

## Article

# Spatiotemporal Hybrid Air Pollution Early Warning System of Urban Agglomeration Based on Adaptive Feature Extraction and Hesitant Fuzzy Cognitive Maps

Xiaoyang Gu, Hongmin Li \* and Henghao Fan

College of Economics and Management, Northeast Forestry University, Harbin 150040, China; xiaoyangu@nefu.edu.cn (X.G.); fanhenghao@nefu.edu.cn (H.F.)

\* Correspondence: lhm@nefu.edu.cn; Tel.: +86-187-4208-2511

**Abstract:** Long-term exposure to air pollution will pose a serious threat to human health. Accurate prediction can help people reduce exposure risks and promote environmental pollution control. However, most previous studies have ignored the spatial spillover of air pollution, i.e., that the current region's air quality is also correlated with that of geographically adjacent areas. Therefore, this paper proposes an innovative spatiotemporal hybrid early warning system based on adaptive feature extraction and improved fuzzy cognitive maps. Firstly, a spatial spillover analysis model based on the Moran index and local gravitational clustering was proposed to capture the diffusion and concentration characteristics of air pollution between regions. Then, an adaptive feature extraction model based on an optimized Hampel filter was put forward to process and correct the outliers in the original series. Finally, a hesitant fuzzy information optimized fuzzy cognitive maps model was proposed to forecast the air quality of urban agglomeration. The experimental results show that the air quality forecasting accuracy of urban agglomerations can be significantly improved when the geographical conditions and other interactions among cities are comprehensively considered, and the proposed model outperformed other benchmarks and can be used as a powerful analytical tool during urban agglomeration air quality management.

**Keywords:** fuzzy theory; machine learning; optimization algorithm; air pollution forecasting; spatiotemporal feature analysis



**Citation:** Gu, X.; Li, H.; Fan, H. Spatiotemporal Hybrid Air Pollution Early Warning System of Urban Agglomeration Based on Adaptive Feature Extraction and Hesitant Fuzzy Cognitive Maps. *Systems* **2023**, *11*, 286. <https://doi.org/10.3390/systems11060286>

Academic Editors: Wendong Yang, Jinpei Liu and Jianzhou Wang

Received: 13 May 2023

Revised: 30 May 2023

Accepted: 31 May 2023

Published: 2 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the advancement of science and technology, the continuous improvement of urban spatial structure has promoted the rapid growth of the economy. But at the same time, rapid urbanization has also caused serious air quality problems [1,2]. In addition, studies have also shown that people's long-term exposure to particulate pollutants and other air pollutants mainly includes SO<sub>2</sub>, NO<sub>2</sub>, O<sub>3</sub>, and CO [3], which will cause harm to the human respiratory system, nervous system, cardiovascular system, and reproductive system [4–7]. Therefore, it is necessary to establish an air quality forecasting system to grasp the development trend of air quality and air pollution, provide technical support for regional air pollution prevention and control, and provide early warning information for daily travel or production activities.

The individual methods commonly used to forecast air quality are summarized into three types: data simulation method, mathematical statistics method, and machine learning method. The data simulation method uses professional knowledge such as atmospheric physical diffusion formulas and chemical reaction equations to simulate and forecast the formation and diffusion of pollutants [8]. This method is simple to calculate, but it needs to follow strict theoretical assumptions. Because the complexity that occurs in the atmosphere cannot be considered, forecasting accuracy is low [9]. Compared with simulation-based methods, statistical methods based on probability theory show better forecasting results.

The most common are classical time series models, such as the auto-regressive (AR) models, moving average (MA) models, and auto-regressive moving average (ARMA) models [10]. However, statistical methods cannot perform nonlinear fitting during the formation and diffusion of pollutants, which leads to deviations. To address the nonlinear relationship between input and output outcomes, scholars were beginning to leverage machine learning. Support vector machine (SVM) [11,12], random foresting (RF) [13,14], and artificial neural network (ANN) [15] are widely applied in air quality forecasting. Saeid et al. used artificial neural networks to forecast PM<sub>10</sub> concentration and applied it to health risk assessment, and the correlation coefficient reached 0.87 [16]. In addition, as the historical data collected by monitoring sites increases, deep learning methods for big data show superior predictive performance. Given that the recurrent neural network (RNN) can take into account the time-dependent effects of air quality [17], it can be applied to forecast air quality by processing series well. However, in highly complex and large-scale monitoring site networks, it is difficult to capture spatiotemporal correlations between sites.

Fuzzy cognitive maps (FCM) were developed in combination with the above advantages. FCM not only has the ability to handle fuzzy logic uncertainty problems but also has the machine learning algorithms used by neural networks. The model has a good predictive effect while explaining causality [18]. For example, Liu et al. established an EMD-HFCM model combining recurrent neural networks and fuzzy logic, which greatly improved the performance of processing large-scale and non-stationary time series [19]. Yang et al. established the Wavelet-FCM model, which decomposed the original nonstationary time series into multiple time series through the wavelet transform, which improved the forecasting performance and could be applied to a variety of forecasting tasks [20]. Therefore, an in-depth study of FCM can effectively improve the performance of the model, which has the potential and advantages for dealing with dynamic, noisy air quality [21].

However, the forecasting effect of individual models is limited. So, people started building hybrid models. The multi-scale decomposition step is introduced in the forecasting system to decompose the complex system into multiple subsystems that are easy to analyze, such as decomposition and recombination methods [22], data reconstruction [23], feature extraction [24], optimization algorithms [25], etc. For example, Wang et al. applied data preprocessing based on different strategies to improve the performance of wind speed prediction systems [26]. Yang et al. added the variational mode decomposition technique based on adaptive parameters to the machine learning model, and the average absolute error of the last four data sets reached 0.5121 [27]. Hybrid models can reduce the difficulty of modeling complex systems and improve the forecasting performance of the model. Therefore, it makes sense to construct a hybrid model based on FCM.

**Based on the combing and summarizing of the literature, the research motivation of this paper can be summarized as follows:**

- (1) Although the hybrid model can serve the purpose of the problem well by combining specific methods, there are still some disadvantages. Most of the current hybrid models only focus on extracting data features through the decomposition of time series, which greatly improves the fitting effect of data but ignores the potential risk of information leakage in the decomposition process [28]. Therefore, the data preprocessing method needs to be further improved.
- (2) In the past, the study of air pollution was based on the division of administrative regions, and the effects of meteorology and time were analyzed separately. However, air pollution is a cross-regional environmental pollution problem, and the spatial spillover effect between urban agglomerations cannot be ignored [29].
- (3) Traditional single models cannot achieve high accuracy requirements, while hybrid models can improve forecasting accuracy.
- (4) The study of time series data only focuses on the order of time and ignores the particularities of the air quality data itself, such as the ambiguity and uncertainty of the data.

In order to solve the above problems, an innovative spatiotemporal hybrid forecasting model based on adaptive feature extraction and improved fuzzy cognitive maps is proposed. It consists of three modules: a spatial correlation analysis module, data preprocessing module, and fuzzy information forecasting module. For the spatial correlation analysis module, a spatial feature extraction model based on the Moran index and local gravitational clustering (LGC) is designed that can combine spatiotemporal features and extract key influencers from complex data to represent a variety of complex issues in a visual way. For the data preprocessing module, a novel adaptive feature extraction model optimized by the squirrel search algorithm (SSA) is proposed to process and correct the outliers of the original series. Finally, novel hesitant fuzzy cognitive maps (HFCM) are proposed for the fuzzy information forecasting module. The fuzzy logic in the original model is optimized by hesitate fuzzy information, and the forecasting at the numerical level is transformed into the forecasting of probability at the interval level.

**In summary, the main contributions of this paper are:**

- Spatial spillover effects are considered the main factors affecting the air quality index (AQI) of different cities. It verifies that the spatiotemporal correlation of the extracted data is necessary to improve the accuracy of air pollution forecasting.
- The Hampel filter algorithm optimized by the squirrel search algorithm is innovatively introduced into the air quality forecasting model to process and correct the data outliers to improve the forecasting accuracy of the hybrid model.
- Hesitant fuzzy cognitive maps are first proposed to forecast air pollution. It can effectively solve the gray information of air quality or fuzzy relationships and uncertainties, thus further improving the accuracy of forecasting.
- The proposed model was comprehensively evaluated with the actual AQI dataset, five model evaluation criteria, and thirteen comparative models collected from the Beijing-Tianjin-Hebei region. The empirical results show that the proposed hybrid method has superior forecasting performance compared with the comparison models and can provide a theoretical basis for air pollution forecasting and early warning.

## 2. Design of Spatiotemporal Hybrid Air Pollution Early Warning System

Air pollution not only has a temporal complexity but also has a spatial spillover effect. Therefore, in order to forecast air pollution scientifically and rationally, this paper proposes a spatiotemporal hybrid air pollution early warning system of urban agglomeration. Figure 1 depicts the framework of the system.

### 2.1. Spatial Correlation Analysis Module

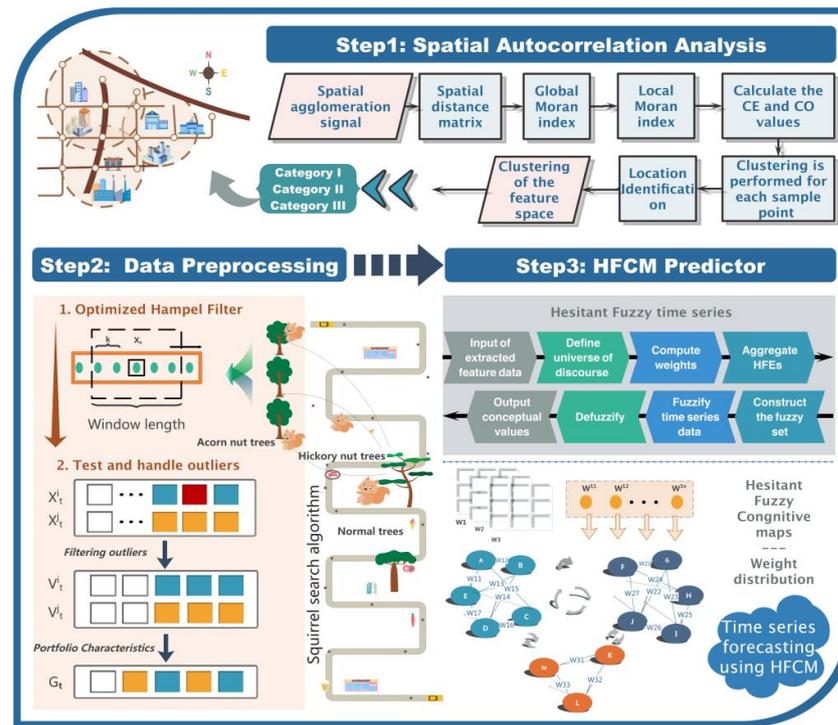
In this section, the Moran I-Local gravitational clustering (ILGC) model is proposed to verify the spatial correlation and capture the diffusion and concentration characteristics of air pollution between regions.

#### 2.1.1. Moran Index

The Moran index is applied to analyze the spatial correlation between different cities. The Moran index is divided into the global Moran index and the local Moran index [30,31]. The global Moran index is used to determine whether there is aggregation or anomaly in space, and if it is judged to be global autocorrelation, the local Moran index is further used to explore the specific manifestations of aggregation or outliers.

Global Moran index

$$I^c = \frac{\sum_{i=1}^{\psi} \sum_{j=1}^{\psi} w_{ij} (X_i^c - \sum_{i=1}^{\psi} X_i^c / \psi) (X_j^c - \sum_{j=1}^{\psi} X_j^c / \psi)}{\sum_{i=1}^{\psi} \sum_{j=1}^{\psi} w_{ij} \sum_{i=1}^{\psi} (X_i^c - \sum_{i=1}^{\psi} X_i^c / \psi)^2 / \psi} \quad (1)$$



**Figure 1.** Framework for the proposed spatiotemporal hybrid air pollution early warning system for urban agglomeration.

Local Moran index

$$I_i^c = \frac{X_i^c - \sum_{i=1}^{\psi} X_i^c / \psi}{(X_i^c - \sum_{i=1}^{\psi} X_i^c / \psi) / \psi} \sum_{j \neq i} w_{ij} (X_j^c - \sum_{i=1}^{\psi} X_i^c / \psi) \quad (2)$$

where  $X_i^c$  is each region ( $i = 1, 2, 3 \dots, \psi$ ),  $c$  represents a different group of cities. The spatial weights between features  $i$  and  $j$  are denoted as  $w_{ij}$  and the total number of features is  $\psi$ , so that the aggregation of all spatial weights can be found.

Most of Moran’s  $I$  belong to  $(-1, 1)$ , but there are also extreme cases, that is, Moran’s value appears outside this range [32]. When Moran’s  $I$  is closer to 1, the more pronounced the positive spatial correlation and the closer the relationship between cities. When Moran’s  $I$  is closer to  $-1$ , the more pronounced the negative spatial correlation and the greater the spatial difference between cities. When Moran’s  $I$  is 0, the space is random, indicating that cities are not correlated.

### 2.1.2. Local Gravitational Clustering

The data are grouped using LGC based on a spatial neighborhood matrix. LGC is a clustering algorithm based on local center metrics. It describes the connectivity of the adjacent regions of each sample point through local gravitational resultant force and centrality. By comparing clustering algorithms such as k-means clustering [33] and fuzzy c-means clustering [34], LGC has the advantage of being able to train singular values to obtain possible impact information [35].

#### Step 1: The local resultant force is calculated for each sample point.

The local resultant force (LRF) is the sum of the forces acting at point  $X_i$  with mass  $m_i$ , denoted as  $F(X_i)$ .

$$F(X_i) = \frac{1}{m_i} \sum_{j=1}^k \hat{D}_{ij} \quad (3)$$

where the reciprocal of all distance points acting on the  $k$  nearest neighbors of the data point  $X_i$  is obtained by summing up  $m_i$ .

$$m_i = \frac{1}{\sum_{j=1}^k D_{ij}} \quad (4)$$

$D_{ij}$  is a unit vector connecting two mass directions and  $k$  is the number of nearest neighbors around  $X_i$ . The closer to the center, the greater the mass of the sample points and the less susceptible they are to other forces.

**Step 2: Calculate the centrality (CE) and coordination (CO) values.**

The relationship between LRF is measured by two metrics: centrality and coordination. Where the centrality of data point  $X_i$  is calculated as follows:

$$CE_i = \frac{\sum_{j=1}^m \cos(\vec{F}_j, \vec{D}_{ij})}{m} \quad (5)$$

where  $\vec{D}_{ij}$  is the displacement vector from  $X_i$ 's neighbor to  $X_i$ ,  $\cos(\theta)$  belongs to  $(-1, 1)$ , and when  $CE_i$  is greater than zero, it is pointed out that most of the neighbors point to it, indicating that it has better centrality.

The CO of data point  $X_i$  is calculated as follows:

$$CO_i = \sum_{j=1}^m (\vec{F}_i \cdot \vec{F}_j) \quad (6)$$

Among them,  $\vec{F}_i$  and  $\vec{F}_j$  are the local resultant forces of points  $X_i$  and  $X_j$ , and their neighbors, and a smaller local gravitational resultant force and a larger  $CE$  value are selected as local proxy points in their own domain, and the local proxy point should preferably appear above the direction pointed by the local gravitational resultant force of the sample point.

**Step 3: Clustering is performed for each sample point  $X_i$ .**

The current local proxy point selected in the previous step connects to communicate with other local proxy points. The current local proxy point looks for the target local proxy point in the neighborhood for connection communication, and if it cannot be found, the best sample is found in the neighborhood sample, and the intermediate communication point continues to complete the work. The intermediate communication point must have a  $CE$  value greater than 0 and be closer to the target proxy point than the origin. The local proxy point itself forms a small cluster, which forms a large cluster by communicating with the target proxy point.

## 2.2. Data Preprocessing Module

This stage introduces a new type of feature extraction model based on an improved Hampel filtering method called SHP. Since the default parameters of the adaptive Hampel filter cannot adapt to the characteristics of all data, we use the squirrel algorithm to optimize the  $DX$  (the scalar of the half-width of the filter window),  $T$  (the threshold used in the equation), and  $Threshold$  (adaptive threshold). The final minimum  $MAPE$  value is the optimization goal, and the data preprocessing algorithm for different urban agglomerations finds the best parameters and then obtains the optimal forecasting effect. The basic theoretical composition of the model is as follows.

### 2.2.1. Squirrel Search Algorithm

The SSA is a naturally inspired optimization paradigm that simulates the dynamic process of finding pinecones in southern flying squirrel gliding. Compared with other

optimizations, SSA considers the seasonal factors of the data to be random factors, which disrupts the search space and improves the randomness of the optimization algorithm.

First, the position of the squirrel in the dimensional search space is represented using the following matrix:

$$FS = \begin{pmatrix} FS_{11} & FS_{12} & \cdots & \cdots & FS_{1d} \\ FS_{21} & FS_{22} & \cdots & \cdots & FS_{2d} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ FS_{n1} & FS_{n2} & \cdots & \cdots & FS_{nd} \end{pmatrix} \tag{7}$$

where  $FS_{nd}$  is the initial position. Next, random initialization is performed using the following formula:

$$FS_i = FS_L + U(0,1) \times (FS_U - FS_L) \tag{8}$$

where  $FS_i$  is the updated position.  $U(0,1)$  denotes the randomness of the flying squirrel population contained in the range 0 to 1.  $FS_U$  and  $FS_L$  are the upper and lower limits, respectively.

Eventually, the squirrel colony locations are updated to the following matrix:

$$f = \begin{pmatrix} f_1(FS_{11} & FS_{12} & \cdots & \cdots & FS_{1d}) \\ f_1(FS_{21} & FS_{22} & \cdots & \cdots & FS_{2d}) \\ & & \vdots & & \\ & & \vdots & & \\ f_1(FS_{n1} & FS_{n2} & \cdots & \cdots & FS_{nd}) \end{pmatrix} \tag{9}$$

The difference between the advantages and disadvantages of searching for food sources will depend on the different fitness levels of each flying squirrel’s location. Therefore, based on the initial position of the squirrel, the position fitness is calculated by determining the variables, namely the best food source (hickory tree  $FS_{ht}$ ), normal food source (acorn tree  $FS_{at}$ ), and no food source (flying squirrel on normal tree  $FS_{nt}$ ). Its position updating for three different foraging scenarios is mathematically modeled as follows.

Case 1: Best food source (hickory tree)

$$FS_{at}^{t+1} = \begin{cases} FS_{at}^t + d_g G_c (FS_{ht}^t - FS_{at}^t) & P_{dp} \leq R_1 \\ \text{Random location} & \text{otherwise} \end{cases} \tag{10}$$

Case 2: Normal food source (acorn tree)

$$FS_{nt}^{t+1} = \begin{cases} FS_{nt}^t + d_g G_c (FS_{at}^t - FS_{nt}^t) & P_{dp} \leq R_2 \\ \text{Random location} & \text{otherwise} \end{cases} \tag{11}$$

Case 3: No food source (flying squirrel on normal tree)

$$FS_{nt}^{t+1} = \begin{cases} FS_{nt}^t + d_g G_c (FS_{nt}^t - FS_{nt}^t) & P_{dp} \leq R_3 \\ \text{Random location} & \text{otherwise} \end{cases} \tag{12}$$

where  $d_g$  is the random glide distance,  $R_1, R_2, R_3$  is a random number in the range of  $[0, 1]$ ,  $FS_{nt}^{t+1}$  is the position of the squirrel reaching the hickory tree, and  $t$  is the current iteration. The sliding constant  $G_c$  is used in the mathematical model to seek a balance between exploration and development. Its value has a non-negligible role in the performance of the proposed algorithm. In the current work, the value of  $G_c$  is 1.9, which is derived from rigorous analysis. The predator presence probability  $P_{dp}$  is 0.1 in all cases.

At the same time, in winter, due to low-temperature environmental conditions, squirrels tend to exercise on a small scale compared to other seasons and are mostly used to store pecans to maintain energy for the winter, so the influence of seasonal factors is considered. It can effectively solve the problem that the algorithm is trapped in the local optimal solution. Seasonal variables are expressed as:

$$S_c^t = \sqrt{\sum_{k=1}^d (FS_{at,k}^t - FS_{ht,k}^t)^2} \tag{13}$$

$$S_{min} = \frac{10H^{-6}}{(365)^T (T_m/2.5)^{-1}} \tag{14}$$

where  $k$  is the current iteration value and  $S_{min}$  is a direct influence on the algorithm's search and development capability. The larger the  $S_{min}$ , the stronger the exploration capability, and the smaller the  $S_{min}$ , the stronger the development capability.

After winter, flying squirrel foraging will be active again, considering that those squirrels that have not been able to find the best food source in winter but still survive will be able to forage again, and there is randomness, which will move in different directions.

The repositioning of the flying squirrel can be expressed by the equation as:

$$FS_{nt}^{new} = FS_L + Levy(n)(FS_U - FS_L) \tag{15}$$

where  $FS_U$  and  $FS_L$  are the upper and lower bounds of variables, *Lévy* is used to generate random solutions.

$$Lévy(x = 0.01 \times \frac{r_a \sigma}{|r_b|^{1/\beta}}) \tag{16}$$

where  $r_a$  and  $r_b$  are random number matrices belonging to (0, 1),  $\beta = 1.5$ . And  $\sigma$  is defined as follows:

$$\sigma = \left( \frac{\Gamma(1 + \beta) \sin(\frac{\beta\pi}{2})}{\Gamma(\frac{1+\beta}{2}) \beta 2^{(\frac{\beta-1}{2})}} \right)^{1/\beta} \tag{17}$$

where  $\Gamma(x) = (x - 1)!$ .

### 2.2.2. Hampel Filter

The Hampel filter is a decision filter used to detect and remove outliers [36]. Like the  $3\sigma$  rule, which distinguishes between the mean and standard deviation, it determines whether the data in the data set are a singular value by the median and the median that deviates from the absolute value of the median, and then replaces the outlier with the median of the short series in the filter movement window.

$$|x_i - x^*| > tS \quad j = 1, 2, \dots, N \tag{18}$$

$$S = 1.4286 median\{|x_i - x^*|\} \tag{19}$$

where  $t$  is a predefined threshold for determining the absolute difference from the median, and  $x^*$  denotes the median of the data sequence of length  $N$ , and  $S$  is the median absolute difference (MAD) scale estimator. The constant 1.4286 guarantees that the expected value of  $S$  is equal to the standard deviation of the normal distribution data [37].

### 2.3. Fuzzy Information Forecasting Module

In this section, the basic theory of hesitant fuzzy time series is first described, followed by the details of the innovative hesitant fuzzy cognitive maps. This paper innovatively proposes hesitant fuzzy cognitive maps (HFCM) and introduces the hesitant fuzzy logic relationship to modify the output results of the fuzzy cognitive maps.

### 2.3.1. The Basic Definition of Hesitant Fuzzy Theory

Before constructing hesitant fuzzy cognitive maps for forecasting, let us review the definitions of fuzzy set, hesitant fuzzy set, and fuzzy time series (FTS). The basic theory is as follows [38].

**Definition 1.** Define  $U = \{u_1, u_2, \dots, u_n\}$  the universe of discourse. A fuzzy set  $A_{ij}$  in  $U$  can be defined by its membership function.

$$A_{ij} = \sum_{i=1}^n \sum_{j=1}^m \frac{\mu A_{ij}(x_{ij})}{x_{ij}} = \frac{\mu A_{11}(x_{11})}{x_{11}} + \frac{\mu A_{12}(x_{12})}{x_{12}} + \dots + \frac{\mu A_{nm}(x_{nm})}{x_{nm}} \quad (20)$$

where  $A_{ij} : U \rightarrow [0, 1]$  represents the membership function of the fuzzy set  $A_{ij}$  and  $\mu A_{ij}$  represents the member degree of  $\mu$  to  $A_{ij}$ .

**Definition 2.**  $H, A_{ij}, X$  over the reference set  $U = \{u_1, u_2, \dots, u_n\}$  is represented by a mathematical expression as follows:

$$H = \{ \langle x_{ij}, h_H(x_{ij}) \rangle \mid \forall x_{ij} \in U \} \quad (21)$$

where  $h_H(x_{ij})$  is the set of possible subsets in  $[0, 1]$  by membership function that computes the possible membership degrees of the elements in  $U$  to the set  $H$ .  $P[0, 1]$  is the set of multiple subsets of  $[0, 1]$ . For the element  $u \in U$ ,  $h_H(u_i)$  is called the hesitant fuzzy element (HFE).

**Definition 3.** Let  $Y(\xi)$  ( $\xi = 0, 1, 2, \dots$ ),  $R$  (a subset of real number) be a universe of discourse. Different fuzzy sets  $f_i(\xi)$  ( $\xi = 0, 1, 2, \dots$ ) are defined on  $Y(\xi)$ . If  $F(\xi)$  is a collection of  $f_i(\xi)$  ( $\xi = 0, 1, 2, \dots$ ), then it is known that  $F(\xi)$  is called an FTS on  $Y(\xi)$  ( $\xi = 0, 1, 2, \dots$ ).

**Definition 4.** Let  $F(\xi)$  be the FTS and  $R(\xi, \xi - 1)$  be a first order model of  $F(\xi)$ . If  $R(\xi, \xi - 1) = R(\xi - 1, \xi - 2)$  for any time  $\xi$ , then  $F(\xi)$  is designated as time-invariant FTS, otherwise  $F(\xi)$  is named as time-variant FTS.

**Definition 5.** The fuzzy logic relation (FLR) can be expressed as  $F(\xi - q) \rightarrow F(\xi)$ , Their relationship can be expressed by the formula:

$$F(\xi) = F(\xi - q) \circ R(\xi - q, \xi) \quad (22)$$

where “ $\circ$ ” is a max-min composition operator,  $R(\xi - q, \xi)$  is fuzzy relationship.

**Definition 6.** The relationship between  $F(\xi)$  and  $F(\xi - q)$  can be denoted as  $A_w \rightarrow A_v$ , where  $A_w$  and  $A_v$  are called the left-hand side and the right-hand side of the FLR, respectively. FLRs with the same left-hand side can be categorized into an ordered fuzzy logic group (FLG).

### 2.3.2. The FCM Framework

FCM consists of a set of nodes and weighted edges, which can be regarded as a weighted directed graph. The nodes represent the variables in the AQI time series, that is, each city in the urban agglomeration, and the edges represent the causal relationship between these nodes. The set of these concept nodes is represented as vector  $C$ , that is, the set of urban agglomerations. Each FCM containing  $N$  concepts can be defined by four elements,  $(C, W, A, F)$ , where  $C = \{C_1, C_2, \dots, C_N\}$  is the collection of concepts that represent all detected air quality indices. These concepts define the state value of  $t$  at any moment as a vector  $A_i$ , that is, the AQI time series data of each city:

$$A_i = \{A_1, A_2, \dots, A_N\} \quad (23)$$

where  $W^c : C \times C \rightarrow [-1, 1]$  is an  $N \times N$  connection matrix representing the relationship from  $C_i$  to  $C_j$ .

The circular relationship between  $A(\xi - 1)$  and  $A(\xi)$  on  $\xi \geq 0$  is expressed as follows:

$$A_i^c[k + 1] = f \left( A_i^c[k] + \sum_{j=1, j \neq i}^m (A_i^c[k] W_{ij}) \right) \tag{24}$$

$A_i^c(\xi)$  denotes the state value of node  $A_i^c$  at time  $\xi$ ,  $A_i^c(\xi + 1)$  denotes the value of node  $A_i^c$  at time  $\xi + 1$ . The state value of the node at the  $(t + 1)$ -th iteration is determined by the weight matrix and state value of all connected nodes at the  $t$ -th iteration.

The  $f$  is the transfer function which simulates the forecasting process of the conceptual state of the  $c$ -th FCM model. It is the activation function that updates the activation state of the node at each iteration to perform a simulation of conceptual state forecasting.

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \tag{25}$$

where  $\lambda (\lambda \geq 0)$  acts as a regulator that determines the appropriate shape of the function and the speed at which the curve converges to the boundary.  $\lambda = 1$  is commonly chosen.

### 2.3.3. Hesitant Fuzzy Processing Time Series

#### Step 1: Define the universe of discourse and divide the interval.

The universe of discourse is defined as  $U = [Da_{\min} - \sigma_1, Da_{\max} + \sigma_2]$ . Among them,  $Da_{\max}$  and  $Da_{\min}$  are the maximum and minimum values of AQI data, respectively,  $\sigma_1$  and  $\sigma_2$  are the standard deviation of the training sample.

Then, the cumulative distribution function (CDF) is used to divide the equal frequency interval.

$$F(x_i^c) = \Pr(X \leq x_i^c) \text{ for } -\infty < x_i^c < \infty \tag{26}$$

To determine the optimal number of intervals, adaptive interval division is performed based on fuzzy  $c$ -means clustering. The cluster generated by clustering is regarded as a fuzzy set; the cluster center is the midpoint of the interval, and the size of the clustered category is the interval size.

$$u_{ij}^c = \frac{1}{\sum_{center=1}^k (d_{ij}^c / d_{ik}^c)^{2/f_z - 1}} \tag{27}$$

where  $u_{ij}^c$  is the membership degree between the sample point  $x_i^c$  and the cluster enter  $v_j^c$ ,  $i$  is the sample,  $c$  is the city category,  $f_z$  is the fuzzy index ( $f_z > 1$ ),  $d_{ij}^c$  is the distance between the sample point  $x_i^c$  and cluster center  $v_j^c$ . The most common method for calculating the distance is Euclidean distance.

$$v_j^c = \frac{\sum_{i=1}^n (u_{ij}^c)^{f_z} x_i}{\sum_{i=1}^n (u_{ij}^c)^{f_z}} \tag{28}$$

#### Step 2: Calculate membership degree.

For the calculation of the weights  $w_e^h, w_u^h$ , this article uses the trigonometric membership function with the following formula:

$$w_e^h = \frac{d_e}{d_e + d_u} \tag{29}$$

$$w_u^h = \frac{d_u}{d_e + d_u} \tag{30}$$

where  $d_e$  and  $d_u$  are the equal and unequal spacing lengths calculated above, respectively.

The triangle membership function is used to calculate the membership degree on the interval divided by different methods, and the  $mbge(i)$ ,  $mbgu(i)$  is obtained. The triangle membership function formula is as follows:

$$triangle(x_i^c; a, b, c) = \begin{cases} 0 & x_i^c \leq a \\ \frac{x_i^c - a}{b - a} & a \leq x_i^c \leq b \\ \frac{c - x_i^c}{c - b} & b \leq x_i^c \leq c \\ 0 & c \leq x_i^c \end{cases} \tag{31}$$

The formula for aggregating hesitant fuzzy elements and constructing fuzzy sets, where the aggregation operator is used to calculate the member rank of the elements, is defined as follows:

$$O\{x_1, x_2, \dots, x_n\} = 1 - \prod_{i=1}^n (1 - x_i)^{w_i} \tag{32}$$

Based on Definition 3, the fuzzy set is determined according to the maximum membership principle. Because the air quality index lags one-stage correlation, the first-order fuzzy logic relationship is determined according to Definition 5. After that, the fuzzy relationship group is established based on Definition 6 by counting the frequency of fuzzy relationships in the training set.

**Step 3: Build fuzzy relationship matrix.**

The weight matrix is established by the elements in the fuzzy relationship matrix with the frequency of “ $A_w \rightarrow A_v$ ” in the training set.

$$\begin{pmatrix} w_{11} & w_{12} & \dots & w_{1m} \\ w_{21} & w_{22} & \dots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mm} \end{pmatrix} = [W_1^c, W_2^c, \dots, W_m^c] \tag{33}$$

where  $w_{ij} \in [0, 1]$ ,  $(i, j = 1, 2, \dots, m)$ . It is the result of weighting the frequency of each fuzzy relationship.

**Step 4: Defuzzify**

Fuzzy output is obtained on the basis of hesitant fuzzy relations. Before that, the combined midpoints of the equal and unequally spaced trigonometric membership functions are calculated as follows:

$$CM = \frac{M_e w_e^h + M_u w_u^h}{e_e^h + e_u^h} \tag{34}$$

where  $M_e$ ;  $w_e^h$  and  $M_u$ ;  $w_u^h$  are the midpoints and weights of the membership functions in the interval of equality and inequality, respectively. Use the following formula to defuzzify them for numerical forecasting.

$$NF = \frac{\sum f_i CM_i}{\sum f_i} \tag{35}$$

where  $f_i$  is the fuzzified output obtained by applying the max–min combination operation on fuzzy logic relationships (FLR).

The pseudo-code of the data hesitant fuzzification process of Algorithm 1 is as follows:

**Algorithm 1: HFCM**

**Objective function:**  $\min(\text{MAPE}) = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i(x) - y_i(x)}{y_i(x)} \right| \times 100\%$

**Input:**  $(x_1^c, x_2^c, x_3^c, \dots, x_n^c)$ —a sequence of sample data.  
 $(h_1^c, h_2^c, h_3^c, \dots, h_n^c)$ —a sequence of FCM output

**Output:** 
$$\begin{cases} \text{MAPE}(r, :) = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i(x) - y_i(x)}{y_i(x)} \right| \times 100\% \\ \text{RMSE}(r, :) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i(x) - \hat{y}_i(x))^2} \\ \text{MAE}(r, :) = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i(x) - y_i(x)| \end{cases}$$

**Parameters:**

$ip$ : number of the intervals

$i$ :  $i$ -th sample

$j$ :  $\begin{cases} 0 & \text{adaptive division interval} \\ 1 & \text{equal frequency division interval} \end{cases}$

$Lo(N)$ : the left endpoint of the adaptive division interval

$Up(N)$ : the right endpoint of the adaptive division interval

$P_{LB}(N)$ : the left endpoint of the equal frequency division interval

$P_{UB}(N)$ : the right endpoint of the equal frequency division interval

$mbge(i)$ : the membership degree of the adaptive division interval

$mbgu(i)$ : the membership degree of the equal frequency division interval

$mn$ : min number of the intervals

$mm$ : max number of the intervals

---

```

1: /* Initialize the data and convert it into growth rate */
2: /* Define the universe of discourse */
3: FOR  $ip = mn: mm$  (number of the intervals);  $N = ip$ ;
4:   /* Calculate  $Lo(N); Up(N)$  by cumulative distribution function. (Equation (26)) */
5:   /* Calculate  $P_{LB}(N); P_{UB}(N)$  by fuzzy c-means clustering (Equation (27)) */
6:   /* Calculate the weights of different intervals */
7:    $de = P_{UB}(N) - P_{LB}(N)$ ,  $du = Up(N) - Lo(N)$ ;
8:    $we = de / (de + du)$ ,  $wu = du / (de + du)$ 
9:   /* Calculate  $mbge(i); mbgu(i)$  */
10:  Calculate: Membership grades  $u(i) = 1 - ((1 - mbge(i))^{we}) \times ((1 - mbgu(i))^{wu})$ ;
11:  /* Judgment fuzzy sets */  $u_{ki} = \max(u_{1i}, u_{2i}, u_{3i}, \dots, u_{ji}), 1 < k < n$ 
12:    IF  $H_{Ak}$  is fuzzy set corresponding to  $u_{ki}$ 
13:      | Assign fuzzy set  $H_{Ak}$  to  $x_i^c$ 
14:    END
15:    IF  $H_{Ai}$  is fuzzy production of day  $n$ , and  $H_{AK}$  is fuzzy production of day  $n + 1$ 
16:      | Fuzzy logic relationship =  $\{H_{A1} - H_{A2}, \dots, H_{Ai} - H_{Aj}\}$ 
17:    END
18:  /* Determine the fuzzy logic relation group */
19:  /* Count the frequency of each logical relationship */
20:  /* Calculate the percentage rate of each occurrence logic */
21:  /* Calculate the weight matrix and normalize the weight. */
22:  Calculate:  $grade(i, j) = trimf((h_1^c, h_2^c, h_3^c, \dots, h_n^c), [Lower(j), mid\_point(j), Upper(j)])$ ;
23:  /* The maximum membership principle determines the fuzzy set to which it belongs. */
24:  Calculate:  $combined\ mid = (we * mid\_point(j) + wu * mid\_point(j)) / (we + wu)$ ;
25:  /* Defuzzification to obtain the predicted value */
26:  /* Turning growth rates into data  $\hat{y}_i(x)$  */
27: END
28:  $[r, c] = find(MAPE == \min(MAPE))$ 
29: Returned:  $r$  (The location of the optimal concept)

```

---

#### 2.4. Error Evaluation Module

To evaluate the superiority of the proposed system, the error evaluation module was constructed, which contained three statistical indicators and two test models.

##### 2.4.1. Error Test

The following three statistical indicators were employed to evaluate the effectiveness of the predictive model: mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE). They are defined as:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i(x) - y_i(x)}{y_i(x)} \right| \times 100\% \quad (36)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i(x) - \hat{y}_i(x))^2} \quad (37)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i(x) - y_i(x)| \quad (38)$$

where  $\hat{y}_i(x)$  is the forecasting data,  $y_i(x)$  is the original data, and  $i$  is the number of samples.

##### 2.4.2. Hypothesis Testing

This section uses the Diebold–Mariano (DM) test [39] and the modified Diebold–Mariano (MDM) test [40] to evaluate hypothesis testing methods for the predictive performance of different models. Its theory is described below.

The raw data is  $\{u_t^n\}$ , and for the two forecasting sequences  $\{\varepsilon_{jt}\}$  and  $\{\varepsilon_{kt}\}$ , their error can be calculated by the following formula:

$$er_{jt}^n = \varepsilon_{jt} - u_t^n \quad (39)$$

$$er_{kt}^n = \varepsilon_{kt} - u_t^n \quad (40)$$

$$\psi(er_{jt}^n) = (er_{jt}^n)^2 \quad (41)$$

$$\psi(er_{kt}^n) = (er_{kt}^n)^2 \quad (42)$$

At this point, the difference between the two error sequences  $\{diff_t\}$  can be defined as follow:

$$diff_t = \psi(er_{jt}^n) - \psi(er_{kt}^n) \quad (43)$$

The null and alternative hypotheses are set as follows:

$$H_0 : E(diff_t) = zero;$$

$$H_1 : E(diff_t) \neq zero;$$

The DM statistic follows a standard normal distribution and can be calculated as:

$$DM = \overline{diff} / \sqrt{S^2/T - 1} \quad (44)$$

Specifically,  $T$  is the number of advance multi-step forecasting. Where  $\overline{diff}$ ,  $S^2$  can be defined as:

$$\overline{diff} = T^{-1} \sum_{t=1}^T diff_t \quad (45)$$

$$s^2 = (T - 1)^{-1} \sum_{t=1}^T \left( diff_t - \overline{diff} \right)^2 \tag{46}$$

However, efficient  $T$ -period forecasting has forecasting errors after the  $MA(T - 1)$  process, so MDM is proposed to solve this problem with the assumption that “no observed” lag has zero covariance. The DM statistic is compared with the cut-off value  $Z_{\alpha/2}$ . If  $-Z_{\alpha/2} \leq DM \leq Z_{\alpha/2}$ , accept the null hypothesis, indicating that there is no difference in the forecasting effect of the two models; If  $DM > Z_{\alpha/2}$  or  $DM < -Z_{\alpha/2}$ , the null hypothesis was rejected, indicating that the difference in forecasting performance between the two models was not caused by accidental factors.

### 3. Results

This section includes three parts: data description, spatial feature capture, and model comparison experiments. What is more, three sets of comparison experiments are established to verify the validity of the forecasting of the proposed spatiotemporal mixture model.

#### 3.1. Study Area and Data Description

To verify the accuracy and effectiveness of the proposed model, the Beijing-Tianjin-Hebei region was selected as the research object. The site of the study is shown in Figure 2.

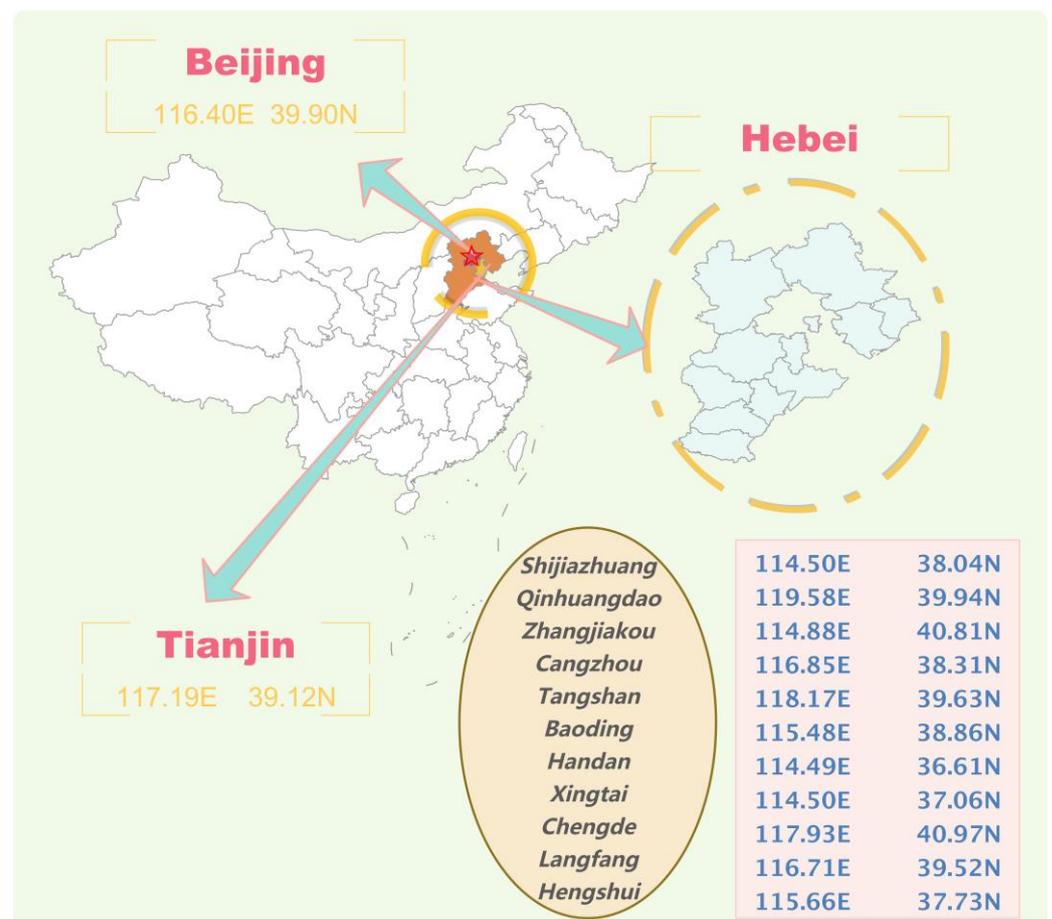


Figure 2. Location of the study area.

China's major urban agglomerations with high levels of economic development are Beijing-Tianjin-Hebei, Yangtze River Delta, and Pearl River Delta. Among them, the Beijing-Tianjin-Hebei region is not only a representative area of China's economic development but also a typical area with serious air pollution [41]. Due to the high population density in the region, the proportion of heavy industry [42], industrial transfer [43], and other factors, the air environment problem has become increasingly prominent. While the Beijing-Tianjin-Hebei region has coordinated economic development [44], air quality between different cities also has a certain mutual influence [45]. The air pollution problem of the Beijing-Tianjin-Hebei urban agglomeration has become a topic of general concern for the Chinese government, the media, and the public. So, it is typical to select the Beijing-Tianjin-Hebei region as the research object.

This study uses daily AQI data from 1 January 2017 to 31 August 2022, for 13 cities with a total of 2069 observations. Daily AQI data were collected from <http://www.tianqihoubao.com> (accessed on 10 September 2022). AQI is a comprehensive indicator of air quality [46], and the higher the value, the more serious the pollution. It was calculated according to the National Ambient Air Quality Standard of China (GB3095-2012) [47], including sulfur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>), PM<sub>2.5</sub>, PM<sub>10</sub> (particulate matter with particle size less than or equal to 10 µm), ozone (O<sub>3</sub>), carbon monoxide (CO), and other six pollutants [18] to comprehensively reflect the regional air quality and pollution degree. Comprehensive AQI can emphasize the possible chronic health effects of air pollution and long-term damage to the environment. Compared with the air pollution index (API) [48], the detection of pollutants is more comprehensive, the grading restriction standards are stricter, and the evaluation results are more objective. Specifically, the air pollution level is divided into seven attribute categories according to the air quality index (AQI), as shown in Table 1.

**Table 1.** Air quality index scale.

AQI	Level	I	Descriptions	Color
0–50	I	0.155	Good	Green
51–100	II	0.207	Moderate	Yellow
101–150	III	0.176	Lightly polluted	Orange
151–200	IV	0.095	Moderately polluted	Red
201–300	V	0.066	Heavily polluted	Purple
>300	VI	0.195	Severely polluted	Maroon

When the air pollution index is less than 100, people can engage in normal activities. When the air quality index reaches mild pollution (100~200), patients with heart disease and respiratory diseases should reduce physical consumption and outdoor activities; when the air quality index reaches 200~300, healthy people ought to reduce outdoor activities. Additionally, the elderly and patients with heart disease and lung disease should stay indoors and reduce physical activity; when heavy pollution (air quality index above 300) is reached, healthy people should also avoid outdoor activities.

Statistical analysis of daily data is shown in Table 2.

From Table 2, in the Beijing-Tianjin-Hebei region, Shijiazhuang and Chengde have the best air quality situation, with mean values of 59.69 and 59.34, respectively, and less fluctuation, with variances of 31.2 and 36.17. In comparison, Handan has the worst air quality and a larger variation, indicating that the air pollution level fluctuates more seriously in the process of economic and social development.

**Table 2.** Statistical analysis of raw data.

No.	City Name	Longitude	Latitude	Training	Testing	Max	Min	Mean	Std
1	Shijiazhuang	114.502461	38.045474	1656	413	442	19	97.85	57.78
2	Hengshui	115.665993	37.735097	1656	413	377	16	89.68	50.88
3	Baoding	115.482331	38.867657	1656	413	476	17	92.46	56.39
4	Xingtai	114.508851	37.068256	1656	413	463	17	97.28	56.61
5	Handan	114.490686	36.612273	1656	413	389	16	99.81	56.43
6	Beijing	116.405285	39.904989	1656	413	454	11	72.05	47.77
7	Cangzhou	116.857461	38.310582	1656	413	346	16	81.99	45.15
8	Langfang	116.713873	39.529244	1656	413	413	13	78.05	46.25
9	Tianjin	117.190182	39.125596	1656	413	365	14	78.29	42.51
10	Tangshan	118.175393	39.635113	1656	413	399	16	84.83	45.85
11	Chengde	117.939152	40.976204	1656	413	409	17	59.69	31.20
12	Qinhuangdao	119.586579	39.942531	1656	413	364	16	66.22	35.44
13	Zhangjiakou	114.884091	40.811901	1656	413	488	19	59.34	36.17

### 3.2. Spatial Feature Extraction Results

For each city, the annual average of daily AQI data is taken as the data for Moran’s index, and the geographical distance matrix is used to describe the static spatial correlation between cities, so as to further use the model to summarize the neighborhood characteristics of the target city for spatiotemporal forecasting.

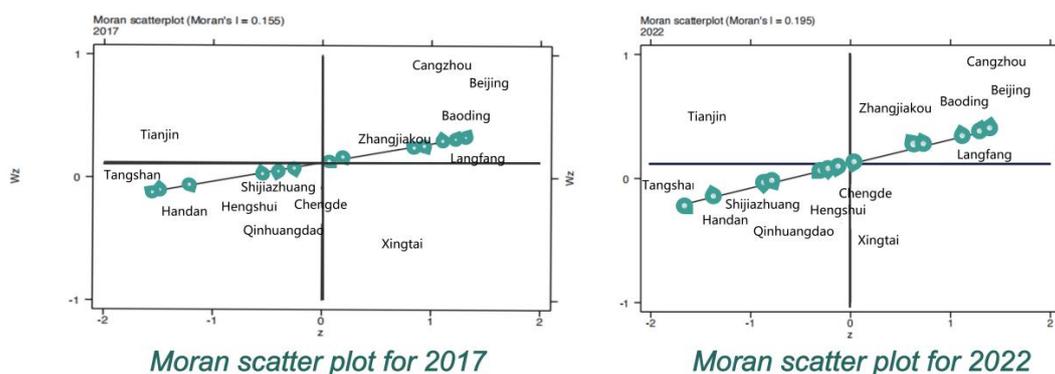
#### 3.2.1. Spatial Autocorrelation

Table 3 shows the calculation results of the global Moran index to verify whether it is spatially correlated. The average daily data for one year was taken as the calculation data for Moran’s index. Figure 3 is the Moran scatter plot, which visually shows the distribution of AQI clusters in various cities and further explores the degree of spatial correlation in a specific area. Table 4 shows the distribution trend of the Moran index, from which there is a temporal and spatial transfer process for the AQI.

**Table 3.** Moran index of the AQI for 2017 and 2022.

Year	I	E(I)	Sd(I)	z	p-Value *
2017	0.155	−0.083	0.092	2.599	0.005
2018	0.207	−0.083	0.092	3.164	0.001
2019	0.176	−0.083	0.09	2.873	0.002
2020	0.095	−0.083	0.09	1.983	0.024
2021	0.066	−0.083	0.09	1.661	0.048
2022	0.195	−0.083	0.091	3.064	0.001

**Note:** I is Moran’s index, reflecting the spatial correlation. E(I) is the mean of I, Var(I) is the standard deviation of I; Z is the standard score, reflecting the degree of dispersion of a data set; p-value represents the probability. When the probability value is small, the null hypothesis will be rejected. \* represents 10 % confidence.



**Figure 3.** Local Moran maps of AQI in 2017 and 2022.

**Table 4.** Spatial transition of Moran scatter of urban AQI (2017–2022).

Cities	2017	2018	2019	2020	2021	2022
1	LL	LL	LL	LL	LL	LL
2	LL	LL	LL	<b>HL</b>	LL	LL
3	HH	HH	HH	HH	HH	HH
4	HL	HL	HL	HL	HL	HL
5	LL	LL	LL	LL	LL	LL
6	HH	HH	HH	HH	HH	HH
7	HH	HH	HH	HH	HH	HH
8	HH	HH	HH	HH	<b>HL</b>	HH
9	LH	LH	LH	LH	LH	LH
10	LL	LL	LL	LL	LL	LL
11	<b>HH</b>	HL	HH	HL	LL	<b>HL</b>
12	LL	LL	LL	LL	LL	LL
13	HH	HH	HH	HH	HH	HH

**Note:** HH: in the first quadrant, areas with high AQI are surrounded by other cities that also have high AQI; LH: in the second quadrant, areas with a low AQI are surrounded by other cities with a high AQI; LL: in the third quadrant, areas with a low AQI are surrounded by other cities with the same low AQI; HL: in the fourth quadrant, areas with a high AQI are surrounded by other cities with a low AQI.

Table 3 shows the AQI of thirteen cities in the Beijing-Tianjin-Hebei region from 2017 to 2022. It can be seen from the table that the Moran index has passed the 5% confidence test and is positive. The results showed that there was a positive spatial correlation, the AQI tended to be agglomerated in space, and the distribution of the air quality index was concentrated and encircled.

As shown in Figure 3, the Moran scatter plot set for each city's AQI is distributed in the first and third quadrants, indicating a significant positive contribution of each city's air quality in the local space. The total number of cities located in the 1,3-quadrant re-mained unchanged at 0.77 in the total sample. Most cities maintain similar cluster characteristics to their neighbors.

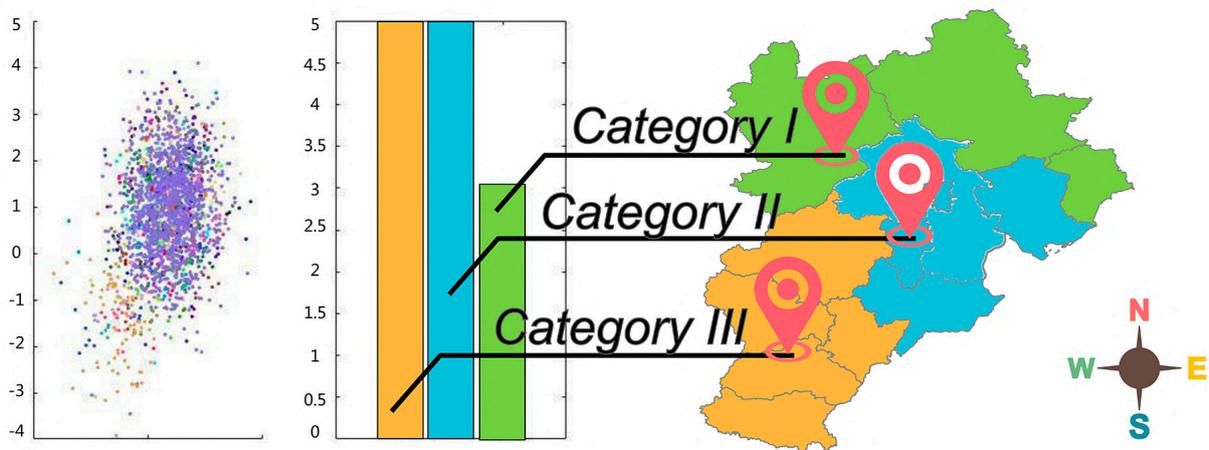
In addition, only three cities have experienced location migration of the Moran scattering point, namely, Chengde, Langfang, and Hengshui. Among them, Hengshui and Langfang return to their original spatial states after being influenced by the surrounding environment, while Cangzhou, which was at low air pollution, completely changes its original state and enters the HL region under the influence of high surrounding air pollution. This indicates that the air quality of Chinese urban agglomerations is highly spatially correlated, that air pollution between urban agglomerations is interacting, and that the magnitude of the effect is not the same.

### 3.2.2. Local Gravitational Clustering

As described in Section 3.2.1, the AQI is spatially dependent. Non-independent air quality may affect clustered data with interdependent observations. Because the AQI is a comprehensive index, it can reflect the air quality of the city. Therefore, the daily AQI data of different cities are used as clustering variables to cluster cities.

It can be seen from Figure 4 that the cities are divided into three different classes of air quality categories, showing good zonal differences and regional similarities.

Category I includes five cities: Shijiazhuang, Hengshui, Baoding, Xingtai, and Handan. This type of urban agglomeration is in the plains and has poor air pollution dispersion conditions. It is a heavy industrial city close to the inland, with many polluting industries superimposed on the transmission of air pollution from western Lu and northern Yu, where air pollution is most serious.



**Figure 4.** Irregular cluster results.

Category II includes five cities: Beijing, Cangzhou, Langfang, Tianjin, and Tangshan. Due to the large population and rapid economic development, in which the steel industry in Tangshan, coal in Cangzhou, and cement in Langfang were coupled, pollution is serious and air quality is deteriorating.

Category III includes three cities: Chengde, Qinhuangdao, and Zhangjiakou. Tourism-driven economic development, a few polluting industries, and mountainous terrain limit the flow and interaction of the atmosphere between the three cities and other cities, thus keeping air quality low.

Therefore, in order to continuously improve the environmental benefits and air quality of the Beijing-Tianjin-Hebei region, the interaction between its cities cannot be ignored, and it is necessary to establish an improvement mechanism according to its geographical location, industrial structure, and economic development level to achieve the coordinated development of the Beijing-Tianjin-Hebei region.

### 3.3. Model Comparison Results

Based on the above experimental purposes, this section establishes three comparative experiments and uses the same evaluation index system to verify the forecasting performance.

#### 3.3.1. Feature Extraction Strategy

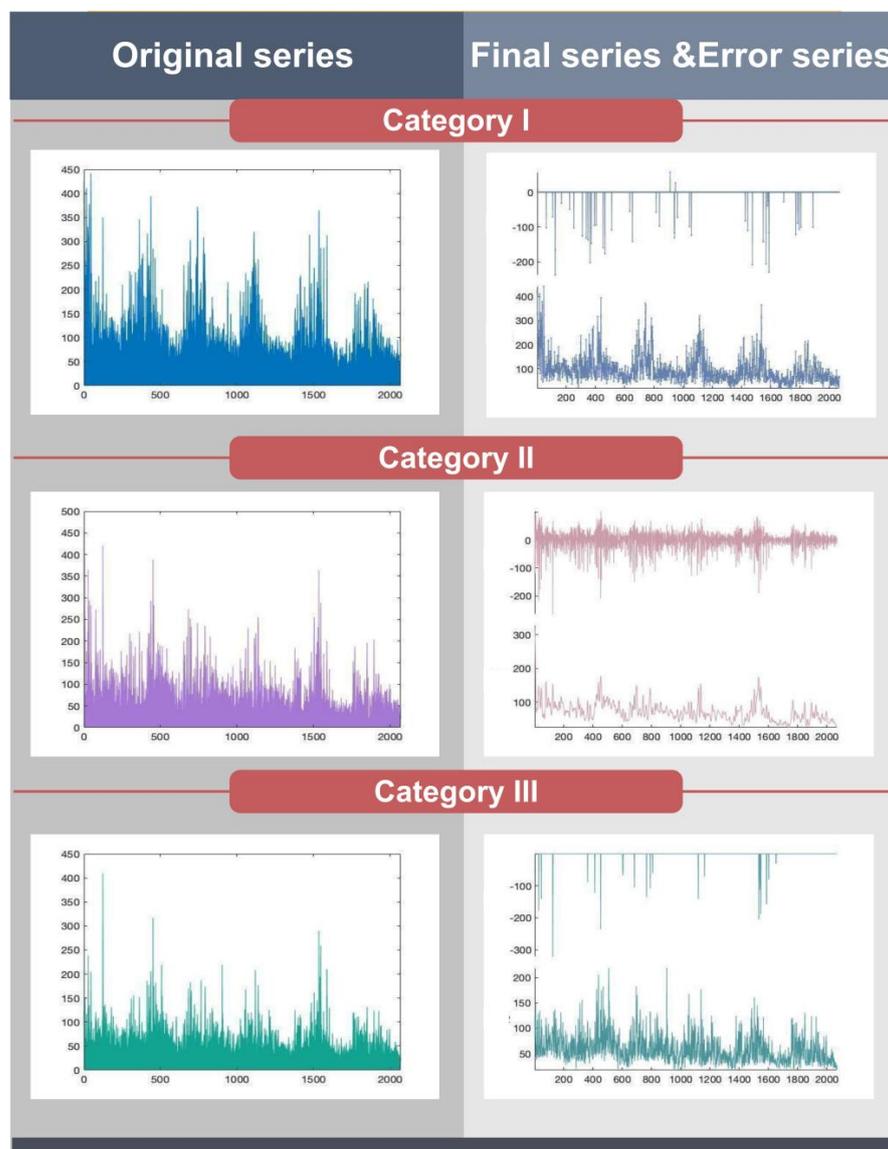
Due to the variability and nonlinearity of AQI in time series changes, it is difficult to accurately predict. In addition, there are many factors affecting air quality, including policy factors, natural factors, etc., which are prone to fluctuations and singular values. Outliers are data with significant deviations from most observations, which interfere with the training efficiency of the model. Removing outliers can improve forecasting accuracy. Therefore, before the experimental forecasting, it is necessary to test and correct the outliers of the data.

To illustrate the excellence of the optimized Hempel filtering in detecting and correcting outliers, comparative models based on different optimization algorithms were developed, including the dragonfly algorithm (DA), SSA, and the honey badger algorithm (HBA). The comparison experiments were FCM and DHP-FCM, FCM and SHP-FCM, and FCM and HHP-FCM. The specific results are shown in Table 5 and Figure 5.

**Table 5.** Comparison of feature extraction strategies based on different optimization algorithms.

Cities	Method	MAPE	RMSE	MAE
Category I Beijing	FCM	25.54	25.15	17.76
	DHP-FCM	39.47	37.52	24.25
	SHP-FCM	<b>25.12</b>	<b>26.34</b>	<b>17.79</b>
	HHP-FCM	25.40	21.94	15.91
Category II Shijiazhuang	FCM	40.51	37.73	21.18
	DHP-FCM	31.96	24.38	15.43
	SHP-FCM	<b>36.27</b>	<b>21.02</b>	<b>15.36</b>
	HHP-FCM	36.25	21.08	15.40
Category III Chengde	FCM	29.50	20.66	14.25
	DHP-FCM	28.85	20.31	13.93
	SHP-FCM	<b>27.78</b>	<b>19.22</b>	<b>13.29</b>
	HHP-FCM	29.56	19.62	13.87

**Note:** HP: Hampel filtering; DHP: Hampel filtering optimized by dragonfly algorithm (DA); SHP: Hampel filtering optimized squirrel search algorithm (SSA); HHP: Hampel filtering optimized honey badger algorithm (HBA).



**Figure 5.** The result of feature extraction from raw data.

From Table 5, it can be seen that the SHP-FCM model has the best performance on MAPE, RMSE, and MAE compared to other optimization models for Category I, Category II, and Category III. In addition, the SSA algorithm has an outstanding advantage in finding the best model parameters to build a better hybrid model. In the three urban clusters, the MAPE metrics were reduced by 1.6%, 10.5%, and 5.8% compared to those optimized using the SSA algorithm. This finding has been confirmed in several application scenarios.

**Remark 1.** SHP not only removes significantly deviated data and reduces interference with the model training rate, but also reduces data variability and preserves original features. Furthermore, from the residual sequence of the right diagram, it can be seen that the processed data and the original data only change slightly, and only the median substitution in the data pane is performed on the extreme data. Therefore, the data preprocessing in this paper will not cause information loss to the data.

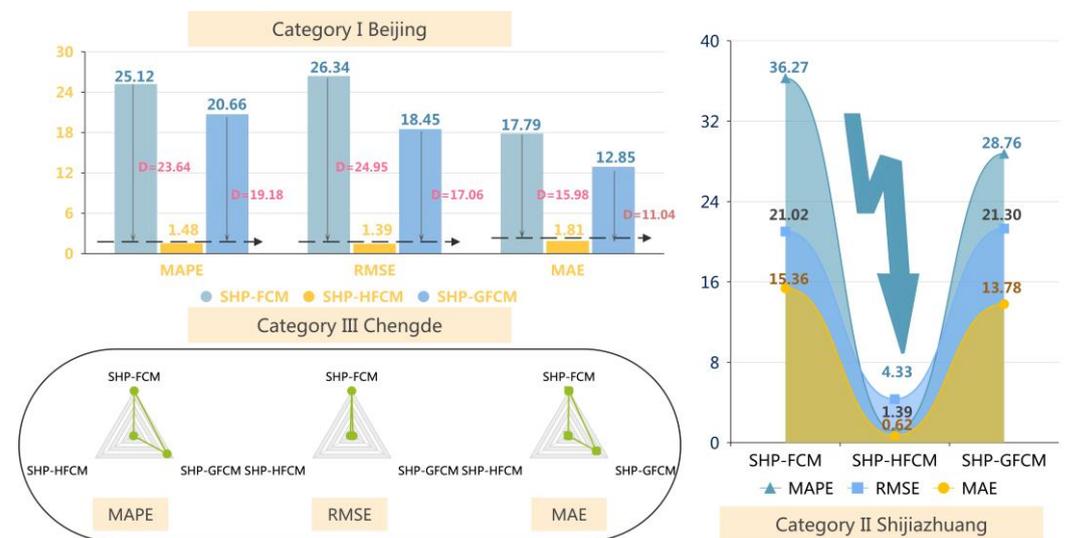
### 3.3.2. Probabilistic Hesitation Fuzzy Set Strategy

In the previous section, it was known that SHP is superior to DHP and HHP through comparative experiments. In the setting of SHP to extract data features, in order to verify the effectiveness of the HFCM prediction system constructed by combining the hesitant fuzzy information set strategy, FCM and Gaussian smooth fuzzy cognitive maps (GFCM) are also constructed for comparative experiments. The experimental results are shown in Table 6, Figures 6 and 7.

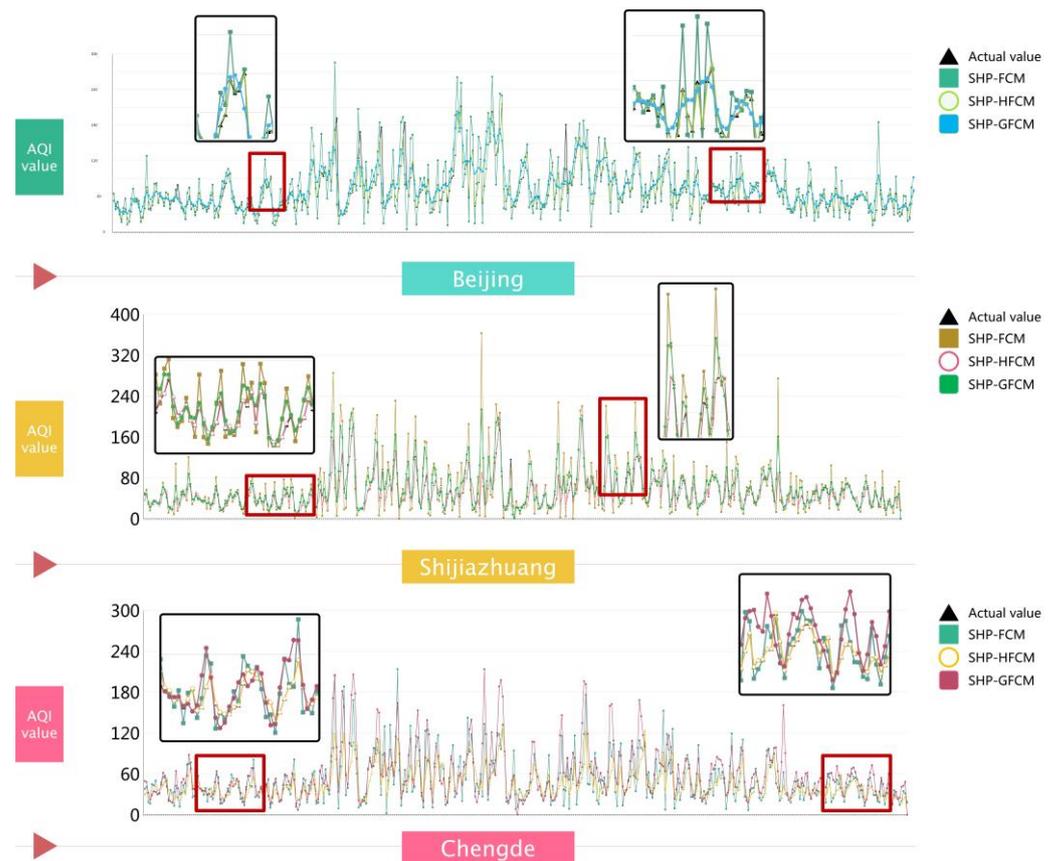
**Table 6.** Comparison of forecasting effects based on different submodules of SHP.

Cities	Method	MAPE (%)	RMSE	MAE
Category I Beijing	SHP-FCM	25.12	26.34	17.79
	SHP-GFCM	20.66	18.45	12.85
	SHP-HFCM	<b>1.48</b>	<b>1.39</b>	<b>1.81</b>
Category II Shijiazhuang	SHP-FCM	36.27	21.02	15.36
	SHP-GFCM	28.76	21.30	13.78
	SHP-HFCM	<b>1.12</b>	<b>4.33</b>	<b>0.69</b>
Category III Chengde	SHP-FCM	27.78	19.22	13.29
	SHP-GFCM	24.14	1.11	9.75
	SHP-HFCM	<b>0.49</b>	<b>0.62</b>	<b>0.23</b>

**Note:** FCM: fuzzy cognitive maps; HFCM: hesitant fuzzy cognitive maps. GFCM: Gaussian smoothed fuzzy cognitive maps.



**Figure 6.** Comparison of performance indicators of hybrid systems based on SHP.



**Figure 7.** The result of a hybrid model built from raw data.

The results in Table 6 showed that adaptive feature extraction using the Hampel filtering algorithm on the original data significantly improved forecasting accuracy compared to the direct use of fuzzy cognitive maps. In the comparison of the first type of model to improve the forecasting effect of the original single model, in the case of MAPE, the proposed model decreased by 94.1083% compared with FCM. And GFCM decreased by 17.7548% compared with FCM. In addition, MAE and RMSE also showed different degrees of improvement in each type of urban cluster. Taking Beijing as an example, SHP-FCM and SHP-GFCM are 24.95 and 17.06 higher than the RMSE of SHP-HFCM, respectively. The proposed model SHP-HFCM in Shijiazhuang has MAPE, RMSE, and MAE of 1.12%, 4.33, and 0.69, respectively, which are significantly less than the other two comparison models.

In particular, the third type of urban agglomeration has the best forecasting accuracy. In all evaluation indicators, the feature extraction strategy based on SHP is significantly better than the hybrid forecasting system based on ordinary fuzziness and Gaussian smoothing.

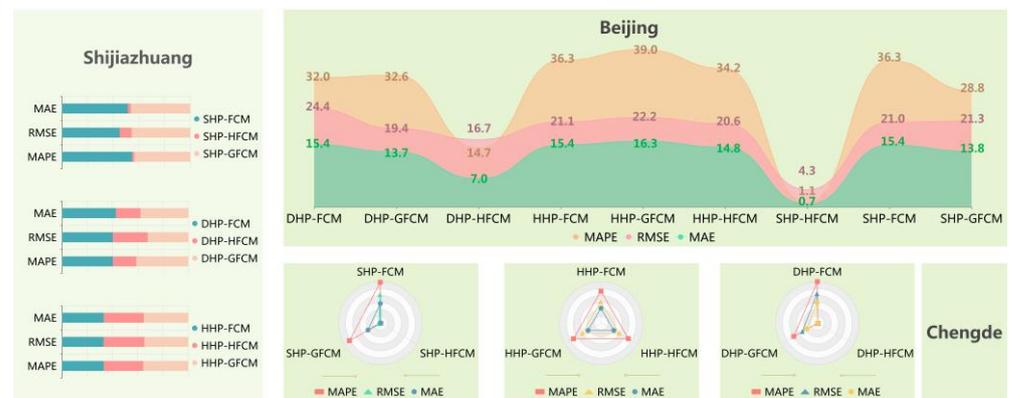
**Remark 2.** HFCM has non-negligible advantages over FCM and GFCM in data forecasting. The proposed novel hesitant fuzzy cognitive maps can not only effectively eliminate uncertain information and unstable elements in time series but also be a promising method for processing the characteristics of time series forecast data itself.

### 3.3.3. Comparison of Mixed Models in Different Data Preprocessing Environments

In addition, to show the scalability of hesitant, ambiguous information, the third comparison model is established. The FCM, HFCM, and GFCM data processing methods in the SHP, DHP, and HHP cases were compared. The results are shown in Table 7 and Figure 8.

**Table 7.** Comparison of hybrid models based on various feature extraction strategies.

Cities	Method	MAPE (%)	RMSE	MAE
Category I Beijing	DHP-FCM	39.47	37.52	24.25
	DHP-GFCM	30.21	26.19	18.98
	DHP-HFCM	21.45	21.74	13.48
	HHP-FCM	25.40	21.94	15.91
	HHP-GFCM	27.24	23.24	16.89
	HHP-HFCM	25.00	21.48	15.58
	SHP-FCM	25.12	26.34	17.79
	SHP-GFCM	20.66	18.45	12.85
Category II Shijiazhuang	DHP-FCM	31.96	24.38	15.43
	DHP-GFCM	32.64	19.40	13.66
	DHP-HFCM	14.66	16.72	7.01
	HHP-FCM	36.25	21.08	15.40
	HHP-GFCM	38.97	22.18	16.30
	HHP-HFCM	34.24	20.64	14.83
	SHP-FCM	36.27	21.02	15.36
	SHP-GFCM	28.76	21.30	13.78
Category III Chengde	DHP-FCM	28.85	20.31	13.93
	DHP-GFCM	18.67	11.87	8.38
	DHP-HFCM	0.53	0.47	0.25
	HHP-FCM	29.56	19.62	13.87
	HHP-GFCM	29.27	19.61	13.82
	HHP-HFCM	29.46	19.63	13.87
	SHP-FCM	27.78	19.22	13.29
	SHP-GFCM	24.14	1.11	9.75
SHP-HFCM	<b>0.49</b>	<b>0.62</b>	<b>0.23</b>	



**Figure 8.** Comparison of performance indicators of hybrid models based on different feature extraction strategies.

Table 7 shows the performance indicator values for all submodules launched for the three considered environments. Especially taking the third type of urban agglomeration as an example,  $SHP - HFCM_{Chengde}^{MAPE} = 0.49$ ,  $SHP - HFCM_{Chengde}^{RMSE} = 0.62$ ,  $SHP - HFCM_{Chengde}^{MAE} = 0.23$ , so for the same data preprocessing environment, the model using the HFCM method is better than the model using the FCM and the GFCM. In addition, the MAPE values for various hybrid forecasting systems are  $SHP - HFCM_{Beijing}^{MAPE} = 1.48$ ,  $SHP - HFCM_{Shijiazhuang}^{MAPE} = 1.12$ ,  $SHP - HFCM_{Chengde}^{MAPE} = 0.62$ , thus it is concluded that SSA plays an important role in HFCM forecasting. After comprehensive comparison, the proposed model has a good effect in the forecasting of pollutant concentration series.

**Remark 3.** In different data preprocessing scenarios, the submodule with the best predictive performance is uniquely identified. The hybrid forecasting system based on HFCM constructed in this paper achieves the best results of all evaluation indicators under different data processing environments. This property has been demonstrated in three different scenarios.

#### 4. Discussion

##### 4.1. Robustness of the Proposed Model

The purpose of the robustness test is to investigate whether the daily data are non-stationary when the air quality receives the impact of high pollution. At this time, whether the model can still work normally and whether the forecasting accuracy fluctuates greatly are evaluated in this test. In this experiment, the data of the training set increases the random number in the range of  $(-2, 2)$ , which is considered to be from random interference; then, observe the changes of each performance index; the comparison results are shown in Table 8.

**Table 8.** The results of robustness test.

	MAPE (%)			RMSE			MAE		
	Random	Proposed	Change	Random	Proposed	Change	Random	Proposed	Change
<b>Beijing</b>									
FCM	32.15	25.54	6.61	30.15	25.15	5.00	20.26	17.76	2.5
SHP-FCM	30.72	25.12	5.6	28.61	26.34	2.27	19.89	17.79	2.1
SHP-HFCM	<b>1.39</b>	<b>1.48</b>	<b>0.09</b>	<b>1.58</b>	<b>1.39</b>	<b>0.19</b>	<b>1.76</b>	<b>1.81</b>	<b>0.05</b>
Mean	21.42	17.38	4.04	20.11	17.63	2.49	13.97	12.45	1.52
Std	17.36	13.77	3.59	16.07	14.07	1.99	10.58	9.22	1.36
<b>Shijiazhuang</b>									
FCM	45.49	40.51	4.98	28.28	37.73	9.45	19.63	21.18	1.55
SHP-FCM	45.71	36.27	9.44	28.16	21.02	7.14	19.57	15.36	4.21
SHP-HFCM	<b>0.83</b>	<b>1.12</b>	<b>0.29</b>	<b>4.17</b>	<b>4.33</b>	<b>0.16</b>	<b>0.52</b>	<b>0.69</b>	<b>0.17</b>
Mean	30.68	25.97	4.71	20.20	21.03	0.82	13.24	12.41	0.83
Std	25.85	21.62	4.23	13.89	16.70	2.81	11.02	10.56	0.46
<b>Chengde</b>									
FCM	32.26	29.50	2.76	20.34	20.66	0.32	13.81	14.25	0.44
SHP-FCM	32.04	27.78	4.26	20.55	19.22	1.33	13.98	13.29	0.69
SHP-HFCM	<b>0.67</b>	<b>0.49</b>	<b>0.18</b>	<b>0.50</b>	<b>0.62</b>	<b>0.12</b>	<b>0.31</b>	<b>0.23</b>	<b>0.08</b>
Mean	21.66	19.26	2.40	13.80	13.50	0.30	9.37	9.26	0.11
Std	18.18	16.28	1.90	11.52	11.18	0.34	7.84	7.83	0.01

**Note:** **Random** represents adding random disturbance to simulate air situation. **Proposed** represents the original data. **Change** is the absolute value of the difference between the data after adding the disturbance and the accuracy data model, that is, the magnitude of the change.

Before and after adding random disturbance, the MAPE values of Beijing were 1.39% and 1.48%, respectively; only 0.09% changed, and the same was true for other cities. After adding random disturbances, the forecasting accuracy of the proposed model changed slightly: Shijiazhuang changed by 0.29%, and Chengde changed by 0.18%. It can also be seen from Table 7 that the average RMSE of the three observation cities of the original data model is 17.63, 21.03, and 13.50, respectively. Compared with the model with increased disturbance, the values of MAPE change by 2.49, 0.82, and 0.3, respectively, which clarifies that the random disturbance will not affect the forecasting performance. In addition, for the original data model, taking the predicted value of Beijing as an example, the average MAE is 12.45, and the standard deviation is 9.22. From the model, after increasing the disturbance, the average MAE increases slightly, the standard deviation increases to 13.97, the change range is 1.52, and the standard deviation change range is 1.36, which indicates that the random disturbance is not significant.

In summary, by observing the fluctuation of the forecasting results of each city, the forecasting performance of the model proposed in this paper has not changed significantly

and still maintains an outstanding forecasting effect. Therefore, there is enough evidence to prove that although the air quality will be affected by other factors, such as high pollution or other abnormal conditions, the proposed model can still have stability.

#### 4.2. Differences of the Proposed Model

The discrepancy test of the forecasting system proposed in this paper will be further discussed in this section. To test whether the forecasting performance of the proposed hybrid forecasting model is significantly different from other forecasting models, the DM test and the MDM test are used to perform statistical tests.

From the experimental results in Table 9, the loss series of the proposed hybrid forecasting system passed the significance test at the 5% confidence level. It denotes that the proposed hybrid forecasting system has higher forecasting efficiency than other hybrid strategies, and it is not a coincidence that the forecasting is significantly different.

**Table 9.** The results of testing model differences.

Category	Method	DM	P	MDM	P
Category I Beijing	DHP-FCM	15.2192	$8.39 \times 10^{-42}$	15.2007	$1.01 \times 10^{-41}$
	DHP-GFCM	9.7186	$3.11 \times 10^{-20}$	9.7068	$3.42 \times 10^{-20}$
	DHP-HFCM	8.0531	$3.88 \times 10^{-25}$	8.0423	$2.63 \times 10^{-25}$
	HHP-FCM	13.6566	$2.83 \times 10^{-35}$	13.6401	$3.31 \times 10^{-35}$
	HHP-GFCM	5.8835	$8.34 \times 10^{-9}$	5.8764	$8.67 \times 10^{-9}$
	HHP-HFCM	5.1544	$3.26 \times 10^{-7}$	5.1478	$3.19 \times 10^{-7}$
	SHP-FCM	5.7104	$2.16 \times 10^{-8}$	5.7034	$2.25 \times 10^{-8}$
Category II Shijiazhuang	SHP-GFCM	6.4946	$2.41 \times 10^{-10}$	6.4867	$2.52 \times 10^{-10}$
	DHP-FCM	16.4863	$3.10 \times 10^{-47}$	16.4663	$3.78 \times 10^{-47}$
	DHP-GFCM	10.0901	$1.55 \times 10^{-21}$	10.0778	$1.71 \times 10^{-21}$
	DHP-HFCM	5.9216	$8.34 \times 10^{-11}$	5.9148	$8.34 \times 10^{-11}$
	HHP-FCM	15.9536	$6.12 \times 10^{-45}$	15.9342	$7.42 \times 10^{-45}$
	HHP-GFCM	9.5553	$1.14 \times 10^{-19}$	9.5437	$1.25 \times 10^{-19}$
	HHP-HFCM	9.1332	$1.25 \times 10^{-13}$	9.1277	$1.36 \times 10^{-13}$
Category III Chengde	SHP-FCM	10.5755	$2.79 \times 10^{-23}$	10.5626	$3.11 \times 10^{-23}$
	SHP-GFCM	11.6521	$2.63 \times 10^{-27}$	11.6380	$2.98 \times 10^{-27}$
	DHP-FCM	13.0401	$4.63 \times 10^{-41}$	13.0242	$5.54 \times 10^{-41}$
	DHP-GFCM	12.9219	$8.93 \times 10^{-34}$	12.9062	$1.04 \times 10^{-33}$
	DHP-HFCM	1.4234	$3.74 \times 10^{-51}$	1.4217	$4.62 \times 10^{-51}$
	HHP-FCM	10.3126	$4.40 \times 10^{-43}$	10.3001	$5.30 \times 10^{-43}$
	HHP-GFCM	10.2920	$1.39 \times 10^{-39}$	10.2795	$1.65 \times 10^{-39}$
HHP-HFCM	10.3339	$2.25 \times 10^{-36}$	10.3214	$2.64 \times 10^{-36}$	
SHP-FCM	13.4234	$9.09 \times 10^{-33}$	13.4071	$1.05 \times 10^{-32}$	
SHP-GFCM	13.2622	$2.72 \times 10^{-32}$	13.2461	$3.14 \times 10^{-32}$	

**Note:** The model SHP-HFCM proposed in this study is used for comparison with the above combinatorial strategy, where the statistic  $p$ -value represents the probability.

Compared with other models, the model proposed in this paper has the best results in any possible data environment. On the one hand, it performs spatial effect analysis and data feature extraction before data forecasting, which can better utilize the spatial and data features of the data for subsequent forecasting, and this is the excellence of the model. On the other hand, the system continuously updates the output conceptual values after  $k$  iterations by combining the geographic location of each data and does not directly output the forecasting results, but it eliminates the uniqueness of the data itself through fuzzy time series, and the forecasting results are more superior.

#### 4.3. Application of the Proposed Model

The proposed model not only has outstanding forecasting performance but also has potential functions in practical applications. The early warning system consists of three

modules: spatial feature analysis module, data preprocessing module, and fuzzy information prediction module. The following analyzes its practical significance from these three modules.

1. Air pollution has spatial spillover effects, so the spatial feature analysis module can comprehensively consider the interaction of urban agglomerations. In addition, it also provides theoretical support for the formulation of pollution control policies among urban agglomerations and facilitates coordinated governance among urban agglomerations.
2. In fact, air quality is affected by many factors, including information ambiguity and uncertainty. Through the data preprocessing module, it can remove the outliers and noise from the data, making the data features more obvious and achieving better forecasting performance.
3. Accurate AQI forecasting results can provide early warning information for actual life and production activities. From the perspective of the public, the forecasting of air quality can let the public understand the air quality, the scope of air quality deterioration, the degree of deterioration, and the development trend; secondly, it guides the daily activities and behaviors of residents, protects the physical and mental health of the people, and reduces the incidence of diseases. From the perspective of social economy, it can not only provide the theoretical basis for pollution control measures, such as strict control of motor vehicle pollution, reduction of coal consumption, shut-down of polluting enterprises, control of construction sites and road dust, and supervision of factories with large pollutant emissions.

## 5. Conclusions

In terms of application, it is profoundly important to make accurate forecasts of air quality. The air environment is an important guarantee for people's health and production safety, and understanding real-time AQI can provide a theoretical basis for establishing an early warning system on the one hand, and a tool for implementing pollution control policies on the other. However, because AQI has the complexity and nonlinearity of time series changes, and has spatial correlation among urban clusters, it is difficult to make accurate forecasting. Coupled with the fact that it is influenced by many factors and prone to fluctuations and singular values, it is even more challenging for the forecasting effect of the model. Therefore, this paper proposes a novel spatiotemporal hybrid forecasting system to realize the time series forecasting of AQI. Three comparative experiments were established using the air quality indices of thirteen cities in the Beijing-Tianjin-Hebei region of China to analyze and demonstrate, and the results are as follows:

1. First, the spatial feature extraction module is built. The module successfully extracted the spatial overflow features, captured the dynamic transition of air quality, and performed cluster analysis with different sizes and weights for irregular data.
2. The adaptive Hampel filtering model improved by SSA is the best data processing sub-module for comparison with FCM, DHP-FCM, and HHP-FCM.
3. The first proposed HFCM forecasting model plays an irreplaceable role in time series forecasting in the same data preprocessing environment. The model reduced MAPE by 94.1083%, 96.9120%, and 98.2361% for three different urban clusters.
4. In the environment of different data preprocessing methods, the model proposed in this paper can still make accurate forecasts for data with large fluctuations and mutations. MAPE, RMSE, and MAE reach the minimum values in the three urban agglomerations.

In summary, the proposed air quality forecasting system has outstanding forecasting performance in handling low-quality and large-noise data. Although the model makes an accurate prediction by analyzing the spatial correlation between the target city and the adjacent city and excavating its internal relationship, this study also has shortcomings, and the stability needs to be improved when optimizing the Hampel parameter values for single objectives. In future research, we can consider a multi-objective optimization algorithm

and select other urban agglomerations for research to further verify the superiority of the forecasting model proposed in this paper for air quality index forecasting.

**Author Contributions:** X.G. conceptualization, methodology, software, writing—original draft preparation; H.L. validation, formal analysis, writing—review and editing, supervision; H.F. methodology, resources, data curation, investigation. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Youth Program of National Natural Science Foundation of China (Grant No. 72201054), Excellent Youth Program of Natural Science Foundation of Heilongjiang Province of China (Grant No. YQ2022G001), Youth Program of Philosophy and Social Science planning project (Grant No. 22GLC278) and Central Universities Basic Research Business Special Funds project (Grant No. 2572021BM01).

**Data Availability Statement:** No additional data are available.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Nomenclature

The abbreviations for each name are as follows:

AR	Auto-regressive model
ANN	Artificial neural network
API	Air pollution index
AQI	Air quality index
ARMA	Auto-regressive moving average model
CDF	Cumulative distribution function
CE	Centrality
CO	Coordination
DA	Dragonfly algorithm
DM	Diebold–Mariano
FCM	Fuzzy cognitive maps
FLG	Fuzzy logic group
FLR	Fuzzy logic relation
FTS	Fuzzy time series
HBA	Honey badger algorithm
HFCM	Hesitant fuzzy cognitive maps
LGC	Local gravitational clustering
LRF	Local resultant force
MA	Moving average model
MAD	Median absolute difference
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MDM	Modified Diebold–Mariano
RF	Random foresting
RMSE	Root mean square error
RNN	Recurrent neural network
SSA	Squirrel search algorithm
SVM	Support vector machine

### References

1. Wang, Y.; Yao, L.; Xu, Y.; Sun, S.; Li, T. Potential Heterogeneity in the Relationship between Urbanization and Air Pollution, from the Perspective of Urban Agglomeration. *J. Clean. Prod.* **2021**, *298*, 126822. [[CrossRef](#)]
2. Ulpiani, G. On the Linkage between Urban Heat Island and Urban Pollution Island: Three-Decade Literature Review towards a Conceptual Framework. *Sci. Total Environ.* **2021**, *751*, 141727. [[CrossRef](#)] [[PubMed](#)]
3. Wei, J.; Guo, X.; Marinova, D.; Fan, J. Industrial SO<sub>2</sub> Pollution and Agricultural Losses in China: Evidence from Heavy Air Polluters. *J. Clean. Prod.* **2014**, *64*, 404–413. [[CrossRef](#)]
4. de Groot, P.; Munden, R.F. Lung Cancer Epidemiology, Risk Factors, and Prevention. *Radiol. Clin. N. Am.* **2012**, *50*, 863–876. [[CrossRef](#)] [[PubMed](#)]

5. Poursafa, P.; Mansourian, M.; Motlagh, M.-E.; Ardalani, G.; Kelishadi, R. Is Air Quality Index Associated with Cardiometabolic Risk Factors in Adolescents? The CASPIAN-III Study. *Environ. Res.* **2014**, *134*, 105–109. [[CrossRef](#)]
6. Rajagopalan, S.; Brook, R.D. Air Pollution and Type 2 Diabetes: Mechanistic Insights. *Diabetes* **2012**, *61*, 3037–3045. [[CrossRef](#)]
7. Chen, P.; Yuan, Z.; Miao, L.; Yang, L.; Wang, H.; Xu, D.; Lin, Z. Acute Cardiorespiratory Response to Air Quality Index in Healthy Young Adults. *Environ. Res.* **2022**, *214*, 113983. [[CrossRef](#)]
8. Soman, S.S.; Zareipour, H.; Malik, O.; Mandal, P. A Review of Wind Power and Wind Speed Forecasting Methods with Different Time Horizons. In Proceedings of the North American Power Symposium 2010, Online, 26–28 September 2010; pp. 1–8.
9. Li, Y.; He, Y.; Su, Y.; Shu, L. Forecasting the Daily Power Output of a Grid-Connected Photovoltaic System Based on Multivariate Adaptive Regression Splines. *Appl. Energy* **2016**, *180*, 392–401. [[CrossRef](#)]
10. You, J.; Yu, C.; Sun, J.; Chen, J. Generalized Maximum Entropy Based Identification of Graphical ARMA Models. *Automatica* **2022**, *141*, 110319. [[CrossRef](#)]
11. Liu, C.-C.; Lin, T.-C.; Yuan, K.-Y.; Chiueh, P.-T. Spatio-Temporal Prediction and Factor Identification of Urban Air Quality Using Support Vector Machine. *Urban Clim.* **2022**, *41*, 101055. [[CrossRef](#)]
12. Murillo-Escobar, J.; Sepulveda-Suescun, J.P.; Correa, M.A.; Orrego-Metaute, D. Forecasting Concentrations of Air Pollutants Using Support Vector Regression Improved with Particle Swarm Optimization: Case Study in Aburrá Valley, Colombia. *Urban Clim.* **2019**, *29*, 100473. [[CrossRef](#)]
13. Yoo, B.H.; Kim, K.S.; Park, J.Y.; Moon, K.H.; Ahn, J.J.; Fleisher, D.H. Spatial Portability of Random Forest Models to Estimate Site-Specific Air Temperature for Prediction of Emergence Dates of the Asian Corn Borer in North Korea. *Comput. Electron. Agric.* **2022**, *199*, 107113. [[CrossRef](#)]
14. Noi, P.T.; Degener, J.; Kappas, M. Comparison of Multiple Linear Regression, Cubist Regression, and Random Forest Algorithms to Estimate Daily Air Surface Temperature from Dynamic Combinations of MODIS LST Data. *Remote Sens.* **2017**, *9*, 398. [[CrossRef](#)]
15. Goudarzi, G.; Hopke, P.K.; Yazdani, M. Forecasting PM2.5 Concentration Using Artificial Neural Network and Its Health Effects in Ahvaz, Iran. *Chemosphere* **2021**, *283*, 131285. [[CrossRef](#)] [[PubMed](#)]
16. Fallahzadeh, S.; Kermani, M.; Esrafil, A.; Asadgol, Z.; Gholami, M. The Effects of Meteorological Parameters on PM10: Health Impacts Assessment Using AirQ+ Model and Prediction by an Artificial Neural Network (ANN). *Urban Clim.* **2021**, *38*, 100905. [[CrossRef](#)]
17. Zhang, J.; Li, S. Air Quality Index Forecast in Beijing Based on CNN-LSTM Multi-Model. *Chemosphere* **2022**, *308*, 136180. [[CrossRef](#)]
18. Zhu, S.; Lian, X.; Liu, H.; Hu, J.; Wang, Y.; Che, J. Daily Air Quality Index Forecasting with Hybrid Models: A Case in China. *Environ. Pollut.* **2017**, *231*, 1232–1244. [[CrossRef](#)]
19. Liu, Z.; Liu, J. A Robust Time Series Prediction Method Based on Empirical Mode Decomposition and High-Order Fuzzy Cognitive Maps. *Knowl.-Based Syst.* **2020**, *203*, 106105. [[CrossRef](#)]
20. Yang, S.; Liu, J. Time-Series Forecasting Based on High-Order Fuzzy Cognitive Maps and Wavelet Transform. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 3391–3402. [[CrossRef](#)]
21. Liang, W.; Zhang, Y.; Liu, X.; Yin, H.; Wang, J.; Yang, Y. Towards Improved Multifactorial Particle Swarm Optimization Learning of Fuzzy Cognitive Maps: A Case Study on Air Quality Prediction. *Appl. Soft Comput.* **2022**, *130*, 109708. [[CrossRef](#)]
22. Korkmaz, D. SolarNet: A Hybrid Reliable Model Based on Convolutional Neural Network and Variational Mode Decomposition for Hourly Photovoltaic Power Forecasting. *Appl. Energy* **2021**, *300*, 117410. [[CrossRef](#)]
23. Wu, Q.; Lin, H. Daily Urban Air Quality Index Forecasting Based on Variational Mode Decomposition, Sample Entropy and LSTM Neural Network. *Sustain. Cities Soc.* **2019**, *50*, 101657. [[CrossRef](#)]
24. Zhu, S.; Lian, X.; Wei, L.; Che, J.; Shen, X.; Yang, L.; Qiu, X.; Liu, X.; Gao, W.; Ren, X.; et al. PM2.5 Forecasting Using SVR with PSO-GSA Algorithm Based on CEEMD, GRNN and GCA Considering Meteorological Factors. *Atmos. Environ.* **2018**, *183*, 20–32. [[CrossRef](#)]
25. Li, H.; Wang, J.; Lu, H.; Guo, Z. Research and Application of a Combined Model Based on Variable Weight for Short Term Wind Speed Forecasting. *Renew. Energy* **2018**, *116*, 669–684. [[CrossRef](#)]
26. Yang, W.; Hao, M.; Hao, Y. Innovative ensemble system based on mixed frequency modeling for wind speed point and interval forecasting. *Inf. Sci.* **2023**, *622*, 560–586. [[CrossRef](#)]
27. Yang, W.; Sun, S.; Hao, Y.; Wang, S. A novel machine learning-based electricity price forecasting model based on optimal model selection strategy. *Energy* **2022**, *238*, 121989. [[CrossRef](#)]
28. Lv, Y.-W.; Yang, G.-H. Centralized and Distributed Adaptive Cubature Information Filters for Multi-Sensor Systems with Unknown Probability of Measurement Loss. *Inf. Sci.* **2023**, *630*, 173–189. [[CrossRef](#)]
29. Liu, X.; Zhong, S.; Li, S.; Yang, M. Evaluating the Impact of Central Environmental Protection Inspection on Air Pollution: An Empirical Research in China. *Process Saf. Environ. Prot.* **2022**, *160*, 563–572. [[CrossRef](#)]
30. Moran, P.A.P. The Interpretation of Statistical Maps. *J. R. Stat. Soc. Ser. B* **1948**, *10*, 243–251. [[CrossRef](#)]
31. Zhang, C.; Luo, L.; Xu, W.; Ledwith, V. Use of Local Moran's I and GIS to Identify Pollution Hotspots of Pb in Urban Soils of Galway, Ireland. *Sci. Total Environ.* **2008**, *398*, 212–221. [[CrossRef](#)]
32. On Extreme Values of Moran's I and Geary's c—Jong—1984—Geographical Analysis—Wiley Online Library. Available online: <https://onlinelibrary.wiley.com/doi/10.1111/j.1538-4632.1984.tb00797.x> (accessed on 8 May 2023).

33. Sadeghi, M.; Naghedi, R.; Behzadian, K.; Shamshirgaran, A.; Tabrizi, M.R.; Maknoon, R. Customisation of Green Buildings Assessment Tools Based on Climatic Zoning and Experts Judgement Using K-Means Clustering and Fuzzy AHP. *Build. Environ.* **2022**, *223*, 109473. [[CrossRef](#)]
34. Setiawan, K.E.; Kurniawan, A.; Chowanda, A.; Suhartono, D. Clustering Models for Hospitals in Jakarta Using Fuzzy C-Means and k-Means. *Procedia Comput. Sci.* **2023**, *216*, 356–363. [[CrossRef](#)] [[PubMed](#)]
35. Kaloni, S.; Singh, G.; Tiwari, P. Nonparametric Damage Detection and Localization Model of Framed Civil Structure Based on Local Gravitation Clustering Analysis. *J. Build. Eng.* **2021**, *44*, 103339. [[CrossRef](#)]
36. Liu, H.; Shah, S.; Jiang, W. On-Line Outlier Detection and Data Cleaning. *Comput. Chem. Eng.* **2004**, *28*, 1635–1647. [[CrossRef](#)]
37. Allen, D.P. A Frequency Domain Hampel Filter for Blind Rejection of Sinusoidal Interference from Electromyograms. *J. Neurosci. Methods* **2009**, *177*, 303–310. [[CrossRef](#)]
38. Song, Q.; Chissom, B.S. Forecasting Enrollments with Fuzzy Time Series—Part II. *Fuzzy Sets Syst.* **1994**, *62*, 1–8. [[CrossRef](#)]
39. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **2002**, *20*, 134–144. [[CrossRef](#)]
40. Harvey, D.; Leybourne, S.; Newbold, P. Testing the Equality of Prediction Mean Squared Errors. *Int. J. Forecast.* **1997**, *13*, 281–291. [[CrossRef](#)]
41. Zheng, D.; Shi, M.; Pang, R. Agglomeration Economies and Environmental Regulatory Competition: Evidence from China. *J. Clean. Prod.* **2021**, *280*, 124506. [[CrossRef](#)]
42. Hao, Y.; Guo, Y.; Li, S.; Luo, S.; Jiang, X.; Shen, Z.; Wu, H. Towards Achieving the Sustainable Development Goal of Industry: How Does Industrial Agglomeration Affect Air Pollution? *Innov. Green Dev.* **2022**, *1*, 100003. [[CrossRef](#)]
43. Tan, G.; Cao, Y.; Xie, R.; Fang, J. Intergovernmental Competition, Industrial Spatial Distribution, and Air Quality in China. *J. Environ. Manag.* **2022**, *310*, 114721. [[CrossRef](#)]
44. Cheng, Y.; Loo, B.P.Y.; Vickerman, R. High-Speed Rail Networks, Economic Integration and Regional Specialisation in China and Europe. *Travel Behav. Soc.* **2015**, *2*, 1–14. [[CrossRef](#)]
45. Zhao, C.; Wang, B. How Does New-Type Urbanization Affect Air Pollution? Empirical Evidence Based on Spatial Spillover Effect and Spatial Durbin Model. *Environ. Int.* **2022**, *165*, 107304. [[CrossRef](#)] [[PubMed](#)]
46. Ruggieri, M.; Plaia, A. An Aggregate AQI: Comparing Different Standardizations and Introducing a Variability Index. *Sci. Total Environ.* **2012**, *420*, 263–272. [[CrossRef](#)]
47. GB3095-2012; National Ambient Air Quality Standard of China. Ministry of Ecology and Environment of the People’s Republic of China: Beijing, China, 2012.
48. Chen, W.; Tang, H.; Zhao, H. Urban Air Quality Evaluations under Two Versions of the National Ambient Air Quality Standards of China. *Atmos. Pollut. Res.* **2016**, *7*, 49–57. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.