Regional Spatiotemporal Collaborative Prediction Model for Air Quality

GUYU ZHAO1,2, GUOYAN HUANG1,2, HONGDOU HE1,2, HAITAO HE1,2, JIADONG REN1,2

1College of Information Science and Engineering, Yanshan University, Qinhuangdao, China
2Computer Virtual Technology and System Integration Laboratory of Hebei Province, Qinhuangdao 066000, China

Corresponding author: Guoyan Huang (e-mail: hgy@ysu.edu.cn).

This work was supported by the National Natural Science Foundation of China under Grant 61772451.

ABSTRACT Numerous countries in the world pay close attention to the change of air quality and the control of air pollution. Air quality prediction has become a challenging issue owing to the complex characteristics with time-space nonlinearity and multi-dimensional feature interaction. This paper inventively proposes a three-dimensional presentation of the air quality data and establishes a deep model using spatiotemporal collaborative strategy to predict regional air quality, which can properly deal with multiple characteristics of air quality. Firstly, a theory of regional air quality prediction is given with reasonable analyses and scientific definitions, for systematically analyzing the closely connected areas and simultaneously predicting multiple locations. Secondly, a three-dimensional data structure called Relevance Data Cube is constructed utilizing a clustering algorithm, time sliding windows and correlation analysis of factors. This structure depicts the whole relationships among multiple dimensions of air quality data clearly and provides a basic analysis foundation for subsequent processing. Thirdly, based on the above theory and data structure, a prediction model is proposed to demonstrate the applicability of our work. Spatial relevance and factor influence are processed using dimension reduction technique provided by CNN, which performs automatic mining of spatial correlation and factors reaction of air quality. Besides, LSTM is combined with CNN to deal with the temporal dependency among the data dynamically. Finally, the proposed model is compared with other neural networks by a large number of experiments. The model is proved to better adapt to the characteristics and predict the air quality more accurately, which is more suitable and reliable for the field of air quality.

INDEX TERMS Air pollution, deep learning, dynamic model, data mining

I. INTRODUCTION

The deterioration of air quality significantly impacts human health, which has arisen an increasing concern in many countries. According to the researches [1-5], fine particulate pollutants, such as \( \text{PM}_{2.5} \), are extremely easy to inhale into the lungs, which poses a serious threat to human health. Additionally, a recent report by the World Health Organization (WHO) [6] shows that air pollution caused about 4.2 million deaths each year from stroke, heart disease, lung cancer, and chronic respiratory tract and other related diseases. Figure 1 is the distribution of the global Environmental Performance Index (EPI) in 2018 [7]. Among the listed 180 countries, the pollution situation is still very unsatisfactory. Government management and early prevention are the basic functions of environmental protection work, which are not only related to scientific decision-making but also the long-term development of the country.

FIGURE 1. 2018 Environmental Protection scores by country.

Note: Darker shades indicate higher scores in overall Environmental Performance.

At present, various regions in many countries have established a wide range of air quality monitoring systems, resulting in a certain amount of monitoring data [8, 9]. However, there are many difficulties to be solved to indicate air quality more scientifically. Firstly, air quality is affected by
multiple factors [10-12], such as pressure, temperature, humidity, rainfall, illumination, etc. These factors will affect each other or produce physical and chemical reactions, making the research of air quality more dynamic, variable and complex. Considering the complicated data collecting procedure of these factors, studies need to be conducted to identify factors that play major roles in affecting air quality and should be further used rationally. Secondly, the amount of existing air quality monitoring stations is not substantial in a city due to the expensive cost of building and maintaining such a station [13], so that air quality monitoring data should be perceived properly before effective application. Finally, basic theoretical guidance and core technical support of dynamic analyses of air quality are still lacking, leading to the accuracy of air quality prediction difficult to guarantee. Due to the dynamism and complexity, a reliable model is needed to accurately characterize the distribution structure and evolutionary behavior of the regional air quality.

According to data analysis, air quality evolution often follows some regularity with the development of time and space. Therefore, the air quality analysis model should be established, taking both spatially-related and temporally-related features into account [13]. From the angle of time, air quality status in a given area usually has no significant change over a consecutive period. Meanwhile, adjacent areas often appear similar variations in the spatial dimension. Accordingly, relying solely on temporal or spatial dimension cannot achieve the realistic requirement of air quality analysis. Consequently, the key to realizing the accurate analysis of air quality is scientifically representing the main influencing factors of spatial and temporal dimensions and characterizing complex spatial and temporal correlations. Besides, the cooperative spatial-temporal training strategy for prediction has already been demonstrated better than training the spatial aspect and temporal aspect separately in prior researches [14, 15]. Hence, this paper takes advantage of the Spatiotemporal Collaborative strategy for air quality prediction.

As for air quality prediction methods, there exist many schemes roughly divided into two categories: numerical simulation method and statistical analysis method.

The numerical simulation method is a kind of deterministic method that relies on pollution source emission data and physical and chemical reactions. These methods simulate pollutant emission, transmission, and diffusion processes, and predict air quality by model-driven strategy. For example, Kindap et al. [10] employed the Community Multiscale Air Quality (CMAQ) model and the fifth-generation Penn State/NCAR mesoscale model (MM5) aiming to identify the contribution of long-range aerosol transport to air pollution. Besides, Eder et al. [11] coupled NOAA’s Eta meteorological model with EPA’s CMAQ model for forecasting air quality. And, Saide et al. [12] proposed a Weather Research and Forecasting/Chemistry (WRF/Chem) model to predict high pollution events. Such models need lots of prior knowledge and are highly dependent on parameter settings. Therefore, the performances of these methods are not ideal for a large amount of data in our work, and lots of researchers search for other solutions to deal with the large data volume.

The statistical method applies a data-driven strategy to predict air quality taking advantage of historical air quality data and associated meteorological data. As a result, this methodology is much easier than the complex theoretical system of the numerical model and provides a solution for big data processing. Early statistical methods rely mainly on linear regression models, including linear regression [16], multiple linear regression (MLR) [17], and Gaussian process regression [18]. Latterly, Yang, et al. [19] also presented a geographically and temporally weighted regression (GTWR) model to generate ground-level $PM_{2.5}$ concentrations. These models are relatively fast and accurate, but they are not suitable for nonlinear and unstable air quality prediction problem. Otherwise, Zheng et al. [13] proposed an assemble model combined a neural network and a hidden Markov model for spatiotemporal analysis of big data. Since then, neural networks and spatiotemporal analysis have appeared in the field of air quality, and have been widely analyzed and applied. Oprea, Sanda, and Marian compared artificial neural network models for time short-term $PM_{2.5}$ forecasting [20]. Spatio-Temporal Extreme Learning Machine (STELM) method was proposed for air quality prediction [21]. And, a hybrid approach using Local mean decomposition and Support Vector Regression-Elman (LSE) was proposed [22]. In contrast, the neural network models can use the data itself to establish the prediction process, which makes the results have a strong generalization ability. Unfortunately, as a data-driven approach, neural networks suffer from problems with the data itself, such as the fact that real-world data is rarely complete and 100% accurate. In consequence, deep models were taken for better solutions in later studies. Ong, et al. proposed a deep recurrent neural network (DRNN) to predict $PM_{2.5}$ concentration in Japan [23]. Qi et al. [24] come up with a Deep Air Learning (DAL) model embedding feature selection and semi-supervised learning in different layers of the deep learning network. From these researches, Deep Neural Network has shown great advantages in dealing with non-linear problems. It can extract useful information from data to the greatest extent and can process and learn time dependence well for data with time attributes. Therefore, our work constructs the prediction method based on deep network to better deal with the air quality features.

Though early deep models always used Auto-Encoder to extract higher features of data, CNN was used in recent years [25]. Some researchers like Du et al. [26], Feng et al. [27]and Huang et al. [28] adopted convolutional neural network (CNN) [29], which adapted convolution kernel to the specific task, to preprocess raw data and then fed them into LSTM. The perfect performances of CNN are proved in those researches. Also, this paper takes CNN for feature extraction in the first stage of air quality prediction.
Furthermore, there are more and more spatial analyses of air quality using cluster conceptions currently. For instance, Soh et al. [30] used CNN to extract terrain information and utilized LSTM and ANN to extract information from the target station and its high related stations selected by the cluster method. At last, they merged all the information for final prediction. Additionally, the work we did before demonstrated that air quality system can be studied by clustering methods [31]. Hence, the regional prediction in our research is built upon the basis of clustering results.

However, none of the above models show and emphasize the importance of regional air quality analysis and prediction. Also, systematic and complete theory for regional air quality analysis and prediction has not been shown or proposed in previous researches. Note that aerosol factors, meteorological factors, geographical factors, and their interactions form the complex relationships and constraints of regional air quality spatial-temporal characteristics. Regional analysis can fully consider the reaction and interaction of air quality system and avoid artificial interventions. Regional prediction of air quality is a forecasting method that accomplishes a prediction on a region, where the air quality status of different sites associating with each other closely and influencing frequently. Hence, the contributions of this paper for solving the above tasks are summarized as follows.

1. Regional air quality theory is constructed based on the investigations and definitions of air quality from a perspective of systematic. This concept can be treated as an innovative strategy for the analyses of the evolution trends of air quality system. For the demonstration of the regional theory, the regional prediction is conducted based on spatial clustering methods, in which the prediction results of each site in the same cluster are gained synchronously. Consequently, the relation of spatial relevance can be detected equally and closely connected and frequently interacting sites can be interdependent for better analysis.

2. A three-dimensional structure named Relevance Data Cube (RDC) is established with the purpose of proposing a measurement approach of air quality characteristics. This structure is built up by clustering algorithm, time sliding windows and correlation computation of factors, which vividly integrates the information of temporal, spatial and factor relevance. Based on this structure, a novel training strategy employing a sliced data feeding is proposed to mine the temporal and spatial and feature influences dynamically. By combining multi-dimensional features, RDC can fully guarantee the close connection between them and further prediction accuracy.

3. A spatiotemporal collaborative prediction model for regional air quality prediction called STCNN-LSTM is established based on the combination of CNN and LSTM, which improves the regional air quality prediction by analyzing spatial and temporal dependencies and influencing factors simultaneously. The temporal correlation and spatial constraint are defined for integrating temporal and spatial features, which treats the three-dimensional properties equally.

4. On the basis of the RDC structure and the spatial and temporal collaborative strategy provided in this paper, the prediction effects of different neural networks are compared. We have done a lot of experiments to show that the STCNN-LSTM neural network proposed in this paper can better process data with factor attributes and spatial-temporal correlations, which greatly improves the prediction accuracy.

II. PRELIMINARIES

The basic networks utilized in our model is firstly introduced to simplify the related description of the framework construction procedures.

A. CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Networks (CNN) is a deep neural network model [31], which has been widely used in speech analysis [32] and image recognition [33]. The CNN model has two characteristics: local receptive field and weight sharing. Local receptive field means that a single neuron in each layer of the network is only connected to neurons in the corresponding neighborhood of its input layer. Weight sharing can greatly reduce the training parameters of the network model and requires relatively few training samples. Figure 2 shows the major process of convolution.

FIGURE 2. Convolution process.

B. LONG SHORT-TERM MEMORY

Long Short-Term Memory (LSTM) [34] is a long-term and short-term memory network and a time-recursive neural network, which is suitable for dealing with and predicting important events with relatively long intervals and delays in time series. LSTM differs from RNN in that the addition of a "processor" to the algorithm is designed to judge whether the information is useful or not. The structure of the processor is called a cell. Particularly, there are three gates in a cell, called input gate, forgetting gate and output gate respectively. Figure 3 shows the basic structure of the LSTM unit. Only information that complies with the algorithm's certification will be left, and the information that does not match will be forgotten through the Forgotten Gate.
The weight training process of the LSTM model in this paper, which can be summarized as the information flow interaction process of these gates, is as follows.

**block input:** $z^t = g(W_z x^t + R_z y^{t-1} + b_z)$  \hspace{1cm} (1)

**input gate:** $i^t = \sigma(W_i x^t + R_i y^{t-1} + p_i c^{t-1} + b_i)$ \hspace{1cm} (2)

**forget gate:** $f^t = \sigma(W_f x^t + R_f y^{t-1} + p_f c^{t-1} + b_f)$ \hspace{1cm} (3)

**cell state:** $c^t = i^t \odot z^t + f^t \odot c^{t-1}$ \hspace{1cm} (4)

**output gate:** $o^t = \sigma(W_o x^t + R_o y^{t-1} + p_o c^t + b_o)$ \hspace{1cm} (5)

**block output:** $y^t = o^t \odot h(c^t)$ \hspace{1cm} (6)

Where, $x^t$ is input vector at time $t$, $W_x$, $W_i$, $W_f$, $W_o$ are the weights matrices connecting $x^t$ to the three gates and block input, $R_z$, $R_i$, $R_f$, $R_o$ are recurrent weight matrices connecting $y^{t-1}$ to the three gates and block input, $b_z$, $b_i$, $b_f$, $b_o$ are the bias vectors. $\sigma$ represents the logistic sigmoid function and $h$ represents hyperbolic tangent function. $\sigma$ is used for activation of the gates and $g$ is used as the block input and output activation function, which is set as hyperbolic tangent function.

### III. PROBLEM AND DEFINITIONS

In this section, some descriptions and definitions are given for further comprehension and construction of the proposed model.

1) **BASIC ANALYSIS OF AIR QUALITY**

Air quality monitoring system covers a collection of monitoring sites, reporting the AQI or main pollutant concentration. As investigated in [31], the monitoring sites record the concentration by hour, resulting in a hierarchical organization as shown in Figure 4, of which each layer represents a circumstance of air quality at a given time.

It is normally that air quality analysis needs to consider both temporal and spatial aspects. Firstly, the concentration of air pollutants within a location varies with time continuously, causing the air quality in some places may stay the same status for a period of time. Moreover, sudden changes of air quality at certain times should also be noticed for further analysis. Therefore, time dependency is an indicator to reflect and predict air quality. Secondly, by the transportation and reaction of pollutants between different places, pollutants interact with each other in a certain region. As a result, the levels of pollutants tend to be similar or have continuous grading changes in nearby locations. So, spatial effect needs to be obtained for further analysis.

In addition, the status of air quality is impacted by numerous meteorological factors, such as wind direction, wind speed, humidity, etc. Researches [16, 17] have shown that meteorological impacts are important in restricting the dilution, diffusion, migration, and transformation of atmospheric pollutants so that they are decisive for the evolution of air quality. Inside the air quality system, geographical factors, meteorological factors, and economic factors always have influence or interaction between each other, making the air quality analysis complicated.

The way that influences the air quality status of a place is usually categorized into two types, which can be termed as local pollution and propagation pollution. Local pollution means that the pollutants produce and disappear in a small range, which is mainly caused by the emission of local sources like automobile exhaust or industrial gases and can be dismissed by dilution settlement. Propagation pollution stands for the regional interaction of pollutants through the spreading process, which is influenced by geographical condition and meteorological situation. Under the effect of local pollution and propagation pollution, air quality shows local similarity in a region of certain size, which will form a distinct local structure called local air quality (LAQ). LAQ is composed of several regional sites, of which the air quality status and changing trends are very similar. The air quality analysis can be divided into several areas, according to tight interaction inside LAQ and limited reciprocal effect between LAQs. Here, some conceptions can be given for better understanding.

![Figure 4. Air quality spatial and temporal relation analysis diagram.](image-url)
There, $v_i$ stands for regional nodes (monitor stations), $G_i$ stands for the LAQ, and $M(t_0)$ respects the regional air quality at time $t_0$.

(2) Temporal correlation of air quality. Considering that air pollutant concentrations of regional nodes change with time, the temporal correlations between nodes can be expressed by the similarity of pollutants concentration and its variation trends over a certain period. Thus, the basis for realizing the temporal correlation between regional nodes is computing correlations of pollutants concentration sequences.

(3) Spatial dependence of air quality. Meteorological factors (such as wind direction) and geographical factors (such as distance) are important constraints for the spatial dependence analysis of regional nodes. According to the propagation procedure, propagation cost of pollutants is computed mainly based on the distance of nodes pair.

(4) Air quality distribution and dynamic evolution. Temporal correlation and spatial constraint of regional air quality can reflect or determine the distribution of air quality and similarity of LAQs over different periods. As shown in Figure 4, under the influence of spatial-temporal relations, the air quality is developed as $M(t_1) \rightarrow M(t_2) \rightarrow \cdots \rightarrow M(t_n)$.

By learning the spatial and temporal association of regional air quality, a three-dimensional presentation model is established shown as Figure 5, whose main goal is rationally characterizing the distribution and interaction process of air quality and supporting further effective prediction.

2) BASIC TERMS OF REGIONAL AIR QUALITY PREDICTION

In this case, this paper is to find a reasonable and scientific model to improve the accuracy of regional prediction integrating spatial-temporal relevance and relevant factors. Based on the above analysis, the basic terms used in this paper is defined as follows.

**Definition 1. Spatial Nodes (SN).** Spatial Nodes is given for conveniently expressing the spatial dimension for air quality analysis and prediction. Usually, SN refers to the locations with monitoring equipment, which can continuously observe and represent the air quality changes in an area. SN is defined as follows.

$$SN = \{s_1, s_2, \ldots, s_K\}$$  \hspace{1cm} (8)

Here, SN stands for the set of spatial node $s$ in the research area, $s_i$ is the $i$th element of SN. And, $K$ means the total number of spatial nodes in SN.

**Definition 2. Time Points (TP).** Time points can be treated as a transformed expression of time series, which represents the time points recording the air quality status of the main research period. The set of all the TP can be expressed as:

$$TP = \{t_1, t_2, \ldots, t_D\}$$  \hspace{1cm} (9)

The number of TP in the set is expressed as $D$. $t_i$ is the $i$th time point of TP.

**Definition 3. Influential Factors (IF).** Influential Factors stands for principal factors impacting on the air quality conditions. The set of IF can be expressed as equation 10.

$$IF = \{f_1, f_2, \ldots, f_M\}$$  \hspace{1cm} (10)

The number of factors in the set is expressed as $M$. The data of factor $f_i$ is recorded by hour.

**Definition 4. Relevance Data Cube (RDC).** Relevance Data cube is the structure of data utilized for the input of air quality analysis and prediction as shown in Figure 5. RDC refers to a 3D cube structure that contains three dimensions including space, time and factor dimensions. RDC is defined as follows.

$$RDC = \{SN \times TP \times IF\}$$  \hspace{1cm} (11)

Actually, RDC is the structure that integrates the three aspects of SN, TP and IF by linking the joints of each feature aspect of air quality. Based on RDC, relationships inside air quality are able to be estimated within one model.

**Definition 5. Air Quality Prediction System (AQPS).** Air Quality Prediction System contains the necessary raw and processed materials for series prediction. As analysis depicted above, AQPS consists of three aspects noted as spatial relevance, temporal relevance, and relevance of factors which are introduced as following definitions.

$$AQPS = \{AQSR \cup AQTR \cup AQFR\mid RDC\}$$  \hspace{1cm} (12)

In this equation, AQSR represents the spatial relevance, AQTR is the temporal relevance, AQFR is the influence of factors.

**Definition 6. Air Quality Spatial Relevance (AQSR).** Air Quality Spatial Relevance reflects the spatial relevance of air quality status. AQSR refers to the interaction relationship and interaction intensity of features between spatial nodes, which is a result of transport and reaction of pollutants. The definition is as follows:

$$AQSR = \begin{pmatrix}
SR_{1,1} & \cdots & SR_{1,D} \\
\vdots & \ddots & \vdots \\
SR_{M,1} & \cdots & SR_{M,D}
\end{pmatrix}$$  \hspace{1cm} (13)

$$SR_{i,j} = \begin{pmatrix}
< s_1, s_1 > & \cdots & < s_1, s_K > \\
< s_2, s_1 > & \ddots & \vdots \\
< s_K, s_1 > & \cdots & < s_K, s_K >
\end{pmatrix}$$  \hspace{1cm} (14)

AQSR is expressed as a spatial relevance matrix. In this matrix, each element represents the spatial relevance of different time points and factors. For instance, $SR_{i,j}$ stands for the spatial relevance of factor $f_j$ at time $t_j$. As for a given $SR_{i,j}$, $< s_a, s_f >$ represents the strength of the correlation between spatial nodes $s_a$ and $s_f$.

**Definition 7. Air Quality Temporal Relevance (AQTR).** Air Quality Temporal Relevance shows the pivotal relationship that causes the variation of air quality status in the time dimension. AQTR stands for the temporal dependency between time points which is defined as follow:
\[
AQTR = \begin{pmatrix}
TR_{1,1} & \ldots & TR_{1,K} \\
\vdots & \ddots & \vdots \\
TR_{M,1} & \ldots & TR_{M,K}
\end{pmatrix}
\] (15)

\[
TR_{ij} = \begin{pmatrix}
< t_{1j}, t_1 > & \ldots & < t_{1j}, t_D > \\
\vdots & \ddots & \vdots \\
< t_{Dj}, t_1 > & \ldots & < t_{Dj}, t_D >
\end{pmatrix}
\] (16)

Here, \( AQTR \) is the matrix holding the temporal relevance of different areas and factors. The subscript of \( TR_{ij} \) stands for factor \( f_i \) and spatial node \( s_j \). For a given \( TR_{ij} \), \(< t_{pj}, t_q >\) represents the strength of the correlation between two temporal points \( t_p \) and \( t_q \) in time series vectors.

**Definition 8. Air Quality Factors Relevance (AQFR).** Air Quality Factors Relevance stands for the physical and chemical reactions between different factors that primarily influence air quality.

\[
AQFR = \begin{pmatrix}
FR_{1,1} & \ldots & FR_{1,D} \\
\vdots & \ddots & \vdots \\
FR_{K,1} & \ldots & FR_{K,D}
\end{pmatrix}
\] (17)

\[
FR_{ij} = \begin{pmatrix}
< f_{1j}, f_1 > & \ldots & < f_{1j}, f_M > \\
\vdots & \ddots & \vdots \\
< f_{Mj}, f_1 > & \ldots & < f_{Mj}, f_M >
\end{pmatrix}
\] (18)

Here, \( AQFR \) is the matrix holding the factor relevance at different areas and points. The subscript of \( FR_{ij} \) stands for spatial node \( s_l \) and time point \( t_j \). For a given \( FR_{ij} \), \(< f_{pj}, f_q >\) represents the strength of the correlation between two factors \( f_p \) and \( f_q \).

**Definition 9. Regional Air Quality Prediction Problem.** Regional Air Quality Prediction Problem is defined as the process of optimizing a multi-objective function \( G \) which predicts multi air quality statuses of spatial nodes under similar environments. Basically, the prediction problem is a sequential forecasting task that can be assumed to solve mainly the time series forecast task. Given a set \( S = \{s_1, \ldots, s_K\} \) including \( K \) elements, in which each element represents a spatial node. The formal expression of RAQPP is shown as Formula 19-21.

\[
Y = \{y_1, y_2, \ldots, y_K\} = G(\{x_1, x_2, \ldots, x_K\}, RDC)
\] (19)

\[
x_z = [c_{z,t-D}, \ldots, c_{z,t-N-1}, c_{z,t}]
\] (20)

\[
y_z = [c_{z,t+1}, \ldots, c_{z,t+N}, c_{z,t+N+1}]
\] (21)

Here, \( x_z \) refers to historical concentrations of target pollutant of spatial node \( s_z \in S \) over \( D \) timestamps and \( t \) is the current time. Regional output \( Y \) is the object need to be studied and predicted, such as \( PM_{2.5} \). \( y_z \) is the pollutant concentration of one spatial node. The prediction \( y_z \) is a sequence of concentrations at time series \( \{t + 1, \ldots, t + N\} \). The subscript of \( c \) stands for the station and the time point that owning a concentration record respectively. Besides, \( N \) is the prediction time domain and \( D \) is the amount of past data used for input. Especially, the particle size of the time series of \( PM_{2.5} \) concentration is set to 1 hour.

**FIGURE 5.** The framework of regional air quality spatial and temporal collaborative prediction.

**IV. MODEL AND STRATEGY**

The process of constructing the spatiotemporal dynamic prediction model of regional air quality based on deep learning is divided into two phrases, as shown in Figure 5. Firstly, through the analysis of meteorological data, pollutant monitoring data and geographic location data, the relationships of air quality system are characterized to form a three-dimension structure called Relevance Data Cube. In this procedure, spatial clustering, temporal sliding, and factor correlation calculation are done to find the major ingredients of \( SN \), \( TP \) and \( IF \) forming the \( RDC \). Secondly, the \( STCNN-LSTM \) model is constructed to extract air quality features effectively and carry out the regional prediction. The \( CNN \) model is utilized for mining the \( AQSR \) and \( AQFR \). Then, the extracted features are put into the \( LSTM \) model. In the \( LSTM \) model, the temporal memory of \( LSTM \) is used to deal with the temporal correlation (\( AQTR \)) of air quality. And, regional air quality prediction results are gained to demonstrate the proposed theory.

**A. THE THREE-DIMENSION OF RDC**

Based on the above analysis, the basic methods for constructing the \( RDC \) are given in the following parts.
1) SPACE DIMENSION OF RDC

To perform regional air quality prediction and investigate the Air Quality Prediction System (AQPS), space dimension is firstly constructed using the clustering algorithm. Traditional methods always preset a target location to perform air quality analysis and prediction. Other places surrounding the set location will be treated as impacting sites. As a result, the deviation caused by human intervention could not be avoided. Instead of treating spatial nodes as different roles, our approach utilizes every spatial node in SN equally. By predicting concentrations of each node simultaneously, adjacent stations can be dynamically and entirely analyzed and predicted along with the change of time and influential factors.

Particularly, the SN set is established with a clustering algorithm according to the distances between them. The distance formula is as follows:

\[
\text{hav}\left(\frac{d_{ij}}{r}\right) = \text{hav}(\text{Lat}_j - \text{Lat}_i) + \cos(\text{Lat}_i) \cos(\text{Lat}_j) \text{hav}(\text{Lng}_j - \text{Lng}_i)
\]  

(22)

Here, hav(\(\beta\)) = \(\sin^2\left(\frac{\beta}{2}\right) = \frac{1-\cos(\beta)}{2}\), \(d_{ij}\) is the spherical distance between two points, \(r\) denotes the spherical radius, \(\text{Lat}_i\) and \(\text{Lat}_j\) are the latitudes of \(s_i\) and \(s_j\), \(\text{Lng}_i\) and \(\text{Lng}_j\) are the longitudes.

Based on the distance formula, the K-means algorithm is utilized for generating clusters. For the purpose of comprehending the distribution of clustering results intuitively, the size of each cluster set is given in Table 1. Cluster Name in Table 1 is given according to the belonging cities of most station nodes in the cluster. As Table 1 shown, the largest group containing 37 members is the Beijing Group, of which the station nodes lie in Beijing area. Additionally, the smallest cluster is the Chengde Group 2, consisting of 3 sites. Other clusters own the sizes around 11 to 27. Particularly, the station nodes in the TangQin Group distribute in both Tangshan and Qinhuangdao cities.

<table>
<thead>
<tr>
<th>Cluster Id</th>
<th>Cluster Size</th>
<th>Cluster Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>22</td>
<td>The Tianjin Group</td>
</tr>
<tr>
<td>2</td>
<td>27</td>
<td>The Shijiazhuang Group</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>The Zhangjiakou Group</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>The Tangshan Group</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>The Baoding Group</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>The Chengde Group 2</td>
</tr>
<tr>
<td>7</td>
<td>22</td>
<td>The Cangzhou Group</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>The TangQin Group</td>
</tr>
<tr>
<td>9</td>
<td>37</td>
<td>The Beijing Group</td>
</tr>
<tr>
<td>10</td>
<td>11</td>
<td>The Chengde Group</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>The Hengshui Group</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>The Beijing Group 2</td>
</tr>
</tbody>
</table>

The clustering result is shown in Figure 6, in which different colors represent clusters gained from the algorithm. Each station in the same cluster will be assigned an ordinal number of the cluster and treated uniformly for prediction. Moreover, the order of elements in each cluster of SN is randomly shuffled for ensuring the homogeneity. Hence, this dimension lays the foundation for further analysis of AQSR.

2) TIME DIMENSION OF RDC

To explain the temporal relevance, the time dimension of RDC is built up. For forming the TP set, this paper defines two sliding windows for the time dimension, which are the past time window and the prediction domain window. The past time window is used to produce the feature period applied for predicting, and the prediction domain window is utilized for corresponding prediction labels. Besides, the reasonable window length is selected in the purpose of full and clear experimental comparison. In general, each factor is sliced according to the two sliding windows and cut into a pair of specific time dimension data sets as shown in Figure 7. The windows both slide in steps of 1 hour to build high-quality features and labels. The two window lengths correspond to \(D\) and \(\text{SN}\) in Definition 9. In this way, the sliding windows can highly enrich the training data quantity and retain the dynamic nature of each feature along constructing the time-dimensional features.

3) FACTOR DIMENSION OF RDC

The establishment of factor dimension in the RDC can extend the feature dimension in time and space. By using the influence factors of adjacent stations and the corresponding influence factors during the past period in the time domain, the change of factors can be analyzed systemically and dynamically. Instead of using a single factor in the current moment for analysis, this dimension can avoid the phenomenon of the weak correlation between feature dimension and prediction accuracy. The influence of the
dynamic changes of IF on the prediction of air quality can be completely preserved and analyzed by using the temporal and spatial correlations.

To select the relevant influential factors and study the factor relevance, correlation analysis is used in this paper. The correlation formula is as follows:

$$\text{Corr} = \frac{\text{Cov}(f_i, f_j)}{\sqrt{\text{Var}[f_i] \times \text{Var}[f_j]}}$$  \hspace{1cm} (23)$$

Here, $f_i$ is the $i$th factor, $\text{Cov}(f_i, f_j)$ denotes the covariance of $f_i$ and $f_j$, $\text{Var}[f_i]$ and $\text{Var}[f_j]$ are the variances of $f_i$ and $f_j$, respectively.

According to the correlation formula, a diagram is drawn to display the result explicitly. Here, different colors represent correlation degree gained based on correlation equation. As shown in Figure 8, the pollutant concentration is strongly correlated with the weather condition, humidity and wind speed. The correlation between pollutant concentration and temperature and pressure is a little week. Besides, wind direction is the most irrelevant. Nevertheless, the wind speed shows a strong relevant relationship with temperature and wind direction, of which temperature has a tight correlation with pressure. Therefore, all the above meteorology influences have been considered in our experiment. Along with these factors, the aerosol data is also utilized in IF set, which has been demonstrated relevant to each other and beneficial to the prediction accuracy [35].

![Figure 8. Correlations of influential factors.](image)

### B. STCNN-LSTM ESTABLISHMENT

As shown in Figure 5, the overall framework contains two major part for air quality prediction: the spatial relational feature extraction part based on CNN and the time-dependent predictor based on LSTM. To use 3D-RDC reasonably for regional system prediction, AQSR and AQFR are firstly integrated by CNN. The AQTR analysis is then performed using the LSTM to make predictions. The detailed structure of STCNN-LSTM is shown in Figure 9.

![Figure 9. Prediction model structure of STCNN-LSTM.](image)

1) COLLABORATIVE SPATIAL AND FACTOR FEATURE EXTRACTOR

As spatial interactions must be analyzed upon the distributions of pollutant data and influential factors, AQSR and AQFR are synergistically extracted using CNN network. Structured as a feature matrix, the input grids of CNN contain factors and stations shown in Figure 9. Consequently, CNN abstracts both spatial and factor features as the filters slide over the grids.

Former prediction models always chose a target site as $s_{\text{Target}}$. And, the prediction result is the concentration of $s_{\text{Target}}$'s pollutant of prediction domain. In the proposed model, each spatial node in the same cluster gets into the training process of the neural networks to train the model simultaneously and predict the future air quality pollutant concentration of all the spatial nodes (SN) together. Moreover, most neural network models evaluate the spatial correlation by only using the object pollutant concentrations in adjacent areas as an input of the model. This paper innovatively uses all the relevant factor data (IF) of sites (SN) to carry out model training. In this way, every feature can be dynamically and spatially detected.

The strategy proposed in this paper uses the weights training process of the neural network itself to analyze the spatial and factor correlation strength so that each element in...
SN and IF can be trained in the STCNN-LSTM model as independent individuals. CNN can continuously reduce the dimensions of data and further extract the core features to make the prediction result stable and reliable.

In this procedure, the RDC is converted into several two-dimensional matrices. The dimensions of each matrix represent stations and factors in RDC. Besides, the number of matrices is equal to the time points number. This processing can be treated as that the RDC is cut apart according to the time dimension.

Then, each matrix is fed into the CNN separately. In the convolutional layer, both dimensions of space and factor are extracted using the peripheral nodes in the network. Therefore, similar information will be processed into an integration. And, the dimension of the network will be reduced in the pooling layer. Through the process of convolution and pooling, the two-dimensional input matrix at time $T$ is compressed to attain the AQSR and AQFR for prediction. After that, the AQSR and AQFR are existing in the form of feature maps. As a result, the multiple dimensional features of air quality are highly extracted for further analysis.

2) TIME-DEPENDENT PREDICTOR BASED ON FEATURE MAPS

The training process of STCNN-LSTM for time series prediction based on AQTR are introduced in this section as shown in Figure 9. In this paper, LSTM model is utilized to analyze and predict the time series, which is called the LSTM layer in the picture. And, the prediction object is the concentration of PM$_{2.5}$ in the model.

AQTR is implied in the sequential variation of features. To mine these evolution patterns, LSTM network is applied. Here, we use the time series of features to characterize the temporal correlation degree. As for each spatial node $s_i$, or each factor $f_j$ in different dimensions contains data from the past $D$ hours. Since CNN and LSTM is combined together, the input will be considered as $D$ two-dimensional matrices before time $t$. The data set including observation data of $K$ stations and $M$ influential factors is transformed in CNN models to feature maps. The outputs of CNN’s last pooling layer are converted into one-dimensional vectors and combined together. Therefore, the two-dimensional matrices are extracted into highly concentrated one-dimensional vectors with time-series characteristics. After that, the vectors will be fed into the LSTM layer.

The LSTM layer selectively forgets some past PM$_{2.5}$ data information and other factors of different stations firstly. In the unit of LSTM, new information is decided to be stored for updating status. At last, the LSTM determines the output values and gives it to the fully connected layer.

The fully connected layer is used to decode the CNN-LSTM output and obtain the final prediction results.

Finally, the prediction target is the hourly PM$_{2.5}$ concentration values for $N$ hours after time $t$ based on $D$ hours of data ($D$ and $N$ are the set time windows). It is worth mentioning that the output of this model is the PM$_{2.5}$ concentration prediction sequence of the given cluster group, containing $K$ predicted concentration series chronologically recorded by the hour for the given stations.

In the neural network training process, the data is divided into three sets: training data, verification data and test data. For a given set of input features and output comparisons, the purpose of the neural network is to minimize the error between the data results and the comparison set. In this paper, RMSE is used as the objective function of neural network training, of which the expression will be given in the Experiment section. The stochastic gradient descent algorithm is used for adjusting the weight of the model to obtain the final trained model.

3) STCNN-LSTM ARCHITECTURE

Several hyperparameters should be preset before building the STCNN-LSTM model, including the number of CNN layers and LSTM layer, the number of fully connected layers, the number of nodes in each layer, the correlation thresholds of data and other basic parameters. The effect of each parameter was examined while keeping the other parameters fixed, and a random search method with 5-fold cross-validation was employed to determine the optimum hyperparameters.

Through a series of contrast experiments, the basic structure of the model was first determined. For CNN, a two-layer structure was used. The first layer was a convolutional layer and the second layer was the pooling layer. Also, as to LSTM, one LSTM layer with 300 nodes is utilized. Besides, one fully connected layer with 100 nodes was used. The above structure configuration achieved the best prediction performance in this experiment. Here, an example of choosing the best nodes size of the LSTM layer is elaborated. As shown in Figure 10, the RMSE value is changed with the size of the nodes of LSTM layer. Evidently, the layer with 300 nodes gets the lowest RMSE value in the scatter diagram.

**FIGURE 10. Comparison of different LSTM cell sizes of STCNN-LSTM.**

After that, RMSE used as indicator determines the impact of different correlation thresholds on prediction accuracy. The thresholds include the time lags, the spatial scale of data. And other basic parameters of the prediction model are also
decided through minimizing the RMSE value. For the better illumination of the process of parameter selection in the prediction model, we give an example of learning rate selection. The RMSE and MAE value varied with the allocation of learning rate of the training process, which is shown in Figure 11. The learning rate is compared by different orders of magnitude. And, there are two values of each magnitude as shown in Figure 11. The setting of learning rate with 0.005 gains the lowest RMSE and MAE value, which is consistent with the results of Figure 11.

The data used in this paper are the actual data monitored and recorded by air quality monitoring stations and weather condition monitoring stations in the Beijing-Tianjin-Hebei region. The data set is detected by hours, that is to say, every record of the data is produced every hour. According to the specific details of the data set listed in Table 2, it can be seen that the data set covers a wide range area and has certain representativeness. At the same time, the data is very suitable to verify the regional analysis proposed in this paper because of the strong regional interaction near the surface recorded in the data set.

### TABLE 2. Some Details of Datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Beijing</th>
<th>Tianjin</th>
<th>Average</th>
<th>Nearby cities</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-city stations</td>
<td>36</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-city instances</td>
<td>278,023</td>
<td>189,604</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ave PM$_{2.5}$</td>
<td>106.4</td>
<td>104.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighbor Sta.</td>
<td>233</td>
<td>267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#. of instances</td>
<td>1,272,979</td>
<td>1,436,051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aqi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-city sources</td>
<td>17</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-city instances</td>
<td>176,847</td>
<td>195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nearby sources</td>
<td>116,847</td>
<td>195</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nearby instances</td>
<td>177</td>
<td>1,108,873</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The continuity feature indicates that the observed data of the feature is continuously changing according to time. The temperature, pressure, humidity, wind speed, and concentration of pollutants (PM$_{2.5}$, PM$_{10}$, NO$_2$, CO, O$_3$, SO$_2$) are the continuity factors of the input features in this paper. The standardization formula is:

$$x_{norm} = \frac{x - \bar{x}}{\sigma_x}$$

(24)

Here, $x$ represents the present value of data. $x_{norm}$ represents the data value after standardization. $x$ is the data set of factors. And, $\bar{x}$ is the mean value of the data set. $\sigma_x$ stands for the standard deviation of the data set.

### FIGURE 11. Comparison of different learning rates of STCNN-LSTM.

V. EXPERIMENTAL ANALYSIS

We evaluate our model with different neural network models to perform comparative analysis, which takes advantage of data of air quality monitor stations in Beijing and Tianjin and the surrounding areas. Meanwhile, a large number of experiments show that our experimental results are reducible. This paper uses the data comes from the reference [13] detailed in Table 2.

- Major cities: the 31 cities within Beijing, Tianjin and around 300 kilometers;
- Air quality data: 220 sites per hour of data;
- Real-time atmospheric data: zone level or city level;

### TABLE 3. Status value of weather and wind direction.

<table>
<thead>
<tr>
<th>Weather</th>
<th>Status Value</th>
<th>Wind Direction</th>
<th>Status Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunny</td>
<td>0</td>
<td>No</td>
<td>0</td>
</tr>
<tr>
<td>Cloudy</td>
<td>1</td>
<td>East</td>
<td>1</td>
</tr>
<tr>
<td>Overcast</td>
<td>2</td>
<td>West</td>
<td>2</td>
</tr>
<tr>
<td>Rainy</td>
<td>3</td>
<td>South</td>
<td>3</td>
</tr>
<tr>
<td>Sprinkle</td>
<td>4</td>
<td>North</td>
<td>4</td>
</tr>
<tr>
<td>Moderate rain</td>
<td>5</td>
<td>Unstable</td>
<td>9</td>
</tr>
<tr>
<td>Heaver rain</td>
<td>6</td>
<td>Southeast</td>
<td>13</td>
</tr>
<tr>
<td>Rain storm</td>
<td>7</td>
<td>Northeast</td>
<td>14</td>
</tr>
<tr>
<td>Thunder storm</td>
<td>8</td>
<td>Southwest</td>
<td>23</td>
</tr>
<tr>
<td>Freeing rain</td>
<td>9</td>
<td>Northwest</td>
<td>24</td>
</tr>
<tr>
<td>Snowy</td>
<td>10</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Light snow</td>
<td>11</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Moderate snow</td>
<td>12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Heavy snow</td>
<td>13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Foggy</td>
<td>14</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sand storm</td>
<td>15</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dusty</td>
<td>16</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The data comes from the actual data monitored and recorded by air quality monitoring stations and weather condition monitoring stations in the Beijing-Tianjin-Hebei region. The data set is detected by hours, that is to say, every record of the data is produced every hour. According to the specific details of the data set listed in Table 2, it can be seen that the data set covers a wide range area and has certain representativeness. At the same time, the data is very suitable to verify the regional analysis proposed in this paper because of the strong regional interaction near the surface recorded in the data set.

### A. DATA PREPROCESSING

The linear interpolation method is used to fill the vacancies of aerosol data and meteorological factors. After that, continuous factors are processed by standardization, and discontinuous factors are treated as continuous encoding. Besides, because of the particularity of air quality analysis and the imperfection of the existing criteria for judging noise data, it is impossible to determine whether the data is noise data or a reasonable mutation. For example, pollution emissions and heavy rainfall can easily cause sudden increases or decreases in data. This paper considers that the mutation fitting of data is also an important part of data fitting analysis so that this article tries to preserve data integrity as much as possible.
Discontinuous data uses specific numbers to represent the current states of the feature. The weather conditions and wind direction characteristics are treated as discontinuous factors. The weather states and the wind direction states are shown in Table 3. In particular, the state values of the wind direction data are not continuous numbers as shown in the table.

B. EVALUATION CRITERION

Some evaluation criteria are introduced in the section. Model reliability is evaluated by $R^2$. Otherwise, three evaluation criteria are used for model accuracy analysis, namely mean absolute error (MAE), mean square error (RMSE) and mean absolute percentage error (MAPE). They are defined as follows.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|$$

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

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{y_i}$$

Here, $y_i$ corresponds to the true value of $PM_{2.5}$, $\hat{y}_i$ is the predicted value of corresponding $PM_{2.5}$, and $\bar{y}$ represents the mean value of sample data. $N$ represents the total number of samples. The larger $R^2$, the higher the reliability of the model. In terms of accuracy, the smaller the values of the three evaluation criteria, the higher the accuracy of the representative model prediction.

C. COMPARISON MODELS

To prove the effectiveness of our proposed model, several existing state-of-the-art spatial-temporal models of air quality are adopted in experiments. Following are the details of these compared models.

(1) The AE-RNN model. The AE-RNN model is proposed in [23] which combines spatial-temporal information of air quality based on auto-encoder and recurrent neural network for prediction of a target station in Japan. The model in our paper is conducted for comparison based on the RDC structure that contains all the relevant data into account. In addition, the model is executed for the prediction of the regional area instead of a single station and the region is gained by the clustering algorithm.

(2) The SVR model. The SVR (Support Vector Regression) model compared in our paper is proposed in [36] applying in air quality prediction. In this work, a space-time support vector regression model is proposed to predict hourly $PM_{2.5}$ concentrations of the stations in Beijing city. Different from the model, our work adopts this method by constructing a data structure named RDC and feed it to the SVR. This model is used for predicting the pollutants of the clustering-based regions of the Beijing-Tianjin-Hebei area.

(3) The GRU model. The GRU (Gated Recurrent Unit) model is proposed for air quality prediction in [37] that conducts only temporal prediction of air quality based on the pollution and meteorological time series AirNet data. In our work, the cooperative spatial-temporal training strategy for air quality prediction is adopted by simultaneously input the Spatiotemporal information into this model.

(4) The CNN-LSTM model. The CNN-LSTM model proposed in [38] is a prediction model combining spatial and temporal analysis, which takes the features of neighbor stations and predicts only the $PM_{2.5}$ of the target station. In our paper, this model is adopted on the basis of our proposed data structure RDC to achieve the purpose of regional prediction.

(5) The CNN-GRU model. The CNN-GRU model is proposed in [39] which conducts spatial-temporal prediction of air quality based on recurrent neural networks. This model preprocesses spatial-temporal data by convolution function before feeding them into predictors. In our paper, this model is implemented by feeding data with the RDC structure to realize air quality prediction of a region.

D. EXPERIMENTS OF STCNN-LSTM BASED ON RDC

In different aspects, STCNN-LSTM is compared and analyzed by using evaluation criteria and running time in this paper.

### TABLE 4. Training and model parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datasets span</td>
<td>20000</td>
</tr>
<tr>
<td>Unit</td>
<td>1 h</td>
</tr>
<tr>
<td>Training set</td>
<td>60%</td>
</tr>
<tr>
<td>Validation set</td>
<td>20%</td>
</tr>
<tr>
<td>Test set</td>
<td>20%</td>
</tr>
<tr>
<td>Station set (K)</td>
<td>Cluster size</td>
</tr>
<tr>
<td>Prediction horizon (N)</td>
<td>1</td>
</tr>
<tr>
<td>Past data (D)</td>
<td>48</td>
</tr>
<tr>
<td>Number of sensors (M)</td>
<td>10</td>
</tr>
<tr>
<td>Training method</td>
<td>Stochastic gradient descent</td>
</tr>
<tr>
<td>Kernel size of convolution layers</td>
<td>5*5</td>
</tr>
<tr>
<td>Kernel number</td>
<td>The same with station set</td>
</tr>
<tr>
<td>Kernel size of pooling layers</td>
<td>2*2</td>
</tr>
<tr>
<td>Number of convolution layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of pooling layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of LSTM layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of LSTM layer nodes</td>
<td>300</td>
</tr>
<tr>
<td>Number of fully connected layers</td>
<td>1</td>
</tr>
<tr>
<td>Number of fully connected layer nodes</td>
<td>100</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.005</td>
</tr>
<tr>
<td>Batch size</td>
<td>24</td>
</tr>
<tr>
<td>Value for cross-validation (k)</td>
<td>10</td>
</tr>
<tr>
<td>Maximum epoch</td>
<td>200</td>
</tr>
</tbody>
</table>

The relevant parameters of this paper are given in Table 4. During the training process, RMSE decreases with the increase of iteration times and finally reaches the convergence state as shown in Figure 12. In our work, each test runs ten times to ensure reliable results.

Experiment results of regional air quality prediction are shown and digested based on the RDC. The major research object to predict is set to be $PM_{2.5}$ in our experiment, which is an important influencing element for human health and main pollutant in the field of air pollution research.
1) THE SPATIAL AND TEMPORAL SCALABILITIES OF STCNN-LSTM BASED ON RDC

During this part, the spatial and temporal scalabilities of STCNN-LSTM are compared and analyzed according to prediction accuracy, reliability, and running time.

According to previous studies, a small lag cannot guarantee enough long-term memory inputs for the model, and large time lags permit an increased number of unrelated inputs. To reveal the temporal scalability of STCNN-LSTM, four kinds of time segments are utilized to verify the prediction performances. The corresponding prediction performance was obtained, and the effects of different time lags were investigated.

### TABLE 5. Prediction results and running time of different periods.

<table>
<thead>
<tr>
<th>Period(h)</th>
<th>6</th>
<th>12</th>
<th>24</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>19.38</td>
<td>19.37</td>
<td>19.35</td>
<td>19.36</td>
</tr>
<tr>
<td>MAE</td>
<td>15.56</td>
<td>15.54</td>
<td>15.56</td>
<td>15.53</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.25</td>
<td>0.24</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.707</td>
<td>0.687</td>
<td>0.699</td>
<td>0.699</td>
</tr>
<tr>
<td>Running time (s)</td>
<td>1515</td>
<td>3994</td>
<td>6073</td>
<td>11431</td>
</tr>
</tbody>
</table>

**TABLE 6. Prediction results and running time of different area sizes.**

<table>
<thead>
<tr>
<th>Station Number</th>
<th>6</th>
<th>11</th>
<th>22</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>18.09</td>
<td>22.96</td>
<td>19.36</td>
<td>25.26</td>
</tr>
<tr>
<td>MAE</td>
<td>15.26</td>
<td>17.94</td>
<td>15.53</td>
<td>18.01</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.27</td>
<td>0.33</td>
<td>0.26</td>
<td>0.30</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.696</td>
<td>0.694</td>
<td>0.699</td>
<td>0.796</td>
</tr>
<tr>
<td>Running time (s)</td>
<td>7656</td>
<td>10893</td>
<td>11431</td>
<td>26572</td>
</tr>
</tbody>
</table>

Besides, spatial scalability is demonstrated by four clustering data groups. Two of them called the Beijing Group 2 and the Tianjin Group cover smaller areas, which are colored by red and blue in Figure 6 respectively. 6 stations are selected from the Beijing Group 2 with the condition of having no block missing values. And the other cover 22 stations in Tianjin, which is a complete cluster. The other two clusters cover larger areas called the TangQin Group and the Beijing Group as shown in Table 1. The sites with block missing values are deleted in the two groups to avoid large errors caused by interpolation. After deletion, there are 35 sites left for the Beijing Group. Notably, the TangQin Group has only 11 station nodes left, though the spatial size of the cluster and geographic distance between station nodes are large.

As shown in Table 6, the RMSE increases from 18.09 to 19.36 as the cluster size grows from 6 to 22. Although the difference between RMSE is 1.27, the increase of loss presents a reasonable fact that larger cluster carrying more unstable distribution makes the increment acceptable. Consistent with the forecasting results of the above two small clusters, the forecasting outcomes of the two larger clusters will also lose part of the forecasting accuracy as the scale increases. The RMSE value of the Beijing Group decreased by 2.3, compared with the TangQin Group. But, the reliability of prediction is basically unchanged of the Beijing Group 2, Tianjin Group and the TangQin Group, and even the best dependability value is obtained in the largest cluster.

As a whole, the capabilities of STCNN-LSTM measured from temporal and spatial perspectives accord with our expectations. The performances of STCNN-LSTM with different temporal and spatial scales demonstrate the scalability of this model. Naturally, the running time will rise up with the scale increased.

The Tianjin Group has no block missing values, so the rest of our work is conducted with this group.

2) PERFORMANCE OF STCNN-LSTM SUB-MODELS BASED ON RDC

Sub-models of STCNN-LSTM are also compared and analyzed by using prediction accuracy and running time. In this section, the CNN and LSTM networks are separately applied to perform the prediction. To obtain the best consequences, the past period of these three models are all 48 hours and the spatial scale is set as the 22-station cluster. The parameters of CNN and LSTM are obtained by adjusting the prediction error RMSE during the training process to access the best results.

Table 7 is the comparison results, of which STCNN-LSTM is the best one and LSTM takes the second place. The proposed model acquires 13% and 33% improvements compared with LSTM and CNN. As for $R^2$, STCNN-LSTM shows the best reliability. Due to the most complexity, our proposed model possesses the longest running time.

Totally, STCNN-LSTM is better performed than CNN and LSTM both in prediction accuracy and reliability. However, the running time of STCNN-LSTM is sacrificed.
TABLE 7. Prediction results and running time of sub-models.

<table>
<thead>
<tr>
<th>Method</th>
<th>STCNN-LSTM</th>
<th>CNN</th>
<th>LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>19.36</td>
<td>28.87</td>
<td>22.26</td>
</tr>
<tr>
<td>MAE</td>
<td>15.53</td>
<td>23.98</td>
<td>18.33</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.26</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>R²</td>
<td>0.70</td>
<td>0.41</td>
<td>0.60</td>
</tr>
<tr>
<td>Running time (s)</td>
<td>11431</td>
<td>5481</td>
<td>8077</td>
</tr>
</tbody>
</table>

3) EXPERIMENTAL RESULTS OF DEEP NEURAL NETWORKS BASED ON RDC

Furthermore, different models are compared and analyzed with STCNN-LSTM. The parameters of these contrast models are adjusted to minimize the prediction error of RMSE separately. Other parameter values will cause the accuracy of the prediction to be equivalent to or lower than the current results. Furthermore, each operation runs 10 times for ensuring the prediction results.

Comparative networks adopt Auto-Encoder (AE) to extract features and utilize Back Propagation (BP), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) neural network to predict. So, the investigated models include AE-BP, AE-RNN and AE-LSTM. In experiments, the input of AE is 22*12 (22 stations and 12 factors per station) and the extracted feature outputs are 50, 100 or 300. Besides, the output of AE is fed into BP, RNN or LSTM network for conducting training procedures.

Additionally, an experiment applying transfer learning technique is conducted based on STCNN-LSTM. We train STCNN-LSTM in two procedures. That is to say, the CNN model is trained firstly. Then, the weights of the CNN network are shared to train the CNN-LSTM part. In Table 8, this transfer learning model is expressed as CNN-LSTM. The SVR and GRU prediction model are also compared in the part because these two models have attracted more and more attention in recent researches. Since GRU is a variant of LSTM, we also combine CNN and GRU for regional air quality prediction. The combination is named CNN-GRU in Table 8.

TABLE 8. Prediction results of deep models.

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>time(s)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE-BP (AE: 22*12→300)</td>
<td>34.69</td>
<td>29.07</td>
<td>0.69</td>
<td>3987</td>
<td>0.196</td>
</tr>
<tr>
<td>AE-BP (AE: 22*12→100)</td>
<td>30.15</td>
<td>26.15</td>
<td>0.52</td>
<td>1851</td>
<td>0.314</td>
</tr>
<tr>
<td>AE-BP (AE: 22*12→50)</td>
<td>27.21</td>
<td>23.46</td>
<td>0.35</td>
<td>1431</td>
<td>0.445</td>
</tr>
<tr>
<td>AE-RNN (AE: 22*12→300)</td>
<td>23.61</td>
<td>19.73</td>
<td>0.31</td>
<td>8507</td>
<td>0.576</td>
</tr>
<tr>
<td>AE-RNN (AE: 22*12→100)</td>
<td>24.71</td>
<td>20.92</td>
<td>0.30</td>
<td>2831</td>
<td>0.535</td>
</tr>
<tr>
<td>AE-RNN (AE: 22*12→50)</td>
<td>25.14</td>
<td>21.26</td>
<td>0.34</td>
<td>2520</td>
<td>0.519</td>
</tr>
<tr>
<td>AE-LSTM (AE: 22*12→300)</td>
<td>23.46</td>
<td>19.55</td>
<td>0.29</td>
<td>14374</td>
<td>0.571</td>
</tr>
<tr>
<td>AE-LSTM (AE: 22*12→100)</td>
<td>24.42</td>
<td>20.59</td>
<td>0.31</td>
<td>5957</td>
<td>0.534</td>
</tr>
<tr>
<td>AE-LSTM (AE: 22*12→50)</td>
<td>23.94</td>
<td>20.08</td>
<td>0.31</td>
<td>5173</td>
<td>0.543</td>
</tr>
<tr>
<td>CNN-LSTM</td>
<td>20.08</td>
<td>16.17</td>
<td>0.27</td>
<td>16355</td>
<td>0.687</td>
</tr>
<tr>
<td>STCNN-LSTM</td>
<td>19.36</td>
<td>15.53</td>
<td>0.26</td>
<td>11431</td>
<td>0.699</td>
</tr>
<tr>
<td>SVR</td>
<td>32.68</td>
<td>23.99</td>
<td>0.61</td>
<td>20271</td>
<td>0.411</td>
</tr>
<tr>
<td>GRU</td>
<td>22.26</td>
<td>18.24</td>
<td>0.29</td>
<td>7923</td>
<td>0.605</td>
</tr>
<tr>
<td>CNN-GRU</td>
<td>21.65</td>
<td>17.48</td>
<td>0.28</td>
<td>22353</td>
<td>0.659</td>
</tr>
</tbody>
</table>

From Table 8, STCNN-LSTM stands out as the best performer, of which the values of criteria RMSE, MAE and MAPE achieve the optimized 19.36, 15.53 and 0.26. The best RMSE of models AE-BP, AE-RNN and AE-LSTM are 27.21, 23.61 and 23.46 respectively. The proposed model STCNN-LSTM outperforms these models with improvements of 29%, 18% and 17%.

Then, SVR receives the poorest results in these models, with the RMSE of 32.68. The value of RMSE is 22.26 for GRU, showing moderate performance. Especially, CNN-LSTM achieves an RMSE of 20.08 which is about 3.7% higher than STCNN-LSTM and 7.2% less than CNN-GRU.

In the reliability aspect, our proposed model obtains the highest score of 0.699 which exceeds the other models at least 1.7% and at most 57%. Among these deep models, AE-BP is the most unreliable one and its best R²-value is 0.445. The reliability of SVR is also worse than others, owning the R² value of 0.411. Additionally, models adopting a recurrent neural network have relatively good reliability with R² over 0.65, which proves that temporal relevance mining makes a crucial contribution.

The criteria in Table 8 also reveals that the feature extraction ability of CNN is more applicable to air quality analysis than AE. In other words, CNN is more suitable for the extraction of AQSR and AQRFR. While for AE, there exits an interesting phenomenon. The output feature number of AE is sensitive for AE-BP, which can promote the performance of prediction remarkably. But the AE-RNN and AE-LSTM barely increase the prediction ability with the auto-encoder extracting the features to a higher level. Although, the prediction accuracy of AE-BP is lower than other deep models.

Further, the training time of STCNN-LSTM is relatively acceptable considering its high computational complexity.

Figure 13 shows the fitting curves of six different models, which are CNN, LSTM, AE-BP, AE-RNN, AE-LSTM, and STCNN-LSTM, respectively. The blue or green curves stand for the real value of PM2.5, and the red color curves are the prediction value. To demonstrate the advantage of the proposed model, fitting trends corresponding to 3 different stations are shown to interpret the forecasting capabilities of different models. These stations are chosen from the Tianjin group, which can represent special changing conditions.

When comparing the curves predicted using CNN and LSTM, we can find that CNN is inclined to capturing the trends and sudden changings of sequences. As shown in Figure 13.a, the red line is tended to capture more peaks and troughs than Figure 13.b. This phenomenon proves that CNN can well capture the trends of variety, which is consistent with its ability to extract feature of edges in image processing.

While in contrast to CNN, LSTM and other basic deep models generating prediction curves that tend to fluctuate within a certain range, ignoring some abnormal changes. From Figure 13.b, c, d and e, the red line ranges as a relatively stable curve which does not approach to the sudden changes of the other line. That is to say, LSTM network is capable of characterizing long and short trends of time-dependent series. For the matching of real data and prediction results, CNN, AE-BP, and AE-RNN are not as good as LSTM and AE-LSTM.
FIGURE 13. Prediction fitting results of different models contrast with STCNN-LSTM.
Hence, by combining the CNN and LSTM, the ability of prediction and dealing with sudden changes can be improved as manifested in the figure. As analyzed from these curves, our proposed model of STCNN-LSTM outperforms the other models in regional air quality collaborative prediction problem.

4) EXPERIMENTAL RESULTS OF DIFFERENT POLLUTANTS BASED ON RDC AND STCNN-LSTM

Finally, different pollutants are compared with STCNN-LSTM based on RDC. The PM$_{2.5}$, NO$_2$, O$_3$, and SO$_2$ are chosen to be compared in this part, taking the Tianjin Group and time lag of 48 hours.

The results with three error standards and one reliability criterion of predicting PM$_{2.5}$, NO$_2$, O$_3$ and SO$_2$ are shown in Figure 14. The RMSE values of the predictions are 19.36, 14.83, 16.19 and 16.18 of PM$_{2.5}$, NO$_2$, O$_3$ and SO$_2$. And, the $R^2$ values are 0.699, 0.659, 0.641 and 0.540 for each of them. Distinctly, the outperform of STCNN-LSTM based on RDC almost stays the same for different pollutants, which further proves the rationality and reliability of the theory and model presented in this paper. But it also should be noticed that the MAPE is increased by 0.2 with the $R^2$ declined for O$_3$. This change is acceptable in terms of overall outcomes.

![Figure 14. Prediction results of different pollutants with STCNN-LSTM.](image)

VI. CONCLUSION

This paper provides a novel and effective strategy for regional air quality prediction and utilizes neural networks to demonstrate the theory, namely STCNN-LSTM model. First of all, regional air quality prediction problem (RAQPP) is a novel conception that meets actual needs and has research prospects. The key point of RAQPP is dealing with each spatial node in a closely relative region equally, instead of traditional artificial collection of target sites and adjacent sites as model input. Secondly, the 3-dimension data structure Relevance Data Cube (RDC) is defined to fuse all relevant data. The basic structure RDC integrates three aspects of air quality prediction system: space dimension, time dimension and factor dimension, and establishes the connection between the procedures of analysis and prediction. Furtherly, a novel deep neural network model named STCNN-LSTM is constructed. STCNN-LSTM takes the RDC as input and employs CNN and LSTM models to extract air quality spatial relevance, temporal relevance, and factor influential relevance. CNN extracts the multi-dimensional features to gain necessary information effectively and reasonably utilized without compromising the accuracy. Synchronously, the time memory function of LSTM deals with the time-dependent air quality characteristics and greatly improves the accuracy of the model. At last, the accuracy of the STCNN-LSTM model proposed in this paper performs best compared with other neural networks.

In the future work, we will strive to find a better prediction model, which can extract the characteristic more reasonably and reliably and shorten the time of STCNN-LSTM model to make it better applied to the field of air quality analysis. Along with the present of Sequence-to-Sequence prediction, longer prediction horizon for regional prediction will be performed. Also, the improvement of spatiotemporal collaboration based on regional analysis is the content that can be studied later.

ACKNOWLEDGMENT

The authors are grateful to valuable comments and suggestions of the reviewers and editors.

REFERENCES


