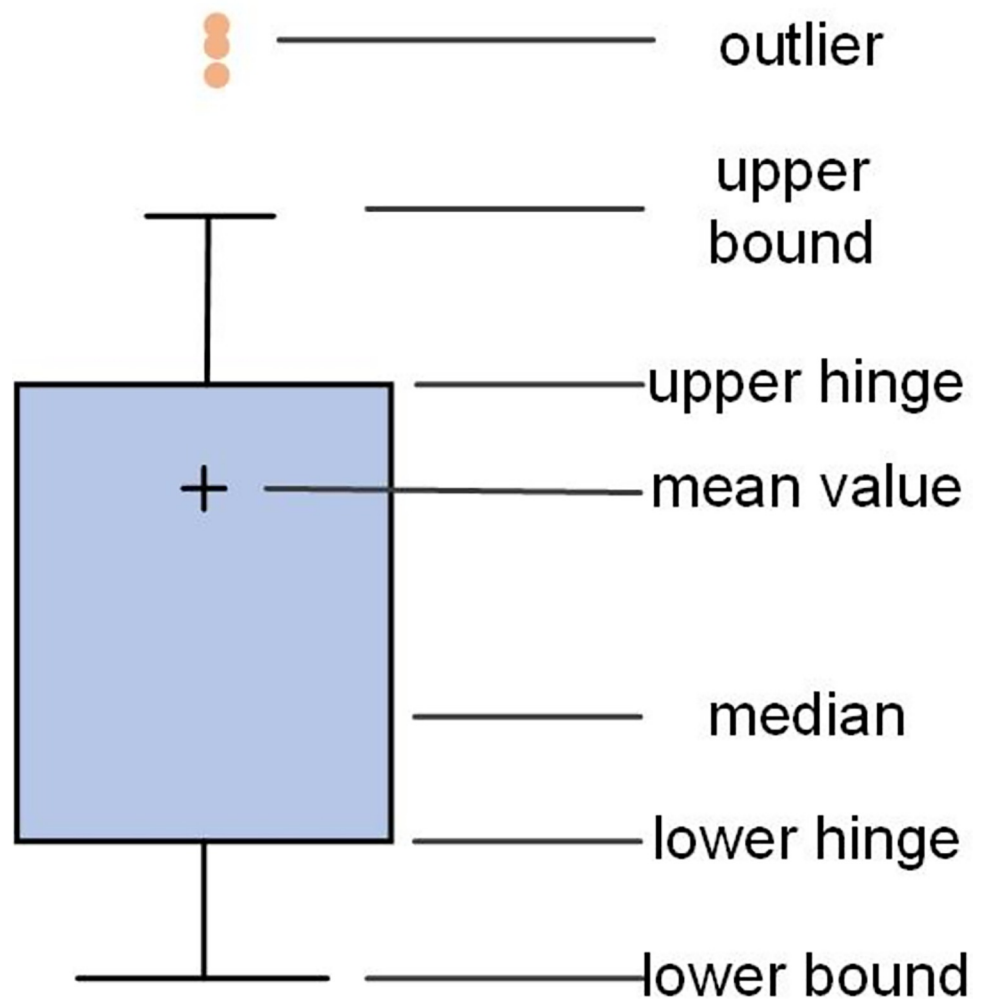


Fig 2. Structure of the box line diagram. A boxplot is a statistical graph showing the distribution of one or more sets of continuous quantitative data, the first quartile, the median, and the third quartile.

<https://doi.org/10.1371/journal.pone.0287423.g002>

monitoring sites in the Aotizhongxin clearly shows that the outliers are significantly reduced and the data are more stable after being processed using the KNN algorithm.

3. Normalization of data

The data are then normalized [33], which is treating each feature equally by removing the variability of values and units between feature parameters and normalizing the initial feature data to increase prediction accuracy. The normalization expression is as follows:

$$x_i = \frac{x_i - \min}{\max - \min} \quad (1)$$

Methods

Spatiotemporal correlation

Because meteorological elements like the wind impact air pollutants, the concentration of pollutants at one site can impact the concentration of pollutants at nearby locations. Because of

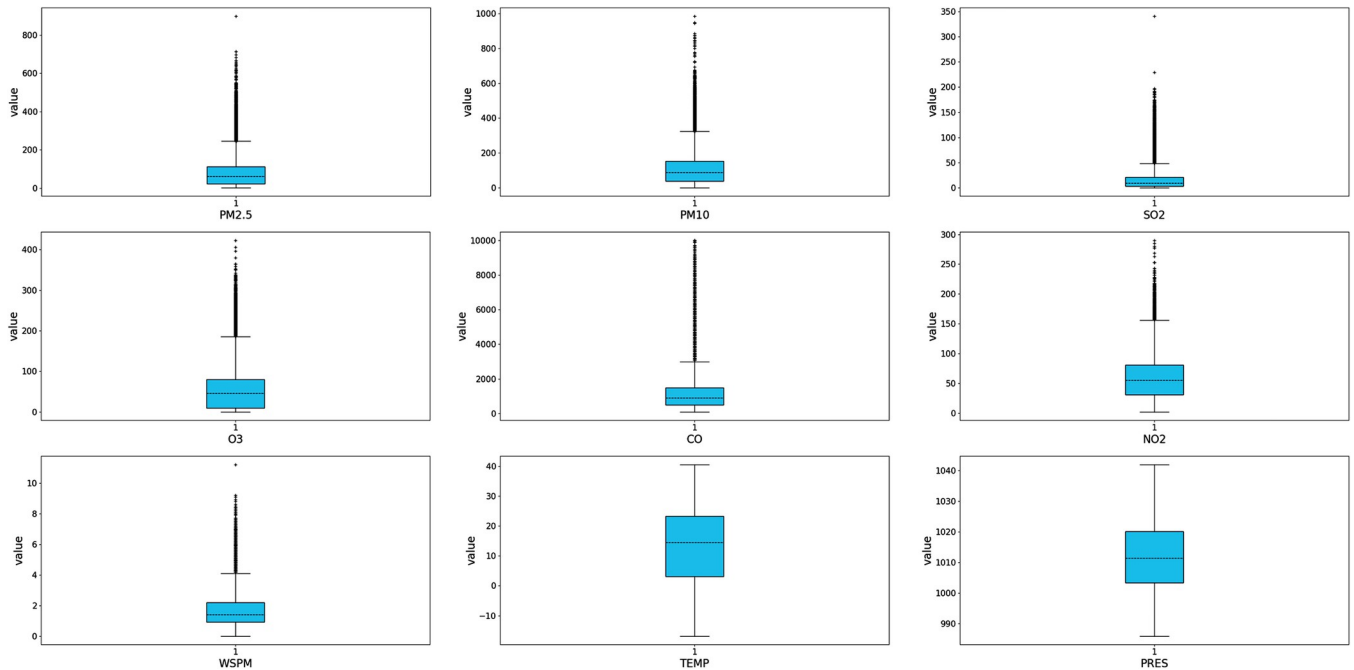


Fig 3. The resulting graph of outliers is tested using a boxplot. Taking the air quality monitoring station of the Aotizhongxin as an example, this research used the boxplot principle to test the outliers of the collected historical data.

<https://doi.org/10.1371/journal.pone.0287423.g003>

this, spatial and temporal correlations between the data are extracted when making predictions of air pollutant concentrations based on data on pollutant concentrations and meteorological data from the central and nearby locations.

To derive elements of spatiotemporal correlation, use the Pearson correlation coefficient. The Pearson correlation coefficient is a statistic used to determine how closely two factors are correlated linearly and to compare traits and categories. A higher absolute number denotes a stronger correlation between the two variables, and the Pearson correlation coefficient ranges from -1 to 1. The formula for calculating the Pearson correlation value is:

$$\begin{aligned} \rho_{X,Y} &= \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} \\ &= \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \end{aligned} \tag{2}$$

Where X and Y represent the two variables, $cov(X, Y)$ represents the covariance between the two variables, and $\sigma_X \sigma_Y$ represents the standard deviation of the two variables.

In terms of extracting temporal correlation analysis, the Pearson correlation coefficient was used to assess the degree of temporal correlation, with X and Y in the Pearson correlation coefficient formula representing the values of historical data before and after a specific time interval, respectively. Fig 5 shows the correlation coefficients of PM_{2.5} concentrations at each station at the current time and 1–24 hours, from which it can be seen that the autocorrelation coefficients of PM_{2.5} concentration values at all stations in 1–24 hours are 0.4 and above. Hence, it is feasible to predict the future PM_{2.5} concentration values to the results with the 24-hour values. However, as time in-creases, the value for the autocorrelation coefficient decreases, and when it exceeds 24 hours, the temporal correlation of the data is weaker.

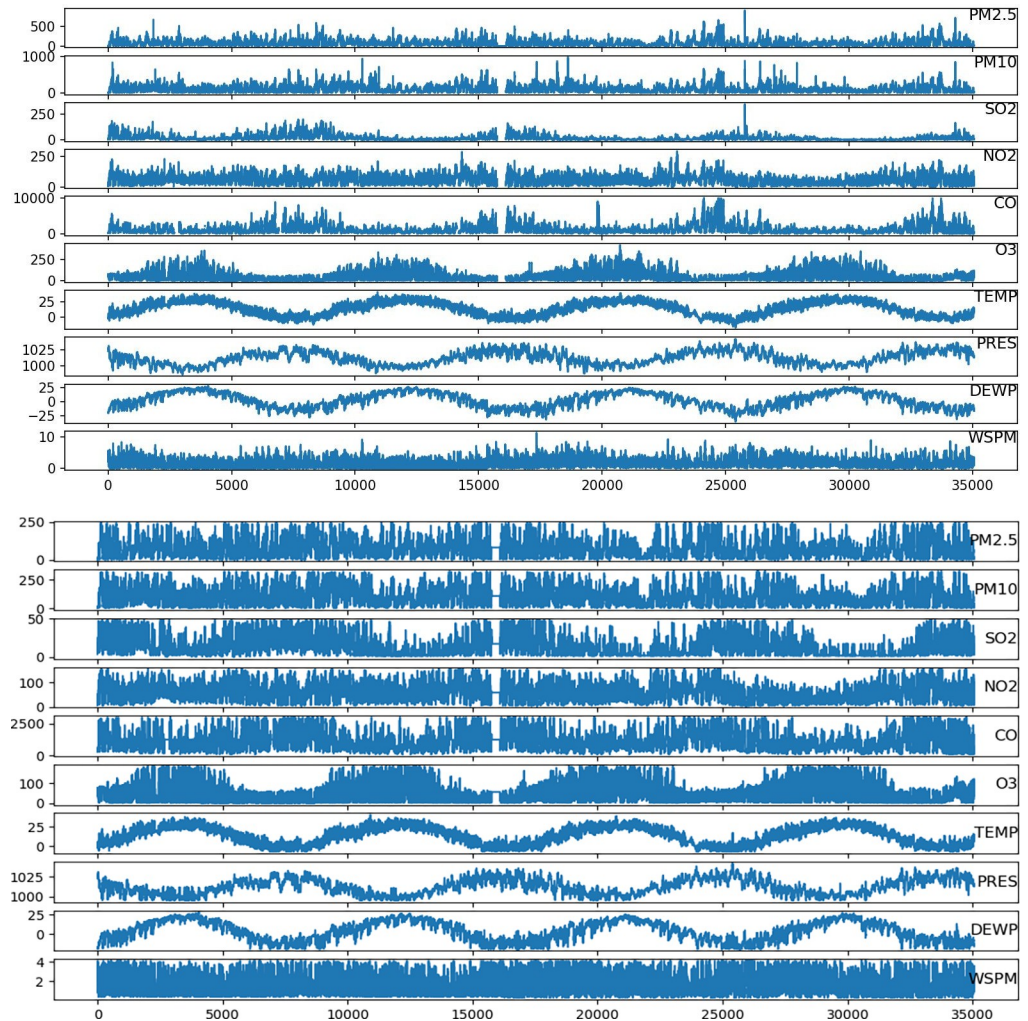


Fig 4. Pre-processed data visualization. Comparison of meteorological data and air pollutant concentration data before and after data preprocessing at the central station. (a) Visualization of unprocessed data. (b) Visualization of pre-processed data.

<https://doi.org/10.1371/journal.pone.0287423.g004>

There is a strong spatial correlation between air quality monitoring stations, and the Pearson correlation coefficient formula can also be used to calculate the correlation coefficients between the monitoring stations in the city. Fig 6 shows the heat map of the correlation coefficients between the central station of Aotizhongxin and the rest of the neighboring stations for each meteorological data and pollutant concentration data. From the heat map, it can be seen that the correlation coefficients between the neighboring stations and the central station are all greater than 0.42, the correlation between the central station and the neighboring stations is stronger, and the stations with larger correlation coefficients are used as neighboring stations to analyze the spatial correlation of the data.

Spatiotemporal prediction model

Informer for forecasting. The Informer model, which effectively enhances the Long Sequence Time-Series Forecasting (LSTF) problem's predictive power and confirms the potential utility of the Transformer-like model, is an improved version of the Transformer model.

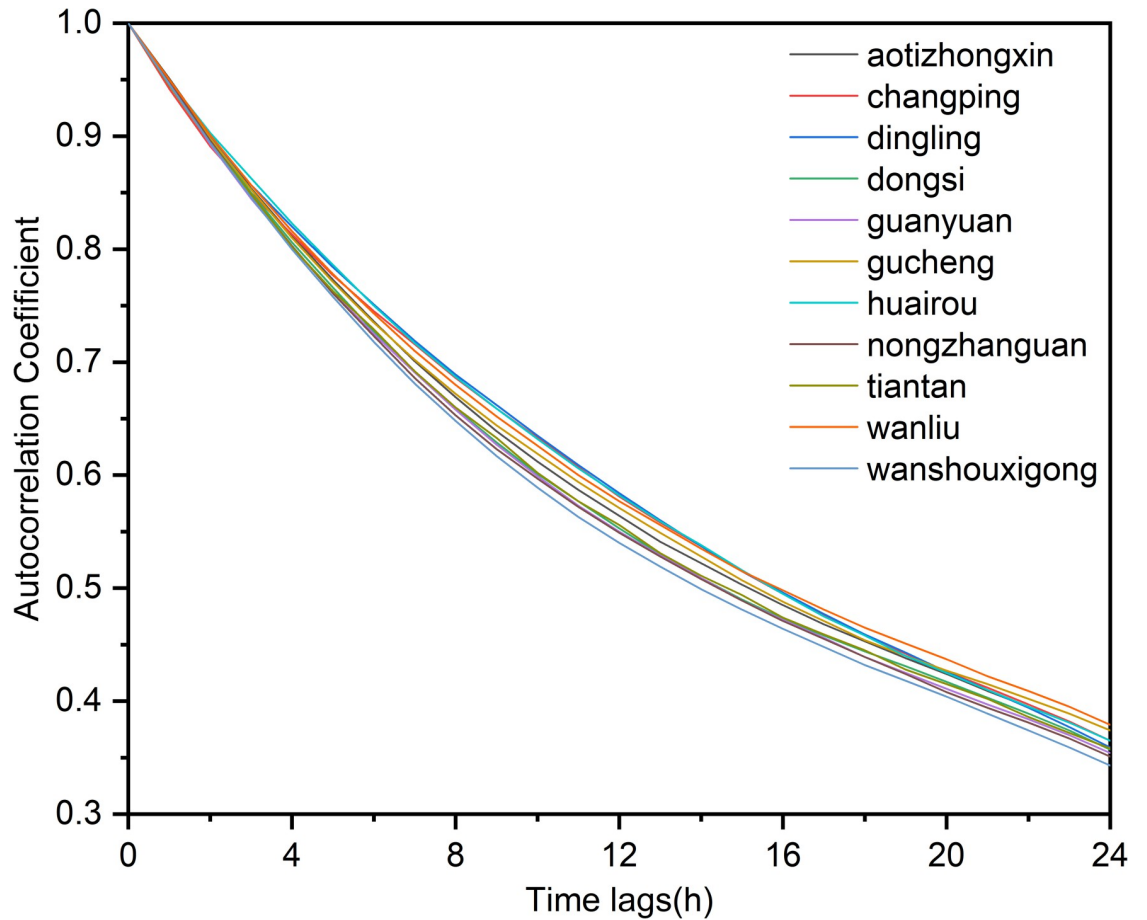


Fig 5. Time autocorrelation analysis. The autocorrelation coefficient of each station changes with time delay of 0–24 hours.

<https://doi.org/10.1371/journal.pone.0287423.g005>

Three significant issues with Transformer are addressed by Informer: complexity in quadratic time and significant memory usage. The answer is suggested by the ProSparse Self-Attention Mechanism, self-attention distilling, and generative style encoder.

Fig 7 depicts the Informer framework. A supervised learning model built on an attention process is known as an informer model. Encoder and decoder are the two components that make up the entire model. To save memory and expedite reasoning, the model enhances the Transformer-based self-attentive mechanism.

In the scrolling prediction setting with fixed window size, the input at time point t is

$$x^t = \{X_1^t, \dots, X_{L_x}^t, X_d^t \in R^{d_x}\} \tag{3}$$

The output predicts the corresponding sequence as

$$y^t = \{Y_1^t, \dots, Y_{L_y}^t, Y_d^t \in R^{d_y}\} \tag{4}$$

1. Encoder

The architecture of the Informer’s encoder for removing reliable remote dependencies from inputs for long-term series. The extra-long incoming data is given to the encoder

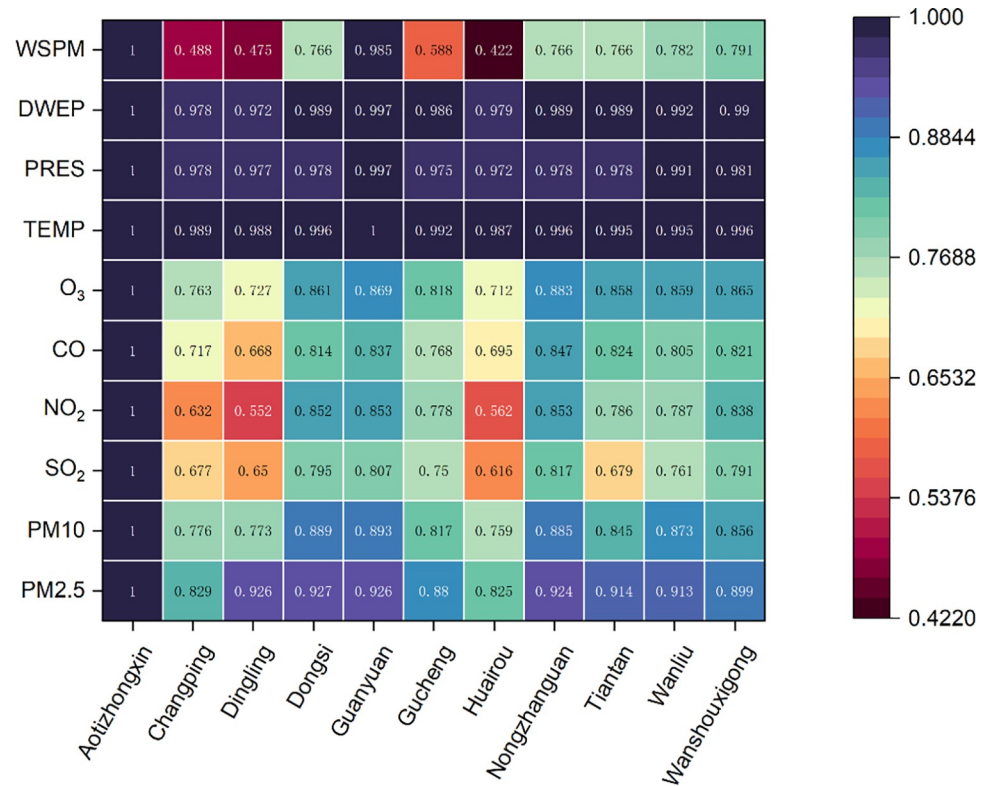


Fig 6. Spatial correlation analysis. Correlation coefficients of meteorological data and pollutant concentration data at each station.

<https://doi.org/10.1371/journal.pone.0287423.g006>

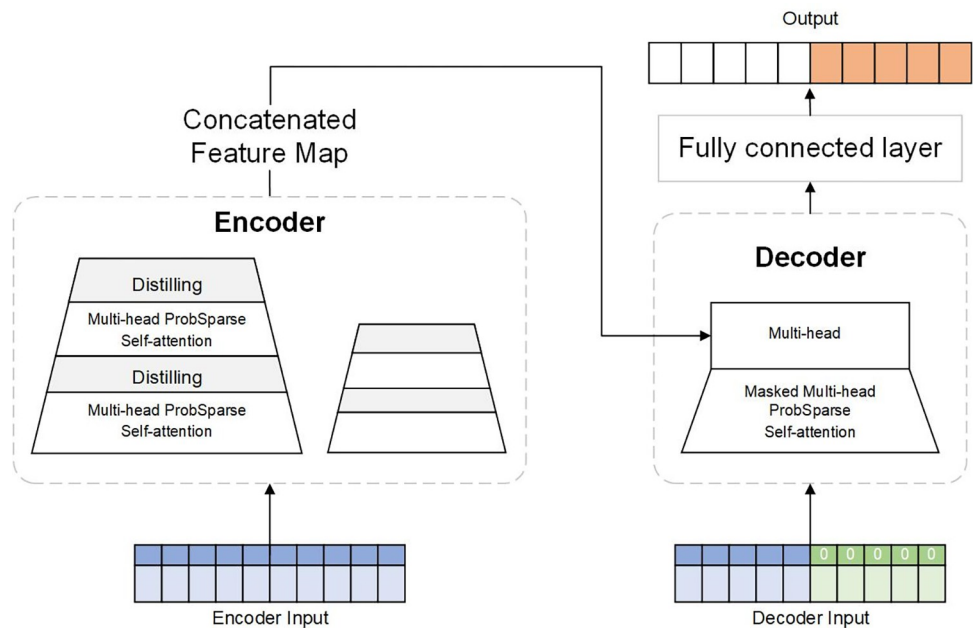


Fig 7. Model structure of Informer. Informer model maintains encoder-decoder structure, and it adds ProbSparse self-attention mechanism, which effectively reduces time complexity and memory usage.

<https://doi.org/10.1371/journal.pone.0287423.g007>

component, as seen in Fig 7. The ProbSparse Self-Attention layer is used in lieu of the traditional Self-Attention layer in the encoder portion, and feature compression is then carried out by the Self-Attention distilling operation. The encoder module increases the robustness of the algorithm by stacking the two processes mentioned above.

The temporal complexity of self-attention dot product computation from optimization is resolved by ProbSparse Self-Attention. Instead of selecting the total dot product, the dot product computation is carried out by sampling each individual point. Allowing each person to concentrate solely on the main allows for ProbSparse self-attention. The equation reads as follows

$$A(Q, K, V) = \text{softmax}\left(\frac{\bar{Q}K^T}{\sqrt{d}}\right)V \quad (5)$$

Where Q , K , and V are three matrices of the same size obtained by a linear transformation of the input feature variables, respectively ($Q \in R^{L_Q \times d}$, $K \in R^{L_K \times d}$, $V \in R^{L_V \times d}$). d is a input dimension; \bar{Q} is obtained from Q by probabilistic sparsification; Softmax is the activation function.

Self-Attention distilling is added as an improvement to the encoder for reducing the size of feature maps. the Self-Attention distilling operation is used for topic compression as a natural consequence of ProbSparse Self-attention, the encoder's feature mapping brings about a V redundant combinations, using distilling to privilege dominant features with dominant features, and generating focus self-attention feature maps at the next layer.

2. Decoder

The decoder gets a long sequence of inputs and places a zero in the predicted target position. The final stage then involves passing through the Attention layer of the Mask to produce the anticipated output. The decoding process is described above, in which each decoder layer conducts feature extraction operations in the target's direction based on the input.

A two-layer stack of Multi-head attention mechanism levels makes up the decoder. Because the future information is unknown at the time of generation, the first layer, Masked Multi-head ProbSparse self-attention, partly masks it out. The current output is then computed using only the prior data using the ProbSparse self-attention method. Multi-head ProbSparse self-attention is the second phase.

The Decoder gets a long sequence of inputs and places a zero in the projected target point. Then the predicted output is generated in the last step by passing through the Attention layer of the Mask. Each decoder layer performs feature extraction operations in the direction of the target based on the given input, and the above process is the decoding process. The decoder uses the encoder's output data to determine what the current decoding should produce in the second layer, which is known as Multi-head ProbSparse self-attention.

ST-Informer. In this paper, an improved model based on Informer is used to capture the spatiotemporal correlation of the data. As shown in Fig 8, the structure of the ST-Informer model is illustrated. In the constructed prediction model, the spatiotemporal data of a region over a while is collected as the input to the prediction task. The input data is a two-dimensional spatiotemporal matrix $X(X \in R^{T \times S \times W})$, where T represents the time lag, and S means the air quality monitoring station information. W represents the air pollutants and meteorological characteristics. The input data is first input to the temporal embedding layer to extract the temporal correlation of the data fully. The temporal embedding layer uses the location embedding in the Transformer model to fully extract the temporal properties of this data. The spatial embedding layer directly captures the spatial characteristics of the input data using the multi-

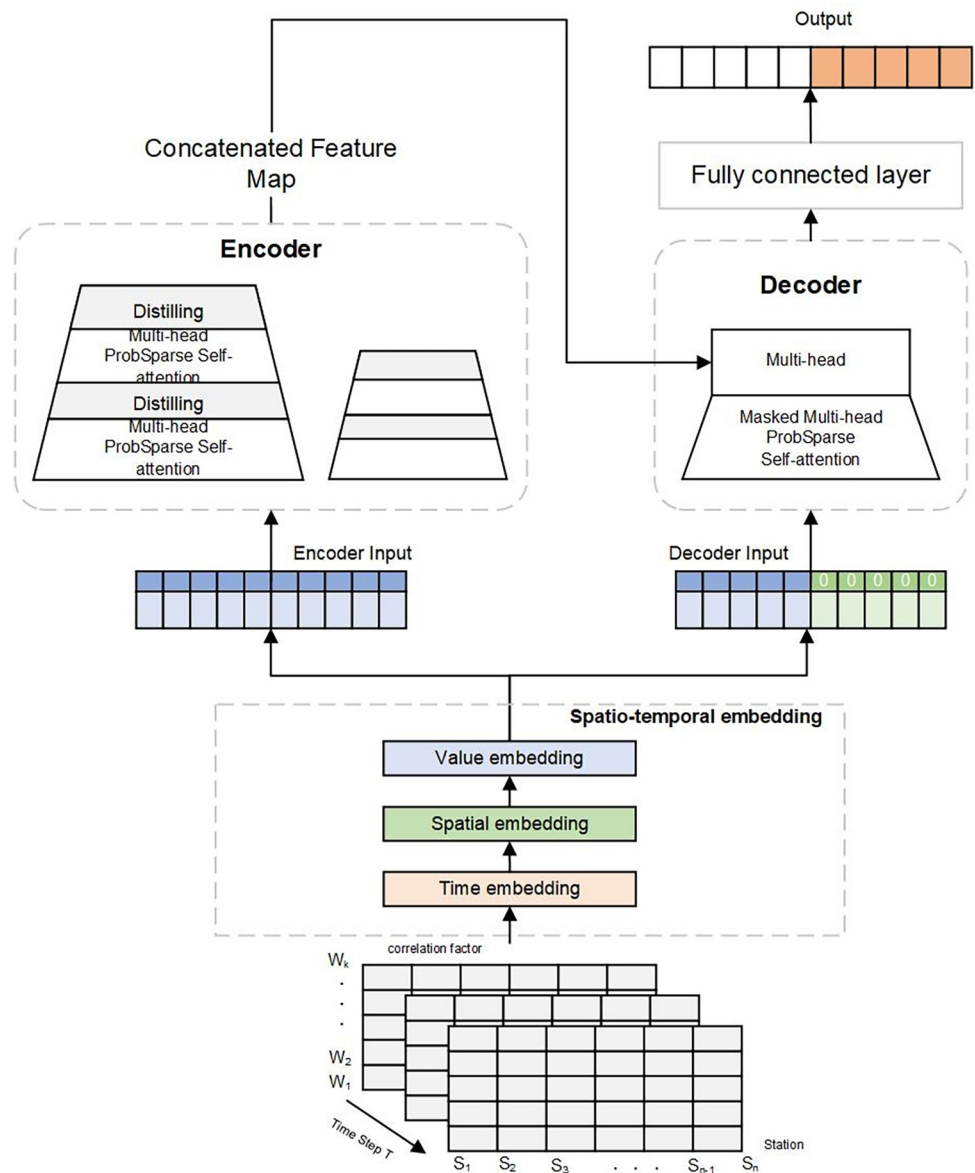


Fig 8. Structure diagram of ST-Informer model. A spatio-temporal embedding layer is added to the original Informer model to fully analyze the spatiotemporal characteristics of the input data.

<https://doi.org/10.1371/journal.pone.0287423.g008>

headed self-attentive mechanism. The addition of the spatial embedding layer captures not only the data’s static elements but also the data’s dynamic spatial dependencies. The value embedding layer separates different timestamps to compute correlations between sites and uses the correlations as weights to adjust the high-dimensional representation of all site input variables. Thus, the complex interdependencies between variables and their dynamic relationships are highlighted in the model. The spatial embedding layer, temporal embedding layer, and value embedding layer data are then linked to represent a comprehensive embedding feature. In the spatiotemporal embedding layer module, the main processes implemented in each layer are.

1. Temporal embedding layer

The model can represent temporal information thanks to the addition of a temporal embedding layer. Since the data’s reliance on time is complex, we must embed not only the sequence’s own temporal characteristics but also its position at the time of the input sequence. Positional coding must be used to embed the data’s temporal properties while taking into consideration the time series’ inherent periodicity.

$$\left\{ \begin{array}{l} PE(pos, 2i) = \sin\left(\frac{pos}{1000d_{model}} \cdot 2i\right) \\ PE(pos, 2i + 1) = \cos\left(\frac{pos}{1000d_{model}} \cdot 2i\right) \end{array} \right. \tag{6}$$

d_{model} refers to the dimensionality of the data after mapping through the input layer; pos refers to the position of a mapped feature in a row of data in the input data; $2i$ and $2i+1$ denotes the number of rows of data; Eq (6) converts the data’s temporal order into a vector by mapping each series’ integer index to a high-dimensional representation. The input to the final time embedding layer is produced by extending each time features into a vectorized index for each new timestamp. The time embedding layer functions as a lookup table, storing embeddings in fixed hours, days, weeks, and months.

2. Spatial embedding layer

Using a multi-headed self-attentive mechanism, the spatial embedding layer immediately captures the spatial characteristics of the input data. Each air quality tracking station’s data has characteristics of spatial and temporal correlation, and by adding a spatial embedding layer, it is possible to record both the static and dynamic spatial dependence of the data.

The input values ($N_S \times N_V$) at each moment are first projected into a high-dimensional subspace utilizing a feedforward neural network.

$$Z_t = w * X_t + b \tag{7}$$

Z_t is a high-dimensional representation of the input variables at moment t . Using the above equation, the site i and site j correlation strength $S_{t,ij}$ can be calculated using the mutual covariance function as follows.

$$S_{t,ij} = xcorr(Z_{t,i}, Z_{t,j}) \tag{8}$$

S_t is obtained as a square matrix of $N_S \times N_V$, representing the strength of the correlation between the two sites at moment t . The value of the spatial embedding layer output E_S at a given moment is calculated as

$$E_S = w * S_t * Z_t + b \tag{9}$$

3. Value embedding layer

The value embedding module uses a one-dimensional convolution of the input dimension d_{model} to represent the input sequence in high dimension as E_V . The value embedding layer calculates the correlation between sites by separating the different timestamps and uses the calculation results as weights to adjust the high-dimensional representation of all site input

variables. The value embedding does not separate timestamps. It makes use of the whole input time series and turns it into a single, highly dimensional representation matrix to illustrate the intricate relationships and dynamics between the model’s variables. A full embedding feature is then represented by connecting the spatial, temporal, and value embedding layer data

$$X_{ST} = E_T \oplus E_S \oplus E_V \tag{10}$$

The encoder-decoder of the Informer model is then fed data from the value encoding layer of the outputs. The Transformer model’s initial self-attention mechanism is replaced by a multi-head probabilistic sparse self-attention mechanism in the encoder, which significantly reduces the computational scale of the model and boosts robustness through layer-by-layer superposition. By padding the target prediction data to zero, measuring the feature map’s weighted attention composition, and then going through a hidden multi-head probability sparse self-attention mechanism layer, the decoder produces the prediction output.

Results

Performance of ST-Informer

The research model in this article was built using the meteorological data and pollutant concentration data introduced in Beijing, with 80% of the data serving as the training set, 10% as the test set, and 10% as the validation set. Following preprocessing, the tested outliers are marked as missing values in the collected dataset, and the KNN algorithm is then used to fill in the missing values.

The constructed prediction model of spatiotemporal data is used to predict the future PM_{2.5} concentration values at the central location using the collected 24 hours of spatiotemporal data of a particular area as the input sequence of the prediction task. By adjusting various batch sizes, learning rates, and iterations, the ST-Informer model’s training was used in this study to identify the ideal set of model parameters. The parameters chosen for this model in this exercise are shown in Table 3. The Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the Coefficient of determination (R²) are then selected as the evaluation indicators of the model in this study. If the MAE and RMSE are smaller and the R² is closer to 1, it means that the model has higher prediction accuracy.

$$MAE = \frac{1}{n} \sum_{i=1}^n |X_{act}(i) - X_{pred}(i)| \tag{11}$$

Table 3. Setting of ST-Informer model parameters.

attribute name	parameter setting
Encoder stack	3\2\1
Batch size	128
training cycle	20
Dropout	0.01
optimizer	Adam
learning rate	0.0001
activation function	Relu
loss function	MSE

<https://doi.org/10.1371/journal.pone.0287423.t003>

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [X_{act}(i) - X_{pred}(i)]^2}{n}} \quad (12)$$

$$R^2 = \frac{\sum_{i=1}^n [X_{pred}(i) - \bar{X}]^2}{\sum_{i=1}^n [X_{act}(i) - \bar{X}]^2} \quad (13)$$

in which n is the sample size of the test set; $X_{act}(i)$, $X_{pred}(i)$ are the true and predicted values of PM_{2.5} concentration at moment i ($i = 1, 2, \dots, n$); \bar{X} is the average of the true values of the predicted samples.

Model comparison

Three models were chosen by the experiment for comparison with the ST-Informer model used in this research in order to assess the models' performance.

1. **LSTM:** LSTM is an RNN model with advantages over traditional neural network frameworks and is often used for time series prediction; and adds a memory cell to determine whether the information is valid, solving the gradient vanishing and gradient explosion problem during training of long sequences. This improvement allows it to perform better in longer sequences.
2. **Transformer:** The Transformer model is essentially a model built based on the Attention mechanism. Attention can solve the long-range dependency problem of RNN and its variants and supports parallelized computation.
3. **Informer:** Informer is an improved model based on Transformer that is more suitable for long sequence prediction. Informer aims to improve self-attentive mechanisms, reduce memory use, and speed up inference.

The above three models were selected in order to verify the advantages of this model in long-term prediction, we selected different times for PM_{2.5} concentration prediction. The above four models were used to predict PM_{2.5} concentrations for the next 4 hours, 8 hours, 12 hours, 16 hours, 20 hours, and 24 hours, respectively. The optimal set of hyperparameters was selected for each prediction model, and the results obtained were calculated for the evaluation indexes shown in Table 4. Fig 9 shows the specific values of each evaluation index for predicting different times in the future, and it can be seen in the figure that all four models' show good results in predicting PM_{2.5} concentrations at 4 hours, with the value of R^2 above 0.85. As the prediction length increases, the MAE and RMSE of all four models increase and the value

Table 4. Compare the performance of different models.

Model	MAE($\mu\text{g}/\text{m}^3$)	RMSE($\mu\text{g}/\text{m}^3$)	R^2
ST-Informer	7.50 \pm 3.16	4.31 \pm 2.04	0.88 \pm 0.05
LSTM	12.50 \pm 7.41	10.34 \pm 6.65	0.80 \pm 0.12
Transformer	9.34 \pm 5.66	7.89 \pm 4.21	0.84 \pm 0.06
Informer	8.91 \pm 3.57	6.52 \pm 3.06	0.86 \pm 0.04

<https://doi.org/10.1371/journal.pone.0287423.t004>

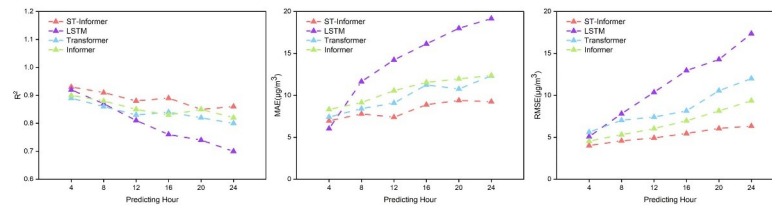


Fig 9. The value of the model evaluation index. The values of the evaluation indicators obtained by predicting PM_{2.5} in different models.

<https://doi.org/10.1371/journal.pone.0287423.g009>

of R^2 decreases. This proves that the longer the prediction time is, the error of the model increases and the prediction becomes worse. the values of *MAE* and *RMSE* of the LSTM model increase significantly with the increase of the prediction time compared to the other models, and the difference between the predicted and the actual values is larger. The prediction accuracy of the Transformer and Informer models is lower than that of the ST-Informer model, and the ST-Informer model does not have a sudden decrease in prediction accuracy due to the long prediction sequences. The ST-Informer model has better prediction results than the other models mentioned above for longer sequences.

To more visually show the advantages of the ST-Informer model in prediction, Fig 10 shows the prediction results of the four models for PM_{2.5} for the next 4 hours, 8 hours, 12 hours, 16 hours, 20 hours, and 24 hours, respectively. The PM_{2.5} concentrations were predicted from 0:00 a.m. on January 1 to 12:00 a.m. on January 3, 2017 (a total of 60 hours), and it can be seen from the figure that the predicted PM_{2.5} concentrations using the ST-Informer model are closer to the true values, and the model can more fully extract the spatial and temporal characteristics of the data. As the predicted future PM_{2.5} concentrations increase in time, the ST-Informer model shows better performance, and the model can capture the sudden changes in the predicted values more accurately and has better prediction results.

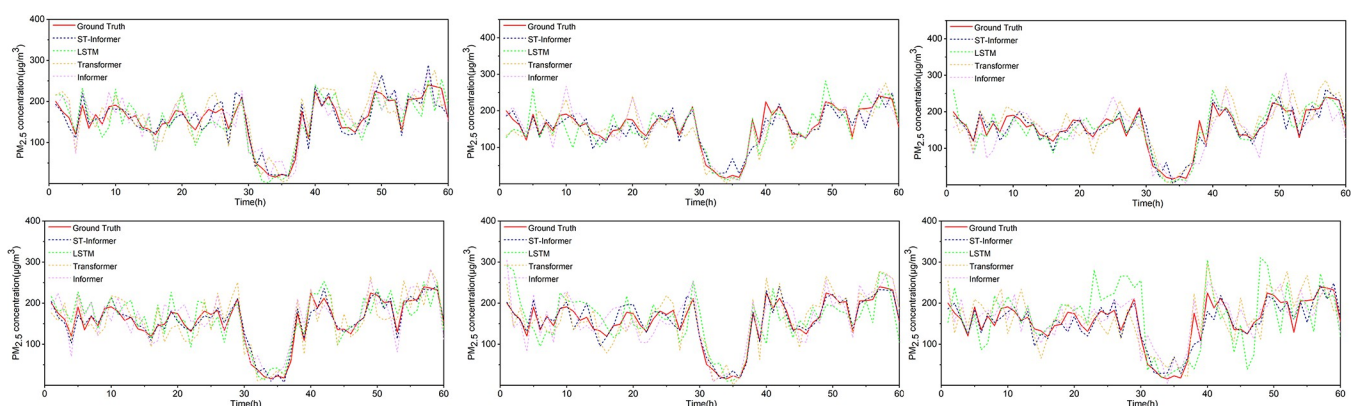


Fig 10. The results of four models to predict PM_{2.5} concentration in different time periods in the future. (a) The results of four models predicting PM_{2.5} concentrations over the next 4 hours; (b) The results of four models predicting PM_{2.5} concentrations over the next 8 hours; (c) The results of four models predicting PM_{2.5} concentrations over the next 12 hours; (d) The results of four models predicting PM_{2.5} concentrations over the next 16 hours; (e) The results of four models predicting PM_{2.5} concentrations over the next 20 hours; (f) The results of four models predicting PM_{2.5} concentrations over the next 24 hours.

<https://doi.org/10.1371/journal.pone.0287423.g010>

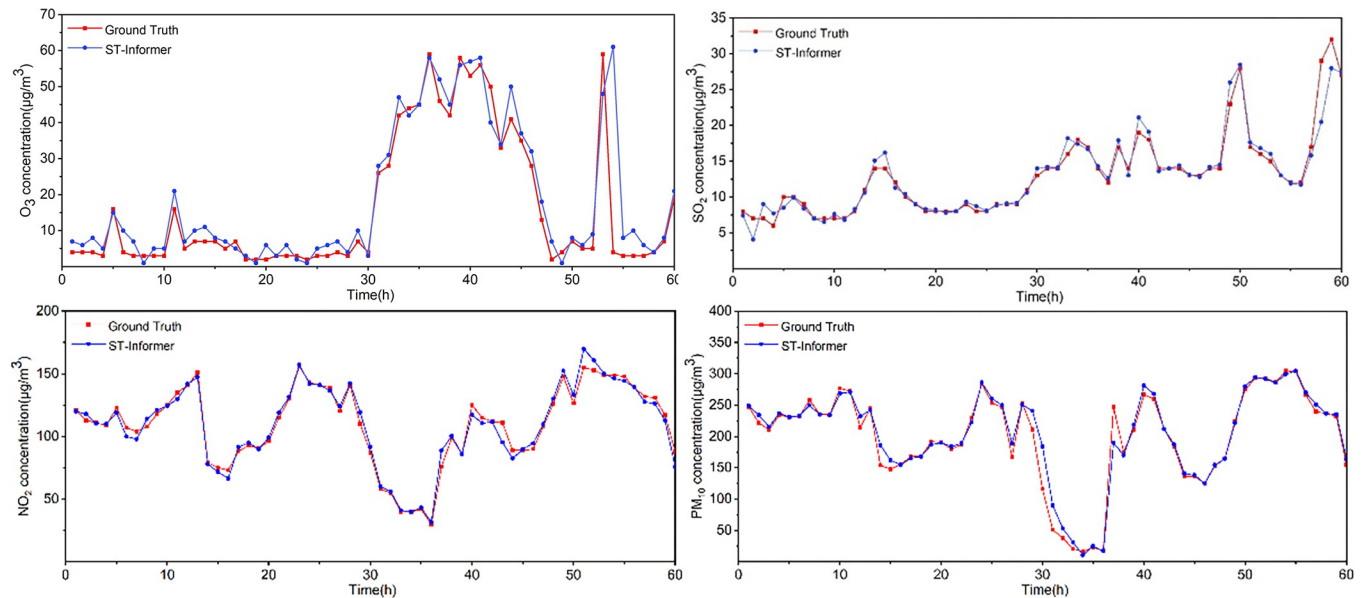


Fig 11. Predicted results of other air pollutant concentrations. The ST-Informer model was used to predict 60-hour other air pollutant concentrations. (a) The results of the O₃ concentration prediction. (b) The results of the SO₂ concentration prediction. (c) The results of the NO₂ concentration prediction. (d) The results of the PM₁₀ concentration prediction.

<https://doi.org/10.1371/journal.pone.0287423.g011>

Application of pollutant concentration prediction

To further analyze the generalizability of the ST-Informer model, the model was used to predict the concentration values of other pollutants. This paper selected four atmospheric pollutants, O₃, SO₂, NO₂, and PM₁₀, as targets for prediction. O₃, SO₂, NO₂, and PM₁₀ are also significant components of atmospheric pollutants, and these pollutants are also extremely harmful to the environment [34–36], so predicting the concentrations of these pollutants also has important practical significance. In past studies, many scholars contended that the prediction of these four pollutants had been studied separately, and all of them have achieved good prediction results. In this paper, in investigating the generalizability of the ST-Informer model, the model was trained with the four pollutant concentrations as the output of the model, and the optimal parameter set of the model was selected. The final prediction results of the model are shown in Fig 11. It can be seen that the ST-Informer model also achieves good results in the prediction experiments of other pollutant concentrations. It is proved that the ST-Informer model can be used to solve various complex spatiotemporal prediction problems.

Discussion

This study aims to combine deep learning models to develop a prediction model for spatial and temporal data, through which PM_{2.5} concentrations can be accurately predicted to provide data support for relevant departments for PM_{2.5} management. Specifically, this paper does the following work: (1) Taking Beijing as an example, meteorological data and pollutant concentration data from several stations in the city are integrated. And the data set is finely pre-processed to make sufficient preparation for the input of the subsequent prediction model and effectively improve the prediction ability of the model. (2) The multidimensional historical meteorological data and pollutant concentration data were analyzed for spatial and temporal correlation, and the dynamic multivariate dependence of the data in space and time was analyzed. (3) Capturing the highly dynamic spatiotemporal dependence in the data. A

SpatioTemporal-Informer (ST-Informer) model is proposed, which utilizes independent temporal and spatial embedding layers to process the input data and capture the complex spatiotemporal properties of the data, and the model's unique ProbSparse Self-Attention mechanism is used to focus on extracting data across time, spatial and multidimensional information of the data. In this study, the research results in the attention mechanism are applied to the spatiotemporal data prediction task, and a novel PM_{2.5} concentration prediction model combining spatial location, temporal variation and contextual information is proposed, which provides a new idea for PM_{2.5} concentration prediction.

In the process of verifying the performance of the ST-Informer model, three more typical deep learning models, LSTM, Transformer and Informer, were selected for comparative analysis. The mean absolute error MAE, root mean square error RMSE and coefficient of determination R² were selected as the evaluation indexes of the models in this study. According to the experimental results, the ST-Informer model has higher prediction accuracy compared with the other three comparison models when predicting PM_{2.5} concentrations of different time durations. In the face of sudden and abrupt changes in PM_{2.5} concentration values, the other comparison models tend to produce erroneous analyses and the prediction results have large errors with the true values, while the ST-Informer model can more sensitively sense large changes in the predicted data. Secondly, to study the generalizability of the ST-Informer model, this paper also applies the model to the prediction of other pollutant concentrations, such as O₃, SO₂, NO₂, and PM₁₀, and compared with the prediction of other pollutant concentrations, the real value of SO₂ is smoother and the model effect is significant, while for the prediction of other pollutant concentrations, the prediction results of ST-Informer model are closer to the prediction results of ST-Informer model are also close to the real results.

The model still needs further improvement in the future: (1) In predicting PM_{2.5} concentrations, the influencing factors are not sufficiently considered, and the influence of factors such as geographical environment is not taken into account. (2) This study only predicted the PM_{2.5} concentration data in a few cities, and the migration analysis of the model is not comprehensive enough. In future work, the ST-Informer model can be applied to predict pollutant concentrations in multiple regions. In addition, the model can also be applied to other studies of spatiotemporal data prediction.

Conclusion

This study proposes a spatiotemporal prediction model: SpatioTemporal-Informer (ST-Informer), which is used to cope with the accurate prediction of fine particulate matter PM_{2.5} in the air. The model introduces a spatiotemporal embedding layer, which calculates the temporal, spatial, and value embedding layers separately and analyzes the data temporal and spatial correlations independently. In addition, the model applies the ProbSparse Self-Attention mechanism, which may contribute to a more focused analysis of the complex dynamic spatial and temporal variation characteristics of PM_{2.5} concentrations. In this paper, the ST-Informer model is applied to predict PM_{2.5} concentration at a site in Beijing. The selected influencing factors include meteorological factors and air pollutant concentration factors. The historical data of adjacent areas are used to predict the future PM_{2.5} concentration values, and the ideal prediction results are achieved. In the experiments of predicting PM_{2.5} concentration, the prediction results of the ST-Informer model proposed in this paper are significantly better than those of the LSTM, Transformer, and Informer models ($MAE \approx 7.50 \mu\text{g}/\text{m}^3$, $RMSE \approx 4.31 \mu\text{g}/\text{m}^3$, $R^2 \approx 0.88$). In addition, the ST-Informer model can more thoroughly analyze the nonlinear relationship between the data and more sharply detect sudden changes in PM_{2.5} concentration values and identify the peaks of the data. In

addition, the model is more generalizable and has been applied to predicting other pollutant concentrations with good results. The ST-Informer model can be applied to prediction problems with complex spatial and temporal characteristics and can also effectively identify peak spikes or pits due to various factors. The results of this study provide data support for the national control of air pollution. Relevant departments can make corresponding air pollution control measures based on the results of this study according to local conditions.

Acknowledgments

We thank the U.S. Embassy in Beijing for the data used in our work. (<http://archive.ics.uci.edu/ml/datasets/Beijing+Multi-Site+Air-Quality+Data>).

Author Contributions

Conceptualization: Zhanfei Ma, Wenli Luo.

Data curation: Wenli Luo.

Formal analysis: Zhanfei Ma, Wenli Luo.

Investigation: Jing Jiang, Bisheng Wang.

Methodology: Wenli Luo.

Project administration: Bisheng Wang, Dongxiang Liu.

Resources: Wenli Luo.

Software: Wenli Luo.

Supervision: Zhanfei Ma.

Visualization: Ziyuan Ma, Jixiang Lin.

Writing – original draft: Wenli Luo.

Writing – review & editing: Zhanfei Ma.

References

1. Cao Z, Zhou JB, Li M, Huang JZ, Dou DJ. Urbanites' mental health undermined by air pollution. *Nature Sustainability*. <https://doi.org/10.1038/s41893-022-01032-1> WOS:000950603100001.
2. Yang H, Huang XY, Westervelt DM, Horowitz L, Peng W. Socio-demographic factors shaping the future global health burden from air pollution. *Nature Sustainability*. 2023; 6(1):58–+. <https://doi.org/10.1038/s41893-022-00976-8> WOS:000871310200003.
3. Aguilar-Gomez S, Dwyer H, Zivin JG, Neidell M. This Is Air: The "Nonhealth" Effects of Air Pollution. *Annual Review of Resource Economics*. 2022; 14:403–25. <https://doi.org/10.1146/annurev-resource-111820-021816> WOS:000865583800018.
4. Yang J, Zhao Y, Cao J, Nielsen CP. Co-benefits of carbon and pollution control policies on air quality and health till 2030 in China. *Environment International*. 2021; 152. <https://doi.org/10.1016/j.envint.2021.106482> WOS:000641419900010. PMID: 33706036
5. Zheng GJ, Liu H, Yu CQ, Li Y, Cao ZJ. A new PM_{2.5} forecasting model based on data preprocessing, reinforcement learning and gated recurrent unit network. *Atmospheric Pollution Research*. 2022; 13(7). <https://doi.org/10.1016/j.apr.2022.101475> WOS: 000865782400003.
6. Yang J, Zhou X. Prediction of PM_{2.5} Concentration Based on ARMA Model Based on Wavelet Transform. 2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)2020. p. 63–6.
7. Wang P, Zhang H, Qin ZD, Zhang GS. A novel hybrid-Garch model based on ARIMA and SVM for PM_{2.5} concentrations forecasting. *Atmospheric Pollution Research*. 2017; 8(5):850–60. <https://doi.org/10.1016/j.apr.2017.01.003> WOS:000417391200006.

8. Zhang LY, Lin J, Qiu RZ, Hu XS, Zhang HH, Chen QY, et al. Trend analysis and forecast of PM_{2.5} in Fuzhou, China using the ARIMA model. *Ecological Indicators*. 2018; 95:702–10. <https://doi.org/10.1016/j.ecolind.2018.08.032> WOS:000456907400068.
9. Lara-Benitez P, Carranza-Garcia M, Riquelme JC. An Experimental Review on Deep Learning Architectures for Time Series Forecasting. *International Journal of Neural Systems*. 2021; 31(3). <https://doi.org/10.1142/S0129065721300011> WOS:000627406100002. PMID: 33588711
10. Dai H, Huang G, Zeng H, Yu R. Haze Risk Assessment Based on Improved PCA-MEE and ISPO-LightGBM Model. *Systems*. 2022; 10(6). <https://doi.org/10.3390/systems10060263>
11. Zhu SL, Lian XY, Wei L, Che JX, Shen XP, Yang L, et al. PM_{2.5} forecasting using SVR with PSO-GSA algorithm based on CEEMD, GRNN and GCA considering meteorological factors. *Atmospheric Environment*. 2018; 183:20–32. <https://doi.org/10.1016/j.atmosenv.2018.04.004> WOS:000432761600003.
12. Huang KY, Xiao QY, Meng X, Geng GN, Wang YJ, Lyapustin A, et al. Predicting monthly high-resolution PM_{2.5} concentrations with random forest model in the North China Plain. *Environmental Pollution*. 2018; 242:675–83. <https://doi.org/10.1016/j.envpol.2018.07.016> WOS:000446150300072. PMID: 30025341
13. Biancofiore F, Busilacchio M, Verdecchia M, Tomassetti B, Aruffo E, Bianco S, et al. Recursive neural network model for analysis and forecast of PM₁₀ and PM_{2.5}. *Atmospheric Pollution Research*. 2017; 8(4):652–9. <https://doi.org/10.1016/j.apr.2016.12.014> WOS:000417388800007.
14. Cabaneros SM, Calautit JK, Hughes B. A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software*. 2019; 119:285–304. <https://doi.org/10.1016/j.envsoft.2019.06.014> WOS:000478965700021.
15. Gao X, Li WD. A graph-based LSTM model for PM_{2.5} forecasting. *Atmospheric Pollution Research*. 2021; 12(9). <https://doi.org/10.1016/j.apr.2021.101150> WOS:000701912100001.
16. Tsai YT, Zeng YR, Chang YS, Ieee, editors. Air pollution forecasting using RNN with LSTM. 16th IEEE Int Conf on Dependable, Autonom and Secure Comp/16th IEEE Int Conf on Pervas Intelligence and Comp/4th IEEE Int Conf on Big Data Intelligence and Comp/3rd IEEE Cyber Sci and Technol Congress (DASC/PiCom/DataCom/ CyberSciTech); 2018 Aug 12–15; Athens, GREECE2018.
17. Kristiani E, Lin H, Lin J-R, Chuang Y-H, Huang C-Y, Yang C-T. Short-Term Prediction of PM_{2.5} Using LSTM Deep Learning Methods. *Sustainability*. 2022; 14(4). <https://doi.org/10.3390/su14042068>
18. Sun J, Di LP, Sun ZH, Shen YL, Lai ZL. County-Level Soybean Yield Prediction Using Deep CNN-LSTM Model. *Sensors*. 2019; 19(20). <https://doi.org/10.3390/s19204363> WOS:000497864700011. PMID: 31600963
19. Huang CJ, Kuo PH. A Deep CNN-LSTM Model for Particulate Matter (PM_{2.5}) Forecasting in Smart Cities. *Sensors*. 2018; 18(7). <https://doi.org/10.3390/s18072220> WOS:000441334300242. PMID: 29996546
20. Qi YL, Li Q, Karimian H, Liu D. A hybrid model for spatiotemporal forecasting of PM_{2.5} based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*. 2019; 664:1–10. <https://doi.org/10.1016/j.scitotenv.2019.01.333> WOS:000460245600001. PMID: 30743109
21. Zhao JC, Deng F, Cai YY, Chen J. Long short-term memory—Fully connected (LSTM-FC) neural network for PM_{2.5} concentration prediction. *Chemosphere*. 2019; 220:486–92. <https://doi.org/10.1016/j.chemosphere.2018.12.128> WOS:000458591200052. PMID: 30594800
22. Zhu JQ, Deng F, Zhao JC, Zheng H. Attention-based parallel networks (APNet) for PM_{2.5} spatiotemporal prediction. *Science of the Total Environment*. 2021; 769. <https://doi.org/10.1016/j.scitotenv.2021.145082> WOS:000626618100107.
23. Ma J, Ding YX, Gan VJL, Lin CQ, Wan ZW. Spatiotemporal Prediction of PM_{2.5} Concentrations at Different time Granularities Using IDW-BLSTM. *Ieee Access*. 2019; 7:107897–907. <https://doi.org/10.1109/access.2019.2932445> WOS:000481980800014.
24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al., editors. Attention Is All You Need. 31st Annual Conference on Neural Information Processing Systems (NIPS); 2017 Dec 04–09; Long Beach, CA2017.
25. Devlin J, Chang MW, Lee K, Toutanova K, Assoc Computat L, editors. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Conference of the North-American-Chapter of the Association-for-Computational-Linguistics—Human Language Technologies (NAACL-HLT); 2019 Jun 02–07; Minneapolis, MN2019.
26. Chu Y, Zhang Y, Wang Q, Zhang L, Wang X, Wang Y, et al. A transformer-based model to predict peptide–HLA class I binding and optimize mutated peptides for vaccine design. *Nature Machine Intelligence*. 2022; 4(3):300–11. <https://doi.org/10.1038/s42256-022-00459-7>

27. Li SY, Jin XY, Xuan Y, Zhou XY, Chen WH, Wang YX, et al., editors. Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. 33rd Conference on Neural Information Processing Systems (NeurIPS); 2019 Dec 08–14; Vancouver, CANADA2019.
28. Zhou H, Zhang S, Peng J, Zhang S, Zhang W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. 2020.
29. Gong MJ, Zhao Y, Sun JW, Han CT, Sun GN, Yan B. Load forecasting of district heating system based on Informer. *Energy*. 2022; 253. <https://doi.org/10.1016/j.energy.2022.124179> WOS: 000808012500002.
30. Zhang Y, Li Y, Zhang GJE. Short-term wind power forecasting approach based on Seq2Seq model using NWP data. 2020; 213.
31. Zhang SX, Yao RM, Du CQ, Essah E, Li BZ. Analysis of outlier detection rules based on the ASHRAE global thermal comfort database. *Building and Environment*. 2023; 234. <https://doi.org/10.1016/j.buildenv.2023.110155> WOS: 000962181000001.
32. Ma J, Cheng J, Lin C, Tan Y, Zhang J. Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. 2019.
33. Cheng ML, Fang FX, Navon IM, Zheng J, Tang X, Zhu J, et al. Spatio-Temporal Hourly and Daily Ozone Forecasting in China Using a Hybrid Machine Learning Model: Autoencoder and Generative Adversarial Networks. *Journal of Advances in Modeling Earth Systems*. 2022; 14(3). <https://doi.org/10.1029/2021ms002806> WOS:000776466100017.
34. Dai H, Huang G, Wang J, Zeng H. VAR-tree model based spatio-temporal characterization and prediction of O3 concentration in China. *Ecotoxicology and environmental safety*. 2023; 257:114960. <https://doi.org/10.1016/j.ecoenv.2023.114960> MEDLINE: PMID: 37116452.
35. Zhu SL, Qiu XL, Yin YR, Fang M, Liu XR, Zhao XJ, et al. Two-step-hybrid model based on data preprocessing and intelligent optimization algorithms (CS and GWO) for NO2 and SO2 forecasting. *Atmospheric Pollution Research*. 2019; 10(4):1326–35. <https://doi.org/10.1016/j.apr.2019.03.004> WOS:000472996900030.
36. Ng KY, Awang N. Multiple linear regression and regression with time series error models in forecasting PM10 concentrations in Peninsular Malaysia. *Environmental Monitoring and Assessment*. 2018; 190(2). <https://doi.org/10.1007/s10661-017-6419-z> WOS: 000424113700009. PMID: 29306973