



Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering[☆]

Rui Yan^{a,b}, Jiaqiang Liao^{a,c}, Jie Yang^{a,d}, Wei Sun^{a,e,f,*}, Mingyue Nong^a, Feipeng Li^a

^a School of Geography and Planning, Sun Yat-Sen University, Guangzhou, Guangdong 510275, China

^b Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

^c Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

^d State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

^e Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China

^f Institute for Energy, Environment and Sustainable Communities, University of Regina, Regina, Saskatchewan S4S 0A2, Canada



ARTICLE INFO

Keywords:

LSTM
CNN
Forecasting
AQI
Spatiotemporal clustering

ABSTRACT

Effective air quality forecasting models are helpful for timely prevention and control of air pollution. However, the spatiotemporal distribution characteristics of air quality have not been fully considered in previous model development. This study attempts to establish a multi-time, multi-site forecasting model of Beijing's air quality by using deep learning network models based on spatiotemporal clustering and to compare them with a back-propagation neural network (BPNN). For the overall forecasting, the performances in next-hour forecasting were ranked in ascending order of the BPNN, the convolutional neural network (CNN), the long short-term memory (LSTM) model, and the CNN-LSTM, with the LSTM as the optimal model in the multiple-hour forecasting. The performance of the seasonal forecasting was not significantly improved compared to the overall forecasting. For the spatial clustering-based forecasting, cluster 2 forecasting generally outperforms cluster 1 and the overall forecasting. Overall, either the seasonal or the spatial clustering-based forecasting is more suitable for the improvement of the forecasting in a certain season or cluster. In terms of model type, both the CNN-LSTM and the LSTM generally have better performance than the CNN and the BPNN.

1. Introduction

With urbanization and industrialization, air pollution in many countries and cities is increasingly serious. Previous studies have shown that long-term inhalation of air pollutants causes adverse effects on human health. Various air pollutants stimulate respiratory mucosa and induce respiratory diseases (Brunekreef & Holgate, 2002). Long-term exposure to atmospheric particles may cause impaired vascular endothelial functions and a series of cardiovascular diseases (N. Bai, Khazaei, van Eeden, & Laher, 2007). Atmospheric fine particles also affect the nervous system, increasing the risk of neurological diseases, such as Alzheimer's and Parkinson's diseases (Calderón-Garcidueñas, et al., 2008). It is estimated that the death toll due to outdoor and indoor air

pollution in China is as high as 2.5 million people each year (Kulmala, 2015). Effective air quality management tools are indispensable for reducing the adverse effects of air pollution.

Forecasting models are such an air quality management tool, classified into numerical and statistical categories. Numerical models are based on atmospheric physical and chemical processes and use meteorological principles and mathematical methods to simulate the air quality, on a large scale, in both horizontal and vertical dimensions (An, et al., 2001). Since the middle of the 20th century, numerical models have been developed to a third generation, including Community Multiscale Air Quality Modeling System (CMAQ) (Streets, et al., 2007), Weather Research and Forecasting Model with Chemistry (WRF-Chem) (Tie, et al., 2007), and Nested Air Quality Prediction Modeling System

[☆] This article has been certified as Reproducible by the CodeOcean: <https://codeocean.com/>

* Corresponding author at: School of Geography and Planning, Sun Yat-Sen University, Guangzhou, Guangdong 510275, China.

E-mail addresses: yanr5@mail2.sysu.edu.cn (R. Yan), liaojq23@mail2.sysu.edu.cn (J. Liao), yangj289@mail2.sysu.edu.cn (J. Yang), sunwei29@mail.sysu.edu.cn (W. Sun), nongmy@mail2.sysu.edu.cn (M. Nong), 577238134@qq.com (F. Li).

(NAQPMS) (Z. Wang et al., 2014). These models depend on detailed data sets of pollutant emissions, require relatively complex calculations, and have high uncertainties in the forecasting results.

Statistical models do not consider atmospheric processes explicitly. Statistical models are used to develop air quality forecasting models from a data-driven perspective. The methods, such as autoregressive integrated moving average (ARIMA) (U. Kumar & Jain, 2010) and multiple linear regression (MLR) (Vlachogianni, Kassomenos, Karppinen, Karakitsios, & Kukkonen, 2011), have been widely used in short-term forecasting. Since Boznar, Lesjak, and Mlakar (1993) first established an SO₂ concentration forecasting model based on the Multi-Layer Perceptron (MLP) in 1993, machine learning has been widely applied, such as the artificial neural network (ANN) (Esen, Inalli, Sengur, & Esen, 2008c), the random forest (RF) (Rubal & Kumar, 2018), the classification and regression tree (CART) (Gocheva-Ilieva, Voynikova, Stoimenova, Ivanov, & Iliev, 2019), and the support vector regression (SVR) (Leong, Kelani, & Ahmad, 2019). The back-propagation neural network (BPNN) was used to predict concentrations of the pollutants and outperformed an MLR (Cai, Yin, & Xie, 2009). The wavelet neural network (WNN) combines the strengths of discrete wavelet transform and neural network processing and has been successfully applied to forecasting (Y. Bai, Li, Wang, Xie, & Li, 2016; Esen, Ozgen, Esen, & Sengur, 2009). The adaptive neuro-fuzzy inference system (ANFIS) outperforms the ANN since the ANFIS inherits the interpretability of the fuzzy inference system and the learning ability of adaptive networks (Esen, Esen, & Ozsolak, 2017; Esen, Inalli, Sengur, & Esen, 2008a; Prasad, Gorai, & Goyal, 2016).

Machine learning is also combined with other methods to further improve forecasting performance. For example, a statistical weighted pre-processing method was developed to optimize ANN to use the minimum data set for prediction (Esen, Inalli, Sengur, & Esen, 2008b). Air quality index (AQI) data transformed by empirical mode decomposition (EMD) were input into SVR to obtain better forecasting performance (S. Zhu, et al., 2017). Evolutionary neuro-fuzzy inference systems, such as the ANFIS with particle swarm optimization (ANFIS-PSO) and genetic algorithm (ANFIS-GA), produce better estimates compared with the classical ANFIS (Alizamir, Kisi, Muhammad Adnan, & Kuriki, 2020). The gravitational search algorithm (GSA) can efficiently search for the optimal values of the unknown parameters of the least-square support vector regression (LSSVR). The integrated LSSVR-GSA enhances LSSVR's ability to estimate and forecast (Muhammad Adnan, et al., 2020). However, these models have drawbacks in feature extraction and generalization ability. Especially when the problem requires accurate prediction with big data, the forecasting ability of traditional statistical models is limited.

Deep learning is a powerful machine learning method suitable for big data, which has been widely applied in image understanding, speech understanding, natural language processing (Hao, Zhang, & Ma, 2016), and recently air quality forecasting. For example, an air quality prediction model based on long short-term memory (LSTM) appeared to have better performance than other traditional models such as ARIMA and SVR (X. Li, et al., 2017). The LSTM was combined with both the secondary decomposition and the optimization algorithm to establish a better hybrid model for AQI prediction in Beijing and Guilin (Q. Wu & Lin, 2019). The convolutional neural network (CNN), with successful applications in image analysis (Dong, Loy, He, & Tang, 2016), can account for the spatial distribution characteristics of air pollution. The CNN has been used to process the historical data of multiple sites in the form of a "class diagram" according to spatial locations and obtained higher prediction accuracy than the factorization-machine-supported neural network (FNN) (C. Wu, Li, Hou, Karimian, & Chen, 2018).

Although deep learning is promising to improve the forecasting accuracy of air quality, the accuracy of a single network may be limited when more complex problems are encountered. Many researchers have examined the combination of multiple deep learning networks (Karim, Majumdar, Darabi, & Chen, 2018; Quang & Xie, 2016). The well-

received CNN-LSTM combines the advantages of both the CNN and the LSTM. The CNN can effectively extract features of grid data and the LSTM has excellent processing ability for time series data. For example, a deep CNN-LSTM model for PM_{2.5} forecasting in smart cities used the LSTM to analyze the features extracted by the CNN (C.-J. Huang & Kuo, 2018). A hybrid model based on the CNN-LSTM for ozone concentration prediction significantly reduced the error compared with the MLP and LSTM (Pak, Kim, Ryu, Sok, & Pak, 2018). However, these CNN-LSTM models lack the data for the spatiotemporal distribution characteristics of air quality before the establishment of the CNN-LSTM.

Cluster analysis is a common method to divide disordered data into several categories with high similarity to analyze the internal patterns (K. Wang, Qi, Liu, & Song, 2018). It can provide training samples with high similarity for a model, which not only reduces the training time but also enhances the generalization ability (D. Li, Jiang, & Cao, 2006). Cluster analysis has been used in air quality forecasting by selecting or dividing the input variables to enhance the data regularity (S. Chen, Wang, & Zhang, 2019; Liu, Long, Duan, & Shi, 2020). However, the spatiotemporal distribution may affect the air quality forecasting. Reports on spatiotemporal clustering-based air quality forecasting are limited. A study on PM_{2.5} forecasting of seven urban agglomerations in China based on the RNN-CNN ensemble model shows that the forecasting accuracy for different regions is diverse (J. Huang, Zhang, Du, Liu, & Cao, 2019). The seasonal nonlinear grey Bernoulli model outperformed the traditional nonlinear grey Bernoulli model in improving the prediction accuracy (Zhou, Wu, Ding, & Cheng, 2020). Therefore, air quality forecasting based on deep learning combined with spatiotemporal clustering is promising.

The objective of this study is to establish multi-time multi-site deep learning models (LSTM, CNN, CNN-LSTM) in hourly forecasting of air quality (AQI) in Beijing, China. It entails (1) spatiotemporal distribution characteristics analysis of AQI, with explanations of the cause and effect relations for air quality; (2) development and comparison of the overall forecasting based on LSTM, CNN, CNN-LSTM models; (3) combination of CNN, LSTM, CNN-LSTM models with spatiotemporal clustering and comparison with the overall forecasting; and (4) application of the proposed models to multi-site hourly AQI forecasting in Beijing. In this study, the spatiotemporal distribution characteristics of AQI and their impact on air quality forecasting are analyzed by coupling spatial or temporal cluster analysis with deep learning.

2. Study area and dataset

2.1. Air quality in Beijing

Beijing was selected as the study area because it has some of the most serious air pollution in the world. It is also the economic, social, and cultural center of China. It is in the north of North China Plain with Yanshan Mountains in the north and Taihang Mountains in the west. Its terrain is high in the northwest and flat in the East. It has shorter springs and autumns, hot and rainy summers, and cold and dry winters, because of the sub-humid warm temperate continental monsoon climate. Owing to the coupling effect of the special terrain and climatic conditions, the air pollutants emitted from the heavy industrial cities around Beijing are not easily diffused. According to the annual report of the Beijing Municipal Economy and Environment Bureau, there were 347 days of obvious air pollution in 2015 and 2016, with PM_{2.5}, O₃, and PM₁₀ being the main pollutants.

To quickly and accurately evaluate air quality, a scientific evaluation method and a reasonable evaluation standard of air pollutants are needed to reflect the current ambient air quality qualitatively or quantitatively (Ning, Guan, Liu, Zhang, & Hare, 2019). The Air Quality Index (AQI) is one of the most commonly used indexes at present. This index considers the concentration of six pollutants (CO, NO₂, O₃, SO₂, PM₁₀, and PM_{2.5}), adopts a piecewise linear function to obtain the sub-index of each pollutant, and selects the highest value of sub-indexes to reflect the

overall air quality. The AQI reflects the current complex air pollution situation comprehensively and is widely concerned and applied in scientific research and daily life. The hourly AQI at 12 national ambient air quality monitoring stations in Beijing between 2015 and 2016 was collected from the China National Environmental Monitoring Center (<http://www.cnemc.cn/ssj/>). The study area and distribution of 12 national ambient air quality monitoring stations are shown in Fig. 1 and Table 1.

2.2. Input variables

Air quality is influenced by pollution emissions, meteorological conditions, terrain, social, and economic factors. The change in the emitted air pollutant concentration in the past can provide a reference for air quality forecasting in the future (A. Kumar & Goyal, 2011). The variation of meteorological conditions would modulate pollutant concentrations and provide the primary driving force for the variation of regional pollution (Zhang, Wang, Hu, Ying, & Hu, 2015). Meteorological indicators have been reported as important input variables for air quality forecasting in both numerical and statistical models (Kurt & Oktay, 2010; L. Li et al., 2011). The impact of these factors on air quality is based on spatiotemporal distribution characteristics (C. Fang, Liu, Li, Sun, & Miao, 2015; Yun, He, Jiang, Dou, & Dai, 2019). Therefore, air quality, meteorological, and spatiotemporal data for 6 h were selected as candidate input variables for the proposed AQI forecasting models (Table 2). The time resolution of all types of data is hourly.

The air quality data, collected from China National Environmental Monitoring Center (<http://www.cnemc.cn/ssj/>), include hourly PM_{2.5}, PM₁₀, SO₂, NO₂, O₃, and CO of 12 national ambient air quality monitoring points in Beijing from January 1, 2015, to December 31, 2016. The meteorological data collected from the China Meteorological Data Service Center (<http://data.cma.cn/site/index.html>) include hourly precipitation, temperature, dew point, pressure, wind speeds, and wind directions (angles). The meteorological data were collected from the national meteorological stations which were closest to each air quality monitoring point. The spatiotemporal data include three types of spatial data (the longitude and the latitude of air quality ground monitoring stations, the distance from stations to Tiananmen Gate in the center of Beijing) and three types of temporal data (the season, the weekday index, and hour time of each monitoring). The structures of the three spatial data of each site are constant. The monitoring season was mapped using one to four; the weekday index was represented using one

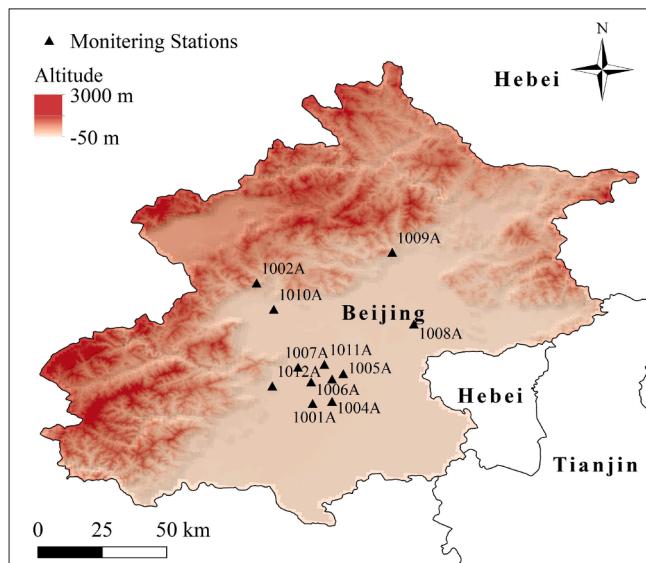


Fig. 1. Study area.

Table 1
Investigated Air Quality Monitoring Sites in Beijing.

Monitoring station	Latitude	Longitude	District
1001A	39.87°N	116.37°E	Xicheng
1002A	40.29°N	116.17°E	Changping
1003A	39.95°N	116.43°E	Dongcheng
1004A	39.87°N	116.43°E	Dongcheng
1005A	39.97°N	116.47°E	Chaoyang
1006A	39.94°N	116.36°E	Xicheng
1007A	39.99°N	116.32°E	Haidian
1008A	40.14°N	116.72°E	Shunyi
1009A	40.39°N	116.64°E	Huairou
1010A	40.20°N	116.23°E	Changping
1011A	40.00°N	116.41°E	Chaoyang
1012A	39.93°N	116.23°E	Shijingshan

Table 2
Candidate input variables for AQI forecasting models.

Type	Variable	Unit	Min	Max	Transformed
Air quality data	CO	µg/m ³	0.1	10.0	[0, 1]
	NO ₂	µg/m ³	2	290	[0, 1]
	O ₃	µg/m ³	1	500	[0, 1]
	PM _{2.5}	µg/m ³	3	999	[0, 1]
	PM ₁₀	µg/m ³	3	999	[0, 1]
	SO ₂	µg/m ³	2	500	[0, 1]
Meteorological data	Precipitation	mm	0.0	52.1	[0, 1]
	Temperature	°C	-19.9	39.3	[0, 1]
	Dew point	°C	-36	27.5	[0, 1]
	Pressure	hPa	982.4	1042.8	[0, 1]
	Wind speed	m/s	0.0	11.2	[0, 1]
Spatiotemporal data	Wind direction	°	0.0	337.5	[0, 1]
	Longitude	°	116.17	116.72	[0, 1]
	Latitude	°	39.87	40.39	[0, 1]
	Distance from Tiananmen Gate	m	4797.95	57311.36	[0, 1]
	Season	-	1	4	[0, 1]
	Weekday	-	0	1	[0, 1]
	Hour	-	0	23	[0, 1]

(Monday to Friday) or zero (Saturday and Sunday).

Since each site has <5% missing data, the data of all sites were retained. As some variables had missing data at individual sites at a certain time, the mean value of the data from other sites at the same time was used instead. If all site data for a variable were missing at a certain time, the mean value of each site data before and after that time was used to fill in the gap. To eliminate the effects of dimensional difference and improve the convergence speed of models, all of the data were transformed to the range of [0,1] by the following Min-Max scaling equation:

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

3. Methods

3.1. Overall framework

In this study, three types of deep learning neural network models (CNN, LSTM, and CNN-LSTM) based on spatiotemporal clustering were developed for the Beijing AQI forecasting for 6 h. After the spatiotemporal analysis of AQI in Beijing was conducted, all data were divided

into several clusters in the temporal or spatial dimension. The AQI forecasting models were established based on either the total data or those in different spatiotemporal clusters. Different models were evaluated and compared in terms of training and validation performance. Also, the BPNN was developed as a control model. The flowchart of this study is illustrated in Fig. 2.

3.2. Spatiotemporal analysis and clustering

In the temporal dimension, the mean values and the standard deviations of AQI at different time scales were calculated. The temporal variation characteristics of AQI in Beijing at the seasonal and monthly scales were analyzed. In the spatial dimension, Kriging interpolation and the Moran's I statistic were used to analyze the spatial distribution features of AQI in Beijing. The ordinary Kriging interpolation using local factors was selected to estimate the unknown points. The expression is as follows:

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (2)$$

where $Z(x_0)$ is the interpolation of unsampled point x_0 , $Z(x_i)$ is the characteristic value of sample point x_i , n is the number of sample points, λ_i is the weight coefficient. Additionally, the Kriging interpolation needs to determine the semivariogram model used. The commonly used models (e.g., circle, spherical, Gaussian models) are selected by their accuracy.

Moran's I statistic represents the spatial correlation of one variable. The calculation formulas for the global and local Moran's I statistics are as follows:

$$\frac{I = \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} \quad (3a)$$

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_{j=1}^n w_{ij} (x_j - \bar{x}) \quad (3b)$$

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (3c)$$

where I is the global Moran's I statistic, I_i is the local Moran's I statistic of sample point i , S^2 is the variance of all values, x_i and x_j are values of sample point i and j , \bar{x} is the mean value of all sample points, w_{ij} is the spatial weight, and n is the number of sample points.

The global Moran's I statistic indicates whether there is agglomeration in the values of all sample points. Its value range is between -1 and 1 . The positive global Moran's I indicates that the values in the region are positively correlated in space. Otherwise, they are negatively correlated. The local Moran's I statistic is used as a local indicator of spatial association. When it passes the significance test, all sample points can be divided into high-high, high-low, low-high and low-low clusters.

According to the spatiotemporal distribution characteristics of AQI in Beijing, the dataset was divided into different clusters. In the temporal dimension, four clusters of data were obtained according to different seasons. In the spatial dimension, 12 stations were divided into multiple clusters by hierarchical clustering of the AQI. The Euclidean equation was used to calculate the distance between points. The Ward linkage method was used to calculate the distance between clusters. As the size of convolution kernels is mostly 3×3 or above, the number of stations in each cluster should be more than 3 to complete the convolution process. Then, the forecasting models based on the CNN, LSTM, CNN-LSTM, and BPNN models were established using different seasonal or spatial clusters of data.

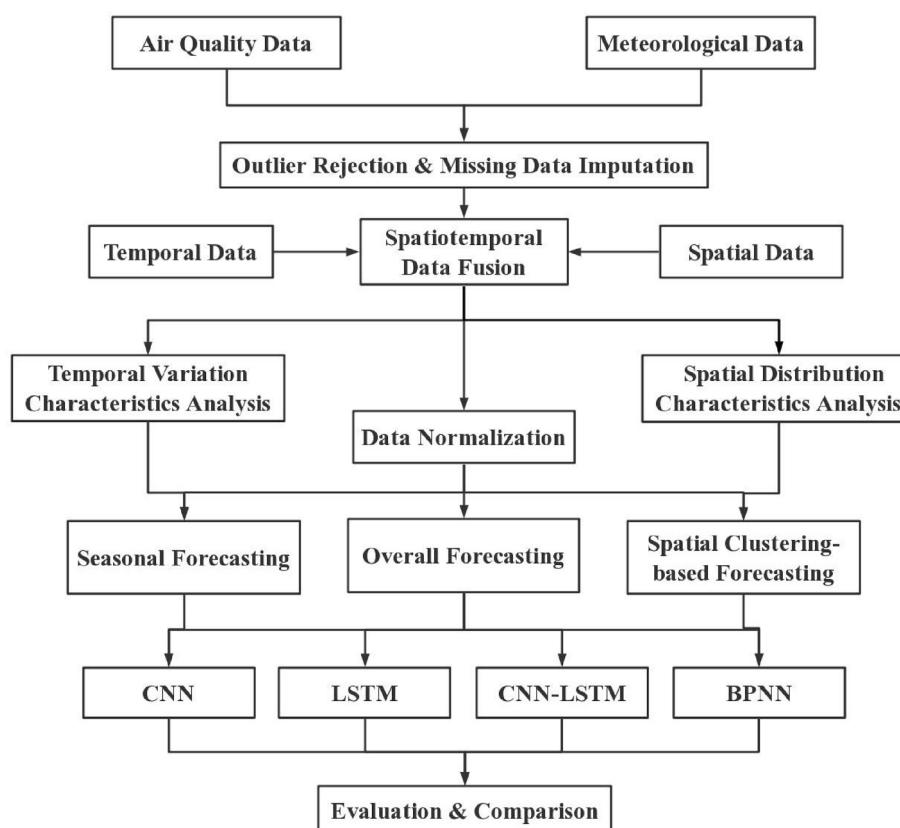


Fig. 2. Flowchart of proposed models.

3.3. CNN

The CNN, with powerful processing ability for grid data, has been widely applied in image analysis (Krizhevsky, Sutskever, & Hinton, 2017; Sun, Wang, & Tang, 2014). The main structure of CNN includes input, convolution, pooling, full connection, and output layers. Information in the input layer is processed through feature transformation and extraction in the convolution and pooling layers. This local information of convolution and pooling layers is further integrated by the fully connected layers and is mapped to output signals through the output layer (Fig. 3a).

The convolution layer is the most important and unique one in the CNN because it can extract features of input variables by convolution kernels. Different from the fully connected neural network, the convolution layer of the CNN only connects part of neurons of the previous layer. The scale of convolution kernels is smaller than that of the input matrix. The convolution layer uses convolution operations instead of general matrix operations to output the feature map. The calculation formula of each element in the feature map is:

$$x_{i,j}^{\text{out}} = f_{\text{cov}} \left(\sum_{m=0}^k \sum_{n=0}^k w_{m,n} x_{i+m,j+n}^{\text{in}} + b \right) \quad (4)$$

where $x_{i,j}^{\text{out}}$ is the output value in row i and column j of the feature map; $x_{i+m,j+m}^{\text{in}}$ is the value in row i and column j of the input matrix; $f_{\text{cov}}(\cdot)$ is the selected activation function; $w_{m,n}$ is the weight in row m and column n for the convolution kernel; and b is the bias of the convolution kernel. The input matrix generally uses multiple kernels for convolution operations. Each convolution kernel will extract a feature from the input matrix and generate a feature map. After that, the pooling layer reduces the length and width of the previous feature map and improves the efficiency of the calculation by down-sampling.

Since the CNN has a good ability to extract grid data features, m variables of each type at n stations were expanded to obtain a matrix of m rows and n columns. A total of c types (air quality, meteorological, and spatiotemporal attribute) of data were input in the multi-channel form. The TimeDistributed layer can transform a layer to every temporal slice of the input, which gains information of more historical time steps of time series data and obtains the long-term characteristics of input variables. Therefore, the TimeDistributed layer was selected to wrap the data of the past t hours. Thus, the input of the CNN was a 4-D array with the dimensions of $t \times m \times n \times c$ ($6 \times 6 \times 12 \times 3$; Fig. 3b). After convolution and pooling layers, the extracted features were flattened into a 1-D array. Finally, the forecasting AQI at n stations in the next t' hours were generated through the fully connected output layers.

3.4. LSTM

The recurrent neural network (RNN) has a recurring mechanism in hidden layers. The input includes the features of certain and past moments. Therefore, it can correlate with contextual information and perform well in many fields, such as for natural language processing and sequence data modeling (LeCun, Bengio, & Hinton, 2015). However, with the increase of network layers and iterations, the subsequent nodes of the RNN will gradually forget the previous information, resulting in gradient diminishing or gradient explosion problems (Hochreiter, 2011). In solving these problems, the LSTM is proposed as an excellent variation of the RNN.

In the LSTM, the memory cell is specially used to save historical information (Fig. 3c). The key to the memory cell is the cell state and three gates. The cell state is the path of information transmission, which enables information to be transmitted in sequence. The gates are used to update or discard historical information. The cell contains three gates, which are input, forgetting, and output gates. The specific formula derivation of the LSTM is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (5a)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5b)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (5c)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad (5d)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5e)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (5f)$$

Considering the processing ability of the LSTM for time series data, the $m \times c$ variables at n stations in the past t hours were expanded to obtain a 2-D array with the dimensions of $t \times v$ ($v = m \times n \times c$), i.e., a dimension of 6×216 ($216 = 6 \times 12 \times 3$) (Fig. 3d), which were added as inputs into the network. Furthermore, the AQIs of n stations in the next t' hours were generated through the fully connected output layers.

3.5. CNN-LSTM

The CNN-LSTM, which combines the advantages of CNN and LSTM, has achieved breakthrough results, especially in the field of natural language processing, as well as video recognition and classification (Bhunia, et al., 2019; Song, Huang, & Ruan, 2019). As a hybrid network, different researchers have proposed various combinations of modes of the CNN-LSTM. The CNN-LSTM adopted in this study includes a few processes. The $m \times c$ variables at n stations in the past t hours were taken as input data. The input dimension was the same as that of the CNN, which was $t \times m \times n \times c$, i.e., $6 \times 6 \times 12 \times 3$. The input, convolution, and pooling layers of the CNN were reserved to extract the features of input data. The obtained features were flattened into 1-D arrays and were input into the LSTM layer as time sequence to analyze the time sequence features of input data. Finally, the AQIs at n stations in the next t' hours were obtained through the fully connected and output layers (Fig. 3e).

3.6. Performance evaluation indices

To validate the generalization ability of models, the data set was divided into the training (first 70%) and validation (last 30%) sets. The root mean squared error (RMSE) and the index of agreement (IA) was used to evaluate the performance of forecasting models. The RMSE is sensitive to the maximum or minimum error between the observed and forecasting values. The lower the RMSE, the smaller the error. The IA indicates the distribution similarity between the observed and forecasting values. The IA varies in the range of [0, 1]. The closer to 1 the IA is, the higher the distribution consistency is. The calculation formulas are illustrated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$\text{IA} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|y_i - \bar{y}| + |\hat{y}_i - \bar{y}|)^2} \quad (7)$$

where n is the number of data points; y_i and \hat{y}_i are the observed value and forecasting value of each sample, respectively; and \bar{y} is the mean value of all the observed values.

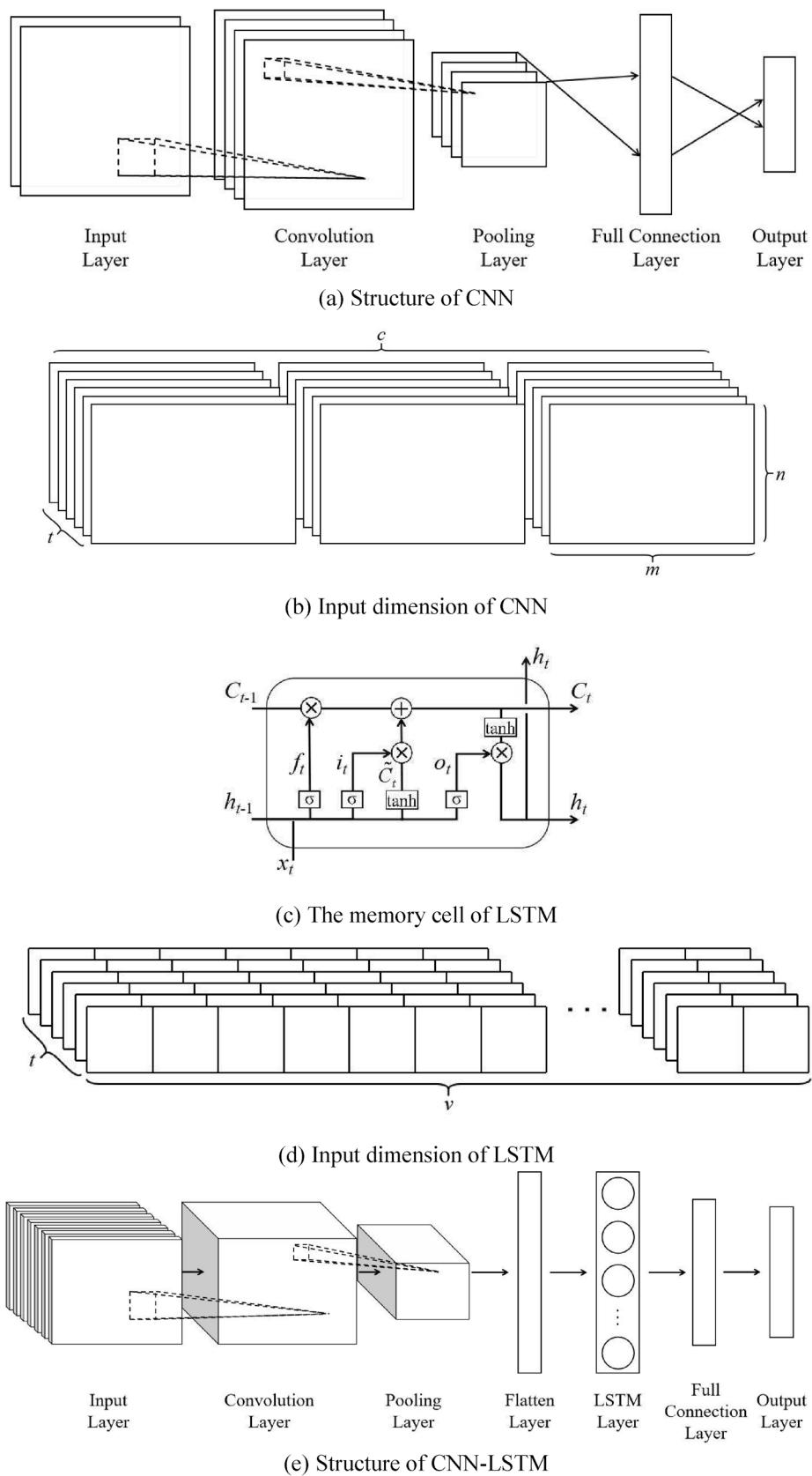


Fig. 3. Structure and input dimension of CNN, LSTM, and CNN-LSTM.

4. Results

4.1. Spatiotemporal distribution characteristics of AQI

In the temporal dimension, the annual seasonal and monthly changes of the AQI were analyzed (Fig. 4a). The average AQI of Beijing from

2015 to 2016 was 105.42. There were obvious seasonal differences ($P < 0.001$) of the AQI in the two years, decreasing in the order of winter, spring, autumn, and summer. The summer average AQI (84.12 ± 52.37) was lower than the overall average. But the winter average AQI (125.05 ± 117.64) was significantly higher than the overall average. The average AQI of each month in 2015 to 2016 varied from 70.72 to

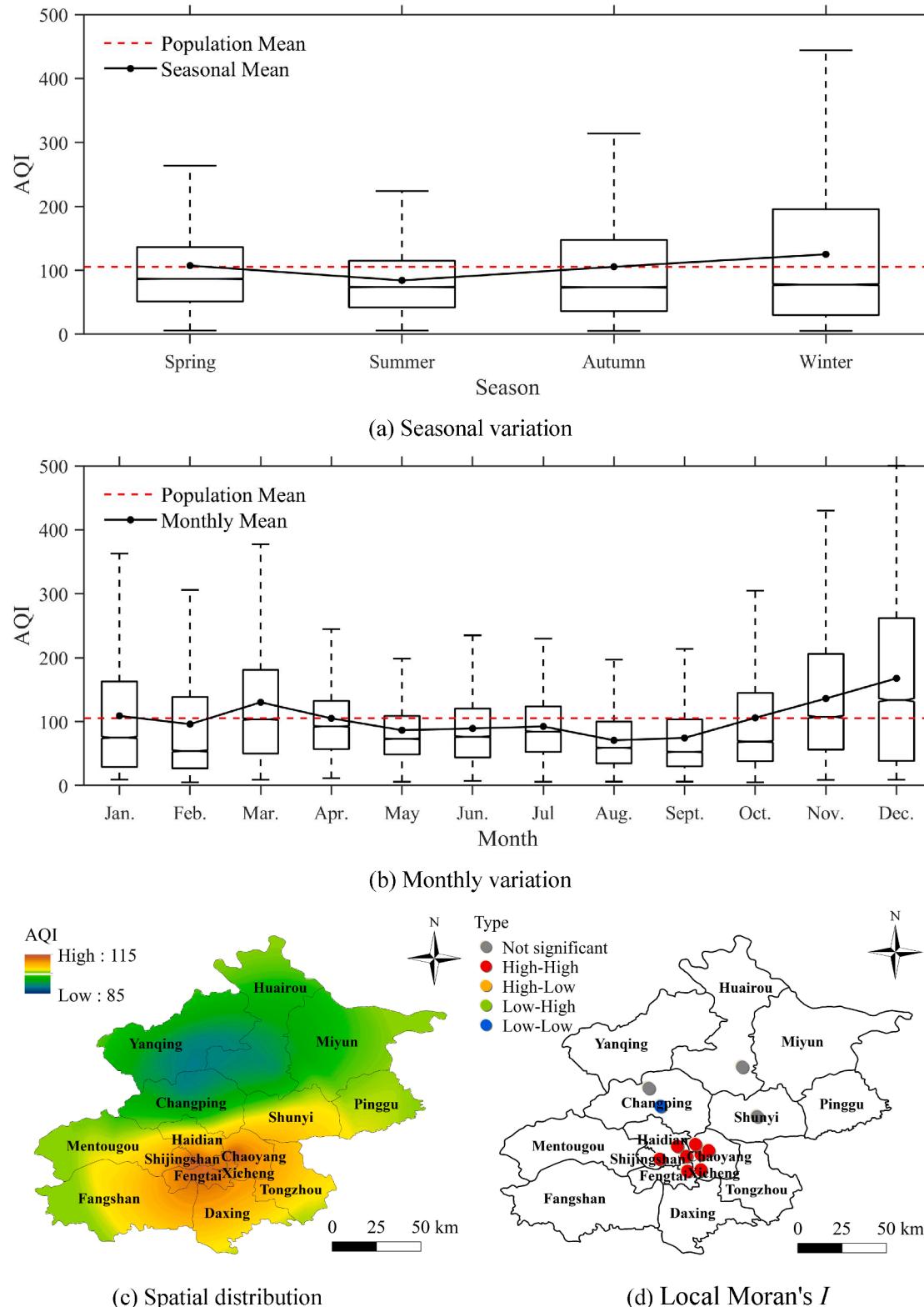


Fig. 4. Spatiotemporal distribution characteristics of AQI (Outliers are not displayed in Fig. 4a and 4b).

167.74. The AQI was higher at both the beginning and end of the year and lower in the middle of the year, showing a U-shaped variation (Fig. 4b). In August, the average AQI was the lowest (70.72 ± 44.45) with high stability. In December, the average AQI was the highest (167.74 ± 160.00) and with a high discrete degree. These variations indicate that the air quality of Beijing in summer months was generally good and stable from 2015 to 2016. The air pollution increased in spring and autumn, and was the worst in winter with large fluctuations and susceptibility to high pollution weather events.

In summer, strong solar radiation causes the surface temperature to rise sharply and heats the near-surface air. This leads to increased convection, unstable atmosphere stratification, and increased precipitation, which are beneficial to the diffusion and deposition of air pollutants (Cui, et al., 2019; Yang, Yuan, Li, Shen, & Zhang, 2017). In spring, the frequent occurrence of sandstorms in Northwest China brings abundant air pollutants to Beijing (Y. Wang, et al., 2005). In autumn, stable weather and biomass combustion are common (Cheng, et al., 2013). These seasonal factors increase air pollutants and decrease the AQI in spring and autumn. In winter, the low surface temperature causes a surface inversion, and the height of the atmospheric mixed layer is low. The meteorological conditions are not conducive to vertical convection (Yang, et al., 2017), therefore the near-surface air pollution is high. Also, coal-fired emissions for heating in winter and the pollutants generated by fireworks and crackers during the Spring Festival have a significant negative impact on air quality (Cui, et al., 2019; Y. Wang, Zhuang, Xu, & An, 2007).

In the spatial dimension, the global Moran's I statistic for the AQI at multiple sites in Beijing between 2015 and 2016 was 0.420 ($P < 0.001$). This indicated that the AQI at multiple sites in Beijing showed significant spatial positive correlation and spatial agglomeration. The ordinary Kriging interpolation with the circular model was conducted for the average AQI from 2015 to 2016 (Fig. 4c). The AQI in Beijing was higher in the southeast and lower in the northwest. The areas with higher AQI were mainly concentrated in Dongcheng, Xicheng, Chaoyang, the southeast of Haidian, and east of Fengtai districts, with a maximum value of 113.46. According to the local Moran's I statistic, the eight stations located in this area belonged to high-high clusters (Fig. 4d). As the central city of Beijing, this area had the most serious air pollution and its AQI showed significant positive spatial correlation. With the increase of distance from the central city, the AQI showed a decreasing trend and the air quality improved (Fig. 4c). The low AQI areas were mainly concentrated in Changping, Yanqing, and Huairou Districts with high altitudes in the northwest and a minimum AQI of 91.05. Station 1010A in Changping district belonged to the low-low cluster, which is the low pollution area in Beijing. The local Moran's I statistics of stations 1002A in Changping, 1009A in Huairou, and 1008A in Shunyi were not statistically significant, indicating that these stations did not show significant spatial correlation with others.

In general, the AQI in Beijing is higher in the southeast and lower in the northwest, because of various reasons. The higher AQI in the southeast may be caused by higher population density, municipal emissions, traffic volume, and automobile exhaust emissions in the

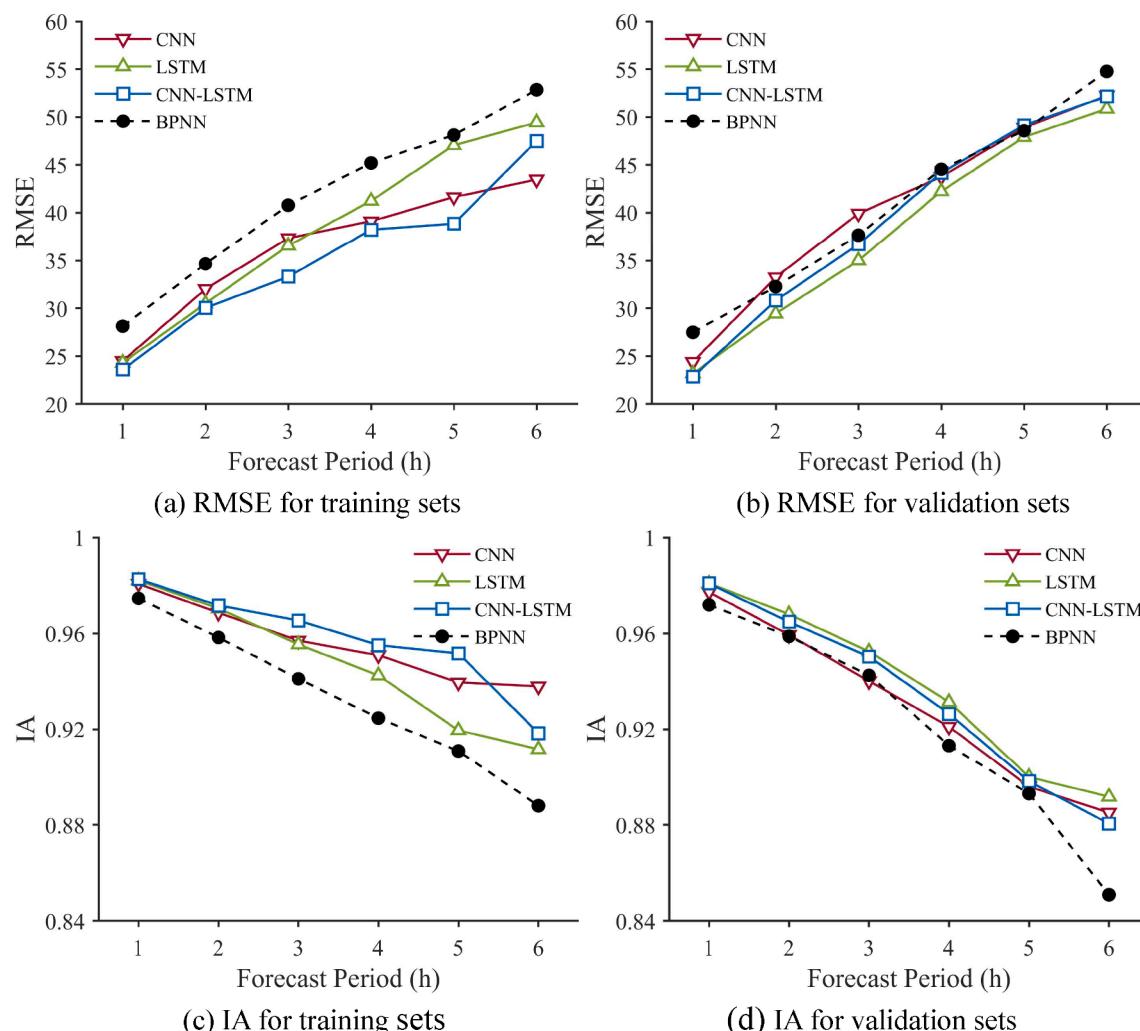


Fig. 5. Overall AQI forecasting.

central city of Beijing (Tian, et al., 2019). Furthermore, there are many industrial parks in Daxing and Tongzhou Districts in the southeast of Beijing. This area is also close to Tianjin and the main industrial cities in Hebei. Industrial production may produce exhaust gas pollution, which lowers air quality in the southeast (W. Chen, Tang, & Zhao, 2015). Additionally, the west, north, and northeast of Beijing are surrounded by mountains. When the pollutants from the southeast of Beijing and surrounding cities spread to the northwest mountains, they are influenced by orographic uplifting (Xu, et al., 2017).

4.2. Overall AQI forecasting model using deep learning

The accuracy of four types of AQI forecasting models based on the two years of data (2015/01/01–2016/12/31) is shown in Fig. 5 and Table S1. In terms of the next-hour forecasting, the training and validation RMSE were ranked in the descending order of the BPNN, CNN, LSTM, and CNN-LSTM, which was the same as the ascending order of the IA. Compared with the BPNN, for the training set, the RMSE of the CNN, the LSTM, and the CNN-LSTM decreased by 12.92%, 13.57%, and 16.11%, respectively; the IA increased by 0.62%, 0.77%, and 0.82%, respectively. For the validation set, the RMSE decreased by 11.25%,

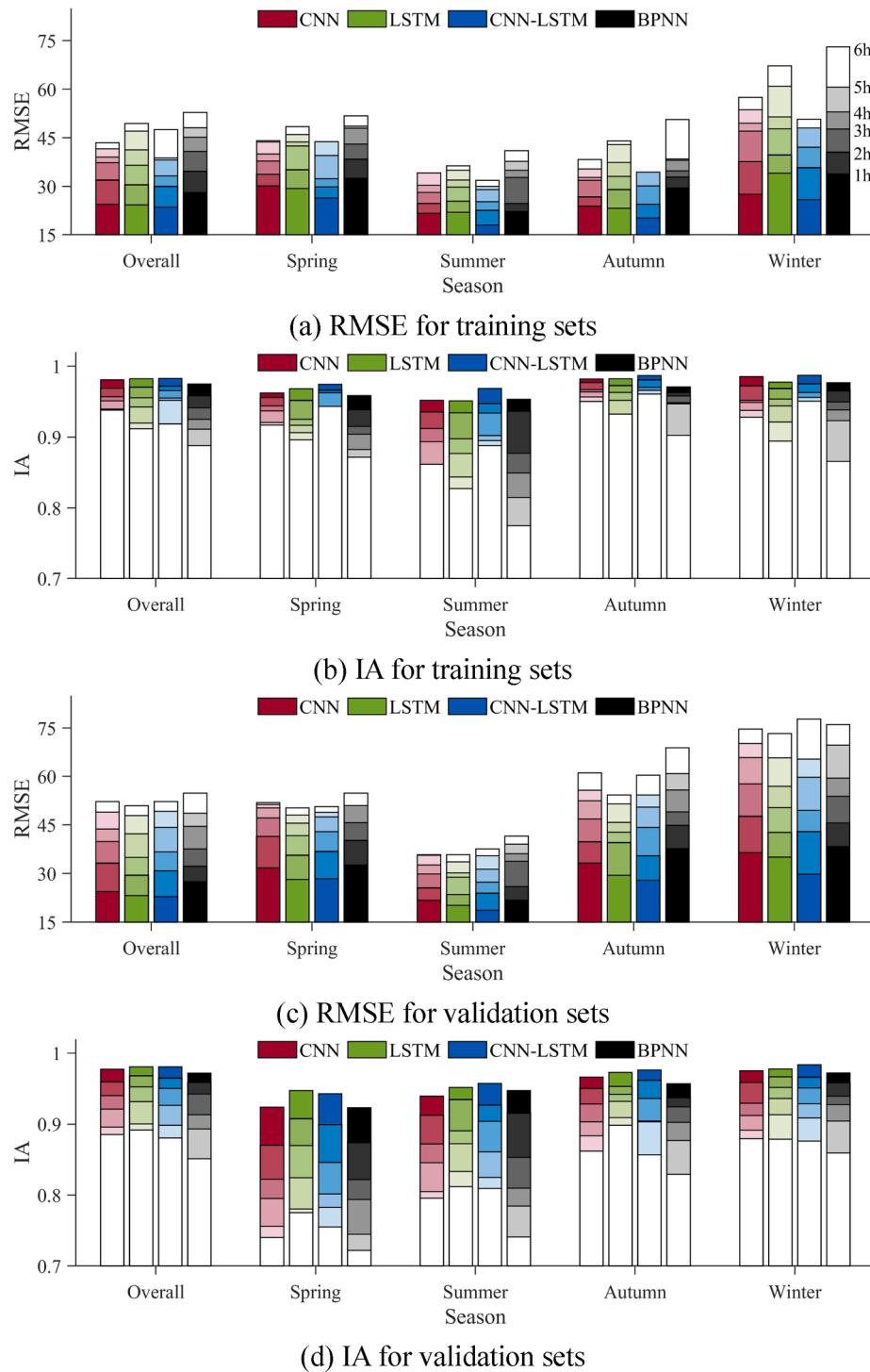


Fig. 6. Seasonal AQI forecasting.

15.65%, 16.86%, and the IA increased by 0.54%, 0.92% and 0.92%, respectively. This indicates that the forecasting errors and the variation consistency of three deep learning models outperform the BPNN. Especially, the CNN-LSTM presents the best forecasting ability for both training ($\text{RMSE} = 23.60$, $\text{IA} = 0.983$) and validation ($\text{RMSE} = 22.86$, $\text{IA} = 0.981$) datasets. Thus, the CNN-LSTM is considered the optimal model for the next-hour AQI forecasting.

In terms of multi-hour forecasting, the forecasting accuracy of all models decreases with the number of hours (Fig. 5). The validation performance of the four models becomes slightly worse than the training performance. The differences in validation performance between the four models become smaller than those in training. For the training sets, the BPNN has the lowest accuracy for multi-hour forecasting; the accuracy of the CNN is higher than the LSTM when the leading time is more than 3 h. The CNN-LSTM has the lowest error and the highest IA, except for the 6-hour forecasting. For the validation sets, the accuracy of either the BPNN or CNN is the lowest for individual-hour forecasting. The CNN-LSTM generally has higher accuracy and the accuracy of the LSTM is the best in terms of the lowest error and the highest IA. As the validation performance is of most concern, the LSTM is considered the optimal model for multi-hour AQI forecasting. However, it is also noted that the difference of validation performance between the LSTM and the CNN-LSTM is relatively small.

4.3. Seasonal forecasting

Because the AQI in Beijing varies seasonally, all the data were divided into four clusters according to the seasons. The seasonal forecasting models based on each cluster were established (Table S2). Fig. 6 shows the comparison between seasonal and overall forecasting models using four types of neural networks. The difference in the accuracy of multiple leading hours for seasonal forecasting models is generally similar to the overall forecasting models. For the training sets, the three deep learning models outperform the BPNN in terms of both the forecasting error and the variation consistency. The forecasting accuracy of the CNN-LSTM is the highest. The CNN has better forecasting performance than the LSTM, which gradually improves with a greater number of leading hours. For the validation sets, the ability of CNN-LSTM in next-hour forecasting is more prominent. With more leading hours, the LSTM maintains better forecasting performance than the CNN, the CNN-LSTM, and the BPNN.

There are some seasonal differences in forecasting accuracy. For the training sets, the autumn forecasting generally outperforms the overall forecasting. The former has lower RMSE and higher IA ranges using four types of models with the leading time of one to six hours. The training performances of the spring, summer, and winter forecasting generally are worse than that of the overall forecasting. For the validation sets, in the next-hour forecasting, the RMSE ranges for different models can be ranked in the descending order of winter ([29.81, 38.32]), autumn ([27.91, 37.71]), spring ([28.18, 32.62]), overall ([22.86, 27.50]), and summer ([18.64, 21.82]). Only the RMSE of the summer forecasting is consistently lower than that of the overall forecasting. The IA ranges of different models can be sorted in ascending order of spring ([0.923, 0.947]), summer ([0.940, 0.957]), autumn ([0.957, 0.976]), overall ([0.972, 0.981]), and winter ([0.972, 0.984]). The IA ranges of CNN-LSTM and BPNN in the winter forecasting are higher than those of the overall forecasting. In the multi-hour forecasting, the seasonal differences of the RMSE and IA ranges for validation are similar and gradually become larger with more leading hours. Moreover, it is noted that the performance of models in different seasons is not identical in terms of reducing error and improving variation consistency. In the spring and summer forecasting, the errors are lower while the distribution consistency is not higher. In the autumn and winter forecasting, the larger errors correspond with the higher distribution consistency. These results are similar to the other studies where the MLR and BPNN models were developed (H. Bai, Shen, Shi, & Dong, 2013; X. Fang, Duan, Hu, & Cai,

2019).

Seasonal forecasting may not always effectively improve forecasting accuracy. The reduced amount of data in different seasons cause the deceasing data patterns for model development, which may limit the ability of the seasonal forecasting. Considering the AQI's annual characteristics, it is found that the forecasting error may be related to the AQI levels and the discrete degrees in different seasons. The higher the AQI level, the greater the discrete degrees, the higher the uncertainty of pollution change, and the more difficult the air quality forecasting (Y. Zhu, Li, Hou, Fan, & Feng, 2016). In summer, the AQI is generally low without numerous high values. Even though the forecasting values cannot reflect the variation in observed values well, the forecasting error is always small. In winter, the AQI is discrete with more fluctuations. The high AQI values limit the forecasting ability of all types of models, causing the winter forecasting error to be the largest of the four seasons. Also, the lowest consistency in spring may be related to the frequent occurrence of dust weather patterns in this season in Beijing (H. Bai, et al., 2013).

4.4. Spatial clustering-based forecasting

According to the spatial characteristics of the AQI in Beijing, the twelve monitoring stations were divided into two clusters through hierarchical clustering. Stations 1001A, 1003A, 1004A, 1005A, 1006A, 1007A, 1011A and 1012A belong to cluster 1, which is located in the central city. Stations 1002A, 1008A, 1009A, and 1010A belong to cluster 2, which is at a certain distance north of the central city (Table S3, and Fig. 7a). The clustered data were individually used to develop four types of models. The results are shown in Fig. 7b–e.

The forecasting performance of the same cluster varies using different models with more leading hours. In terms of the next-hour forecasting, the accuracy of cluster 1 and overall forecasting for both training and validation is in the ascending order of the BPNN, CNN, LSTM, and CNN-LSTM. The training accuracy of cluster 2 forecasting is in the ascending order of the BPNN, CNN, CNN-LSTM, and LSTM; its validation accuracy is slightly different, with the ascending order of the CNN, BPNN, CNN-LSTM, and LSTM. In terms of the multi-hour forecasting, the performances of cluster 1, 2, and overall forecasting models have similar patterns. The forecasting accuracies for training and validation using the 3 deep learning models are generally better than the BPNN. The CNN-LSTM has the highest forecasting accuracy for most of the training sets, and the LSTM outperforms the other models in the forecasting performance for validation.

There are also differences in the forecasting performance of the same model for different clusters. In the next-hour forecasting, for training and validation sets, the RMSE values of CNN and CNN-LSTM in cluster 1 ([24.41, 24.69] and [21.44, 22.42]) are lower than those in cluster 2 ([24.52, 26.00] and [22.72, 24.89]), while the RMSE values of LSTM and BPNN in cluster 1 ([22.07, 22.77] and [28.31, 29.38]) are larger than those in cluster 2 ([21.47, 23.28] and [24.35, 27.59]). The IA values of all models are ranked in the descending order of cluster 1 ([0.972, 0.985]), the overall ([0.972, 0.983]), and cluster 2 ([0.972, 0.981]). With the leading time rises, both the increase of RMSE and the decrease of IA of cluster 1 are greater. In the two- to five-hour forecasting, the RMSE of cluster 1 is higher than that of cluster 2 and the overall. In the four- to five-hour forecasting, the IA of cluster 1 is lower than that of both cluster 2 and the overall. In the six-hour forecasting, the RMSE values of all models are ranked in the descending order of cluster 1, the overall, and cluster 2. The IA generally shows the ascending order of the overall, cluster 1, and cluster 2. Therefore, cluster 1 forecasting shows certain advantages of improving variation consistency in one- to three-hour forecasting; cluster 2 forecasting, which has improved the ability to reduce errors and increase variation consistency, generally outperforms cluster 1 and the overall forecasting.

The poorer performance of cluster 1 than cluster 2 forecasting may be for multiple reasons. The AQI in cluster 1 is generally higher with

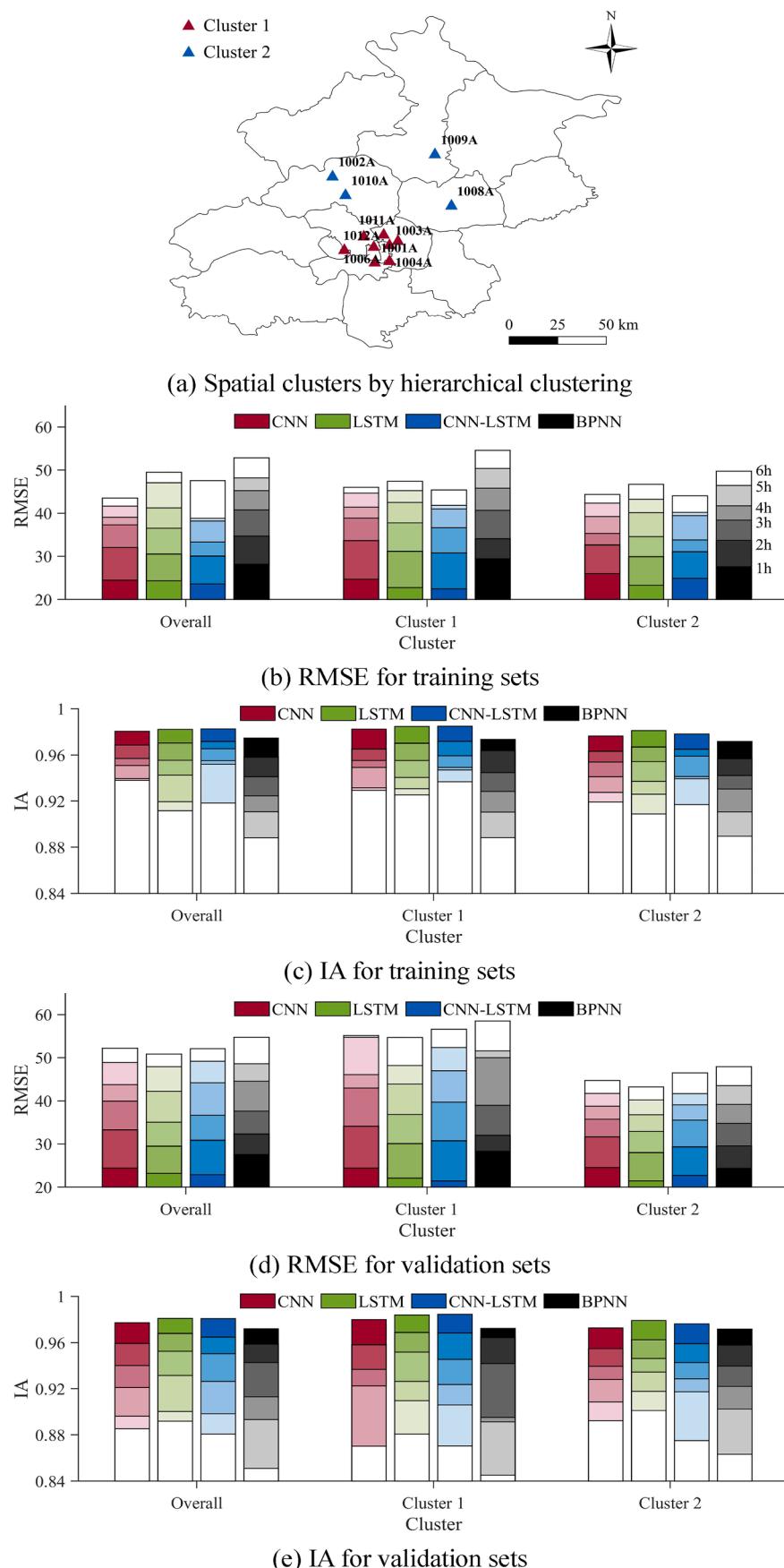


Fig. 7. Spatial clustering-based AQI forecasting.

larger fluctuations, which may cause the forecasting models to underestimate high values or overestimate low values. This may cause the RMSE increase and the IA decrease in cluster 1 forecasting. Also, the distances between these stations in cluster 1 are shorter, resulting in a

smaller AQI difference and stronger spatial autocorrelation. The forecasting difficulty of spatial variation characteristics in cluster 1 is relatively low. By contrast, the AQI in cluster 2 is generally lower, which may be the main reason leading to smaller forecasting errors. The spatial

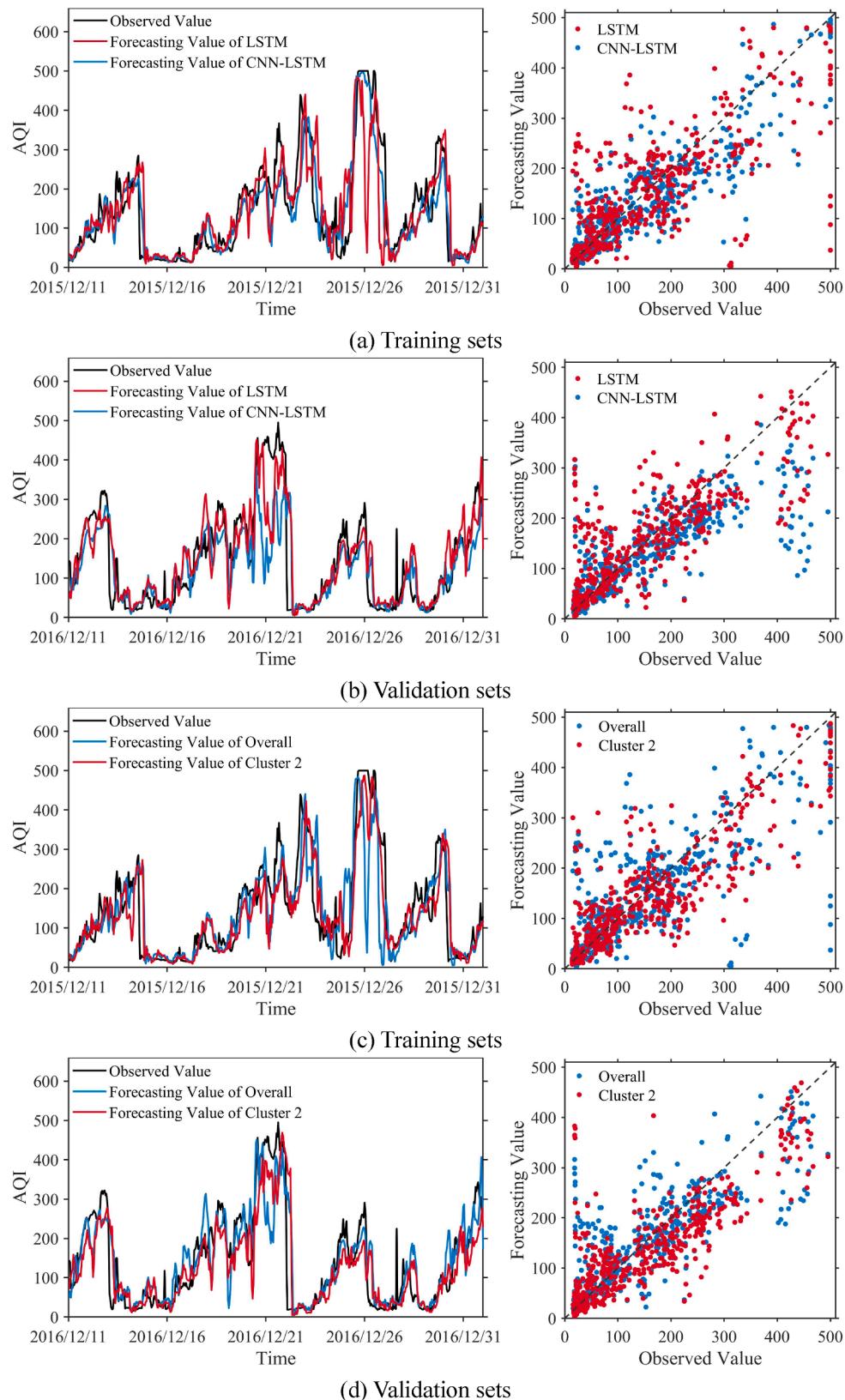


Fig. 8. Comparison of AQI forecasting models for station 1002.

autocorrelation between the four stations in cluster 2 is low, resulting in a larger AQI difference. This may limit the difficulty of the spatial variation characteristics in cluster 2 forecasting. With more lead time (hours), the effect of spatial variation characteristics on the forecasting difficulty in each cluster may increase.

4.5. Optimal forecasting

The performance of the seasonal forecasting is not significantly improved compared with the overall forecasting. Similarly, only the performance of the cluster 2 forecasting is better than the overall forecasting. Either the seasonal or the spatial clustering-based forecasting is more suitable for improvement of the forecasting in a certain season or cluster. In terms of the model type, the CNN-LSTM and the LSTM are the optimal model for the next-hour and multiple-hour forecasting, respectively. Thus, the overall forecasting based on the LSTM is considered the optimal model for all the stations in Beijing.

To further illustrate the performance difference between the model types, Fig. 8a and 8b show the comparison of the LSTM and the CNN-LSTM in the six-hour overall forecasting at Station 1002. The forecasted AQI values using the LSTM are closer to the observed ones than using the CNN-LSTM. To further illustrate the performance difference between the clusters, Fig. 8c and 8d show the comparison of cluster 2 and overall forecasting at Station 1002 with the six hours. The forecasted AQI values in Cluster 2 are closer to the observed ones than in the overall dataset.

5. Discussion

In this study, the AQI forecasting models based on three deep learning methods, one shallow learning method, and spatiotemporal clustering were developed and compared. The results show that the LSTM and the CNN-LSTM generally perform better in multi-hour forecasting, whereas spatiotemporal clustering-based forecasting is more suitable for a particular season or cluster. This study indicates that deep learning is a more effective tool to process big data, especially spatiotemporal data, compared with shallow learning. The spatiotemporal clustering and model combination can improve the performance in spatiotemporal data forecasting to some extent.

The proposed methodology is feasible for multi-hour and multi-site AQI forecasting. In terms of input variables, the available data of routine monitoring variables from national monitoring stations were used. In terms of the modeling method, the deep learning methods were combined and coupled with spatial or temporal clustering. In terms of network combination, different from ensemble learning in previous studies (J. Huang, et al., 2019), the flatten layer converts the features extracted by the CNN into the array dimension for the LSTM, which combined the CNN and the LSTM. The TimeDistributed layer in this study was applied as the contained layer to handle multiple time steps of the input (Ru, Li, Liu, & Chai, 2018), which expanded the input of the CNN to the historical time series data. This solved the problem that the CNN-based AQI forecasting models only used past data at a single time point as the input (C. Wu, et al., 2018).

In previous studies, the CNN-LSTM was considered as the optimal forecasting model of air quality, compared with the CNN and the LSTM (C.-J. Huang & Kuo, 2018; T. Li, Hua, & Wu, 2020; Pak, et al., 2018). However, the LSTM generally performs better than the CNN-LSTM in this study. It is because only the air quality for the next hour or day was forecasted in the previous studies. The CNN-LSTM demonstrates the best ability for the next-hour/day forecasting in previous studies and this study. Due to the complexity of the CNN-LSTM, however, the generalization ability of the CNN-LSTM decreases with the increase of leading hours. The LSTM is generally considered as the optimal model in multi-hour AQI forecasting.

The performance of spatiotemporal clustering-based forecasting is not improved in spatial and temporal clusters. The performance

difference among these models may originate from the relations between the AQI and the driving factors, the spatiotemporal clustering methods, the model type, the model structure, and the model development methods. Some other factors, such as the data volume, the magnitude of the AQI, and the dispersion degree and the spatial correlation between clusters, may also result in different forecasting performance. Further analysis of possible reasons for the difference is warranted. Furthermore, the forecasting models were combined with temporal and spatial clustering separately in this study, lacking a combination with the joint of spatiotemporal clustering. The performance of forecasting models combined with spatiotemporal joint clustering is also worth exploring in further study. Additionally, more signal processing methods (e.g., EMD) can be used to pre-process the input data so that the prediction accuracy improves.

6. Conclusions

In this study, based on the spatiotemporal clustering analysis of the AQI in Beijing from 2015 to 2016, the CNN, the LSTM, and the CNN-LSTM were developed as multiple-hour and multiple-site forecasting models. The performance of these models was compared with the BPNN. The major findings are as follows:

- In the temporal dimension, there were obvious seasonal differences of the AQI in the two years, decreasing in the order of winter, spring, autumn, and summer, showing a U-shaped variation on annual seasonal and monthly scales. The seasonal forecasting models were established based on four data clusters in the seasons. In the spatial dimension, the AQI was higher in the southeast and lower in the northwest, with a gradual decrease from city centers to the suburbs. The AQI at different monitoring stations showed significant autocorrelation. The spatial clustering-based forecasting models were developed based on two clusters through hierarchical clustering.
- Using the entire dataset, in the next-hour forecasting, the performance of the overall forecasting was ranked in the ascending order of the BPNN, CNN, LSTM, and CNN-LSTM. In the multiple-hour forecasting, as the validation performance is of most concern, the LSTM is considered the optimal model. The difference in validation performance between the LSTM and the CNN-LSTM is relatively small. The overall forecasting based on the LSTM is considered as the optimal model for all stations in Beijing.
- For the seasonal forecasting, the performance was not significantly improved compared to the overall forecasting. For spatial clustering-based forecasting, cluster 2 forecasting generally outperforms cluster 1 and the overall forecasting. Overall, either the seasonal or the spatial clustering-based forecasting is more suitable for a certain degree of improvement for forecasting in a certain season or cluster. In terms of the model type, both the CNN-LSTM and the LSTM generally have better performance than the CNN and the BPNN.

CRediT authorship contribution statement

Rui Yan: Conceptualization, Methodology, Software, Writing - original draft, Project administration. **Jiaqiang Liao:** Methodology, Data curation. **Jie Yang:** Methodology, Software. **Wei Sun:** Writing - review & editing, Funding acquisition, Supervision. **Mingyue Nong:** Investigation, Resources. **Feipeng Li:** Investigation, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study is financially supported by Top-Notch Young Talents of Pearl River Talents Plan (2019QN01G106), the Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (99147-42080011), the Hundred Talents Program of Sun Yat-Sen University (37000-18841201), National Undergraduate Training Programs for Innovation and Entrepreneurship (201901211), and the National Program on Key Research Projects of China (2017YFC1502706).

Appendix A. Supplementary data

Tables S1 to S3 are in the supplementary material. The data used in this study and the developed Python code are available from the corresponding author upon request. Supplementary data to this article can be found online at <https://doi.org/10.1016/j.eswa.2020.114513>.

References

- Alizamir, M., Kisi, O., Muhammad Adnan, R., & Kuriqi, A. (2020). Modelling reference evapotranspiration by combining neuro-fuzzy and evolutionary strategies. *Acta Geophysica*, 68(4), 1113–1126.
- An, J., Huang, M., Wang, Z., Zhang, X., Ueda, H., & Cheng, X. (2001). Numerical regional air quality forecast tests over the mainland of China. *Water, Air, and Soil Pollution*, 130, 1781–1786.
- Bai, H., Shen, R., Shi, H., & Dong, Y. (2013). Forecasting model of air pollution index based on BP neural network. *Environmental Science & Technology*, 36, 186–189.
- Bai, N.i., Khazaei, M., van Eeden, S. F., & Laher, I. (2007). The pharmacology of particulate matter air pollution-induced cardiovascular dysfunction. *Pharmacology & Therapeutics*, 113(1), 16–29.
- Bai, Y., Li, Y., Wang, X., Xie, J., & Li, C. (2016). Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric Pollution Research*, 7(3), 557–566.
- Bhunia, A. K., Konwer, A., Bhunia, A. K., Bhowmick, A., Roy, P. P., & Pal, U. (2019). Script identification in natural scene image and video frames using an attention based Convolutional-LSTM network. *Pattern Recognition*, 85, 172–184.
- Bognar, M., Lesjak, M., & Mlakar, P. (1993). A neural network-based method for short-term predictions of ambient SO₂ concentrations in highly polluted industrial areas of complex terrain. *Atmospheric Environment. Part B. Urban Atmosphere*, 27(2), 221–230.
- Brunekeef, B., & Holgate, S. T. (2002). Air pollution and health. *The Lancet*, 360(9341), 1233–1242.
- Cai, M., Yin, Y., & Xie, M. (2009). Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transportation Research Part D: Transport and Environment*, 14(1), 32–41.
- Calderón-Garcidueñas, L., Solt, A. C., Henríquez-Roldán, C., Torres-Jardón, R., Nuse, B., Herritt, L., ... Reed, W. (2008). Long-term air pollution exposure is associated with neuroinflammation, an altered innate immune response, disruption of the blood-brain barrier, ultrafine particulate deposition, and accumulation of amyloid β-42 and α-Synuclein in children and young adults. *Toxicologic Pathology*, 36(2), 289–310.
- Chen, S., Wang, J.-Q., & Zhang, H.-y. (2019). A hybrid PSO-SVM model based on clustering algorithm for short-term atmospheric pollutant concentration forecasting. *Technological Forecasting and Social Change*, 146, 41–54.
- Chen, W., Tang, H., & Zhao, H. (2015). Diurnal, weekly and monthly spatial variations of air pollutants and air quality of Beijing. *Atmospheric Environment*, 119, 21–34.
- Cheng, Y., Engling, G., He, K. B., Duan, F. K., Ma, Y. L., Du, Z. Y., Liu, J. M., Zheng, M., & Weber, R. J. (2013). Biomass burning contribution to Beijing aerosol. *Atmos. Chem. Phys.*, 13, 7765–7781.
- Cui, J., Lang, J., Chen, T., Mao, S., Cheng, S., Wang, Z., & Cheng, N. (2019). A framework for investigating the air quality variation characteristics based on the monitoring data: Case study for Beijing during 2013–2016. *Journal of Environmental Sciences*, 81, 225–237.
- Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2), 295–307.
- Esen, H., Esen, M., & Ozsolak, O. (2017). Modelling and experimental performance analysis of solar-assisted ground source heat pump system. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(1), 1–17.
- Esen, H., Inalli, M., Sengur, A., & Esen, M. (2008a). Artificial neural networks and adaptive neuro-fuzzy assessments for ground-coupled heat pump system. *Energy and Buildings*, 40(6), 1074–1083.
- Esen, H., Inalli, M., Sengur, A., & Esen, M. (2008b). Forecasting of a ground-coupled heat pump performance using neural networks with statistical data weighting pre-processing. *International Journal of Thermal Sciences*, 47(4), 431–441.
- Esen, H., Inalli, M., Sengur, A., & Esen, M. (2008c). Performance prediction of a ground-coupled heat pump system using artificial neural networks. *Expert Systems with Applications*, 35(4), 1940–1948.
- Esen, H., Ozgen, F., Esen, M., & Sengur, A. (2009). Artificial neural network and wavelet neural network approaches for modelling of a solar air heater. *Expert Systems with Applications*, 36(8), 11240–11248.
- Fang, C., Liu, H., Li, G., Sun, D., & Miao, Z. (2015). Estimating the impact of urbanization on air quality in China using spatial regression models. *Sustainability*, 7, 15570–15592.
- Fang, X., Duan, H., Hu, M., & Cai, J. (2019). The seasonal differential effects of meteorological parameters on atmospheric pollutants and the prediction model comparison: A case study of Shenzhen. *Environmental Pollution & Control*, 41, 541–546.
- Gocheva-Ilieva, S. G., Voynikova, D. S., Stoimenova, M. P., Ivanov, A. V., & Iliev, I. P. (2019). Regression trees modeling of time series for air pollution analysis and forecasting. *Neural Computing and Applications*, 31(12), 9023–9039.
- Hao, X., Zhang, G., & Ma, S. (2016). Deep learning. *International Journal of Semantic Computing*, 10(03), 417–439.
- Hochreiter, S. (2011). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 06, 107–116.
- Huang, C.-J., & Kuo, P.-H. (2018). A deep CNN-LSTM model for particulate matter (PM_{2.5}) forecasting in smart cities. *Sensors*, 18, 2220.
- Huang, J., Zhang, F., Du, Z., Liu, R., & Cao, X. (2019). Hourly concentration prediction of PM_{2.5} based on RNN-CNN ensemble deep learning model. *Journal of Zhejiang University(Science Edition)*, 46, 370–379.
- Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2018). LSTM fully convolutional networks for time series classification. *IEEE Access*, 6, 1662–1669.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.
- Kulmala, M. (2015). Atmospheric chemistry: China's choking cocktail. *Nature*, 526 (7574), 497–499.
- Kumar, A., & Goyal, P. (2011). Forecasting of daily air quality index in Delhi. *Science of The Total Environment*, 409(24), 5517–5523.
- Kumar, U., & Jain, V. K. (2010). ARIMA forecasting of ambient air pollutants (O₃, NO, NO₂ and CO). *Stochastic Environmental Research and Risk Assessment*, 24(5), 751–760.
- Kurt, A., & Oktay, A. B. (2010). Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert Systems with Applications*, 37 (12), 7986–7992.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Leong, W. C., Kelani, R. O., & Ahmad, Z. (2019). Prediction of air pollution index (API) using support vector machine (SVM). *Journal of Environmental. Chemical Engineering*, 8(3), 103208. <https://doi.org/10.1016/j.jece.2019.103208>
- Li, D., Jiang, Q., & Cao, Y. (2006). Clustering based on support vector regression model and its application in power system transient stability prediction. *Transactions of China Electrotechnical Society*, 21, 75–80.
- Li, L., Chen, C. H., Fu, J. S., Huang, C., Streets, D. G., Huang, H. Y., Zhang, G. F., Wang, Y. J., Jang, C. J., Wang, H. L., Chen, Y. R., & Fu, J. M. (2011). Air quality and emissions in the Yangtze River Delta, China. *Atmospheric Chemistry and Physics*, 11, 1621–1639.
- Li, T., Hua, M., & Wu, X. u. (2020). A Hybrid CNN-LSTM Model for Forecasting Particulate Matter (PM_{2.5}). *IEEE Access*, 8, 26933–26940.
- Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, 231, 997–1004.
- Liu, H., Long, Z., Duan, Z., & Shi, H. (2020). A new model using multiple feature clustering and neural networks for forecasting hourly PM_{2.5} concentrations, and its applications in China. *Engineering*.
- Muhammad Adnan, R., Chen, Z., Yuan, X., Kisi, O., El-Shafie, A., Kuriqi, A., & Ikram, M. (2020). Reference Evapotranspiration Modeling Using New Heuristic Methods. *Entropy*, 22, 547.
- Ning, M., Guan, J., Liu, P., Zhang, Z., & M. P. O Hare, G. (2019). GA-BP air quality evaluation method based on fuzzy theory. *Computers, Materials & Continua*, 58, 215–227.
- Pak, Unjin, Kim, Chungsong, Ryu, Unsook, Sok, Kyongjin, & Pak, Sungnam (2018). A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air Quality, Atmosphere & Health*, 11(8), 883–895.
- Prasad, Kanchan, Gorai, Amit Kumar, & Goyal, Pramila (2016). Development of ANFIS models for air quality forecasting and input optimization for reducing the computational cost and time. *Atmospheric Environment*, 128, 246–262.
- Quang, D., & Xie, X. (2016). DanQ: a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences. *Nucleic Acids Research*, 44, e107–e107.
- Ru, Yunian, Li, Bo, Liu, Jianbo, & Chai, Jianping (2018). An effective daily box office prediction model based on deep neural networks. *Cognitive Systems Research*, 52, 182–191.
- Rubal, & Kumar, Dinesh (2018). Evolving Differential evolution method with random forest for prediction of Air Pollution. *Procedia Computer Science*, 132, 824–833.
- Song, Shengli, Huang, Haitao, & Ruan, Tongxiao (2019). Abstractive text summarization using LSTM-CNN based deep learning. *Multimedia Tools and Applications*, 78(1), 857–875.
- Streets, David G., Fu, Joshua S., Jang, Carey J., Hao, Jiming, He, Kebin, Tang, Xiaoyan, ... Yu, Carolyne (2007). Air quality during the 2008 Beijing Olympic Games. *Atmospheric Environment*, 41(3), 480–492.
- Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1891–1898).
- Tian, Yulu, Jiang, Yuan, Liu, Qi, Xu, Dingxue, Zhao, Shoudong, He, Lihuan, ... Xu, Hui (2019). Temporal and spatial trends in air quality in Beijing. *Landscape and Urban Planning*, 185, 35–43.

- Tie, Xuexi, Madronich, Sasha, Li, GuoHui, Ying, Zhuming, Zhang, Renyi, Garcia, Agustin R., ... Liu, Yubao (2007). Characterizations of chemical oxidants in Mexico City: A regional chemical dynamical model (WRF-Chem) study. *Atmospheric Environment*, *41* (9), 1989–2008.
- Vlachogianni, A., Kassomenos, P., Karppinen, Ari, Karakitsios, S., & Kukkonen, Jaakko (2011). Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM10 in Athens and Helsinki. *Science of The Total Environment*, *409*(8), 1559–1571.
- Wang, Kejun, Qi, Xiaoxia, Liu, Hongda, & Song, Jiakang (2018). Deep belief network based k-means cluster approach for short-term wind power forecasting. *Energy*, *165*, 840–852.
- Wang, Ying, Zhuang, Guoshun, Tang, Aohan, Yuan, Hui, Sun, Yele, Chen, Shuang, & Zheng, Aihua (2005). The ion chemistry and the source of PM2.5 aerosol in Beijing. *Atmospheric Environment*, *39*(21), 3771–3784.
- Wang, Ying, Zhuang, Guoshun, Xu, Chang, & An, Zhisheng (2007). The air pollution caused by the burning of fireworks during the lantern festival in Beijing. *Atmospheric Environment*, *41*(2), 417–431.
- Wang, ZiFa, Li, Jie, Wang, Zhe, Yang, WenYi, Tang, Xiao, Ge, BaoZhu, ... Su, DeBin (2014). Modeling study of regional severe hazes over mid-eastern China in January 2013 and its implications on pollution prevention and control. *Science China Earth Sciences*, *57*(1), 3–13.
- Wu, C., Li, Q., Hou, J., Karimian, H., & Chen, G. (2018). PM2.5 concentration prediction using convolutional neural networks. *Science of Surveying and Mapping*, *43*, 68–75.
- Wu, Qunli, & Lin, Huaxing (2019). A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. *Science of The Total Environment*, *683*, 808–821.
- Xu, W., Tian, Y., Xiao, Y., Jiang, W., Tian, L., & Liu, J. (2017). Study on the spatial distribution characteristics and the drivers of AQI in North China. *Acta Scientiae Circumstantiae*, *37*, 3085–3096.
- Yang, Q., Yuan, Q., Li, T., Shen, H., & Zhang, L. (2017). The Relationships between PM2.5 and Meteorological Factors in China: Seasonal and Regional Variations. *International Journal of Environmental Research and Public Health*, *14*, 1510.
- Yun, G., He, Y., Jiang, Y., Dou, P., & Dai, S. (2019). PM2.5 spatiotemporal evolution and drivers in the Yangtze River delta between 2005 and 2015. *Atmosphere*, *10*, 55.
- Zhang, Hongliang, Wang, Yungang, Hu, Jianlin, Ying, Qi, & Hu, Xiao-Ming (2015). Relationships between meteorological parameters and criteria air pollutants in three megacities in China. *Environmental Research*, *140*, 242–254.
- Zhou, W., Wu, X., Ding, S., & Cheng, Y. (2020). Predictive analysis of the air quality indicators in the Yangtze River Delta in China: An application of a novel seasonal grey model. *Science of The Total Environment*, *141428*.
- Zhu, Suling, Lian, Xiuyuan, Liu, Haixia, Hu, Jianming, Wang, Yuanyuan, & Che, Jinxing (2017). Daily air quality index forecasting with hybrid models: A case in China. *Environmental Pollution*, *231*, 1232–1244.
- Zhu, Y., Li, Q., Hou, J., Fan, J., & Feng, X. (2016). Spatio-temporal modeling and prediction of PM2.5 concentration based on Bayesian method. *Science of Surveying and Mapping*, *41*, 44–48.