Journal of Urban Development and Management
https://www.acadlore.com/journals/JUDM

Spatio-Temporal Dynamics and Haze Agglomeration Analysis in the Beijing-Tianjin-Hebei Region: A WOA-LSTM Approach

Yuanyuan Wang[1], Zhuang Wu[1*], Jiaqi Du[1], Wanshu Fu[1], Yi Zhang[2]

[1] School of Management and Engineering, Capital University of Economics and Business, 100070 Beijing, China
[2] School of Environment, Education and Development, The University of Manchester, M13 9PL Manchester, UK

* Correspondence: Zhuang Wu (wuzhuang@cueb.edu.cn)

Received: 08-07-2023 **Revised:** 09-12-2023 **Accepted:** 09-18-2023

Citation: Y. Y. Wang, Z. Wu, J. Q. Du, W. S. Fu, and Y. Zhang, "Spatio-temporal dynamics and haze agglomeration analysis in the Beijing-Tianjin-Hebei region: A WOA-LSTM approach," *J. Urban Dev. Manag.*, vol. 2, no. 3, pp. 145–159, 2023. https://doi.org/10.56578/judm020304.

Abstract: The intensification of industrial and urban growth has precipitated a significant increase in atmospheric pollutant emissions, thereby exacerbating air quality deterioration. This phenomenon is particularly pronounced within the Beijing-Tianjin-Hebei urban agglomeration, where haze events have manifested with increasing frequency. Prior investigations have predominantly concentrated on temporal trends, often overlooking the critical impact of geographical factors on haze development. This research delves into the spatio-temporal distribution traits of haze within the Beijing-Tianjin-Hebei region, employing a Whale Optimization Algorithm-Long Short-Term Memory (WOA-LSTM) model. Findings indicate a pronounced spatial concentration of urban air pollution in the region's southern sector. In terms of temporal distribution, the Air Quality Index (AQI) demonstrates distinct seasonal fluctuations, with the highest pollution levels recorded in winter and notably lower levels observed during summer. The study's innovation lies in the development of a WOA-LSTM model, which not only predicts the AQI - a comprehensive haze pollution index - but also offers early warnings pertinent to public travel. By integrating extensive datasets and applying advanced analytical techniques, the study contributes significantly to understanding the complex interplay between urban dynamics and haze distribution. The research underscores the necessity for regional policies tailored to specific spatiotemporal characteristics, thereby aiding in effective air quality management and mitigation strategies within urban agglomerations.

Keywords: Beijing-Tianjin-Hebei urban agglomeration; Spatio-temporal distribution; Spatial agglomeration

1 Introduction

The deleterious effects of haze are well-documented and far-reaching. In the 2009 report *Comprehensive Scientific Assessment Report on Air Particles* by the United States Environmental Protection Agency, it was highlighted that inhalable particulate matter in the atmosphere is capable of adsorbing a multitude of harmful carcinogens and genotoxic mutagens. This results in adverse health outcomes, including increased mortality rates, exacerbation of chronic diseases, and deterioration in respiratory [1] and cardiac [2] health. Furthermore, changes in lung function and structure [3], impacts on fertility [4], and alterations in human immune structure [5] have been observed. Beyond its impact on human health, haze incurs increased medical expenses due to the aforementioned health issues and directly affects traffic safety. Poor air quality and reduced visibility contribute to frequent traffic accidents, leading to congestion, inconvenience, and economic losses [6]. The Beijing-Tianjin-Hebei region, as China's economic core and a barometer of national competitiveness, is simultaneously plagued by some of the most severe air pollution in the country. The air quality issue in this region has garnered widespread attention from government entities, academic circles, the public, and media outlets.

This study focuses on the urban conglomerations comprising 13 cities in the Beijing-Tianjin-Hebei region. It aims to delineate the spatial characteristics of haze distribution, understand the temporal patterns and seasonal variations of haze formation, and provide timely predictions and warnings of severe pollution events.

The structure of this paper is organized as follows: Section 2 presents a review of pertinent literature, Section 3 discusses the research methodology and data preprocessing techniques employed, Section 4 details the model construction, Section 5 analyzes the results, and Section 6 offers concluding remarks.

2 Literature Review

The investigation into air pollution's determinants has attracted global scholarly attention. In the 1960s, the inception of air quality prediction research was observed in developed Western nations. Approximately two decades later, Japan initiated the practical application of theoretical research in this field, marking a significant progression. Post-1990s, the incorporation of neural network theory in predicting air pollutant concentration emerged, noted for its efficacy in handling multi-factorial, nonlinear, and random research subjects, yielding notable results. In the 21st century, advancements in neural network theory have significantly bolstered the study of air pollutant concentration prediction. Dun et al. [7] introduced a dynamic graph convolutional neural network, predicated on spatio-temporal correlation, to predict air quality, demonstrating its effectiveness and superiority. Aarthi et al. [8] developed a deep learning model termed "Balanced Spider Monkey Optimization with Bi-LSTM" for sustainable air quality prediction, showing enhanced accuracy and robustness compared to traditional methods. Chen et al. [9] presented a prediction model utilizing the BP neural network, focusing on haze concentration prediction influenced by meteorological factors. This model employs meteorological factors as inputs, processing them through a multi-layer neural network optimized via backpropagation algorithms. Xu et al. [10] conducted an air quality study in Fujian Province, China, employing a B-spline functional linear model. This model elucidated the relationship between pollutant concentration, temporal factors, and meteorological conditions, providing a scientific basis for targeted air quality improvement strategies. Xiong et al. [11] undertook a comprehensive study on the interplay between air pollutant concentrations and meteorological factors in Chengdu, China. Their research revealed the varying impacts of meteorological factors on air quality across different time scales, offering valuable insights for the scientific prediction and management of air quality. Prado-Rujas et al. [12] proposed a sensor-agnostic spatio-temporal air quality forecasting framework based on deep learning. This model, capable of handling multivariate, high-dimensional data and capturing spatial-temporal correlations, proved to be more accurate in predicting future air quality than existing models.

Building on the insights gained from previous research, it is observed that a predominant focus has been placed on time series analysis, often overlooking the critical influence of geographical location on haze distribution. In addressing this research gap and the urgent requirement for effective haze management in the Beijing-Tianjin-Hebei urban agglomeration, this study embarks on an exploration of the spatial distribution characteristics of haze within this region. The approach encompasses both global and local perspectives. Central to this study is the proposal of the WOA-LSTM model. This model is designed to forecast the temporal sequence of the AQI, recognized as a comprehensive evaluation index for haze pollution. Furthermore, it facilitates the issuance of severe air pollution warnings, triggered when AQI levels surpass thresholds indicative of significant pollution. Upon reaching these critical levels, public notifications are set to be disseminated, advising on reduced travel to mitigate exposure. The outcomes of this research are expected to contribute substantially to air pollution control strategies in the Beijing-Tianjin-Hebei urban agglomeration, tailored to its unique spatio-temporal characteristics. The implications of this study extend beyond local environmental and public health improvements, offering invaluable insights into the spatio-temporal analysis and management of air pollution applicable to other urban contexts.

3 Research Methods and Data Preprocessing

3.1 Research Methods

3.1.1 Spatial autocorrelation analysis

Tobler [13] articulated what is known as "the first law of geography," stating that "everything is related to something else, but things near are more relevant than things far away." Spatial autocorrelation, in essence, involves examining the extent of spatial correlation between a particular spatial unit and its surrounding units. This is achieved through statistical methods, which are employed to calculate and analyze the spatial autocorrelation, thereby elucidating the characteristics of the spatial distribution of these units. In this study, both global and local autocorrelation indices are utilized to assess the aggregation or dispersion of spatial feature attribute values. Global autocorrelation analysis primarily focuses on the overall distribution of a specific phenomenon within the research area. Its objective is to ascertain whether this phenomenon exhibits spatial correlation and agglomeration. To quantitatively appraise the spatial dependence of air pollution in adjacent areas, the Global Moran's I statistic is applied in this research. This statistic serves to delineate the global spatial correlation of haze, with its specific definition presented as follows:

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

where, I is the the Moran index, variance $S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$, average value $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, x_i is the observation value of the study city, n is the number of cities and takes 13, w_{ij} represents elements in the weight

matrix. The value range of I is [-1,1], a value of I above 0 implies a positive spatial correlation for the studied attribute, with proximity to 1 indicating a stronger positive spatial correlation.

Local spatial autocorrelation analysis provides a more granular view, enabling the examination of varying spatial aggregations at distinct spatial locations. This approach is instrumental in identifying hotspots or local spatial aggregations, uncovering areas of local instability, and revealing spatial heterogeneity among data. This heterogeneity is characterized by differing levels of spatial autocorrelation across various regions. For local spatial autocorrelation analysis, this study employs the Local Moran's I statistic and Moran's I scatter plot. Additionally, the Local Indicators of Spatial Association (LISA) are utilized for significance analysis. The computation formula for the Local Moran's I statistic is:

$$I_i = z_i \sum_{j \neq i}^{n} W_{ij} z_j \tag{2}$$

where, I_i is the local Moran index, analogous in meaning to the Moran index, falls within the range of [-1, 1]. w_{ij} represents elements in the weight matrix, and z_i signifies the standardization of the attribute under examination.

3.1.2 Spatial autocorrelation analysis

LSTM, a neural network architecture first introduced in 1997, addresses the limitations commonly encountered in traditional Recurrent Neural Network (RNN), such as vanishing and exploding gradients. The pivotal enhancement in LSTM, compared to its RNN predecessor, is the incorporation of memory units within the network architecture. These units are adept at retaining and storing past information. Moreover, LSTM demonstrates a robust learning capacity, making it exceptionally suited for time series tasks. It excels in discerning underlying patterns in sample data and adeptly manages nonlinear challenges, such as air quality prediction. Therefore, LSTM has been selected as the foundational model for predicting air quality time series in this study.

The accuracy of LSTM model predictions is heavily contingent on the selection of key parameters. It is crucial to determine these parameters with precision. The WOA is favored over other optimization algorithms due to its simpler principles, fewer parameter requirements, and superior global search capabilities. These attributes render WOA particularly effective for complex, nonlinear problems, making it an optimal choice for parameter selection in LSTM models.

3.1.3 WOA

The WOA, introduced by Mirjalili and Lewis [14], is a novel swarm intelligence algorithm. Characterized by its simulation of the hunting behavior of whales, the WOA algorithm stands out with its simplicity, relatively few parameters, and powerful global search capabilities.

The operational mechanism of WOA encompasses several key stages:

(1) Initialization: Each whale is assigned an initial position, thereby generating an initial population.

(2) Search: In this phase, whales explore the search space, guided by predefined rules that mimic behaviors like encirclement, pursuit, and attack on prey.

(3) Evaluation: As whales relocate, their fitness values are recalculated. When a whale's current fitness value exceeds the previous one, it is regarded as the new optimal solution.

(4) Update: Following the movement and evaluation of all whales, their positions are updated, and the process is reiterated.

(5) Iteration: The algorithm iteratively seeks the optimal solution, employing multiple iterations as necessary.

The WOA algorithm's merits include ease of operation, minimal requirements for parameter tuning, and an enhanced capacity to avoid local optima. This makes it highly efficient in identifying optimal solutions across a diverse array of optimization problems. Furthermore, the algorithm demonstrates commendable convergence and stability, even in addressing complex problems.

3.2 Data Preprocessing

For this research, a comprehensive dataset has been collated, comprising conventional meteorological observation data, daily meteorological reports, PM2.5 mass concentration data, and surface weather maps. These datasets have been sourced from esteemed institutions and platforms, including the National Oceanic and Atmospheric Administration (NOAA) of the United States, the National Aeronautics and Space Administration (NASA) website, and the Tianjin Environmental Statistics Bureau.

The data utilized in this study are characterized by the following attributes:

The classification of haze pollution levels in this study is based on the concentration of particulate matter in the air, typically denoted by the AQI. Accordingly, AQI, as presented in Table 1, is chosen as the target variable for this research. The classification and encoding of AQI are further detailed in Table 2:

Table 1. Data characteristics and annotations

Feature Name	Annotations
Time	Year/month/day, for example: 2014/1/1
Quality Grade	Optimal, Good, Light pollution, Moderate pollution, Heavy pollution, Severe pollution
City	The Beijing-Tianjin-Hebei Urban Agglomeration is the "Capital Economic Circle" of China, which 13 cities including Beijing, Tianjin, Baoding, Tangshan, Langfang, Shijiazhuang, Qinhuangdao, Zhangjiakou, Chengde, Cangzhou, Hengshui, Xingtai, and Handan in Hebei Province.
PM2.5	PM2.5 fine particles, also known as fine particles, fine particles, or PM2.5. Fine particulate matter refers to particles with an aerodynamic equivalent diameter of 2.5 micrometers or less in the ambient air. It can be suspended in the air for a long time, and the higher its concentration in the air, the more serious the air pollution.
PM10	PM10 inhalable particulate matter refers to particulate matter suspended in the air with an aerodynamic equivalent diameter of ≤ 10 microns.
SO_2	Sulfur dioxide (SO_2) is a common and important atmospheric pollutant, a colorless and irritating gas.
NO_2	The chemical formula NO_2 of nitrogen dioxide is a brownish-red, highly active gaseous substance. Nitrogen dioxide plays an important role in the formation of ozone. The anthropogenic nitrogen dioxide mainly comes from the release of high-temperature combustion processes, such as emissions from motor vehicles and power plants.
CO	Carbon monoxide (CO) is a product of incomplete combustion of carbonaceous substances such as coal and petroleum. It is a colorless, odorless, and nonirritating toxic gas that is almost insoluble in water. It is difficult to react with other substances in the air and can remain in the atmosphere for 2 to 3 years. If local pollution is severe, it may pose a certain risk to the health of the population.
O_3	Ozone is an allotrope of oxygen, with the chemical formula O_3. Ozone exists in the atmosphere and can absorb harmful short-wave (below 30 nm) light from sunlight, preventing it from reaching the ground and protecting organisms from UV damage.
AQI	The air quality index is a dimensionless relative value that comprehensively represents the degree of air pollution or the level of air quality.

Table 2. AQI numerical level classification and coding

AQI	Air Quality Index Level Coding	Air Quality Index
0-50	1	Optimal
51-100	2	Good
101-150	3	Light pollution
151-200	4	Moderate pollution
201-300	5	Heavy pollution
$>$300	6	Severe pollution

Encoding the AQI streamlines the process for subsequent machine learning training and prediction. This study involves normalizing key influencing factors such as PM2.5, PM10, SO_2, NO_2, CO, and O_3, transforming dimensional data into dimensionless forms. Such normalization is crucial for enabling the comparison and weighting of indicators across different units or magnitudes, and it simplifies the computational process. The formula applied for normalization in this research is presented in Eq. (3):

$$X_{new\_i} = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{3}$$

4 Model Building

The LSTM model is adept at capturing temporal dependencies in time series data, making it highly suitable for analyzing AQI values which are often influenced by historical data. This characteristic allows LSTM to effectively learn and utilize these dependency relationships. The unit structure of LSTM is illustrated in Figure 1.

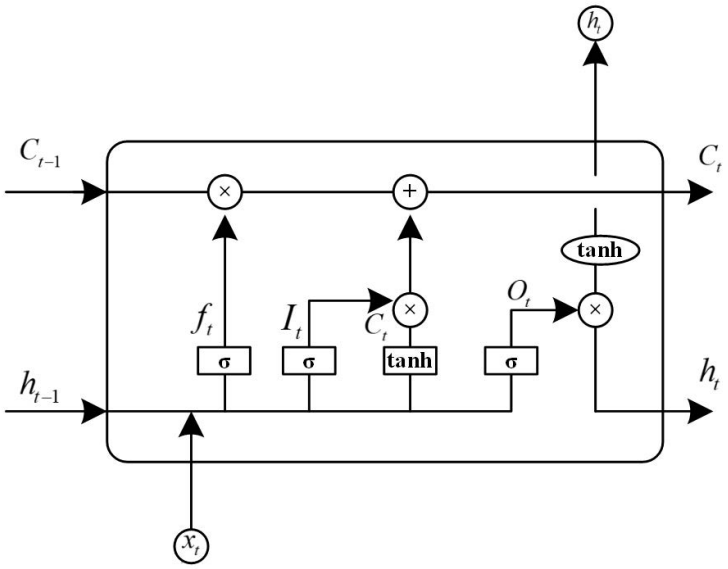


Figure 1. LSTM network unit structure diagram

The neural unit structure of the LSTM network is composed of three key components: the input gate, output gate, and forgetting gate. Where x_t is the input time series data at time t. The computation formulae for each memory unit in the LSTM are as follows.

The input of the forgetting gate mainly consists of two parts, namely the input of the current time and the output of the previous time h_{t-1}. The information to be discarded is determined by the function σ, as shown in the Eq. (4):

$$f_t = \sigma\left(W_f \times [h_{t-1}, x_t] + b_f\right) \tag{4}$$

where, W represents the weight and b represents the bias vector.

Following the forgetting gate, the input gate primarily generates the output vector I_t and the current time state C_t, as described in Eqs. (5) and (6):

$$I_t = \sigma\left(W_i \times [h_{t-1}, x_t] + b_i\right) \tag{5}$$

$$C_t = \tanh\left(W_C \times [h_{t-1}, x_t] + b_C\right) \tag{6}$$

The state of C_t is updated, as shown in Eq. (7):

$$C_t = f_t C_{t-1} + I_t C_t \tag{7}$$

The output gate takes the previous hidden state h_{t-1} as its input. The new output O_t is computed and then C_t is mapped to obtain h_t, as outlined in Eqs. (8) and (9):

$$O_t = \sigma\left(W_o \times [h_{t-1}, x_t] + b_o\right) \tag{8}$$

$$h_t = O_t \times \tanh\left(C_t\right) \tag{9}$$

The LSTM model developed in this study employs the WOA for optimizing parameters such as the number of hidden layer nodes, learning rate, iteration epochs, and batch size. A fitness function based on the target variable AQI is designed, as depicted in Eq. (10). This fitness function, centered on the mean square deviation, aims for minimization, with smaller values indicating better model performance.

$$R = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{\text{pre }\_i} - y_i\right)^2}{n}} \tag{10}$$

The focus of the WOA algorithm is on the basic principles of the iterative update process, as follows:

(1) Update location:

The new position X_{new} of a whale individual can be calculated from the current position X and the positions X_{rand1} and X_{rand2} of two random whale individuals to obtain the Eq. (11):

$$X_{\text{new}} = X - A \times D \tag{11}$$

where, A is a gradually decreasing linear decreasing parameter, and D is the distance from the current position to X_{rand1}, calculated by the Eq. (12):

$$D = \left|X_{\text{rand1}} - X\right| \tag{12}$$

(2) Update speed:

The speed V_{new} of whales is updated as Eq. (13):

$$V_{\text{new}} = A \times V + C\left|X_{\text{rand 1}} - X_{\text{rand 2}}\right| \tag{13}$$

where, *C* is a random weight.

(3) Update optimal solution:

The process involves comparing the current solution with the historically optimal solution, and subsequently updating the optimal solution as per Eq. (14):

$$X_{\text{best}} = \min\left(X_{best}, X\right) \tag{14}$$

These formulas are used to update the position and velocity of whale individuals during the iteration process, gradually optimizing the solution to approach the optimal solution.

5 Result Analysis

In the era of rapid internet industry expansion, numerous sectors are inundated with vast data quantities, leading to diverse and sometimes conflicting methodologies in data generation and interpretation. This proliferation of data, while offering an abundance of information, also leads to a considerable volume that remains unanalyzed, effectively resulting in substantial resource wastage. To counteract this issue, data visualization emerges as a critical tool. In the current study, data visualization techniques, specifically bar charts and scatter plots, were utilized to analyze the temporal distribution of haze. This approach provided clearer insights into the patterns of haze occurrence over time. Additionally, GeoDa software was employed for the processing of monitoring data and for conducting comprehensive spatial evaluations of air quality within the Beijing-Tianjin-Hebei urban agglomeration. By depicting different pollution levels through varying color intensities, an intuitive visual representation of air quality was achieved. Consequently, this study incorporated both temporal and spatial dimensions in the visualization and analysis of haze distribution, offering a more holistic understanding of the phenomena.

5.1 Temporal and Spatial Distribution of Haze

5.1.1 AQI interannual variation

The distribution of air pollutants within the region is subject to influences from geographical, meteorological, and economic factors. These elements render the temporal and spatial distribution characteristics of pollutants notably similar. The AQI serves as a pivotal indicator, reflecting the air's cleanliness or level of pollution. It encompasses a range of pollution factors, thus providing a comprehensive evaluation of air quality.

In this context, AQI is utilized as a representative metric to illustrate the temporal and spatial distribution characteristics of haze in the Beijing-Tianjin-Hebei urban agglomeration. The interannual mean variation of the AQI in this region, spanning from 2014 to 2017, is depicted in Figure 2.

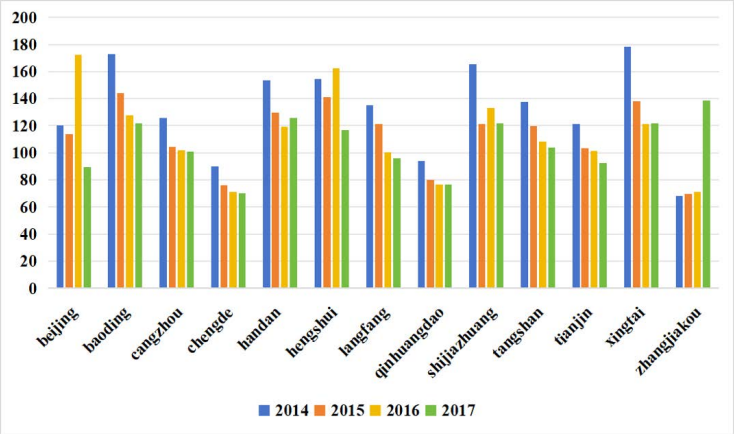


Figure 2. Average AQI histogram of the Beijing-Tianjin-Hebei urban agglomeration from 2014 to 2017

As illustrated in Figure 2, a clear trend emerges in the average AQI values across various cities in the Beijing-Tianjin-Hebei urban agglomeration. Cities such as Baoding, Cangzhou, Chengde, Langfang, Qinhuangdao, Tangshan, Tianjin, and Xingtai demonstrate a consistent decrease in their AQI levels. In contrast, other cities exhibit some variability in their AQI trends. Notably, Zhangjiakou experienced a marked increase in AQI values in 2017, an anomaly in the general downward pattern. Further elucidating these trends, Figure 3 presents a line chart that analyzes the overarching decline in AQI levels across the region.

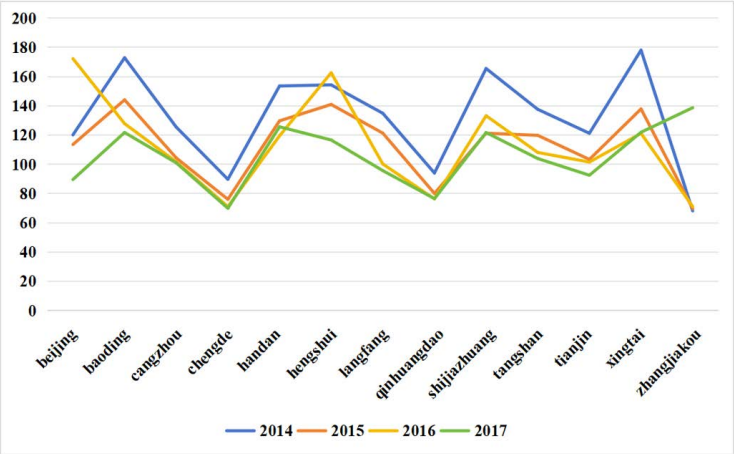


Figure 3. Line chart of average AQI in the Beijing-Tianjin-Hebei urban agglomeration from 2014 to 2017

As depicted in Figure 3, the trend line for the year 2017 is noticeably positioned below those of the previous years. This pattern indicates that, despite individual fluctuations across various cities, the Beijing-Tianjin-Hebei urban agglomeration as a whole is experiencing a declining trend in air pollution. This observation suggests that measures implemented for air pollution control are having a positive impact. However, it is imperative to acknowledge the significant and unexpected increase in AQI observed in Zhangjiakou during 2017.

5.1.2 AQI daily changes

To delve deeper into the daily distribution and fluctuations of AQI in the Beijing-Tianjin-Hebei urban agglomeration, four cities were selected for a detailed analysis. These include Chengde City, representing northern cities generally characterized by lower AQI indices; Beijing and Langfang City, representing transitional zone cities; and Xingtai City, typifying southern cities with generally higher AQI indices. Scatter plots were created to visualize the AQI distribution for each of these cities over the years 2014 to 2017, as illustrated in Figure 4, Figure 5, Figure 6 and Figure 7.

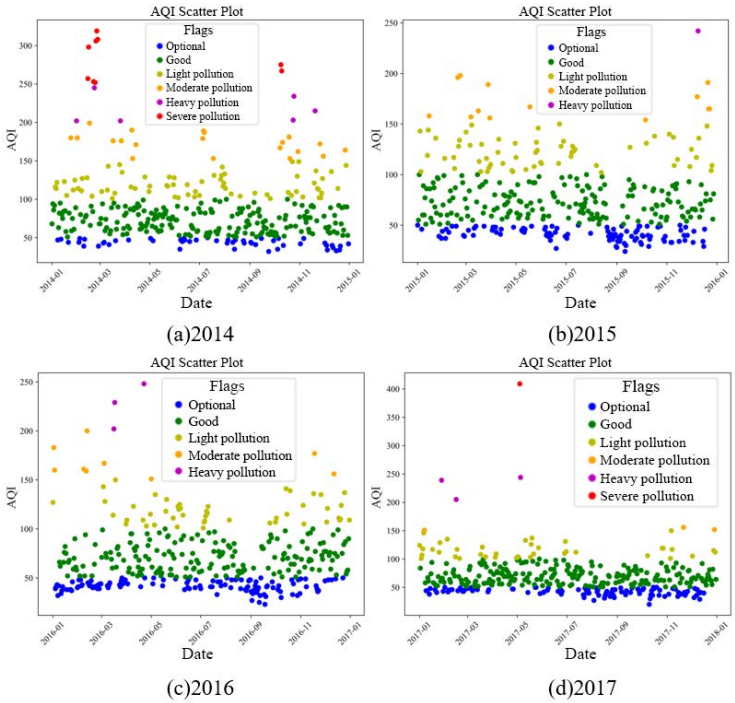


Figure 4. Scatter plot of AQI time distribution in Chengde City from 2014 to 2017

Analyzing Figure 4 reveals notable improvements in the air quality of Chengde City from 2014 to 2017. In 2014, Chengde experienced several instances of severe pollution as indicated by the AQI readings. However, a marked shift is observed in the subsequent years. There were no occurrences of severe pollution in both 2015 and 2016, and only a single instance in 2017. While the frequency of severe pollution days did not show a significant change, remaining at approximately 3 to 4 days of severe pollution haze annually, the instances of moderate pollution have notably decreased. This trend suggests that the environmental governance measures implemented in recent years have had a substantial positive impact on reducing moderate pollution levels in Chengde City.

When contrasting Figure 5 with Figure 4, it becomes evident that the pollution levels in Beijing are markedly higher compared to those in Chengde. Despite a notable reduction in the frequency of moderate pollution events in Beijing, as indicated by the AQI readings, occurrences of severe pollution are still prominent, particularly from December to February. This pattern in Beijing points to a seasonal cyclical trend, where the winter months exhibit a heightened intensity of pollution, manifesting as severe or extremely severe air quality conditions.

An analysis comparing Figure 6 (Langfang City) with Figure 5 (Beijing) reveals similarities in the temporal distribution of AQI between these two cities. Both cities experience periods of severe pollution, indicating comparable pollution patterns. However, a notable distinction is observed in the frequency of days classified as 'excellent' in terms of AQI levels. Langfang City exhibits this 'excellent' AQI category less frequently compared to Beijing, suggesting a disparity in the occurrence of days with particularly low pollution levels.

Figure 7 illustrates that the haze pollution in Xingtai City was alarmingly severe in 2014, with moderate, severe, and very severe pollution levels persisting throughout the year. However, from 2015 to 2017, notable progress in haze pollution control was observed. While instances of the pollution level being classified as 'excellent' were rare, there was a consistent presence of days with mild pollution and good AQI levels. This stability indicates effective management of pollution levels. Nonetheless, it remains crucial to reduce the instances of critical severe pollution (AQI values around 200) to moderate levels as much as possible.

The visual analysis of the daily AQI scatter plot reveals that AQI pollution level changes exhibit seasonal periodic patterns. Consequently, AQI is segmented into seasonal time nodes for further analysis, allowing for a spatial

distribution mapping and examination of these patterns.

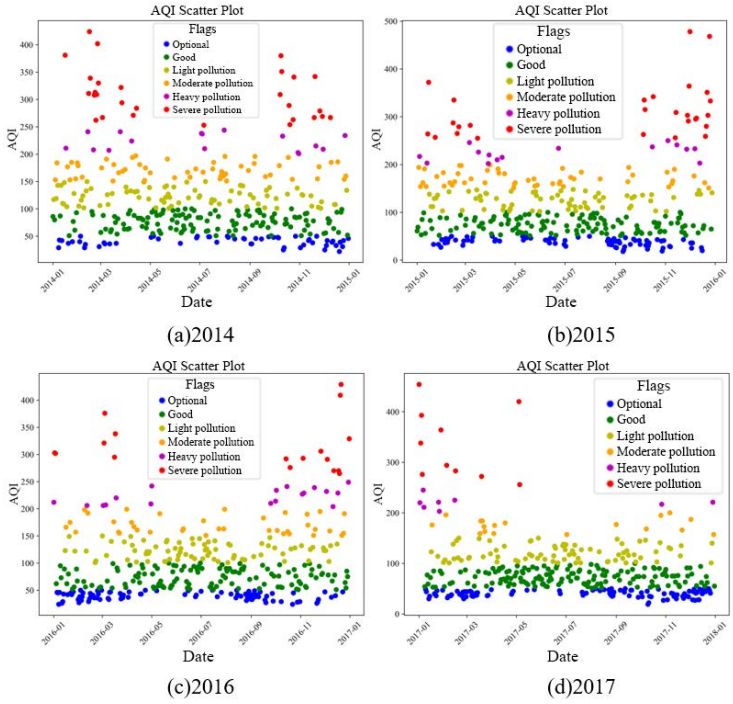


Figure 5. Scatter plot of AQI time distribution in Beijing from 2014 to 2017

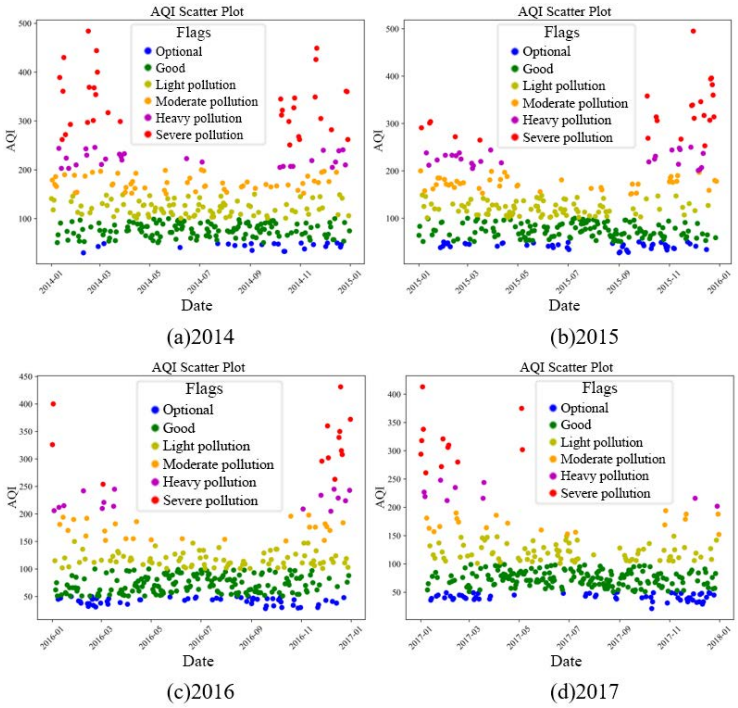


Figure 6. Scatter plot of AQI time distribution in Langfang City from 2014 to 2017

5.1.3 Seasonal distribution characteristics of AQI

To delve deeper into the spatial distribution of air quality in the Beijing-Tianjin-Hebei region and to investigate its underlying causes, this study conducts a seasonal analysis of the AQI index. The seasons are categorized as follows: winter (December of the previous year to February), spring (March to May), summer (June to August), and autumn (September to November). Seasonal AQI distribution maps are drawn in the order of spring, summer,

autumn, and winter. The analysis focuses on the seasonal distribution of the AQI index in the Beijing-Tianjin-Hebei urban agglomeration for the year 2017, as exemplified in Figure 8.

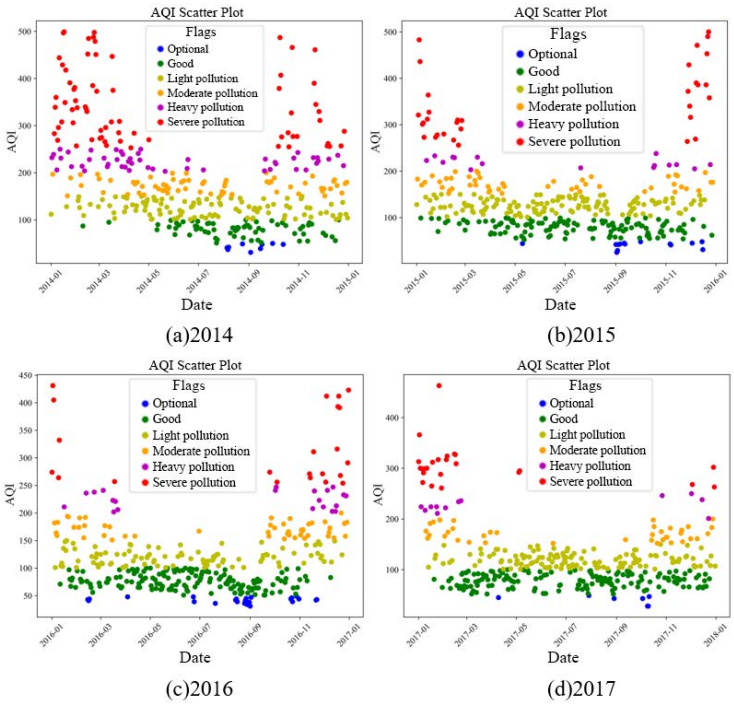


Figure 7. Scatter plot of AQI time distribution in Xingtai City from 2014 to 2017

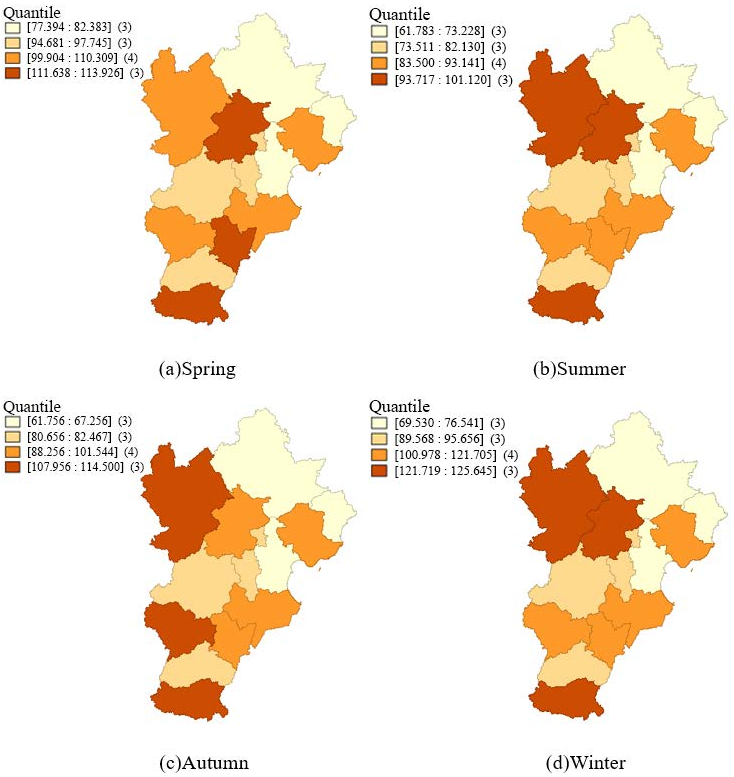


Figure 8. Seasonal AQI distribution map of the Beijing-Tianjin-Hebei urban agglomeration in 2017

Figure 8 presents a composite analysis of the AQI mean values and geographical locations for each quarter in 2017, categorized into four distinct groups. The map utilizes darker shades to represent higher AQI values. Notably, the AQI classification threshold for winter is markedly higher than in other seasons. This is particularly evident in

Beijing and Zhangjiakou at the center and Handan at the bottom, where the AQI indices are higher than in other cities, leading to severe pollution. The clustering results suggest a strong correlation with geographical location, exhibiting patterns of high-high and low-low clustering.

The seasonal variations in AQI distribution within the Beijing-Tianjin-Hebei region are largely attributable to meteorological conditions. During winter, factors such as sustained low winds, weak ground pressure, and diminished air mobility reduce the atmosphere's capacity to disperse pollutants. This results in weaker atmospheric diffusion capabilities compared to summer, exacerbating air pollution in the colder months.

5.1.4 Global spatial auto-correlation analysis

The Moran's I method serves as a widely recognized statistical approach for assessing spatial autocorrelation. This involves calculating the Moran's I Index value and its corresponding z-score. A z-score lying outside the range of -1.96 to 1.96 is considered significant. Specifically, a positive z-score exceeding 1.96 indicates a clustered distribution. This study computes the Global Moran's I random distribution map, utilizing the AQI values from the Beijing-Tianjin-Hebei urban agglomeration during the winter of 2017, as depicted in Figure 9.

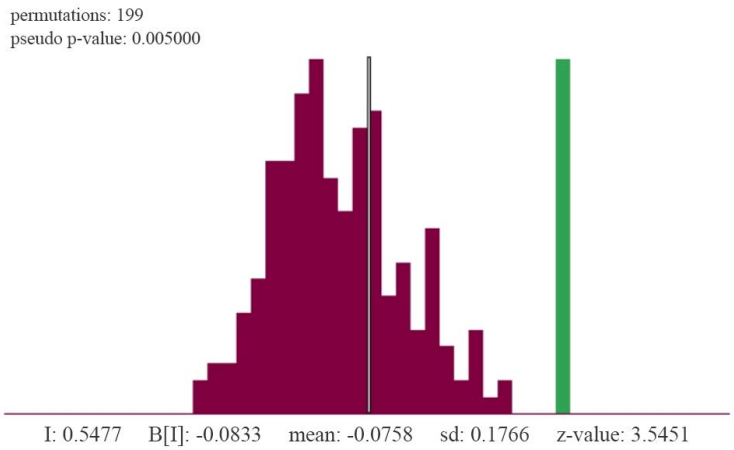


Figure 9. Global Moran's I random distribution of AQI in winter 2017

Analyzing Figure 9 reveals that the global Moran's index for AQI in the winter of 2017 was 0.5477, accompanied by a Z score of 3.5451 and a probability of 0.005. This statistically significant result, at the 1% level, indicates a high degree of confidence (99%) that the AQI values were spatially clustered during this period. Such a positive global Moran's I underscores the presence of substantial spatial autocorrelation in AQI distribution across the Beijing-Tianjin-Hebei urban agglomeration in the winter of 2017. Following this observation, the focus will now shift to a local clustering analysis.

5.1.5 Local spatial auto-correlation analysis

In the context of Moran's I scatter plot, the origin represents the global Moran index for the assessed year. The spatial significance of clustering is inferred from the distance of sample points from this origin within the plot. Greater distances from the origin imply higher levels of clustering significance, indicating more pronounced spatial autocorrelation within the data.

Figure 10 utilizes four quadrants to delineate the relationships between a region and its neighboring areas in terms of their AQI values. The quadrants are defined as follows:

- The first quadrant (HH) represents an area with high AQI values surrounded by areas with similarly high AQI values.
- The second quadrant (LH) indicates an area with low AQI values, yet surrounded by areas with high AQI values.
- The third quadrant (LL) signifies an area with low AQI values surrounded by areas with also low AQI values.
- The fourth quadrant (HL) describes an area with high AQI values but surrounded by areas with low AQI values.

While the HH and LL distribution patterns are prominently observable in the plot, identifying specific cities based on these patterns alone is challenging. To enhance the understanding of these spatial relationships and to correlate them with geographical locations more effectively, a LISA cluster distribution map is utilized. This map, shown in Figure 11, provides a more detailed and geographically contextualized view of the AQI clustering patterns within the Beijing-Tianjin-Hebei urban agglomeration.

Figure 11 provides a clear visualization of the spatial clustering of AQI values in various cities within the Beijing-Tianjin-Hebei urban agglomeration. Chengde City and Beijing exhibit the LL clustering phenomenon, indicating lower AQI values within these areas and their neighboring regions. Conversely, Shijiazhuang, Hengshui

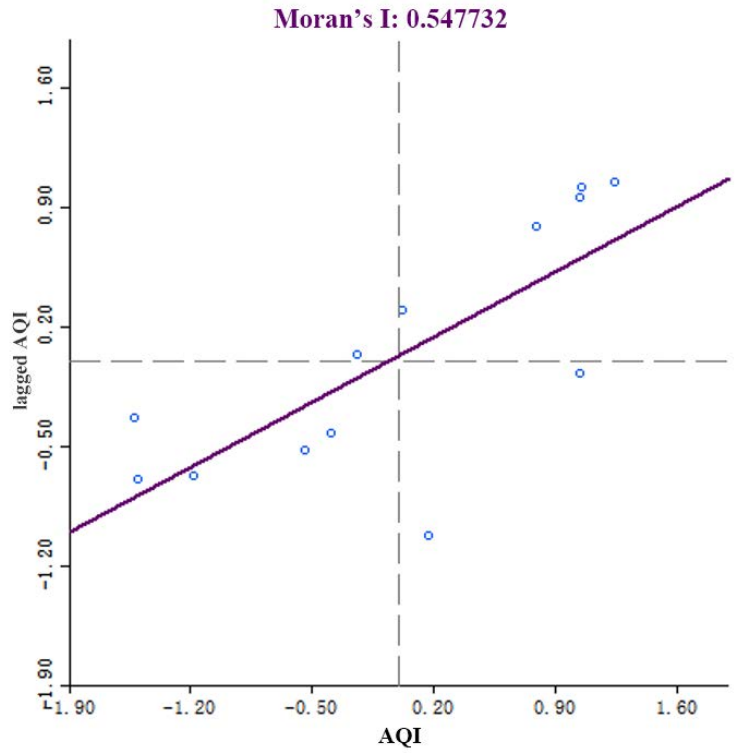
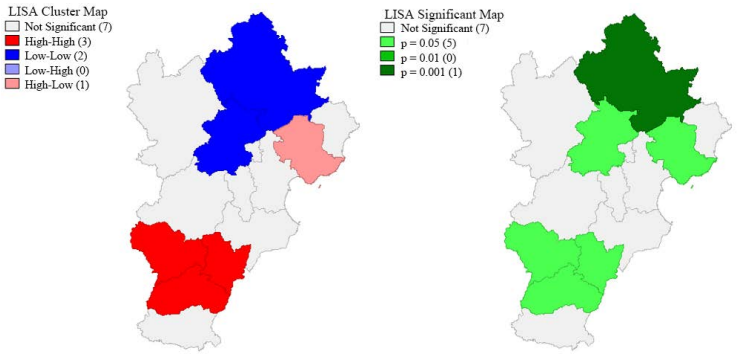


Figure 10. Moran's I scatter chart of AQI in Winter 2017



(a) LISA Cluster Map (b) LISA Signidfiance Map

Figure 11. LISA cluster distribution in winter 2017

City, and Xingtai display HH clustering, characterized by higher AQI values in both these cities and their surrounding areas. Notably, the AQI distribution in Chengde City is significant at the 0.1% level, whereas in Beijing, Tangshan, Shijiazhuang, Hengshui, and Xingtai, it is significant at the 5% level. This pattern reveals a pronounced HH and LL aggregation in spatial distribution, underscoring the significance of these phenomena. Consequently, effective air pollution governance should not only focus on individual administrative regions but also consider cooperative strategies with clustered areas, fostering joint efforts to enhance the impact of pollution control measures.

The differences in AQI spatial distribution are intimately linked to factors such as industrial structure [15], energy structure [16], and geographical location. In the southern region of Beijing-Tianjin-Hebei, heavy industries like steel, chemicals, and building materials predominate, significantly impacting air quality. Additionally, the region's reliance on coal as a primary energy source contributes to substantial emissions of pollutants like sulfur dioxide and nitrogen oxides, exacerbating air quality issues. The topography of the southern region, characterized by mountains, hills, and plains, further complicates matters. Located at the eastern foot of the Taihang Mountains, this area experiences limited air mobility, leading to the accumulation of pollutants and consequent declines in air quality.

5.2 AQI prediction based on WOA-LSTM model

Utilizing the WOA-LSTM model, the study trains data from the Beijing-Tianjin-Hebei urban agglomeration on an annual basis. The iterative mean square deviation chart, segmented by year, is presented in Figure 12.

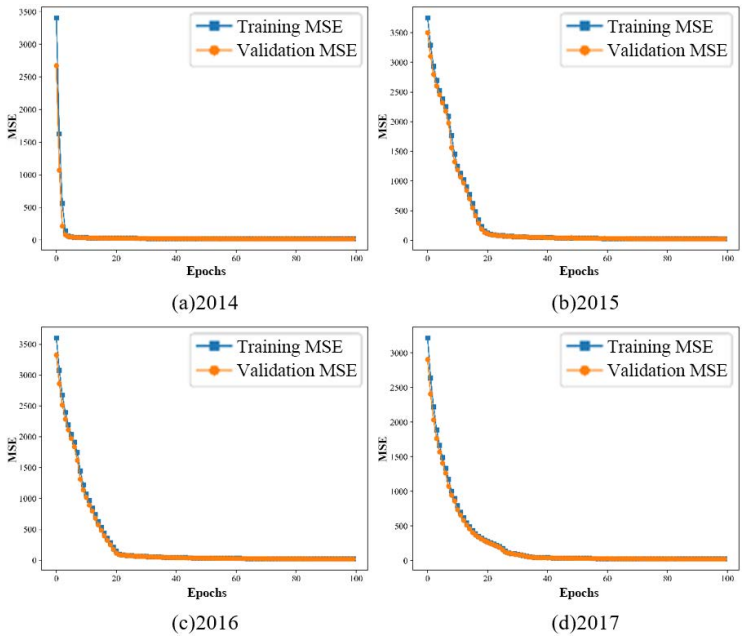


Figure 12. Iterative chart of AQI mean square error for WOA-LSTM model from 2014 to 2017

The training result graph of the mean square deviation, as depicted in Figure 12, demonstrates that the prediction accuracy of the WOA-LSTM model stabilizes at a satisfactory level after 20 iterations. This indicates the model's robust predictive capability. To further evaluate the effectiveness of the model, a comparison chart was created juxtaposing the AQI prediction results for randomly selected dates in Beijing in 2017 against the actual AQI values.

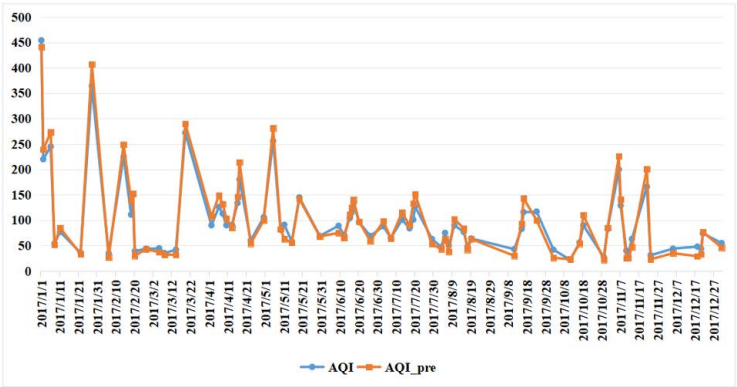


Figure 13. Comparison of partial AQI predictions in Beijing in 2017

The comparison, illustrated in Figure 13, reveals a high degree of congruence between the predicted and actual AQI values. Both sets of values not only align closely but also fall within the same pollution level range, validating the accuracy of the WOA-LSTM model in predicting AQI. Such precise predictions are crucial for implementing early warning systems and providing timely advisories, thereby enhancing the effectiveness of air quality management and public health measures.

6 Style Settings

This research endeavor focused on analyzing the spatiotemporal distribution characteristics of haze AQI within the Beijing-Tianjin-Hebei urban agglomeration and employed numerical methods for its prediction. The investigation embraced both temporal and spatial dimensions. Temporally, the study analyzed annual averages, daily variations, and seasonal periodicity. Spatially, it utilized global and local spatial autocorrelation analyses. The key findings are summarized as follows:

(1) Temporally, the overall annual average AQI of the Beijing-Tianjin-Hebei urban agglomeration demonstrated a downward trend from 2014 to 2017. However, there were notable fluctuations in certain cities, such as a significant increase in AQI in Zhangjiakou in 2017. The daily AQI distribution revealed distinct seasonal patterns, with heightened pollution levels typically occurring in the southern region. This was particularly evident in Xingtai City, which consistently experienced moderate to severe pollution in 2014 and recorded numerous instances of critical severe pollution between 2015 and 2017, highlighting the necessity for ongoing environmental improvements.

(2) Spatially, the study's analysis, based on the Moran index and scatter plots, indicated a significant spatial correlation in AQI distribution at the 5% level. The results pointed towards a trend of higher AQI values in the southeast and lower values in the northwest, along with notable HH and LL clustering phenomena. In the winter of 2017, for example, Chengde and Beijing exhibited LL clustering, signifying lower AQI values, whereas Shijiazhuang, Hengshui, and Xingtai demonstrated HH clustering, indicative of higher AQI values.

In conclusion, this study's findings accentuate the need for enhanced haze pollution control, especially during winter. The seasonal analysis reveals that winter's atmospheric conditions, characterized by weak pressure, low winds, and a stagnant atmospheric structure, significantly impede the dispersion of pollutants. This leads to heightened air pollution levels during the colder months. Therefore, targeted strategies to mitigate haze pollutants in winter are essential. Spatially, the study highlights the importance of transcending administrative boundaries in air pollution management. The AQI data analysis illustrates a positive interaction in air pollution extending up to 200 kilometers between cities, indicating that neighboring regions greatly influence the spread and transfer of pollutants. This observation of spatial significance correlation and aggregation in AQI underscores the importance of cooperative governance among closely linked regions to enhance air quality control effectively.

Technologically, the study successfully implemented the WOA-LSTM model. This model leveraged the strengths of the LSTM neural network, known for its proficiency in time series prediction, and further refined its network parameters using the WOA algorithm. Through meticulous parameter optimization and rigorous training, the WOA-LSTM model demonstrated exceptional accuracy in predicting the AQI index for the Beijing-Tianjin-Hebei urban agglomeration from 2014 to 2017. Its predictive precision was remarkable, proving highly effective in identifying and issuing warnings for severe pollution incidents. These results significantly contribute to the proactive management of air quality, enabling timely alerts and reminders to relevant stakeholders. Overall, the study provides valuable insights into the spatio-temporal dynamics of haze pollution and introduces an advanced predictive model that can serve as a crucial tool in environmental management and policy-making.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. B. Peden, "Respiratory health effects of air pollutants," *Immunol. Allergy Clin. North Am.*, vol. 44, no. 1, pp. 15–33, 2024. https://doi.org/10.1016/j.iac.2023.07.004

[2] T. Abdul-Rahman, P. Roy, Z. S. B. Bliss *et al.*, "The impact of air quality on cardiovascular health: A state of the art review," *Curr. Probl. Cardiol.*, vol. 49, no. 2, p. 102174, 2024. https://doi.org/10.1016/j.cpcardiol.2023.102174

[3] M. Wang, U. Gehring, G. Hoek, M. Keuken, S. Jonkers, R. Beelen, M. Eeftens, D. S. Postma, and B. Brunekreef, "Air pollution and lung function in dutch children: A comparison of exposure estimates and associations based on land use regression and dispersion exposure modeling approaches," *Environ. Health Perspect.*, vol. 123, no. 8, pp. 847–851, 2015. https://doi.org/10.1289/ehp.1408541

[4] D. Ruiz-Sobremazas, M. R. Coca, M. Morales-Navas, R. Rodulfo-Cárdenas, C. López-Granero, M. T. Colomina, C. Perez-Fernandez, and F. Sanchez-Santed, "Neurodevelopmental consequences of gestational exposure to particulate matter 10: Ultrasonic vocalizations and gene expression analysis using a bayesian approach," *Environ. Res.*, vol. 240, no. 1, p. 117487, 2024. https://doi.org/10.1016/j.envres.2023.117487

[5] G. Adami, M. Pontalti, G. Cattani, M. Rossini, O. Viapiana, G. Orsolini, C. Benini, E. Bertoldo, E. Fracassi, D. Gatti, and A. Fassio, "Association between long-term exposure to air pollution and immune-mediated diseases: A population-based cohort study," *RMD Open*, vol. 8, p. e002055, 2022. http://doi.org/10.1136/rmdopen-2021-002055

[6] H. Jaafar, A. Azzeri, M. Isahak, and M. Dahlui, "The impact of haze on healthcare utilizations for acute respiratory diseases: Evidence from Malaysia," *Front. Ecol. Evol.*, vol. 9, p. 764300, 2021. https://doi.org/10.3389/fevo.2021.764300

[7] A. Dun, Y. Yang, and F. Lei, "Dynamic graph convolution neural network based on spatial-temporal correlation for air quality prediction," *Ecol. Inform.*, vol. 70, p. 101736, 2022. https://doi.org/10.1016/j.ecoinf.2022.101736

[8] C. Aarthi, V. J. Ramya, P. Falkowski-Gilski, and P. B. Divakarachari, "Balanced spider monkey optimization with Bi-LSTM for sustainable air quality prediction," *Sustainability*, vol. 15, no. 2, p. 1637, 2023. https://doi.org/10.3390/su15021637

[9] J. Chen, Z. Liu, Z. Yin, X. Liu, X. Li, L. Yin, and W. Zheng, "Predict the effect of meteorological factors on haze using BP neural network," *Urban Clim.*, vol. 51, p. 101630, 2023. https://doi.org/10.1016/j.uclim.2023.101630

[10] Y. Xu, T. You, Y. Wen, J. Ning, Y. Xiao, and H. Shen, "Air quality research based on B-spline functional linear model: A case study of Fujian Province, China," *Appl. Sci.*, vol. 13, no. 20, p. 11206, 2023. https://doi.org/10.3390/app132011206

[11] J. Xiong, J. Li, Y. Zhang, and G. Mao, "Meteorological impact on dynamic air pollutant concentrations in different timescales: Typical case in Chengdu megacity, China," *Water Air Soil Pollut.*, vol. 234, p. 688, 2023. https://doi.org/10.1007/s11270-023-06711-z

[12] I.-I. Prado-Rujas, A. García-Dopico, E. Serrano, M. L. Córdoba, and M. S. Pérez, "A multivariable sensor-agnostic framework for spatio-temporal air quality forecasting based on Deep Learning," *Eng. Appl. Artif. Intell.*, vol. 127, no. A, p. 107271, 2024. https://doi.org/10.1016/j.engappai.2023.107271

[13] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Econ. Geogr.*, vol. 46, no. sup1, pp. 234–240, 1970. https://doi.org/10.2307/143141

[14] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, 2016. https://doi.org/10.1016/j.advengsoft.2016.01.008

[15] H. Roderick, "Industry and environment," *Pure Appl. Chem.*, vol. 45, no. 3-4, pp. 135–139, 2009. https://doi.org/10.1351/pac197645030135

[16] B. Wang, Y. Wang, X. Cheng, and J. Wang, "Green finance, energy structure, and environmental pollution: Evidence from a spatial econometric approach," *Environ. Sci. Pollut. Res.*, vol. 30, pp. 72 867–72 883, 2023. https://doi.org/10.1007/s11356-023-27427-x