

Received October 13, 2020, accepted October 16, 2020, date of publication October 20, 2020, date of current version October 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3032420

An Unsupervised PM_{2.5} Estimation Method With Different Spatio-Temporal Resolutions Based on KIDW-TCGRU

CANYANG GUO¹, GENGGENG LIU¹, (Member, IEEE), LINGJUAN LYU², (Member, IEEE), AND CHI-HUA CHEN¹, (Senior Member, IEEE)

¹College of Mathematics and Computer Sciences, Fuzhou University, Fujian 350116, China

²Department of Computer Science, National University of Singapore, Singapore 117417

Corresponding authors: Genggeng Liu (liugenggeng@fzu.edu.cn), Lingjuan Lyu (dcslyl@nus.edu.sg), and Chi-Hua Chen (chihua0826@gmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61906043, Grant 61877010, Grant 11501114, and Grant 11901100; in part by the Fujian Natural Science Funds under Grant 2019J01243; in part by the Funds of Education Department of Fujian Province under Grant JAT190026; and in part by the Fuzhou University under Grant 0330/50009113, Grant 510872/GXRC-20016, Grant 510930/XRC-20060, Grant 510730/XRC-18075, Grant 510809/GXRC-19037, Grant 510649/XRC-18049, and Grant 510650/XRC-18050.

ABSTRACT The interpolation of fine-grained air quality has significant prospects in the area of air quality monitoring. The solution to this problem can effectively monitor the air quality of the areas by sparse air quality monitoring stations, so as to reduce the monitoring cost. Most of the existing researches are to solve the problem of air quality monitoring in the areas without stations by different interpolation methods. However, most of them are unable to verify the reliability of the proposed interpolation methods and can't provide a feasible range of interpolation methods at spatial resolutions. Therefore, this paper proposes an effectively unsupervised PM_{2.5} estimation method based on time distributed convolutional gated recurrent unit (TCGRU) and an interpolation method based on k -nearest neighbor inverse distance weighted (KIDW) to solve these problems. The model is trained by the data obtained by the interpolation method with missing one station in turn to get the monitoring capability of the areas without monitoring stations. Inverse distance weighted is combined with k -nearest neighbor to improve the performance of interpolation. In addition, time distributed convolutional neural network extracts the spatial features of air quality and storage time information for extracting temporal features by the gated recurrent units. A large number of experiments are carried out to evaluate the performance of the method by using the air quality dataset of Hubei Province, China. The experimental results show that the proposed model is effective for the monitoring of PM_{2.5} in the unsupervised areas and the performance (i.e. MAE=8.23, RMSE=11.27 and MAPE=19.14%) of estimating PM_{2.5} concentration is better than the comparative methods.

INDEX TERMS Air quality modeling, spatio-temporal analysis, deep learning, inverse distance weighted, k -nearest neighbor.

I. INTRODUCTION

In recent years, the deterioration of air quality has caused widespread concern by researchers and the government has spent a lot of energy on air quality monitoring. It is a significant topic in the calculation of air quality to realize the monitoring of air quality in unsupervised areas. The interpolation of fine-grained air quality solves the problem of a limited

number and uneven distribution of air quality monitoring stations [1]. In addition, accurate monitoring of air quality can protect people from the hazards of air pollution and reduce the incidence of cardiovascular diseases [2], [3]. However, there are several challenges for air quality calculation as the complexity of air quality monitoring. First of all, since the high cost of air quality sensors, air quality monitoring stations are hard to cover every corner of the target areas, resulting in limited access to relevant air quality data and difficultly monitoring the vast areas. Moreover, the current researches only

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang¹.

propose different interpolation methods, which is difficult to verify its effectiveness. Finally, there is no recognized method to select the range of interpolation and the monitoring stations participating in the interpolation. Therefore, it is difficult to design a reasonable interpolation and monitoring method, which is also an important factor for air quality monitoring in the areas without monitoring stations.

Most of the existing researches provides interpolation methods, which are unable to verify the estimation performance of interpolation methods for areas without monitoring stations. Moreover, the models cannot process the spatial and temporal characteristics of air quality well. This paper solves these problems by developing a k -nearest neighbor inverse distance weighted (KIDW) for interpolation and proposing a time distributed convolutional gated recurrent unit (TCGRU) to estimate air quality of unknown areas. This paper utilizes k -nearest neighbor method to select feasible station data to improve the performance of interpolation. Two-dimensional convolutional layer is employed to analyze the correlation of air quality distribution in different areas and GRU is used to analyze the creep characteristics of PM2.5 respectively. In addition, the paper analyzes the effects of different spatial and temporal resolutions on the performance of the model. In summary, the main contributions are listed as follows.

- This paper proposes an unsupervised air quality estimation method with different spatio-temporal solutions, which provide estimations of the areas without stations and provided suggestions of stations distribution.
- This paper proposed a KIDW interpolation method for unknown areas and the experimental results show that the interpolation performance of the model is improved.
- This paper proposed a TCGRU model to extract the temporal and spatial characteristics of air quality and a large number of experiments verify the effectiveness of the model.
- In this paper, a training method of missing one is proposed to solve the problem of verifying the performance of interpolation.

II. LITERATURE REVIEWS

In this section, related works including interpolation method and modeling of air quality will be discussed. With the development of hardware equipment, the researches in air quality modeling showed an increasing trend and the focus of research has changed from traditional machine learning including support vector machine (SVM) and decision tree regression (DTR) to ensemble learning and deep learning [4]–[6]. Recent studies have shown that deep learning performs better than other machine learning methods in air quality modeling [7], [8]. El Alaoui El Abdallaoui *et al.* [9] proposed a method combining artificial neural network (ANN) and big data to predict air quality. In this study, the pollutant data, traffic data and weather data collected by smart phone sensors were utilized to verify the performance

of the model and the experimental results showed that the proposed model could predict the concentration of air pollution to a certain extent. Although this method can be applied in predicting the concentration of pollutants, it does not take into account the spatial diffusion and long-term dependence of pollutant concentration. Krishan *et al.* [10] considered that the existing air quality prediction model couldn't solve the long-term dependence of air quality and proposed an air quality prediction model based on long short-term memory (LSTM) to predict air quality. They verified the effectiveness of the LSTM model via relevant data and experimental results showed the superiority of LSTM compared with ANN. On this foundation, Ma *et al.* [8] proposed a bidirectional long short-term memory (BLSTM) to predict air quality using the data including and the results showed better performance than LSTM. Considering the spatial characteristics of air quality, a convolutional neural network (CNN) is applied to extract the local features of air quality. CNN is a weight sharing network, which has achieved great success in computer vision and natural language processing [11], [12]. Huang and Kuo *et al.* [7] employed a one-dimensional convolutional layer to extract the correlation between PM2.5 and meteorological data and proposed a hybrid neural network combined with CNN and LSTM called 1DCNNLSTM to predict the concentration of air pollution. The experimental results showed that the proposed model had better performance than the LSTM and ANN. Similarly, considering the spatiotemporal characteristics of PM2.5, Tao *et al.* [13] proposed a PM2.5 prediction model combining CNN and BGRU called 1DCNNBGRU. Both of them utilized a one-dimensional convolutional to get the local features to improve prediction performance. However, what they obtained are the spatial characteristics between PM2.5 and auxiliary data (e.g. meteorological data) instead of PM2.5 concentration distribution.

Since the sensors of air quality are expensive, interpolation of fine-grained air quality has significant prospects in the areas without stations. The air quality in the unsupervised area is estimated by using the air quality concentration in the monitored area, which can not only reduce the cost of air quality monitoring, but also has better performance. Zhao *et al.* [14] proposed a land-use regression model for interpolating air pollution measurements in California. Jerrett *et al.* [15] proposed an O3 exposure estimation method with an inverse distance weighting (IDW) interpolation. Beckerman *et al.* [16] proposed a multistep prediction model to predict Beijing PM2.5 concentration combining the spatial information of surrounding stations and the data of Beijing stations according to the geographical distance. Huang *et al.* [17] proposed a machine learning method to estimate daily PM2.5 concentrations in New York City at the neighborhood scale. Qi *et al.* [1] proposed a deep air model to realize the interpolation and prediction of PM2.5 concentrations in areas without monitoring stations. They utilized the fully connected neural network as the model to interpolate the PM2.5 concentration in the unsupervised areas by

two nearest temporal neighborhoods and spatial neighborhoods. Although this method can effectively monitor the concentration of PM2.5 in unsupervised areas, it lacks of the specific analysis of selecting a spatial and temporal neighborhood and the long-term dependence of PM2.5 concentration. Considering that PM2.5 concentration is not only related to historical data but also related to future data. Tong *et al.* [18] proposed an interpolation method that takes into account both past and future data. However, PM2.5 concentration has effectiveness for a given period of time and the time delay of obtaining PM2.5 data will lose most of its monitoring significance. In 2019, Ma *et al.* [8] proposed a method combining IDW and BLSTM to fill the areas without monitoring stations. IDW took into account the influence of distance on PM2.5 to realize spatial interpolation. However, there are limited researches on selecting monitoring station data on the performance of interpolation. K -nearest neighbor (KNN) algorithm provides an idea that if most of the k nearest samples of a sample in the feature space belong to a certain category, then the sample also belongs to this category and has the characteristics of samples in this category [19]. Although these studies provide methods to monitor the air quality concentration in areas without monitoring stations, they are unable to verify the performance of the proposed methods because their experimental results are based on the estimation of known station data, which lacks of verifications of unknown stations. This paper proposed a missing one training method to realize the performance validation of interpolation. For each round of training, it regards one known station as unknown station (i.e. target station) so that proposed method has the ability to estimate unknown station. In addition, the combination of KNN and IDW provides a feasible method for the selection of monitoring data and improves the performance of interpolation.

III. MODEL FORMULATION

This section presents the inverse distance weighting KNN (IDW-KNN) method in Subsection III.A, the convolutional-based bidirectional gated recurrent unit (CBGRU) neural network in Subsection III.B, and the learning method in Subsection III.C.

A. IDW-KNN

Previous studies employed interpolation methods to monitor the areas without monitoring stations, but limited work discussed how to select the known monitoring station data. If all known monitoring station data are taken into account, the amount of interpolation calculation will be greatly increased [20]. In addition, PM2.5 is closely related in space, the correlation will decrease with the increase of distance, which affects the performance of interpolation. This paper develops a reasonable method of selecting monitoring data for interpolation based on IDW and KNN to improve the performance of interpolation. KNN is utilized to select k nearest neighbors of the target station to participate in interpolation.

Figure 1 is the distribution of monitoring stations in Hubei Province, China. In this paper, a rectangular monitoring network is established by dividing different longitudes and latitudes, as shown in Figure 2. The longitude of Hubei is ranging from $108^{\circ}21'E$ to $116^{\circ}07'E$ and latitudes is ranging from $29^{\circ}01'N$ to $33^{\circ}6'N$. A monitoring station is set up at the intersection of each vertical line segment and each horizontal line segment. There is a fixed longitude and latitude between the two monitoring stations (e.g. 0.1 longitude and 0.1 latitudes). The influence of different spatial resolution on interpolation performance will be discussed in detail in subsequent experiments. If a monitoring station already exists in the area, it is called an actual monitoring station; otherwise, it is called a virtual monitoring station. The purpose of interpolation is to fill the data of the virtual monitoring station to achieve the purpose of monitoring PM2.5 comprehensively. The distribution of monitoring stations is shown in Equation (1).

$$D = \begin{bmatrix} v_{11} & s_{12} & v_{13} & \cdots & v_{1n} \\ v_{21} & v_{22} & v_{23} & \cdots & v_{2n} \\ v_{31} & v_{32} & s_{33} & \cdots & v_{3n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ s_{m1} & v_{m1} & v_{m1} & \cdots & v_{mn} \end{bmatrix}, \quad (1)$$

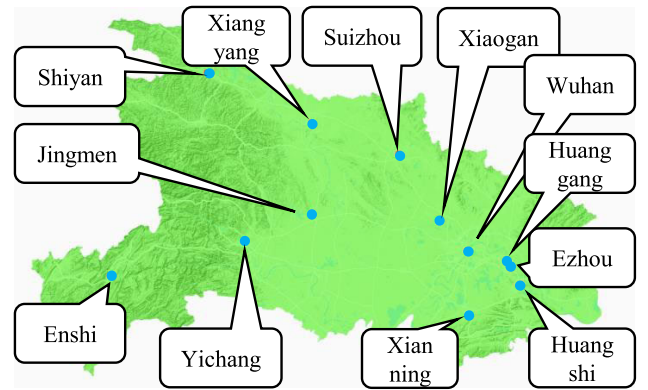


FIGURE 1. The distribution of monitoring stations in Hubei.

where v_{ij} represents the virtual monitoring station of i -th horizontal and j -th vertical area and s_{ij} represents the actual monitoring station of i -th horizontal and j -th vertical area. The data of virtual monitoring station are obtained by IDW using the data of actual monitoring stations. Taking the virtual Station A as an example in Figure 3, the distances between the target monitoring Station A and all the actual monitoring stations are calculated one by one, shown in Equation (2).

$$dis_i = R * ar \cos[\cos(y_A) * ar \cos(y_i) * \cos(x_A - x_i) + \sin(y_A) * \sin(y_i)], \quad (2)$$

where dis_i is the distance between target monitoring Station A and i -th actual monitoring station and R is radius of the earth. x and y are the Latitude and longitude coordinates of the monitoring station. The k actual monitoring stations nearest to the target monitoring Station A are selected by distance and

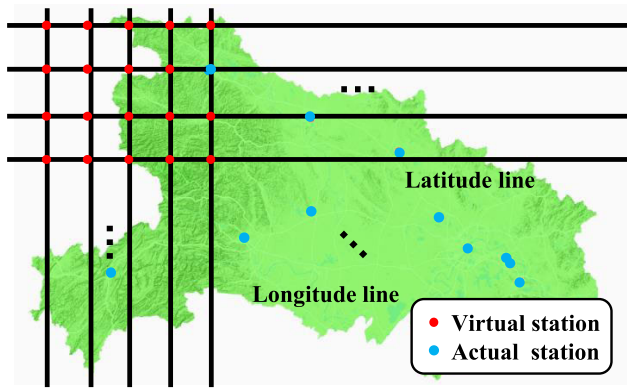


FIGURE 2. The monitoring network including virtual stations and actual stations.

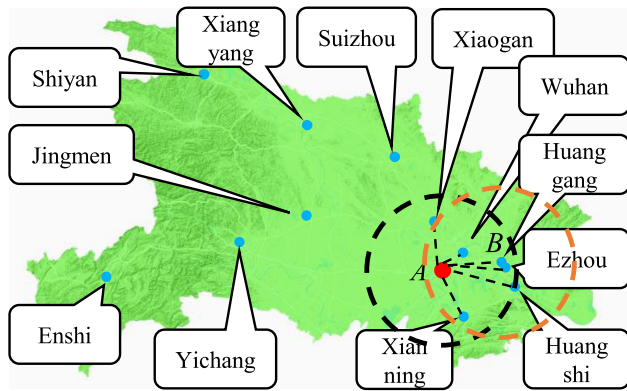


FIGURE 3. An example of interpolation based on IDW-KNN.

the air quality of the target monitoring station is calculated by the weighted method. The weights of selected actual stations and the air quality of target Station A are calculated as Equation (3) and Equation (4).

$$w_i = \frac{\frac{1}{dis_i}}{\sum_{i=1}^k \frac{1}{dis_i}}. \quad (3)$$

$$v_A = \sum_{i=1}^k w_i s_i. \quad (4)$$

It is worth noting that the obtained results of interpolation are not the final estimation results of unsupervised areas. Taking the Station B as an example in Figure 3. There are 12 actual monitoring stations and Station B is one of them. Suppose Station B is the target station and it is regarded as a virtual station. K nearest stations of the station needed to interpolate are selected from the remaining 11 stations to interpolate the station. The input of proposed network (i.e. the PM2.5 concentration distribution of the whole area) is obtained by traversing all the virtual monitoring stations (including Station B). On the other hand, K nearest stations of the station needed to interpolate are selected from all the actual stations (including Station B) to interpolate the

stations. The output of proposed network (i.e. the PM2.5 concentration distribution of the whole area) is obtained by traversing all the virtual monitoring stations. All the 12 actual stations play the part of Station B to get 12 group of training set to train the network in turn. When predicting the PM2.5 concentration of actual Station B, the network gets the input by interpolation method without using Station B, and the output of proposed network the PM2.5 concentration distribution of whole area. The performance of proposed network is measured by comparing the PM2.5 concentration of actual Station B and the output of network with regard to PM2.5 concentration of station B. Similarity, the prediction process is repeated 12 times.

B. CBGRU NEURAL NETWORK

Figure 4(a) shows a typical ANN, which is composed of input layer, hidden layer and output layer and the neurons between layers are fully connected. ANN transmits information layer by layer by forward-propagation and modifies model parameters iteratively by back-propagation. However, since the fully connected characteristics of ANN, it loses the ability to extract local features. Air quality has the local characteristics of space, hence ANN is not competent for air quality modeling effectively. CNN adopts the way of sharing weight and the convolutional kernel in the convolutional layer transmits the regional information to the neurons in the next layer by local calculation, as shown in Figure 4(b). Although the calculation way of CNN meets the spatial characteristics of air quality, it lacks of consideration of historical data the same as ANN. In order to solve this problem, this paper proposes a time distributed convolutional neural network (TCNN). On the basis of CNN, the time distributed layer is added to which is able to storage time information, as shown in Figure 4(c). Although TCNN considers the historical data, it fails to connect the data of the previous moment with the data of the next moment and solve the problem of long-term dependence [21]. Therefore, this paper proposed an air quality estimation model combining TCNN with GRU, where TCNN is utilized to extract the local characteristics of air quality simulating the image processing and storage time information analyzing the temporal creep of air quality by GRU as shown in Figure 5. Both LSTM and GRU can handle long-term dependency problems and have similar performance. GRU is a variant of LSTM and it is chosen in the experiments because it is computationally cheaper [22].

The input of TCGRU is three-dimensional matrixes, which are represented by x^c . Each dimension of the matrix represents the time step, the length of the rectangle and the width of the rectangle where the rectangle represents the monitoring area. The time distributed CNN is utilized to extract local features and storage temporal information as shown in Equation (5).

$$y_t^c = \text{selu}(f(w_t^c x_t^c) + b_t^c), \quad (5)$$

where y^c is the output of the convolutional layer and t is the timestep. w^c denotes the weight matrix and b^c denotes the

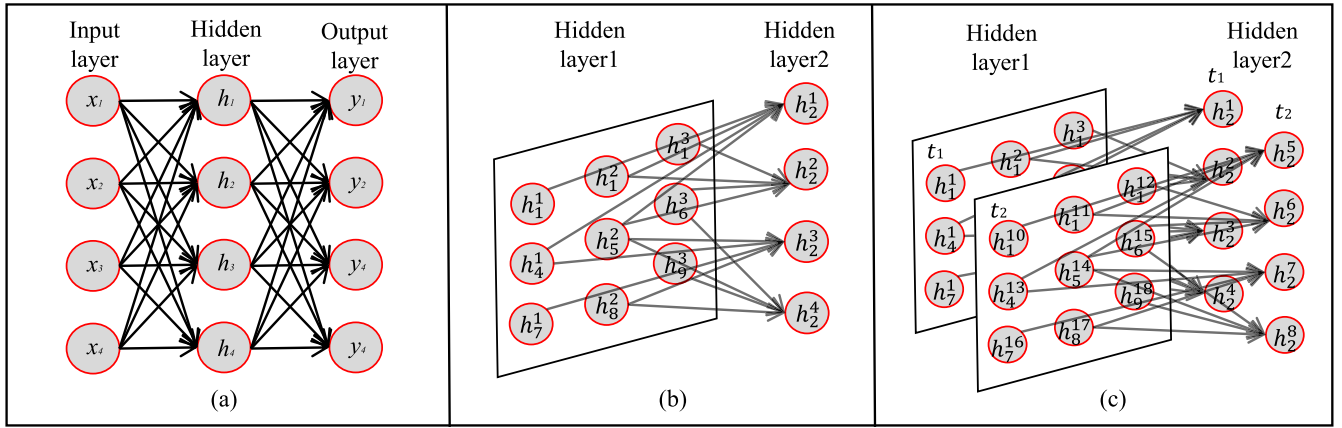


FIGURE 4. Typical structure of neural network.

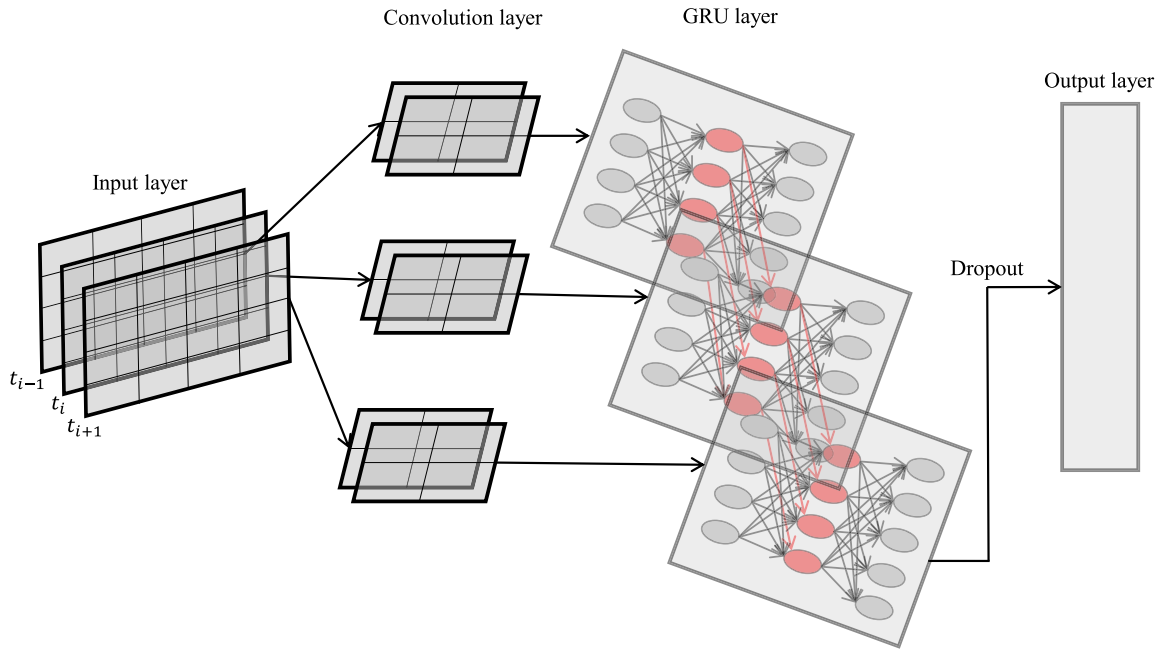


FIGURE 5. The structure of time distributed convolutional gated recurrent unit.

bias matrix of the convolutional layer. selu is the activation function and f is the convolutional computation. The outputs of the convolutional layer are four-dimensional matrixes and each dimension represents the timestep, the length of the rectangle, the width of the rectangle, and the number of channels where the number of channels depends on the number of convolutional kernels. The outputs are compressed into two-dimensional matrixes and each dimension represents the timestep and the number of extracted features respectively by the flatten function. The two-dimensional matrixes are inputted to the GRU layer to extract temporal feature and the details of GRU are shown in Figure 6. As a variant of LSTM, GRU simplifies the structure while ensuring the performance of processing long-term dependence. The structure of LSTM contains three gates: input gate, forgetting gate, and output

gate, while GRU has only two gates, namely update gate and reset gate. The update gate is utilized to control the influence of the previous state on the current state. The larger the value of the update gate, the greater the impact. The reset gate is utilized to control the influence of the previous state on the candidate h'_t . Similarly, the larger the value of the reset gate, the greater the impact. The formulas of the reset gate and update gate are shown in Equation (6) and (7).

$$r_t = \sigma(w_r[x_t, h_{t-1}]), \quad (6)$$

$$z_t = \sigma(w_z[x_t, h_{t-1}]), \quad (7)$$

where w_r and w_z are the weight matrixes of reset gate and update gate. x_t is the input of GRU at time t and h_{t-1} is the hidden state of GRU at time $t-1$. σ denotes the *sigmoid*

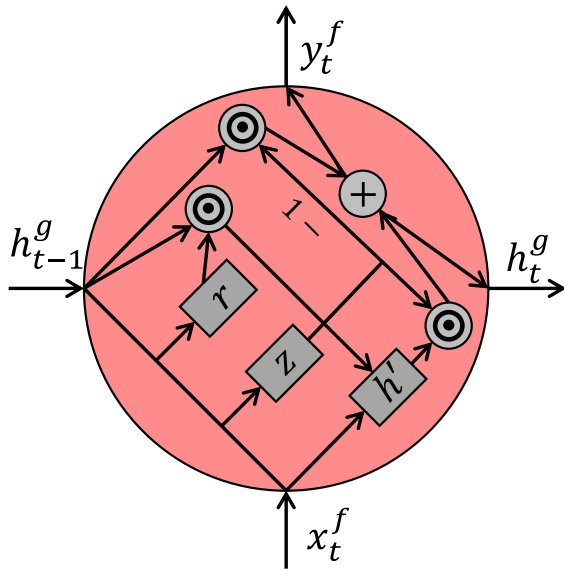


FIGURE 6. The details of gated recurrent unit.

function. h_t' is the candidate containing the input at time t and hidden state at time $t-1$ processed by reset gate, which is calculated as Equation (8).

$$h_t' = \tanh(w_{h'}[r_t \odot h_{t-1}, x_t]), \quad (8)$$

where $\tanh$ denotes the function and $\odot$ represents Hadamard product which is the multiplication of the corresponding elements in a matrix. The formula of hidden state at time t is shown in Equation (9).

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot h_t' \quad (9)$$

where h_t is the hidden state of GRU at time t which is ranging from 0 to 1 and the larger the value, the more information is memorized. The output of GRU at time t is shown in Equation (10).

$$y_t = \sigma(w_o h_t). \quad (10)$$

The dropout layer is added between GRU and the fully connection layer to prevent overfitting and improve model generalization ability. Finally, the output of GRU is obtained by fully connected calculating as shown in Equation (11).

$$y^o = \text{selu}(w^o x^o + b^o), \quad (11)$$

where x^o and y^o are the input and output of the fully connected layer. w^o is the weight matrix and b^o is the bias matrix.

C. LEARNING METHOD

Although the previous works provide some interpolation methods to monitor the air quality in areas without monitoring stations, they can only verify the estimated performance of existing monitoring stations. Since lacking of air quality data in the unsupervised areas, there are limited researches on how to verify the performance of interpolation methods.

Therefore, this paper proposes a method of training in turn with miss one station data, which uses the data of lacking of a monitoring station for each training and obtains the final estimation model by multiple training. For each round of training, it regards one known station as unknown station (i.e. target station) so that proposed method has the ability to estimate unknown station. This method not only achieves the purpose of verifying the performance of the interpolation method but also makes the estimation model have better generalization performance.

Assuming that there are n air pollution monitoring stations, the training data are obtained by the method of KIDW. The output data Y of TCGRU is obtained by interpolation with the data of n monitoring stations while the input data X is obtained by interpolation with the data of $n-1$ monitoring stations. The missing monitoring station data is played by n data stations in turn and n sets of input and output $\{< X_1, Y >, < X_2, Y >, \dots, < X_n, Y >\}$ are obtained. The loss function of GRU corresponds to the minimum square error, shown in Equation (12), which means that the greater the distance between the predicted value y^o and the actual value y , the greater the loss.

$$\begin{aligned} J(w, b) &= \frac{1}{m} \sum_{i=1}^m L(w, b; x_i, y_i) \\ &= \frac{1}{m} \sum_{i=1}^m [y_i^o - y_i]^2, \end{aligned} \quad (12)$$

where m denotes the number of data. The purpose of the training is to minimize the loss function, which is a function with w and b . Adaptive moment estimation (Adam) [23] is used to optimize the model until the parameters converge according to the n sets of data obtained.

IV. EXPERIMENTS

This section presents parameter setting and practical results in experiments.

A. DATA AND PARAMETER SETTING

This paper retrieved the daily air quality data from weather report website [24], which are collected by 12 real monitoring stations in Hubei Province in China from January 2018 to March 2020. This paper used 70% data as training data and 30% data as testing data. The data is processed by the max-min normalization method and compressed from 0 to 1 for a better training effect as shown in Equation (13).

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \quad (13)$$

where z is the data after normalization ranging from 0 to 1 and x is the actual data.

In order to verify the performance of the model in different spatial and temporal resolutions, five spatial resolutions and seven temporal resolutions were set up. The spatial resolutions are $(0.1^\circ \times 0.1^\circ)$, $(0.2^\circ \times 0.2^\circ)$, $(0.3^\circ \times 0.3^\circ)$, $(0.4^\circ \times 0.4^\circ)$ and $(0.5^\circ \times 0.5^\circ)$ respectively. where the left side of the

multiplier is longitude and the right side is latitude. The temporal resolution is ranging from 1 to 7 where the unit is a day. In addition, different values of k are selected to analyze the influence of data selecting on interpolation and the k value of is ranging from 2 to 11. The input and output of model vary according to the spatial-temporal resolution and different value of k . Supposed that there are n monitoring stations, $n - 1$ stations are selected to participate in interpolation and the other one is regard as the station of unknown station so that the input of network can be obtained. The output of network is obtained by interpolation with all station data. Repeat the above operation n times would obtain n sets of input and output of network. The input of the model is a three-dimensional matrix and each dimension represents the timestep, the length and width of the monitoring area respectively. The output of the model is a two-dimensional matrix and each dimension represents the length and width of the monitoring area respectively (i.e the estimation of all areas). The proposed model and the comparative model including ANN, CNN, time distributed neural network (TNN), TCNN, LSTM, BLSTM, 1-DCNNLSTM, 1-DCNNGRU are trained on Keras framework with Tensorflow backend. The number of neurons in all hidden layers is set to 100, the number of hidden layers is set to 3, the number of convolutional kernels is 8 and the activation function of each layer of the neural network is *selu*. In the process of model optimization, the learning rate, epoch, and batch size are set to 0.01, 100, and 16 respectively.

In the area of air quality modeling, the commonly used evaluation indicators are mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE). MAE measures the mean absolute error between the predicted value and actual value. RMSE measures the deviation between the predicted value and the actual value. Therefore, this paper adopts above metrics as the evaluation indicator of the model and their formulas are shown in Equation (14)-(16).

$$MAE = \frac{1}{m} \sum_{i=1}^m |o_i - p_i|, \quad (14)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (o_i - p_i)^2}, \quad (15)$$

$$MAPE = \frac{1}{m} \sum_{i=1}^m \frac{|o_i - p_i|}{o_i}, \quad (16)$$

where p_i donates predicted value of the i -th predicted data and o_i is the value of the i -th observed data.

B. RESULTS AND DISCUSSIONS

The traditional interpolation method estimates the PM2.5 concentration in the unsupervised areas by the known monitoring station data. It is well known that within a certain range, the correlation of PM2.5 concentration of the two stations decreases with the increase of distance. Therefore, it is

obviously not objective that all monitoring station data are used for interpolation calculation. Therefore, this paper compare the influence of different values of k on interpolation performance and the results are shown in Table 1. Line charts are utilized to represent the MAEs, RMSEs and MAPEs of CNN and NN to show the relationship between k values and interpolation performance more intuitively. Figure 7-9 showed the MAPEs of CNN and NN at different k values. The blue line, orange line, and red line represent the MAEs RMSEs and MAPEs of CNN, NN and average with the value ranging from 2 to 11 respectively. The experimental results showed that compared with selecting the appropriate data, using all the monitoring stations ($k = 11$) for interpolation calculation not only increases the computational complexity but also reduces the interpolation performance. Especially, when $k = 6$, it achieve best performance (i.e. MAE=8.56, RMSE=11.85 and MAPE=20.78%).

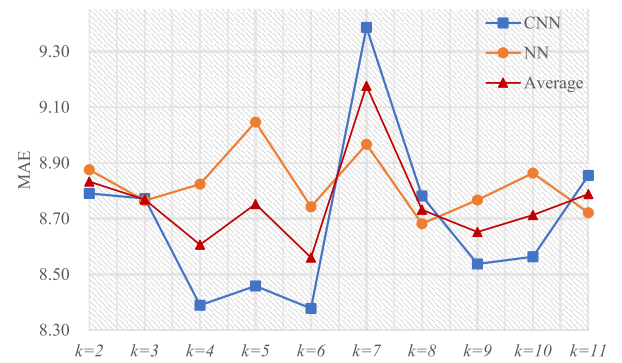


FIGURE 7. The MAEs of different k value on interpolation performance.

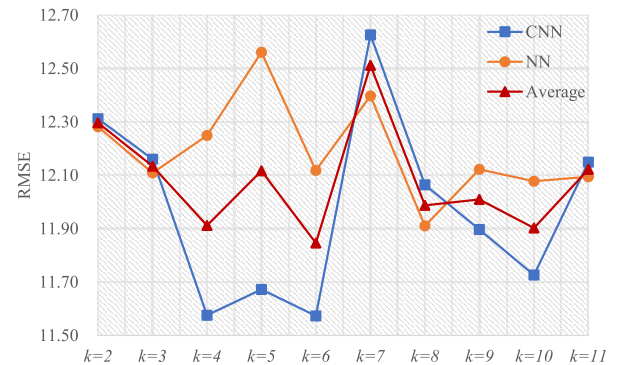
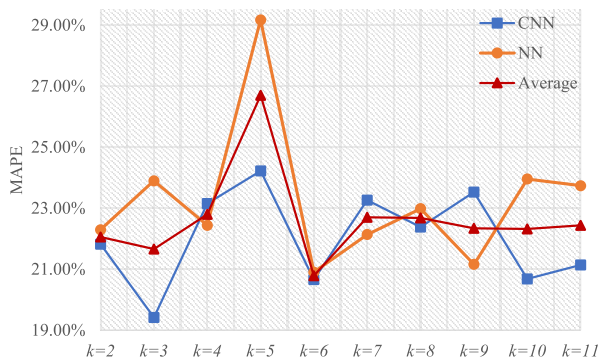
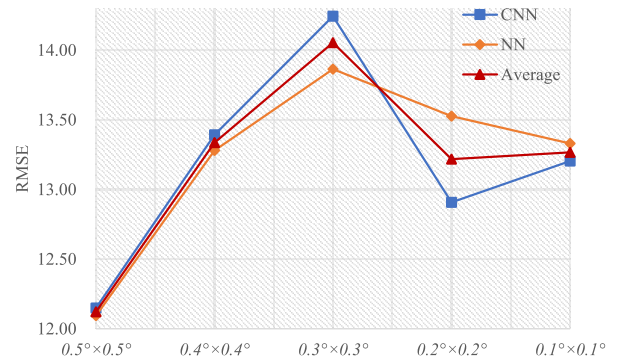
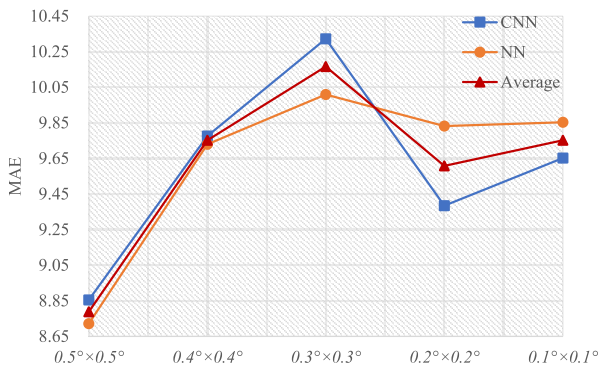
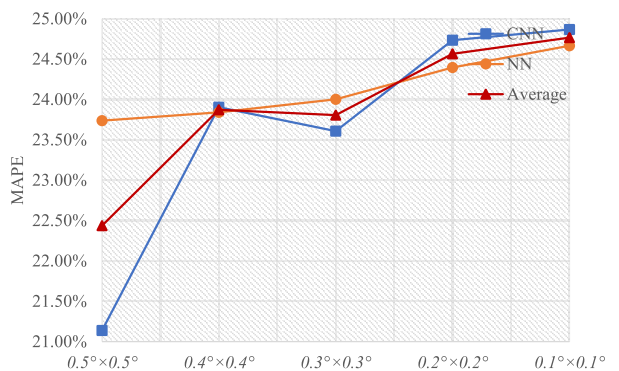


FIGURE 8. The RMSEs of different k value on interpolation performance.

In order to verify the influence of different spatial resolutions on interpolation performance, this paper tests the performance of CNN and NN with different spatial resolutions. Table 2 shows the MAEs, RMSEs and MAPEs of CNN and NN at different spatial resolutions. Line charts are used to represent the relationship between spatial resolutions and interpolation performance more intuitively. Figure 10-12 show the MAEs, RMSEs and MAPEs of CNN and NN at different spatial resolution. The blue line, orange line and

TABLE 1. The MAEs, RMSEs and MAPEs of different k values on interpolation performance.

Metrics	Model	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	$k=11$
MAE	CNN	8.79	8.77	8.39	8.46	8.38	9.39	8.78	8.54	8.56	8.85
	NN	8.88	8.76	8.82	9.05	8.74	8.97	8.68	8.77	8.86	8.72
	Average	8.83	8.77	8.61	8.75	8.56	9.18	8.73	8.65	8.71	8.79
RMSE	CNN	12.31	12.16	11.58	11.67	11.57	12.63	12.06	11.90	11.73	12.15
	NN	12.28	12.11	12.25	12.56	12.12	12.40	11.91	12.12	12.08	12.09
	Average	12.30	12.13	11.91	12.12	11.85	12.51	11.99	12.01	11.90	12.12
MAPE	CNN	21.82%	19.42%	23.15%	24.22%	20.66%	23.26%	22.38%	23.53%	20.68%	21.14%
	NN	22.29%	23.90%	22.44%	29.17%	20.91%	22.14%	22.98%	21.16%	23.95%	23.74%
	Average	22.05%	21.66%	22.80%	26.70%	20.78%	22.70%	22.68%	22.34%	22.32%	22.44%

**FIGURE 9.** The MAPEs of different k value on interpolation performance.**FIGURE 11.** The MAPEs of different spatial resolutions on interpolation performance.**FIGURE 10.** The MAE of different spatial resolutions on interpolation performance.**FIGURE 12.** The MAPEs of different spatial resolutions on interpolation performance.

red line represent the MAEs, RMSEs and MAPEs of CNN, NN and average with interpolation resolution ranging from $(0.1^\circ \times 0.1^\circ)$ to $(0.5^\circ \times 0.5^\circ)$ respectively. The experimental results show that the estimation performance of the model is improved with the increase of interpolation resolution under a certain interpolation resolution. When the spatial resolution is equal to $(0.5^\circ \times 0.5^\circ)$, it obtains best interpolation performance (i.e. MAE=8.79, RMSE=12.12 and MAPE=22.44%). In other words, if better interpolation performance is wanted, the interpolation density will be sacrificed. Conversely, if larger interpolation density is wanted, the interpolation performance will be sacrificed.

This section discusses the influence of different temporal resolutions on the model and the choice of a reasonable

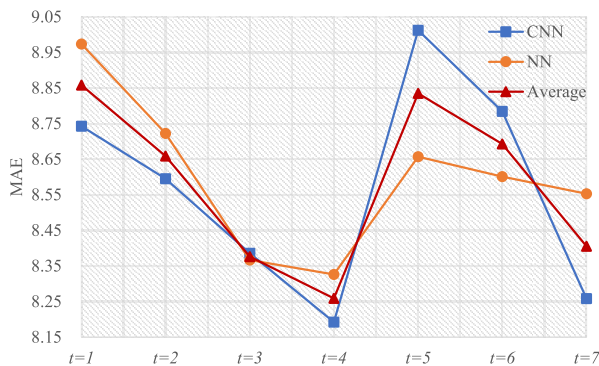
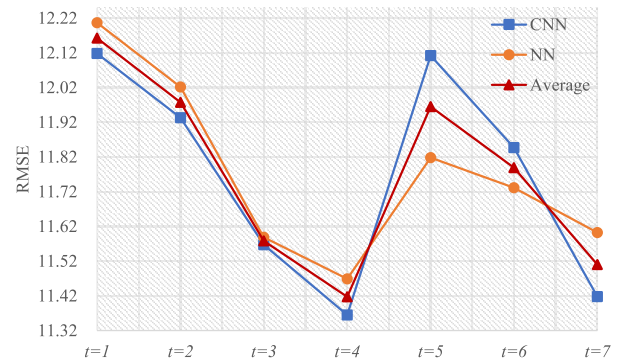
time window. Table 3 shows the performance of CNN and NN at different temporal resolutions. In order to show the relationship between temporal resolutions on model performance more intuitively, line charts is used to represent their MAEs, RMSEs and MAPEs. Figure 13-15 shows the MAEs, RMSEs and MAPEs of CNN and NN at different temporal resolutions. The blue line, orange line, and red line represent the MAEs, RMSEs and MAPEs of CNN, NN, and average with interpolation resolution ranging from 1 day to 7 days respectively. The experimental results show that adding historical data properly can improve the performance of the estimation model. The MAPEs of the model is best (i.e. MAPE=19.76%) when the time window

TABLE 2. The MAEs, RMSEs and MAPEs of different spatial resolutions on interpolation performance.

Metrics	Model	$(0.5^\circ \times 0.5^\circ)$	$(0.4^\circ \times 0.4^\circ)$	$(0.3^\circ \times 0.3^\circ)$	$(0.2^\circ \times 0.2^\circ)$	$(0.1^\circ \times 0.1^\circ)$
MAE	CNN	8.85	9.78	10.32	9.38	9.65
	NN	8.72	9.73	10.01	9.83	9.85
	Average	8.79	9.75	10.17	9.61	9.75
RMSE	CNN	12.15	13.39	14.24	12.91	13.20
	NN	12.09	13.28	13.86	13.53	13.33
	Average	12.12	13.34	14.05	13.22	13.27
MAPE	CNN	21.14%	23.90%	23.61%	24.74%	24.87%
	NN	23.74%	23.84%	24.00%	24.40%	24.67%
	Average	22.44%	23.87%	23.81%	24.57%	24.77%

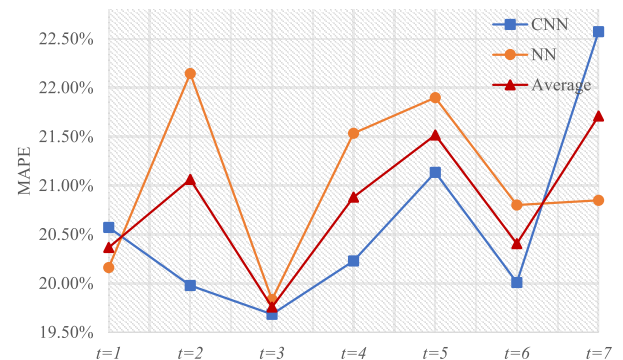
TABLE 3. The MAEs, RMSEs and MAPEs of different temporal resolutions on interpolation performance.

Metrics	Model	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$
MAE	CNN	8.74	8.59	8.39	8.19	9.01	8.78	8.26
	NN	8.97	8.72	8.37	8.33	8.66	8.60	8.55
	Average	8.86	8.66	8.38	8.26	8.83	8.69	8.41
RMSE	CNN	12.12	11.93	11.57	11.37	12.11	11.85	11.42
	NN	12.21	12.02	11.59	11.47	11.82	11.73	11.60
	Average	12.16	11.98	11.58	11.42	11.97	11.79	11.51
MAPE	CNN	20.57%	19.98%	19.68%	20.23%	21.13%	20.01%	22.57%
	NN	20.16%	22.14%	19.83%	21.53%	21.90%	20.80%	20.85%
	Average	20.37%	21.06%	19.76%	20.88%	21.52%	20.40%	21.71%

**FIGURE 13.** The MAE of different temporal resolutions on interpolation performance.**FIGURE 14.** The RMSEs of different temporal resolutions on interpolation performance.

is 3. The MAEs and RMSEs of the model are best (i.e. MAE=8.19, RMSE=11.42) when the time window is 4. Larger time window will make the model more complex and reduce the generalization ability while a smaller time window is not conducive to explore the time dependence of PM2.5.

Table 4 shows the MAEs, RMSEs and MAPEs of different models on the test set, including NN, CNN, TNN, TCNN, 1-DCNNLLSTM, 1-DCNNGRU, LSTM, BLSTM, and BGRU. It is noteworthy that the 1-DCNNLLSTM and 1-DCNNGRU models have larger MAPEs. As mentioned above, since the monitoring area is a rectangle in a two-dimensional plane, one-dimensional convolutional cannot extract local features, which leads to large errors. The performance of CNN is better than that of NN and that of TCNN is better than that of TNN, indicating that PM2.5 concentration has a local correlation.

**FIGURE 15.** The MAPEs of different temporal resolutions on interpolation performance.

The performance of TCNN is better than that of CNN and that of TNN is better than that of NN, which indicates

TABLE 4. The MAEs, RMSEs and MAPEs on the testing set of NN, CNN, TNN, TCNN, 1-DCNNLSTM, 1-DCNNGRU, LSTM, BLSTM and TCGRU.

	NN	CNN	TNN	TCNN	LSTM	BLSTM	1-DCNNLSTM	1-DCNNGRU	TCGRU
MAE	8.74	8.38	8.60	8.37	8.42	8.54	10.91	10.99	8.23
RMSE	12.12	11.57	11.82	11.51	11.71	11.77	14.67	14.68	11.27
MAPE	20.91%	20.66%	20.80%	19.83%	19.82%	19.67%	31.22%	31.61%	19.14%

TABLE 5. The MAEs, RMSEs, MAPEs of different stations of CNN with different k values (temporal resolution=1, spatial resolution= $(0.5^\circ \times 0.5^\circ)$).

k	Metrics	Site_1	Site_2	Site_3	Site_4	Site_5	Site_6	Site_7	Site_8	Site_9	Site_10	Site_11	Site_12	Average
2	MAE	6.65	16.61	3.85	16.96	4.62	4.55	8.09	6.58	7.85	12.90	4.81	12.01	8.79
	RMSE	9.35	24.17	4.87	23.94	5.93	5.88	11.17	8.89	10.33	19.68	6.90	16.63	12.31
	MAPE	17.67%	31.49%	11.59%	37.68%	14.70%	12.35%	16.54%	16.35%	22.52%	26.71%	12.65%	41.53%	21.82%
3	MAE	5.59	16.36	2.51	19.10	3.65	5.23	7.45	6.76	9.34	12.87	4.72	11.68	8.77
	RMSE	7.23	23.75	3.34	26.40	4.57	6.62	10.45	8.85	12.48	19.32	6.45	16.47	12.16
	MAPE	13.69%	28.56%	6.57%	38.87%	10.03%	13.72%	15.08%	16.83%	22.98%	26.02%	12.06%	28.62%	19.42%
4	MAE	4.58	15.58	2.56	18.31	3.35	4.73	7.87	6.50	7.82	13.06	4.10	12.19	8.39
	RMSE	6.06	22.17	3.35	25.83	4.36	5.98	10.84	8.14	10.35	19.33	5.74	16.75	11.58
	MAPE	12.18%	53.11%	7.45%	39.77%	10.02%	13.11%	21.87%	16.74%	22.71%	32.40%	11.50%	36.99%	23.15%
5	MAE	5.11	12.53	3.30	18.36	3.93	4.45	8.32	6.06	7.86	13.85	4.40	13.33	8.46
	RMSE	6.89	17.83	4.06	26.10	4.91	5.76	11.77	8.10	10.21	20.50	6.07	17.88	11.67
	MAPE	13.07%	27.18%	12.41%	41.21%	14.93%	12.20%	19.26%	16.64%	24.45%	31.13%	12.62%	65.59%	24.22%
6	MAE	4.87	13.22	3.18	17.15	3.85	5.79	8.21	6.80	7.80	13.20	4.40	12.05	8.38
	RMSE	6.53	18.78	3.99	24.23	4.85	7.10	11.64	8.81	10.26	20.13	6.09	16.47	11.57
	MAPE	13.87%	28.84%	5.95%	39.21%	9.74%	13.70%	17.15%	18.37%	21.85%	26.18%	12.57%	40.51%	20.66%
7	MAE	5.86	14.47	3.63	20.33	4.33	5.78	8.29	8.54	9.56	14.68	4.65	12.52	9.39
	RMSE	7.41	19.04	4.75	27.44	5.53	7.13	11.09	11.25	12.60	22.02	6.02	17.24	12.63
	MAPE	14.13%	28.08%	9.74%	38.53%	11.95%	14.98%	16.61%	19.17%	22.44%	28.30%	12.01%	63.19%	23.26%
8	MAE	5.58	14.17	2.69	17.67	3.60	5.05	8.91	6.89	8.91	15.12	4.30	12.48	8.78
	RMSE	7.58	19.32	3.34	24.25	4.45	6.37	12.60	9.37	11.61	22.70	5.96	17.21	12.06
	MAPE	13.80%	28.17%	8.96%	37.37%	11.95%	13.57%	19.19%	17.93%	22.58%	33.05%	11.62%	50.39%	22.38%
9	MAE	5.69	12.87	3.05	16.28	3.72	5.43	9.49	6.55	7.85	15.06	4.61	11.84	8.54
	RMSE	7.59	17.95	4.01	23.11	4.80	6.80	13.62	9.20	10.33	22.78	6.31	16.26	11.90
	MAPE	14.74%	26.78%	9.05%	37.08%	11.45%	14.29%	19.07%	16.95%	21.03%	30.37%	12.39%	69.15%	23.53%
10	MAE	5.09	11.49	3.68	18.27	4.30	5.01	9.17	6.72	8.38	14.36	4.64	11.63	8.56
	RMSE	6.91	15.55	4.81	24.73	5.64	6.41	13.07	8.60	10.85	21.71	6.27	16.17	11.73
	MAPE	12.98%	25.35%	10.39%	37.28%	12.84%	13.64%	18.31%	17.13%	22.13%	30.14%	11.93%	36.04%	20.68%
11	MAE	5.10	12.66	3.76	17.07	4.35	5.76	9.68	6.58	8.43	15.92	5.18	11.76	8.85
	RMSE	6.53	17.37	4.81	23.27	5.52	7.36	13.75	8.73	11.14	24.02	7.00	16.29	12.15
	MAPE	13.18%	26.67%	10.87%	37.17%	13.00%	14.77%	19.19%	16.78%	22.12%	32.03%	13.06%	34.79%	21.14%

TABLE 6. The MAEs, RMSEs, MAPEs of different stations of CNN with different spatial resolution (temporal resolution=1, $k = 11$).

d	Metrics	Site_1	Site_2	Site_3	Site_4	Site_5	Site_6	Site_7	Site_8	Site_9	Site_10	Site_11	Site_12
$(0.5^\circ \times 0.5^\circ)$	MAE	5.10	12.66	3.76	17.07	4.35	5.76	9.68	6.58	8.43	15.92	5.18	11.76
	RMSE	6.53	17.37	4.81	23.27	5.52	7.36	13.75	8.73	11.14	24.02	7.00	16.29
	MAPE	13.18%	26.67%	10.87%	37.17%	13.00%	14.77%	19.19%	16.78%	22.12%	32.03%	13.06%	34.79%
$(0.4^\circ \times 0.4^\circ)$	MAE	6.80	11.73	2.02	16.93	3.84	8.17	11.77	8.63	12.24	16.42	6.51	12.25
	RMSE	8.76	15.75	2.84	22.98	4.88	11.27	16.78	11.21	15.94	24.80	8.67	16.85
	MAPE	16.36%	25.36%	5.21%	36.29%	11.22%	18.24%	24.55%	19.06%	26.17%	33.95%	15.60%	54.84%
$(0.3^\circ \times 0.3^\circ)$	MAE	7.94	11.17	2.91	17.40	4.49	9.77	11.82	8.60	13.56	17.60	6.78	11.85
	RMSE	10.43	15.47	4.17	23.68	5.74	13.31	16.72	12.16	17.90	26.11	8.99	16.25
	MAPE	18.36%	24.38%	7.42%	36.85%	13.04%	22.68%	26.43%	19.33%	27.89%	37.21%	16.40%	33.32%
$(0.2^\circ \times 0.2^\circ)$	MAE	6.63	10.26	1.72	17.15	3.30	7.66	11.70	6.90	10.52	17.41	5.88	13.49
	RMSE	8.52	14.53	2.16	23.63	4.16	10.15	16.47	9.73	13.75	25.62	7.86	18.34
	MAPE	16.27%	24.23%	4.84%	37.28%	10.00%	18.48%	25.33%	17.63%	24.83%	38.63%	15.38%	63.93%
$(0.1^\circ \times 0.1^\circ)$	MAE	5.26	13.14	3.89	17.54	3.40	7.35	12.56	6.84	8.28	18.42	5.13	14.02
	RMSE	7.01	17.50	4.57	23.71	4.28	10.85	17.94	9.09	10.75	26.55	7.37	18.84
	MAPE	13.94%	26.67%	5.13%	37.02%	10.15%	11.85%	33.37%	17.23%	20.38%	39.32%	13.20%	70.14%

that PM2.5 concentration is time-dependent. LSTM and BLSTM have similar performance because the experiments

only adopted historical data and did not use future data to estimate the PM2.5 concentration in the unsupervised

TABLE 7. The MAEs, RMSEs, MAPEs of different stations of NN with different k values (temporal resolution=1, spatial resolution=($0.5^\circ \times 0.5^\circ$)).

k	Metrics	Site_1	Site_2	Site_3	Site_4	Site_5	Site_6	Site_7	Site_8	Site_9	Site_10	Site_11	Site_12
2	MAE	8.16	14.95	2.31	17.68	3.69	5.60	9.01	7.21	9.45	10.40	5.22	12.83
	RMSE	11.23	21.19	2.87	24.52	4.61	7.17	13.00	9.10	12.24	16.45	7.37	17.63
	MAPE	21.96%	28.25%	6.97%	36.30%	10.15%	14.04%	17.68%	18.20%	22.96%	20.93%	13.23%	56.83%
3	MAE	7.79	12.97	3.25	17.54	3.27	5.79	10.56	7.30	7.44	10.65	6.38	12.23
	RMSE	10.38	18.56	3.53	25.11	4.32	7.09	14.07	10.47	9.83	16.23	8.23	17.47
	MAPE	22.77%	27.21%	6.22%	37.38%	11.32%	14.41%	35.35%	18.68%	22.26%	20.74%	21.22%	49.23%
4	MAE	5.81	12.89	3.37	17.71	4.01	4.75	8.09	7.31	9.45	14.31	5.41	12.79
	RMSE	8.04	17.86	4.32	24.77	5.31	6.30	11.50	10.58	12.37	21.29	7.45	17.20
	MAPE	14.83%	25.90%	9.39%	39.11%	12.24%	13.05%	16.42%	18.89%	23.66%	31.63%	14.20%	49.94%
5	MAE	6.96	12.58	2.29	18.30	3.40	4.74	9.41	6.62	8.69	16.57	5.81	13.20
	RMSE	9.67	18.01	3.04	25.92	4.35	6.22	13.69	8.97	11.41	23.59	8.11	17.74
	MAPE	17.93%	39.02%	6.18%	41.79%	9.84%	14.22%	19.56%	16.40%	24.47%	42.47%	14.86%	103.36%
6	MAE	6.00	14.73	1.47	19.06	3.06	4.96	8.45	6.56	8.80	14.50	5.04	12.27
	RMSE	8.15	20.69	1.85	26.12	3.92	6.42	12.18	8.39	11.45	22.29	7.05	16.90
	MAPE	15.39%	27.76%	4.13%	38.63%	8.98%	13.30%	17.48%	17.01%	21.61%	29.55%	12.87%	44.16%
7	MAE	6.25	11.73	2.44	18.61	3.72	5.50	10.11	6.74	9.95	14.82	4.92	12.81
	RMSE	8.23	16.03	3.11	25.71	4.69	7.09	14.32	9.49	12.70	22.57	6.87	17.94
	MAPE	16.55%	24.69%	7.00%	39.41%	11.59%	13.93%	20.97%	17.37%	23.08%	30.61%	12.73%	47.72%
8	MAE	5.43	12.50	1.57	18.06	3.38	5.54	9.05	7.21	9.57	14.45	4.81	12.62
	RMSE	7.18	17.00	1.99	24.65	4.24	7.20	12.76	9.22	12.37	21.87	6.64	17.82
	MAPE	14.03%	25.62%	4.53%	37.70%	9.66%	13.79%	18.82%	17.58%	22.94%	30.51%	12.31%	32.32%
9	MAE	5.64	12.35	2.28	17.19	3.67	6.12	9.20	6.77	9.63	15.61	4.89	11.85
	RMSE	7.36	17.14	2.74	24.00	4.60	7.98	13.08	9.37	12.49	23.55	6.65	16.49
	MAPE	15.24%	25.38%	7.07%	38.49%	12.16%	14.90%	18.98%	17.02%	23.39%	33.28%	12.72%	35.25%
10	MAE	5.40	13.21	1.35	17.73	3.53	5.81	9.34	7.69	11.02	14.39	4.80	12.11
	RMSE	7.01	17.44	1.94	24.26	4.40	7.55	13.28	9.99	13.99	21.90	6.51	16.68
	MAPE	13.96%	27.02%	3.51%	37.38%	10.33%	14.18%	20.51%	18.24%	24.30%	29.83%	12.33%	75.87%
11	MAE	5.68	11.93	1.61	16.52	3.44	5.56	9.71	6.67	9.74	16.63	4.97	12.20
	RMSE	7.56	16.45	2.05	22.99	4.26	7.42	13.84	9.34	12.80	24.80	6.92	16.69
	MAPE	14.71%	25.24%	4.49%	37.13%	10.58%	15.83%	20.63%	17.24%	24.88%	36.97%	13.49%	63.66%

TABLE 8. The MAEs, RMSEs, MAPEs of different stations of NN with different spatial resolution (temporal resolution=1, $k = 11$).

d	Metrics	Site_1	Site_2	Site_3	Site_4	Site_5	Site_6	Site_7	Site_8	Site_9	Site_10	Site_11	Site_12
$(0.5^\circ \times 0.5^\circ)$	MAE	5.68	11.93	1.61	16.52	3.44	5.56	9.71	6.67	9.74	16.63	4.97	12.20
	RMSE	7.56	16.45	2.05	22.99	4.26	7.42	13.84	9.34	12.80	24.80	6.92	16.69
	MAPE	14.71%	25.24%	4.49%	37.13%	10.58%	15.83%	20.63%	17.24%	24.88%	36.97%	13.49%	63.66%
$(0.4^\circ \times 0.4^\circ)$	MAE	6.48	11.03	1.48	17.53	3.54	7.49	11.37	10.20	13.16	15.49	6.33	12.64
	RMSE	8.26	14.36	1.96	23.67	4.38	10.28	16.76	13.06	16.93	23.59	8.43	17.67
	MAPE	15.45%	25.29%	4.01%	37.28%	9.88%	16.99%	24.26%	21.24%	26.98%	31.31%	14.69%	58.71%
$(0.3^\circ \times 0.3^\circ)$	MAE	8.03	10.77	2.47	17.59	3.84	8.27	11.11	9.43	13.09	15.96	7.96	11.59
	RMSE	10.55	14.64	3.24	23.48	4.88	11.53	16.31	13.50	17.29	24.58	10.10	16.28
	MAPE	17.68%	24.34%	6.81%	37.05%	10.24%	18.07%	22.30%	20.33%	27.16%	32.62%	17.86%	53.56%
$(0.2^\circ \times 0.2^\circ)$	MAE	7.22	10.35	2.82	18.31	4.27	8.61	12.18	7.60	11.77	16.92	5.64	12.30
	RMSE	9.41	14.68	3.51	24.69	5.37	11.40	17.14	10.45	15.28	25.52	7.69	17.18
	MAPE	16.38%	24.13%	6.88%	37.59%	10.87%	18.49%	26.40%	18.06%	25.12%	34.73%	14.30%	59.82%
$(0.1^\circ \times 0.1^\circ)$	MAE	7.50	13.16	3.38	18.72	4.13	6.45	12.55	7.22	10.23	17.08	5.56	12.25
	RMSE	9.10	16.51	4.37	25.48	5.22	8.16	17.51	9.46	13.20	26.88	7.04	17.04
	MAPE	17.72%	28.50%	7.61%	38.27%	11.05%	15.15%	27.84%	17.90%	23.81%	35.57%	14.01%	58.56%

areas considering the timeliness of PM2.5 monitoring. Compared with other models, TCGRU has the best performance (i.e. MAE=8.23, RMSE=11.27 and MAPE=19.14%), which shows that the model can effectively extract the temporal and spatial characteristics of PM2.5 and achieve the purpose of monitoring PM2.5 concentration in the areas without monitoring stations.

V. CONCLUSION AND FUTURE WORK

This paper focuses on the interpolation and estimation of air quality which can provide a reference for air quality in unsupervised areas and has a significant impact on the government and the public. In order to improve the efficiency of interpolation, this paper proposes an interpolation method based on KIDW, which can effectively select the data used for interpolation. In addition, this paper puts forward a training

TABLE 9. The MAEs, RMSEs, MAPEs of different stations of CNN with different temporal resolutions ($k = 1$, spatial resolution= $(0.5^\circ \times 0.5^\circ)$).

t	Metrics	Site_1	Site_2	Site_3	Site_4	Site_5	Site_6	Site_7	Site_8	Site_9	Site_10	Site_11	Site_12
2	MAE	4.87	13.22	3.18	17.15	3.85	5.79	8.21	6.80	7.80	13.20	4.40	12.05
	RMSE	6.53	18.78	3.99	24.23	4.85	7.10	11.64	8.81	10.26	20.13	6.09	16.47
	MAPE	13.87%	28.84%	5.95%	39.21%	9.74%	13.70%	17.15%	18.37%	21.85%	26.18%	12.57%	40.51%
3	MAE	6.70	12.89	1.54	18.77	3.35	5.60	8.55	7.17	8.11	15.45	5.05	11.49
	RMSE	9.13	17.72	1.87	26.16	4.22	7.44	12.04	9.46	10.61	22.73	6.99	15.90
	MAPE	18.78%	25.69%	4.57%	40.02%	11.00%	15.67%	19.42%	16.89%	22.59%	36.02%	13.51%	41.57%
4	MAE	5.26	11.06	2.12	18.11	3.65	5.72	8.72	7.29	9.38	14.83	4.72	9.56
	RMSE	7.11	14.91	2.91	25.28	4.61	7.47	12.31	9.68	12.17	22.65	6.54	13.42
	MAPE	14.42%	23.95%	5.54%	36.30%	10.22%	14.61%	18.20%	16.52%	21.80%	30.91%	12.11%	33.41%
5	MAE	6.58	11.02	1.95	17.09	3.30	5.17	9.07	6.70	7.35	16.12	5.23	10.34
	RMSE	9.09	14.87	2.43	24.00	4.15	6.80	12.59	8.76	9.57	23.75	7.17	14.45
	MAPE	17.29%	23.92%	5.52%	37.96%	9.33%	14.41%	19.70%	16.53%	20.97%	36.07%	13.47%	43.24%
6	MAE	5.19	14.27	1.78	19.64	3.26	5.49	8.08	8.41	8.77	13.54	4.68	10.78
	RMSE	7.21	18.63	2.22	26.93	4.07	7.35	11.33	10.44	11.36	21.01	6.41	14.86
	MAPE	13.31%	28.10%	5.12%	38.26%	10.15%	14.36%	16.78%	18.70%	20.99%	27.97%	11.67%	57.38%
7	MAE	5.34	12.48	2.08	18.89	3.38	5.58	8.98	7.19	8.60	15.40	4.57	10.73
	RMSE	7.12	16.77	2.55	25.97	4.35	7.38	12.52	9.29	11.24	22.28	6.21	15.09
	MAPE	13.69%	25.54%	5.81%	37.80%	10.31%	13.72%	19.20%	16.63%	20.15%	34.89%	11.58%	40.27%
8	MAE	5.55	13.66	1.96	17.71	3.40	5.37	9.21	7.63	8.18	13.66	5.59	10.69
	RMSE	7.45	17.28	2.46	24.33	4.30	7.13	12.75	9.59	10.71	20.94	7.13	15.15
	MAPE	14.18%	28.55%	6.02%	37.24%	9.38%	13.90%	19.26%	17.67%	19.98%	26.45%	14.06%	43.49%

TABLE 10. The MAEs, RMSEs, MAPEs of different stations of NN with different temporal resolutions ($k = 1$, spatial resolution= $(0.5^\circ \times 0.5^\circ)$).

t	Metrics	Site_1	Site_2	Site_3	Site_4	Site_5	Site_6	Site_7	Site_8	Site_9	Site_10	Site_11	Site_12
2	MAE	6.00	14.73	1.47	19.06	3.06	4.96	8.45	6.56	8.80	14.50	5.04	12.27
	RMSE	8.15	20.69	1.85	26.12	3.92	6.42	12.18	8.39	11.45	22.29	7.05	16.90
	MAPE	15.39%	27.76%	4.13%	38.63%	8.98%	13.30%	17.48%	17.01%	21.61%	29.55%	12.87%	44.16%
3	MAE	6.37	14.55	1.37	17.84	3.27	4.99	9.08	6.79	8.77	14.49	5.57	10.04
	RMSE	8.77	20.37	1.85	24.69	4.15	6.56	12.65	9.04	11.32	21.89	7.75	14.15
	MAPE	16.25%	27.24%	4.01%	36.49%	9.07%	12.99%	19.17%	17.13%	21.16%	29.93%	14.38%	31.92%
4	MAE	5.53	13.77	1.97	17.90	3.25	5.20	8.41	6.94	9.14	13.30	5.27	9.96
	RMSE	7.52	18.88	2.58	24.51	4.16	6.83	11.81	9.31	11.88	20.35	7.35	13.63
	MAPE	13.98%	27.01%	5.32%	36.80%	9.92%	13.79%	17.92%	16.95%	21.70%	27.77%	13.29%	31.72%
5	MAE	5.22	13.43	1.96	17.07	3.42	4.93	8.29	7.12	8.12	13.62	5.25	9.88
	RMSE	6.96	18.59	2.65	23.57	4.44	6.52	11.73	9.25	10.61	21.17	7.25	13.63
	MAPE	13.33%	26.09%	5.30%	36.39%	9.70%	13.14%	16.74%	16.94%	20.01%	26.89%	13.37%	44.84%
6	MAE	5.42	16.82	1.90	19.11	3.38	5.36	8.15	7.63	10.00	14.53	5.20	10.64
	RMSE	7.42	21.56	2.34	25.96	4.20	6.94	11.46	9.37	12.44	22.57	7.11	13.97
	MAPE	13.75%	31.45%	5.55%	37.80%	9.37%	13.58%	17.04%	18.44%	22.59%	29.19%	12.98%	41.87%
7	MAE	5.36	14.22	1.84	19.91	3.31	5.33	9.40	7.46	9.38	13.08	5.38	10.75
	RMSE	7.15	18.62	2.34	26.66	4.17	6.93	13.26	9.29	11.83	20.44	7.15	14.32
	MAPE	13.58%	28.23%	5.85%	38.04%	9.25%	13.38%	18.13%	17.88%	22.24%	24.88%	13.69%	34.94%
8	MAE	6.02	11.59	1.66	17.59	3.13	4.53	9.40	6.39	8.30	14.37	5.90	10.20
	RMSE	8.26	15.82	2.14	24.11	4.07	6.04	13.14	8.68	10.86	21.67	8.29	13.91
	MAPE	15.62%	24.77%	4.66%	36.99%	9.66%	12.34%	21.85%	15.90%	20.19%	30.39%	14.93%	63.56%

method with missing one in turn to verify that the interpolation method is conducive to the air quality monitoring in unsupervised areas. Finally, this paper proposes a deep learning network namely TCGRU to estimate the air quality in unsupervised areas. A large number of experiments show that the proposed model is better than the comparative models (i.e. MAE=8.23, RMSE=11.27 and MAPE=19.14%). Through the analysis of the experimental results, this study finds that the interpolation performance is better than that of sparse monitoring stations in the area with dense

monitoring stations, which indicates that the proposed method needs a certain number of monitoring stations. Moreover, the estimation performance of the model is improved with the increase of interpolation resolutions under a certain interpolation resolution. In other words, if better interpolation performance is wanted, the interpolation density will be sacrificed. Conversely, if larger interpolation density is wanted, the interpolation performance will be sacrificed.

In general, the unsupervised air quality monitoring method proposed in this paper can effectively monitor the air quality

TABLE 11. The MAEs, RMSEs, MAPEs of different models (temporal resolution=3, $k = 6$, spatial resolution=($0.5^\circ \times 0.5^\circ$)).

Metrics	Model	Site_1	Site_2	Site_3	Site_4	Site_5	Site_6	Site_7	Site_8	Site_9	Site_10	Site_11	Site_12
MAE	NN	6.00	14.73	1.47	19.06	3.06	4.96	8.45	6.56	8.80	14.50	5.04	12.27
	CNN	4.87	13.22	3.18	17.15	3.85	5.79	8.21	6.80	7.80	13.20	4.40	12.05
	TNN	5.34	12.48	2.08	18.89	3.38	5.58	8.98	7.19	8.60	15.40	4.57	10.73
	TCNN	5.26	11.06	2.12	18.11	3.65	5.72	8.72	7.29	9.38	14.83	4.72	9.56
	LSTM	4.78	12.72	3.37	18.24	3.95	4.42	10.35	6.28	7.13	14.83	5.73	9.27
	BLSTM	5.70	12.63	2.47	19.47	3.46	4.67	9.04	6.54	8.24	13.18	5.19	11.83
	1-DCNNLSTM	7.99	9.29	9.48	13.43	9.44	6.61	12.56	10.38	8.01	20.45	9.18	14.12
	1-DCNNGRU	7.47	9.03	8.17	14.06	8.17	6.16	13.39	10.87	8.00	22.18	9.59	14.76
RMSE	TCGRU	5.70	13.11	1.94	17.58	3.42	4.66	8.67	6.84	7.78	13.33	5.47	10.29
	NN	8.15	20.69	1.85	26.12	3.92	6.42	12.18	8.39	11.45	22.29	7.05	16.90
	CNN	6.53	18.78	3.99	24.23	4.85	7.10	11.64	8.81	10.26	20.13	6.09	16.47
	TNN	7.12	16.77	2.55	25.97	4.35	7.38	12.52	9.29	11.24	22.28	7.21	15.09
	TCNN	7.11	14.91	2.91	25.28	4.61	7.47	12.31	9.68	11.17	22.65	6.54	13.42
	LSTM	6.33	18.06	4.14	25.56	5.06	5.83	14.93	8.85	9.71	21.07	7.82	13.21
	BLSTM	7.61	18.68	3.20	27.63	4.44	6.01	12.67	8.87	11.58	19.62	7.24	13.71
	1-DCNNLSTM	10.76	12.95	11.21	20.12	11.27	8.85	17.86	13.85	10.67	26.23	11.91	20.30
MAPE	1-DCNNGRU	10.02	12.29	9.81	20.53	9.91	8.05	19.29	14.62	10.57	28.04	12.53	20.47
	TCGRU	7.68	17.36	2.63	24.02	4.31	6.13	12.23	9.06	9.98	20.41	7.55	13.82
	NN	15.39%	27.76%	4.13%	38.63%	8.98%	13.30%	17.48%	17.01%	21.61%	29.55%	12.87%	44.16%
	CNN	13.87%	28.84%	5.95%	39.21%	9.74%	13.70%	17.15%	18.37%	21.85%	26.18%	12.57%	40.51%
	TNN	13.69%	25.54%	5.81%	37.80%	10.31%	13.72%	19.20%	16.63%	20.15%	34.89%	11.58%	40.27%
	TCNN	14.42%	23.95%	5.54%	36.30%	10.22%	14.61%	18.20%	16.52%	21.80%	30.91%	12.11%	33.41%
	LSTM	13.84%	25.43%	7.82%	39.38%	13.38%	12.20%	20.98%	16.35%	20.62%	28.16%	14.63%	24.99%
	BLSTM	14.94%	25.34%	7.21%	40.03%	10.65%	12.61%	18.03%	18.07%	20.47%	27.91%	13.17%	27.59%
MAPE	1-DCNNLSTM	22.10%	29.83%	32.64%	35.93%	33.03%	18.59%	24.64%	29.67%	29.48%	58.94%	25.28%	34.52%
	1-DCNNGRU	20.10%	30.96%	25.53%	35.80%	26.03%	17.34%	26.88%	31.65%	29.47%	68.44%	26.52%	40.58%
	TCGRU	14.06%	26.61%	5.21%	38.81%	10.04%	12.28%	17.78%	17.16%	19.93%	27.08%	14.19%	26.49%

in areas without monitoring stations. This method is conducive to government decision-making, people travel and saves the cost of monitoring stations. However, the study area is assumed to be a two-dimensional plane, ignoring the influence of mountain, river, and other geographical factors. However, there are some limitations of proposed method including lacking of consideration of meteorological data, altitude and human activities. It is known that above factors are significant for air quality and thus, the future work should pay more attentions to these factors. Furthermore, the human behavior can be considered for improving the accuracy of air quality prediction [25].

APPENDIX

Considering the spatial constraints, the MAEs, RMSEs and MAPEs of different stations of CNN, NN with different k values, different temporary resolutions, different spatial resolutions and different models are shown in Table 5 to Table 11.

REFERENCES

- [1] Z. Qi, T. Wang, G. Song, W. Hu, X. Li, and Z. Zhang, "Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 12, pp. 2285–2297, Dec. 2018.
- [2] S. Heft-Neal, J. Burney, E. Bendavid, and M. Burke, "Robust relationship between air quality and infant mortality in Africa," *Nature*, vol. 559, no. 7713, pp. 254–258, Jul. 2018.
- [3] F. Launay, "7 million deaths annually linked to air pollution," *Central Eur. J. Public Health*, vol. 22, no. 1, pp. 53–59, 2014.
- [4] Y. Rybarczyk and R. Zalakeviciute, "Machine learning approaches for outdoor air quality modelling: A systematic review," *Appl. Sci.*, vol. 8, no. 12, p. 2570, Dec. 2018.
- [5] E. Pekel, "Estimation of soil moisture using decision tree regression," *Theor. Appl. Climatol.*, vol. 139, nos. 3–4, pp. 1111–1119, Feb. 2020.
- [6] S. Ding, N. Zhang, X. Zhang, and F. Wu, "Twin support vector machine: Theory, algorithm and applications," *Neural Comput. Appl.*, vol. 28, no. 11, pp. 3119–3130, Nov. 2017.
- [7] C.-J. Huang and P.-H. Kuo, "A deep CNN-LSTM model for particulate matter ($PM_{2.5}$) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, Jul. 2018.
- [8] J. Ma, Y. Ding, V. J. L. Gan, C. Lin, and Z. Wan, "Spatiotemporal prediction of $PM_{2.5}$ concentrations at different time granularities using IDW-BLSTM," *IEEE Access*, vol. 7, pp. 107897–107907, 2019.
- [9] H. El Alaoui El Abdallaoui, A. El Fazziki, J. Ouarzazi, F. Z. Ennaji, and M. Sadgal, "A crowdsensing-based framework for urban air quality decision support," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 27, no. 6, pp. 4298–4313, Nov. 2019.
- [10] M. Krishan, S. Jha, J. Das, A. Singh, M. K. Goyal, and C. Sekar, "Air quality modelling using long short-term memory (LSTM) over NCT-delhi, india," *Air Qual., Atmos. Health*, vol. 12, no. 8, pp. 899–908, Aug. 2019.
- [11] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.
- [12] M. Lan, Y. Zhang, L. Zhang, and B. Du, "Global context based automatic road segmentation via dilated convolutional neural network," *Inf. Sci.*, vol. 535, pp. 156–171, Oct. 2020.
- [13] Q. Tao, F. Liu, Y. Li, and D. Sidorov, "Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU," *IEEE Access*, vol. 7, pp. 76690–76698, 2019.
- [14] J. Zhao, F. Deng, Y. Cai, and J. Chen, "Long short-term memory—Fully connected (LSTM-FC) neural network for $PM_{2.5}$ concentration prediction," *Chemosphere*, vol. 220, pp. 486–492, Apr. 2019.
- [15] M. Jerrett, R. T. Burnett, B. S. Beckerman, M. C. Turner, D. Krewski, G. Thurston, R. V. Martin, A. van Donkelaar, E. Hughes, Y. Shi, and S. M. Gapstur, "Spatial analysis of air pollution and mortality in California," *Amer. J. Respiratory Crit. care Med.*, vol. 188, no. 5, pp. 593–599, 2013.
- [16] B. S. Beckerman, M. Jerrett, R. V. Martin, A. van Donkelaar, Z. Ross, and R. T. Burnett, "Application of the deletion/substitution/addition algorithm to selecting land use regression models for interpolating air pollution measurements in California," *Atmos. Environ.*, vol. 77, pp. 172–177, Oct. 2013.

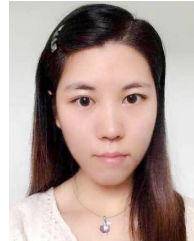
- [17] K. Huang, J. Bi, X. Meng, G. Geng, A. Lyapunov, K. J. Lane, D. Gu, P. L. Kinney, and Y. Liu, "Estimating daily PM_{2.5} concentrations in New York City at the neighborhood-scale: Implications for integrating non-regulatory measurements," *Sci. Total Environ.*, vol. 697, Dec. 2019, Art. no. 134094.
- [18] W. Tong, L. Li, X. Zhou, A. Hamilton, and K. Zhang, "Deep learning PM_{2.5} concentrations with bidirectional LSTM RNN," *Air Qual., Atmos. Health*, vol. 12, no. 4, pp. 411–423, Apr. 2019.
- [19] Y. Zhang, J. Wu, J. Wang, and C. Xing, "A transformation-based framework for KNN set similarity search," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 3, pp. 409–423, Mar. 2020.
- [20] G. Y. Lu and D. W. Wong, "An adaptive inverse-distance weighting spatial interpolation technique," *Comput. Geosci.*, vol. 34, no. 9, pp. 1044–1055, Sep. 2008.
- [21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [22] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: <http://arxiv.org/abs/1406.1078>
- [23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [24] *Weather Report Website of China*. Accessed: Aug. 28, 2020. [Online]. Available: <http://www.tianqihoubao.com/aqi/>
- [25] H. Chen, H. Jin, and S. Wu, "Minimizing inter-server communications by exploiting self-similarity in online social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 4, pp. 1116–1130, Apr. 2016.



CANYANG GUO is currently pursuing the master's degree with the College of Mathematics and Computer Sciences, Fuzhou University, Fuzhou, China. His research interests include machine learning and big data.



GENGGENG LIU (Member, IEEE) received the B.S. degree in computer science and the Ph.D. degree in applied mathematics from Fuzhou University, Fuzhou, China, in 2009 and 2015, respectively. He is currently an Associate Professor with the College of Mathematics and Computer Science, Fuzhou University. His research interests include computational intelligence and its application.



LINGJUAN LYU (Member, IEEE) received the Ph.D. degree from the University of Melbourne, Australia. She is currently a Research Fellow with the Department of Computer Science, National University of Singapore, Singapore. Her work was supported by an IBM PhD Fellowship. Her current research interests include machine learning, privacy, fairness, and edge intelligence.



CHI-HUA CHEN (Senior Member, IEEE) received the Ph.D. degree in information management from National Chiao Tung University, in 2013. He is currently a Full Professor with the College of Mathematics and Computer Science, Fuzhou University. He has published over 300 journal articles, conference papers, and patents. His contributions were published in the *IEEE INTERNET OF THINGS JOURNAL*, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *IEICE Transactions on Information and Systems*, the *Journal of Enterprise Information Management*, *Physica A: Statistical Mechanics and its Applications*, *Transactions on Emerging Telecommunications Technologies*, *Wireless Communications and Mobile Computing*, *WWW'20*, *SIGIR 2020*, and so on. Some of his publications have been recognised as highly cited papers on Web of Science using data from Essential Science Indicators. His research interests include the Internet of Things and machine learning. He also serves as a Chair for several international conferences, such as the IEEE APNOMS 2020, the IEEE ICC 2020, the IEEE BIBM 2020 Workshop, the IEEE TrustCom 2020 Workshop, the IEEE BigData 2020 Workshop, the IEEE ICCE-TW 2020, the IEEE IS3C 2020, and so on. He serves as an Editor for several international journals, such as *IEEE ACCESS*, the *Journal of Database Management*, the *Journal of Applied Statistics*, *IEICE Transactions*, *PLOS ONE*, the *International Journal of Ad Hoc and Ubiquitous Computing*, the *Journal of Medical Imaging and Health Informatics*, *Intelligent Automation & Soft Computing*, *Wireless Communications and Mobile Computing*, *Mobile Information Systems*, *EURASIP Journal on Wireless Communications and Networking*, the *International Journal of Distributed Sensor Networks*, *ISPRS International Journal of Geo-Information*, *Discrete Dynamics in Nature and Society*, *Mathematical Problems in Engineering*, and so on.

...