



Article Monitoring the Influence of Industrialization and Urbanization on Spatiotemporal Variations of AQI and PM_{2.5} in Three Provinces, China

Hu Chen D, Guoqu Deng * and Yiwen Liu

School of Management, Henan University of Science and Technology, Luoyang 471023, China * Correspondence: dengguoqu@haust.edu.cn

Abstract: With the rapid development of industrialization and urbanization, atmospheric pollution research is vital for regional sustainable development and related policies formulated by the government. Previous studies have mainly studied a single evaluation method to analyze the air quality index (AQI) or single air pollutant. This research integrated the Spearman coefficient (SC) correlation analysis, a random search (RS) algorithm and an excellent extreme gradient boosting (XGBoost) algorithm to evaluate the air pollution influence of industrialization and urbanization (APIIU). Industrialization, urbanization and meteorological indicators were used to measure the influence degree of APIIU on AQI and particulate matter 2.5 (PM_{2.5}), respectively. The main findings were: (1) the APIIU-AQI and APIIU-PM_{2.5} of Henan Province, Hubei Province and Hunan Province had significant changes from 2017 to 2019; (2) the value of square of determination coefficient of real value (R^2), the root mean square error (RMSE) and the mean absolute percentage error (MAPE) of APIIU-AQI and APIIU-PM_{2.5} in three provinces predicted by the SC-RS-XGBoost were 0.945, 0.103, 4.25% and 0.897, 0.205, 4.84%, respectively; (3) the predicted results were more accurate than using a SC-XGBoost, RS-XGBoost, traditional XGBoost, support vector regression (SVR) and extreme learning machine (ELM).

Keywords: air pollution; industrialization and urbanization; SC-RS-XGBoost; air quality; particulate matter

1. Introduction

Air quality has become a more significant problem with the development of industrialization and urbanization [1–4]. The remarkable economic growth has resulted in serious environmental issues [5,6]. The emissions of air pollutants from industrial and motor vehicles are currently the most important environmental risk to human health [7]. According to an assessment report from the World Health Organization, urban air pollution has resulted in more than two million deaths every year [8,9]. In the past 20 years, the European Union has made great policy progress in atmospheric emissions and air quality. However, air pollution continues to have a serious influence on the health of people who live in Europe's urban areas [10,11]. As one of the rapidly developing countries, China has experienced rapid economy growth in the past several decades. As a result, industrial activities, urban expansion and human engineering activities have produced a high intensity of air pollutant emissions in China [12–16]. Hence, reducing air pollution is a critical measure for environmental protection and sustainable development [17].

From 2010, China has actively implemented the Clean Air Action to deal with air pollution, and many pollutants' emissions have decreased since then [18]. To effectively assess air quality and provide guidance for outdoor activities, the Chinese Ministry of Environmental Protection (MEP) has adopted and developed an air quality index (AQI) system in 2012 [19–21]. The existing related research works elaborately documented that the six major air pollutants were particulate matter 2.5 (PM_{2.5}), particulate matter 10 (PM₁₀), sulfur oxides (SO₂), nitrogen oxides (NO₂), carbon mono oxide (CO) and ozone (O₃). The



Citation: Chen, H.; Deng, G.; Liu, Y. Monitoring the Influence of Industrialization and Urbanization on Spatiotemporal Variations of AQI and PM_{2.5} in Three Provinces, China. *Atmosphere* 2022, *13*, 1377. https:// doi.org/10.3390/atmos13091377

Academic Editor: Kumar Vikrant

Received: 19 July 2022 Accepted: 25 August 2022 Published: 28 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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/). AQI converted six pollutants into a comparable and easily understandable index [22–24]. Despite improvements over several years, China, with a sizeable economic aggregate and dense population, continues to cause air pollution. Additionally, the economic development model with high energy consumption and low efficiency is an important factor that leads to air pollution in China [25–28]. The MEP released the 2019 China Ecological and Environmental Bulletin in June 2020, which reported that only 157 cities reached the air quality standards among the 337 cities. Accordingly, a total of 1666 days of heavy pollution and 452 days of severe pollution were reported in 337 cities during 2019. Air pollution has caused an obviously serious impact on residents' daily lives.

Considering the above existing problems, the prevention and prediction of air pollution have become the focus of researchers all over the world. Nowadays, the related research works have mainly focused on the precise prediction of AQI and other air pollutant levels, which was important for the early air quality warning and policymakers' work. Many AQI forecasting models have been presented in recent years, including physical models, statistical models and hybrid models. Among them, the physical methods were more complicated and consuming too much time, and statistical models have been proven to outperform the physical methods [29–33]. The hybrid models combined data decomposition, feature extraction and optimization methods to improve the forecasting performance. Munawar et al. [34] adopted a neuro-fuzzy inference system to forecast AQI and presented a case study of Lahore city of Pakistan. Wang et al. [35] used a hybrid forecasting model combining a two-phase decomposition technique and extreme learning machine (ELM) optimized by differential evolution (DE) algorithm to forecast AQI. Li et al. [36] used hybrid models based on the auto regressive integrated moving average model (ARIMA), optimized extreme learning machine (OELM) and fuzzy time series model (FTSM) to forecast the reconstructed series in the dynamic integration forecasting module. By employing time varying parameters to dynamically combine the forecasting results, the accuracy was superior to other forecasting models. Phruksahiran [37] adopted an ensemble prediction methodology and used the additional predictor variables for predicting AQI at the hourly level, which combined the geographically weighted predictor method (GWP) and the related machine-learning algorithms. The results showed that the hybrid model enhanced the accuracy of AQI prediction. Despite the better accuracy of the above models compared to the traditional prediction methods, they lacked the prediction of concentration of a single pollutant [38–42].

With the rapid development of artificial intelligence, more accurate models have been developed. Due to the advantages of robustness and self-adaptability, these models have been regarded as effective and outstanding approaches for forecasting air pollutant concentrations. Feng et al. [43] analyzed and accurately forecasted the atmospheric pollutants in Hangzhou using a hybrid model based on recurrent neural network (RNN) and random forest (RF). Franceschi et al. [44] adopted a hybrid model based on artificial neural networks (ANN), principal component analysis (PCA) and K-means clustering to forecast the concentrations of $PM_{2.5}$ and PM_{10} in Bogotá, Colombia. The above model had the ability to accurately predict air pollutants, which could be used for early warnings of high air pollution. Gao et al. [45] investigated the feasibility of using a hybrid model, which was based on ANN and Monte Carlo simulations (MCS) with meteorological parameters to predict the concentration of O_3 in the urban area of Jinan, China. Li et al. [46] designed a novel long short-term memory neural network (LSTM) extended model to predict air pollutant concentration. It considered historical air pollutant data, meteorological data and time stamp data, which were merged into the model to enhance the accuracy. Liu et al. [47] proposed a hybrid neural network model to forecast the concentration of $PM_{2.5}$. To obtain a much better performance and forecasting outputs, ELM with weighted regularization, whose parameters were optimally determined by the ANN algorithm, was used. Pak et al. [48] used a hybrid model based on convolutional neural networks (CNN) and LSTM (CNN-LSTM) to efficiently extract the inherent features of huge air quality and meteorological data; the model was proposed and used for predicting the concentration of O_3 in Beijing

City and exhibited a better stability and prediction performance. Zhou et al. [49] adopted the Gaussian process mixture (GPM) model based on an iterative learning algorithm to predict the air pollutants' concentrations. Yu et al. [50] developed a dynamic model based on ELM to forecast the concentrations of SO_2 and NO_2 . Using the quantum genetic algorithm (QGA) to optimize the connection weight threshold of the ELM, it contributed to increasing the prediction performance. Middya et al. [51] found that several models, such as the bi-directional LSTM (Bi-LSTM) and convolutional LSTM (Conv-LSTM), achieved a high forecasting accuracy for the majority of air pollutants. Sun et al. [52] established a new model based on the stacking-driven ensemble (SDE) and two kinds of input selection methods to forecast the hourly concentration of PM_{2.5} and found that the proposed model had a higher predicting accuracy, much better performance and more robust forecasting ability. Ribeiro [53] used Bayesian, econometrics and machine-learning models applied to predict the future concentration of SO_2 emissions, respectively. Finally, the machinelearning models had better generalization power than traditional methods. Although previous research works had superiority in predicting the concentrations of air pollutants, there were few studies of air quality based on analyzing the influence of relevant indicators of air quality in the process of industrialization and urbanization. Rapid economic growth and industrialization have decreased air quality in developing countries. The urbanization scale and high population density have increased air pollutant emissions through various human daily travel and life activities [54–58].

The above literature demonstrates the research on AQI and air pollutant concentrations with different methods. Currently, there is no universally agreed method of constructing an indicators system for the assessment of air pollution influence of industrialization and urbanization (APIIU). The definition of an APIIU indicators system is variable, and it can hold multiple dimensions and present different specificities depending on the province, time span or assessment target variable for air pollution. This uncertainty makes it difficult to identify consistent research methods for analysis and prediction. In this paper, considering the availability of data and the cumulative effect of development of industrialization and urbanization, our approach was to choose annual data. Since the study area and time span were selected from 49 cities in three provinces from 2017 to 2019, the above data belonged to a small sample. Extreme gradient boosting (XGBoost), support vector regression (SVR) and ELM have been proven to possess good prediction performance for small samples in past research [59–65]. Additionally, since the construction of the indicators system was mainly based on previous research, the Spearman coefficient (SC) was used for analyzing the constituent indicators of APIIU of AQI and PM_{2.5} to select indicators with a large correlation coefficient. When using XGBoost and SVR to predict data, internal parameters affect the prediction accuracy. Hence, the parameters need to be optimized to improve the performance results. XGBoost includes the general, booster and learning task parameters. Many types and quantities of parameters needed to be optimized in this research, and we adopted a random search (RS) algorithm to optimize the parameters of XGBoost and SVR [66–68]. On the one hand, the algorithm was selected to consider the efficiency of optimizing XGBoost. On the other hand, it was used for comparing the prediction performance of RS-XGBoost and RS-SVR.

This study firstly considered the industrialization, urbanization, population, regional gross domestic product (GDP) and meteorological indicators as the APIIU indicators. Then, we constructed an APIIU indicators system from the aspects of economy, society and natural environment. This research was divided into four sections. The first section introduces the research review. The Materials and Methods section introduces the research region, data sources, indicators system and the principle of the SC-RS-XGBoost. The Analytical Results section introduces the determination of each indicator's weight and calculation of APIIU, and the final section offers a discussion, our conclusions and the limitations of this research.

2. Materials and Methods

2.1. Data

2.1.1. Research Region

The research domain covers 49 cities composed of 18 cities in Henan Province, 17 cities in Hubei Province and 14 cities in Hunan Province, which is shown in Figure 1. Henan Province is located in the central part of eastern China and the middle and lower reaches of the Yellow River. It is bounded by latitude $31^{\circ}23'-36^{\circ}22'$ N and longitude $110^{\circ}21'-116^{\circ}39'$ E, with a total area of 167,000 square kilometers. Hubei Province is located in the central part of China with a total area of 185,900 square kilometers, between latitude 29°0'53"-33°03' N and longitude 108°21′42″-116°07′50″ E, in the middle reaches of the Yangtze River. Hunan Province is located in the central part of China with a total area of 211,800 square kilometers, between latitude $24^{\circ}38'-30^{\circ}08'$ N and longitude $108^{\circ}47'-114^{\circ}15'$ E. At the end of 2021, the resident population of Henan Province was 98.8 million, and the regional GDP was over CNY 58,887 billion. The resident population of Hubei Province was 58.3 million, and the regional GDP was over CNY 50,012 billion. The resident population of Hunan Province was 66.2 million, and the regional GDP was over CNY 46,063 billion. In these three provinces, the proportion of secondary industry is 41.3%, 37.9% and 39.4%, respectively. The related research works showed that the economic scale and high proportion of secondary industry had a significant impact on air quality and environmental pollution. In the 2021 census data, the population of Henan Province, Hubei Province and Hunan Province, respectively, ranked third, seventh and tenth in thirty-four provinces in China. The construction of urbanization has resulted in a decrease in urban green vegetation coverage, which directly contributed to air pollution. The higher proportion of secondary industry and the larger population have resulted in more air pollutant emissions and AQI exceeding the standard. Hence, it is necessary to study air quality in the three provinces. The geographic locations of the research regions are shown in Figure 1.



Figure 1. Regional geographic locations distribution map of APIIU research.

2.1.2. Data Sources

The data on the regional development of industrialization and urbanization used in this study include regional GDP, the GDP of secondary industry, coal consumption, exhaust emissions, population of city jurisdiction, total city population, city jurisdiction areas, administrative land area, density of population and so on. Due to yearbook data generally publishing the annual data of the previous year, the above data range was from 31 December 2018 to 31 December 2020 and was sourced from the National Statistics Administration and the China City Statistical Yearbook (https://data.stats.gov.cn/index.html/ accessed on 2 May 2022). The meteorological data included annual relative humidity, average temperature, rainfall data and time length of sunshine (http://data.cma.cn/dataService/ cdcindex/datacodel/ accessed on 3 May 2022). The average annual AQI in 18 cities from 2017 to 2019 were calculated from the daily data from the air quality online monitoring platform (https://www.aqistudy.cn/historydata/ accessed on 5 May 2022). The annual average PM_{2.5} were calculated from the daily data, which were from the US Embassy's Air Quality Report (https://www.airnow.gov/ accessed on 5 May 2022). In this paper, by using PyCharm 2021 in Python, Anaconda 3, SPSS Statistics 26.0 and ArcGIS 10.2, the average annual AQI changing trend is shown in Figure 2a-c. In addition, the relevant statistical analysis, including maximum value and standard deviation of the annual concentrations of PM_{2.5} in every city in three provinces from 2017 to 2019, is shown in Tables S1–S3.



Figure 2. Cont.



Figure 2. Annual average AQI changing trend in every province from 2017 to 2019: (**a**) Henan Province, Hubei Province and Hunan Province in 2017; (**b**) Henan Province, Hubei Province and Hunan Province in 2018; (**c**) Henan Province, Hubei Province and Hunan Province in 2019.

On the whole, the annual average AQI of Henan Province, Hubei Province and Hunan Province have slowly decreased year by year from 2017 to 2019 in some cities. However, the AQI of some cities in Henan Province have not decreased significantly. For instance, the annual average AQI in Zhengzhou, Anyang and Xinxiang in Henan Province have consistently exceeded 100 in the past three years. The annual average AQI in Xiangyang and inmen in Hubei Province have consistently exceeded 90 in the past three years. The annual average AQI in Changsha, Yueyang, Xiangtan and Zhuzhou in Hunan Province have consistently exceeded 70 in the past three years. Additionally, the annual average AQI in Yiyang in Hunan Province increased year by year from 2017 to 2019. To clearly observe the changing trend of $PM_{2.5}$ in three provinces from 2017 to 2019, the annual average $PM_{2.5}$ data of every city were collected and calculated. The annual average $PM_{2.5}$ changing trend is shown in Figure 3a–c.

The annual average concentrations of $PM_{2.5}$ of Henan Province, Hubei Province and Hunan Province have decreased year by year from 2017 to 2019. For Henan Province, the past three years have witnessed the annual average concentrations of $PM_{2.5}$ decrease by more than 15% in Luoyang, Sanmenxia and Xinyang. However, other cities have consistently exceeded 40 µg/m³ in 2019. For Hubei Province, Huangshi, Enshi and Shennongjia, with relatively pleasant natural environment and low level of urbanization, had a concentration less than 30 µg/m³ in 2019, which was the same as the change trend of AQI in recent years. For Hunan Province, Zhangjiajie, Chenzhou and Xiangxi had a concentration less than 30 µg/m³ in 2019. Among the three cities, the annual average concentration of Zhangjiajie had dropped by more than 18% compared with 2017. Although China has issued related policies to protect the environment and manage air pollutant emissions in recent years, with the development of industrialization and urbanization and the increase in the population scale, the annual average concentrations of PM_{2.5} of most cities are still more than 30 μ g/m³, and even some cities have consistently exceeded 60 μ g/m³ in 2019. Hence, it is very necessary to carry out the relevant environmental research of AQI and particulate matter, providing the government with a feasibility analysis report.



Figure 3. The changing trend of APIIU-AQI from 2017 to 2019: (a) Henan Province: Zhengzhou—ZZ, Kaifeng—KF, Luoyang—LY, Pingdingshan—PDS, Anyang—AY, Hebi—HB, Xinxiang—XX, Jiaozuo—JZ, Puyang—PY, Xuchang—XC, Luohe—LH, Sanmenxia—SMX, Nanyang—NY, Shangqiu—SQ, Xinyang—XY, Zhoukou—ZK, Zhumadian—ZMD, Jiyuan—JY; (b) Hubei Province: Wuhan—WH, Huangshi—HS, Shiyan—SY, Yichang—YC, Xiangyang—XY, Ezhou—EZ, Jingmen—JM, Xiaogan—XG, Jingzhou—JZ, Huanggan—HG, Xianning—XN, Suizhou—SZ, Enshi—ES, Xiantao—XT, Qianjiang—QJ, Tianmen—TM, Shennongjia—SNJ; (c) Hunan Province: Changsha—CS, Zhuzhou—ZZ, Xiangtan—XT, Hengyang—HY, Shaoyang—SY, Yueyang—YUY, Changde—CD, Zhangjiajie—ZJJ, Yiyang—YIY, Chenzhou—CZ, Yongzhou—YZ, Huaihua—HH, Loudi—LD, Xiangxi—XX.

2.1.3. Indicators System

The regional assessment indicators system of APIIU constructed from the perspective of four aspects is shown in Table 1.

Primary Indicators	Secondary Indicators	Reference Source	
	X ₁ : Regional GDP (CNY 100 million)	_ [1,2,4,54,57]	
	X ₂ : Regional GDP of secondary industry (CNY 100 million)		
	X ₃ : Proportion of secondary industry (%)		
	X ₄ : Square of proportion of secondary industry (%)		
	X ₅ : Coal consumption (100 million ton)		
Industrialization indicators	X_6 : Coal consumption per land area (100 million ton/km ²)		
	X ₇ : Square of coal consumption per land area (100 million ton/km ²)	— [1,2,4,54,57]	
	X ₈ : Exhaust emissions (ton)	-	
	X ₉ : Density of exhaust emissions (ton/km ²)		
	X_{10} : Square of density of secondary industry (ton/km ²)		
	X_{11} : Population of city jurisdiction (10 million people)		
	X ₁₂ : Total city population (10 million people) X ₁₃ : Proportion of population (%) X ₁₄ : Square of proportion of population (%)	-	
	X ₁₅ : City jurisdiction areas (km ²)		
Urbanization indicators	X ₁₆ : Density of population (10 million people/km ²)	[1,12,33,58,68]	
	X ₁₇ : Square of density of population (10 million people/km ²)		
	X_{18} : Per capita GDP (CNY 10,000/people) X_{19} : Administrative land areas (km ²)		
Meteorological indicators	X ₂₀ : Annual average relative humidity (%)	[45.46.48.58-66]	
Wetcorological intercators	X ₂₁ : Annual average temperature (°C)		
	X_{22} : Annual average rainfall (mm) X_{23} : Time length of sunshine (hour)	- [45,46,48,58–66]	
Meteorological indicators	X ₂₄ : Wind speed (m/s: Annual average wind speed at 70 m–80 m altitude)		
Type of APIIU	X ₂₅ : APIIU of AQI X ₂₆ : APIIU of PM _{2.5}	[22-24,35-37,52]	

Table 1. Primary and secondary indicators.

Based on the existing related research results, considering the current situation regarding air pollution and the availability of data, this study constructed an indicators system of APIIU assessment using the industrialization, urbanization and meteorological indicators. The industrialization indicators were directly responsible for air pollutant emissions. Previously published studies have shown that the major air pollutant emissions were from the daily industrial operation of the machines. Therefore, the regional GDP, GDP of the secondary industry, coal consumption and exhaust emissions were used as direct or indirect influencing indicators of APIIU. Urbanization indicators included the city's development level, scale and population. The occurrence of air pollution was closely related to the city's development level and scale. With population growth and urban development, more and more building construction areas contributed to the decrease in the city's green vegetation areas, which increased the concentrations of dust and particulate matter in the air and reduced the adsorption capacity of green plants for air pollutants. The higher density of the regional population living in urban areas caused more environmental pollution problems and air pollutant emissions. The annual average air humidity and the annual average rainfall could decrease air pollution to some extent. Higher emissions of industrial and residential pollutants caused higher annual average temperature. PM_{2.5} refers to particulate matter with aerodynamic equivalent diameter of less than or equal to 2.5 microns. Owing to long residence time and conveying distance, it has a greater impact on human health and the quality of the atmospheric environment. PM_{10} is inhalable particulate matter with aerodynamic equivalent diameter of less than or equal to 10 microns, which refers to the general term for solid and liquid particles floating in the air. PM_{10} can enter the upper respiratory tract directly and endanger human lung health. PM_{10} in many cities were positively correlated with PM_{2.5}, but O₃ was negatively correlated with PM_{2.5}. Related studies conducted in other regions also showed that the concentrations of SO_2 , NO_2 and CO have increased year by year with the formulation of relevant policies. For example, the most obvious feature was that the occurrence of acid rain has reduced in the past many years [69,70]. Hence, this research selected the AQI and PM_{2.5} as the objectives of APIIU; industrialization indicators, urbanization indicators and meteorological indicators were selected to evaluate the influence degree of APIIU on AQI and PM_{2.5}.

2.2. The Hybrid SC-RS-XGBoost Model

2.2.1. The Principle of SC

In statistics, the SC is a nonparametric measure of the dependence of two variables. It uses a monotonic equation to evaluate the correlation of two statistical variables. If there are no repeated values in experimental data, the value of SC is 1, or -1 when the two variables are in a perfect monotonical correlation. SC is also referred to as level correlation, which is the observed data being replaced by the level. The SC indicates the direction of correlation between independent variable *x* and dependent variable *y*. If *y* tends to increase with *x* increases, the SC is positive. If *y* tends to decrease with *x* increases, the SC is negative. When SC is zero, it indicates that *y* exhibits no trend with *x* increases. The absolute value of SC increases when *x* and *y* become closer and closer [66–68]. The SC can reflect the direction and degree of the change trend between the 2 random variables, which is calculated as the difference of equal magnitude. The most notable feature is that it need not consider the sample size or overall distribution characteristics of the variables; the SC of two random variables can be expressed as Equation (1):

$$\rho(x,y) = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 (y_i - \overline{y})^2}}$$
(1)

where ρ is the SC between *x* and *y*, *n* is datasets sample size, \overline{x} and \overline{y} is the average value of *x* and *y*.

In fact, the connection between the variables is irrelevant. The principle is to sort the data of two variables and calculate the linear correlation analysis using the value of rank difference obtained after sorting, as shown in Equation (2):

$$p = \frac{6\sum_{i=1}^{n} d^2}{n(n^2 - 1)}$$
(2)

where *n* is datasets sample size, *d* is the rank difference value sorted by 2 variables.

1

2.2.2. The Principle of RS

The grid search (GS) is an exhaustive optimization algorithm, and the searched parameters are defined in the space of the uniform grids. All the nodes in the grid are then evaluated to identify the global minimum. Finally, the grid search finds the global minimum of all the nodes in the parameter grid. By approaching the optimal point in the next step, a finer grid is defined and gradually approaches the optimal point in the searching parameter space. However, the optimization process is more time consuming and inefficient. The RS algorithm selects the specific number of 1 random value per 1 hyperparameter to reduce the computation of the hyperparameter search, shorten the optimization time and improve the model performance. It is performed in a random manner in the spatial distribution, and the RS algorithm samples it as 1 distribution [59–61]. In this research, RS was applied to optimize XGBoost.

2.2.3. The Principle of XGBoost

The integrated learning method refers to combining multiple learning models to obtain better results and a stronger generalization ability. XGBoost is a scalable tree boosting system that is an ensemble method that aims to aggregate weak learning models to form a stronger and more robust estimator in an iterative fashion. The residual of the previous estimator will be used to learn and optimize the loss function during each iteration. A binary decision tree called a classification and regression tree (CART) is selected as a basic learner, and regularization is added to the loss function for improvement and to avoid overfitting in XGBoost [62–65]. XGBoost is a highly flexible and versatile tool that can solve most regression problems and objective functions created by users. The SC-RS-XGBoost algorithm used in this paper is shown in Algorithm 1.

Algorithm 1: SC-RS-XGBoost

Input: $D_{m \times n} = \{(X_i)\}$ ($X_i \in \mathbb{R}^m$, i = 1, 2, ..., n), original data with *n* samples and *m* feature variables **Output:** $D_{l \times n} = \{(X_i)\}$ ($X_i \in \mathbb{R}^l$, i = 1, 2, ..., n), using SC to screen correlation coefficients greater than 0.3 based on $D_{m \times n}$ (l < n) **Input:** Objective function of RS, f(X,Y) = g(X) h(Y)Set default values and value ranges of parameters to be optimized Set the threshold of mean squared error (MSE) **Output:** Every parameter value when f(X,Y) reaches the maximum value **Input:** $D_{l \times n} = \{(X_i)\} (X_i \in \mathbb{R}^l, i = 1, 2, ..., n)$ *I*, instance set of current node *d*, feature dimension *Gain* \leftarrow 0 $G \leftarrow \sum_{i \in I} g_i, H \leftarrow \sum_{i \in I} h_i$ for k = 1 to T do $G_L \leftarrow 0, H_L \leftarrow 0$ **for** *j* in sorted(*I*, by x_{ik}) **do** $G_L \leftarrow G_L g_j, H_L \leftarrow H_L h_j$ $G_R \leftarrow G - G_L, H_R \leftarrow H - H_L$ score $\leftarrow max$ (score, $\frac{G_L^2}{H_L \lambda} \frac{G_R^2}{H_R \lambda} - \frac{G^2}{H \lambda}$) end end Output: Split with max score

The XGBoost algorithm includes three types of parameters: general, booster and learning task. There are relatively many types of parameters, and the selection of parameter combinations directly affects the accuracy of boosting tree regression prediction. In machine learning, a reasonable parameter selection can improve the prediction accuracy of the model to a large extent. In this study, the RS optimization algorithm was used to optimize some parameters of XGBoost, and the other parameters selected appropriate default values according to the training process. The range of parameter values and final parameter determination optimized by the RS algorithm are shown in Table 2.

Parameter Name	Parameter Type	Parameter Definition	Parameter Default Value	Value Range
P ₁ : max_depth P ₂ : learning_rate	Booster Booster	Maximum depth of tree Learning rate	5 0.2	[1,20] [0,1]
P ₃ : reg_gamma	Booster	Adjusting the penalty term, specifying the minimum loss function decreases when the node is divided	0.01	[0,0.5]
P ₄ : reg_alpha	Booster	Regularization coefficient used to adjust <i>L1</i>	0.01	[0,1]
P ₅ : reg_lambda	Booster	Regularization coefficient used to adjust L2	0.1	[0,2]
P ₆ : min_child_weight	Booster	Minimum leaf node weight	1	[1,10]
P7: subsample	Booster	The sampling scale used for the training set	1	[0,1]
P ₈ : colsample_bytree	Booster	The random sampling ratio of features used to construct each tree	1	[0,1]
P ₉ : n_estimators	Learning Task	Controlling the number of trees	70	[30,200]

Table 2. Optimized parameters and value range of XGBoost.

In this study, the objective function of RS was set to the mean squared error (MSE) as the calculation error; the maximum number of iterations was set to 50; and the maximum value of MSE was set to 3.00. In the general parameters of XGBoost, the parameter "booster" was set to the "gbtree"; the parameter "n_thread" was set to not exceed the maximum possible number of threads of the processor; and other parameters were set to the default values. After the above optimized parameters were given the initialized default values, the processed data set was input into the optimization function of the RS, and the partial results of the optimization are shown in Table 3.

Table 3. Partial results of optimization of XGBoost.

Iteration Number	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P 9	MSE
1	5	0.2	0.01	0.01	0.1	1	1	1	70	2.54
2	12	0.009	0.526	0.635	0.137	6.23	0.972	0.892	102	2.23
3	9	0.417	0.391	0.172	0.935	5.33	0.391	0.913	61	1.78
4	16	0.103	0.153	0.229	0.182	6.19	0.672	0.992	94	1.52
5	8	0.319	0.203	0.381	0.286	4.27	0.259	0.715	105	0.936
10	11	0.821	0.337	0.156	0.107	8.31	0.113	0.836	82	0.783
16	7	0.113	0.082	0.096	0.132	4.92	0.463	0.651	103	0.497
22	10	0.016	0.128	0.193	0.092	1.37	0.991	0.192	75	0.585
50	3	0.432	0.11	0.308	0.087	1.86	0.723	0.274	91	0.647

By calculating the MSE, a more suitable parameter combination was found when the iteration number was 16th. However, the effect of the parameter combination gradually deteriorated with the number of iterations continuing to increase. Therefore, the parameters of the XGBoost were selected as the parameter combination in the 16th iteration to forecast APIIU of AQI and PM_{2.5} in this research.

3. Analytical Results of APIIU

Before using SC analysis, based on the influence path of industrialization and urbanization on AQI and $PM_{2.5}$, this research constructed the structural equation modeling (SEM) using AMOS to evaluate the reasonableness of the model and preliminarily screen the relevant indicators from the indicators system [71,72]. Additionally, the least square

estimation method was selected to perform the parameter estimation of SEM. In this research, the comparative fit index (CFI), Akaike information criterion (AIC) and root mean square error of approximation (RMSEA) were selected as evaluation indicators. The model had better performance when CFI was higher than 0.8, RMSEA was lower than 0.1, and AIC was relatively small. Based on the structural equation and evaluation results, this research finally deleted the indicators with a moderating effect, including the square of coal consumption per land area, square of proportion of population and square of density of population. Hence, the above three indicators were no longer considered when using SC analysis. The SC correlation analysis results are shown in Figures 4 and 5.



Figure 4. The SC correlation analysis results of APIIU of AQI.



Figure 5. The SC correlation analysis results of APIIU of PM_{2.5}.

Before using SC to analyze the data, this research used the maximum and minimum standardized method in data processing, and the value of each indicator was controlled between 0 and 1. In order to obtain the weights of relevant indicators affecting the APIIU of AQI (APIIU-AQI) and APIIU of $PM_{2.5}$ (APIIU- $PM_{2.5}$), taking AQI and $PM_{2.5}$ as target variables, respectively, the correlation coefficients are shown in Figures 4 and 5. The indicators were from the indicators system shown in Table 1. In this research, the correlation coefficients of AQI and $PM_{2.5}$ were greater than 0.3, respectively. The correlation coefficients

in the SC analysis results reflected the degree of importance of the role of each indicator in the APIIU-AQI and AQPIIU-PM_{2.5}. Hence, the correlation coefficient of each indicator could be used to express the indicator weight. Using the weighting equation calculated using the air pollution index according to principal indicators, the APIIU could be expressed as in Equation (3):

$$I_i = \sum S_j X_{ij} \tag{3}$$

where I_i is the APIIU of the i_{th} unit with different principal indicators, X_{ij} is standardized values of the j_{th} indicator of the i_{th} unit, and S_j is the SC value of the j_{th} indicator.

The APIIU-AQI and APIIU-PM_{2.5} in every city calculated from the SC results showed that there were significant temporal and regional differences among the three provinces from 2017 to 2019. The changing trends in the APIIU-AQI and APIIU-PM_{2.5} are shown in Figures 6a–c and 7a–c over the past 3 years, respectively. Considering the development of industrialization and urbanization, the APIIU-AQI and APIIU-PM_{2.5} showed a significant changing trend from 2017 to 2019.

Based on the changing trend of APIIU-AQI and APIIU-PM_{2.5} of every city within three provinces over the past 3 years, the following observations can be made. Hubei Province was an area with higher APIIU-AQI among the three provinces, indicating that the development of industrialization and urbanization had a more serious impact on AQI. From 2017 to 2019, the APIIU-AQI in 2019 in every city increased by 20–30% compared to 2017. Hunan Province was an area with higher APIIU-PM_{2.5} among the three provinces, indicating that the development of industrialization and urbanization had a more serious impact on PM_{2.5}. From 2017 to 2019, the APIIU-PM_{2.5} of every city within the three provinces over the past 3 years were higher compared to APIIU-AQI. In the past three years, the APIIU of some cities in Hunan Province has declined; however, some cities in Henan Province and Hubei Province have shown a trend of initial decline and then rise. The data showed a slowly decreasing trend and indicated that Hunan Province was paying attention to the protection and governance of the environment when developing its economy. Additionally, the industry has transformed into a new energy and environmentally friendly direction; the population density has decreased; and the air pollutant emissions emitted by human activities have decreased in Hunan Province. With an increasing trend in Henan Province and Hubei Province, this phenomenon showed that the proportion of secondary industry in the two provinces still occupied a major position. In addition, the increasing population and the government's control with a slow rate of environmental governance made the industrial and human activities aggravate environmental pollution. Through the assessment of APIIU-AQI and APIIU-PM_{2.5} of regional air pollution and the descriptive analysis, the changing results of AQI and PM_{2.5} were generally consistent with the findings published by the China Environmental Statistical Yearbook in Henan Province, Hubei Province and Hunan Province, respectively.

Air pollution has caused a serious impact on the natural environment, human society and residents' health in recent years. Hence, it is vital to evaluate the influence of the development of regional industrialization and urbanization on AQI and PM_{2.5}, which could help the government plan reasonable policies to a certain degree. For Henan Province, with a large population and agricultural industry, importance should be attached to increasing the development of environmental protection enterprises while transforming to heavy industrialization. It is necessary to avoid excessive population concentration during urbanization. For Hubei Province, with a high GDP, the pace of development of traditional industries can be slowed down while developing industrialization. The transformation to new energy enterprises can be intensified, which can not only protect the environment but prevent more population from outflowing. For Hunan Province, with the promotion of cleaner production technology and increasing financial investment in environmental protection, the environmental quality has been greatly improved. However, it should avoid repeating the same predicament. Environmental governance should be processed simultaneously while developing the economy and urbanization.





(b)





(c)

Figure 7. The changing trend of APIIU-PM_{2.5} from 2017 to 2019: (a) Henan Province: Zhengzhou—ZZ, Kaifeng—KF, Luoyang—LY, Pingdingshan—PDS, Anyang—AY, Hebi—HB, Xinxiang—XX, Jiaozuo—JZ, Puyang—PY, Xuchang—XC, Luohe—LH, Sanmenxia—SMX, Nanyang—NY, Shangqiu—SQ, Xinyang—XY, Zhoukou—ZK, Zhumadian—ZMD, Jiyuan—JY; (b) Hubei Province: Wuhan—WH, Huangshi—HS, Shiyan—SY, Yichang—YC, Xiangyang—XY, Ezhou—EZ, Jingmen—JM, Xiaogan—XG, Jingzhou—JZ, Huanggan—HG, Xianning—XN, Suizhou—SZ, Enshi—ES, Xiantao—XT, Qianjiang—QJ, Tianmen—TM, Shennongjia—SNJ; (c) Hunan Province: Changsha—CS, Zhuzhou—ZZ, Xiangtan—XT, Hengyang—HY, Shaoyang—SY, Yueyang—YUY, Changde—CD, Zhangjiajie—ZJJ, Yiyang—YIY, Chenzhou—CZ, Yongzhou—YZ, Huaihua—HH, Loudi—LD, Xiangxi—XX.

4. Discussion

4.1. Evaluation Indicator

To reasonably evaluate the prediction model, the square of determination coefficient of real value (R^2), the root mean square error (RMSE) and the mean absolute percentage error (MAPE) were selected to test the prediction result. In statistical experiments, RMSE actually describes a degree of dispersion, which reflects the size of the average prediction error. MAPE can be used to evaluate different models on the same data source, which actually reflects the median relative error. The evaluation indicators can be shown as follows:

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i}^{*} - y_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \overline{y})^{2}}$$
(4)

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (y_i^* - y_i)^2}$$
(5)

$$MAPE = \sum_{i=1}^{n} \left| \frac{y_i^* - y_i}{y_i} \right| \times \frac{100}{n} \tag{6}$$

where *n* represents the total number of test samples, y_i^* , y_i and \overline{y} represent the predicted value, the real value and the average real value in Equations (4)–(6), respectively.

4.2. Result Analysis Based on SC-RS-XGBoost

To verify the performance of the SC-RS-XGBoost model in predicting APIIU-AQI and APIIU-PM_{2.5}, this model was compared with the SC-XGBoost model, RS-XGBoost model and XGBoost model. For the SC-RS-XGBoost model and SC-XGBoost model, this research used SEM to screen the indicators and adopted the maximum and minimum method to standardize the indicators as the training data. Additionally, the original data were multiplied by the weights after SC analysis as the final training data. For the RS-XGBoost model and XGBoost model, this research used SEM to screen the indicators as the training data. Additionally, the original data were multiplied by the weights after SC analysis as the final training data. For the RS-XGBoost model and XGBoost model, this research used SEM to screen the indicators as the training data. Data were collected from 49 cities from 2017 to 2019, and the number of samples was 147. After the data were preprocessed, the feature data and target data were input into the different models, and the ratio of training data and testing data was set to 3:1. APIIU-AQI and APIIU-PM_{2.5} were selected as the target variables, respectively. Industrialization, urbanization and meteorological indicators were selected as the characteristic variables. The predicted values of different models are shown in Figures A1 and A2 in Appendix A. Collecting the prediction results and calculating the evaluation indicators, the results are shown in Table 4.

Variable	S	C-RS-XGBoo	ost	SC-XGBoost			
	R ²	RMSE	MAPE	R ²	RMSE	MAPE	
APIIU-AQI	0.945	0.103	4.25%	0.886	0.149	6.08%	
APIIU-PM _{2.5}	0.897	0.205	4.84%	0.515	0.443	6.34%	
Veri alti a	RS-XGBoost			XGBoost			
variable	R ²	RMSE	MAPE	R ²	RMSE	MAPE	
APIIU-AQI	0.856	0.167	6.36%	0.753	0.218	13.7%	
APIIU-PM _{2.5}	0.823	0.269	5.74%	0.366	0.646	16.8%	
¥7	SC-RS-SVR			SC-SVR			
variable	R ²	RMSE	MAPE	R ²	RMSE	MAPE	
APIIU-AQI	0.503	0.285	7.13%	0.377	0.514	8.83%	
APIIU-PM _{2.5}	0.419	0.912	10.7%	0.283	2.63	14.6%	
X7	RS-SVR			SVR			
variable	R ²	RMSE	MAPE	R ²	RMSE	MAPE	
APIIU-AQI	0.412	0.393	13.5%	0.461	0.427	17.2%	
APIIU-PM _{2.5}	0.318	1.51	7.32%	0.252	2.74	16.6%	

Table 4. Evaluation indicators of APIIU of each model.

Using SC, reducing the input variables and reducing the target function convergence value effectively solved the correlation between the input variables and the defection of excessive input data. In this paper, RF was used to optimize the parameters of XGBoost to reduce local overfitting, showing better prediction performance. It can be seen from the results that SC-RS-XGBoost performed better than SC-XGBoost. As shown in Table 4, we found that the traditional XGBoost model had the largest prediction error. As seen from the comparison of R^2 , *RSME* and MAPE of APIIU-AQI, the prediction accuracy of the SC-RS-XGBoost model improved by 6.24%, 30.9% and 30.1% upon the SC-XGBoost

model on the R², RMSE and MAPE, respectively. It improved by 8.90%, 38.3% and 33.1% upon the RS-XGBoost model on the R^2 , RMSE and MAPE, respectively, and it improved by 20.3%, 52.8% and 68.9% upon the traditional XGBoost model on the R^2 , RMSE and MAPE, respectively. For a comparison of R^2 , RSME and MAPE of APIIU-PM_{2.5}, the prediction accuracy of the SC-RS-XGBoost model improved by 42.6%, 53.7% and 23.7% upon the SC-XGBoost model on the R^2 , RMSE and MAPE, respectively. It improved by 8.24%, 23.8% and 15.7% upon the RS-XGBoost model on the R^2 , RMSE and MAPE, respectively, and it improved by 59.2%, 68.3% and 71.2% upon the traditional XGBoost model on the R², RMSE and MAPE, respectively. Based on these findings, the prediction accuracy was significantly improved, which verified the effectiveness and feasibility of the SC-RS-XGBoost model compared with the other three models. To further verify the performance and accuracy, the evaluation indicators with the SC-RS-SVR model, SC-SVR model, RS-SVR model and SVR model are shown in Table 4. The comparisons of R^2 , RMSE and MAPE of the SC-RS-SVR model, SC-SVR model, RS-SVR model, SVR model and ELM model are shown in Figures S1–S3. It can be seen from the calculation results of evaluation indicators that other models did not perform as well as the SC-RS-XGBoost model.

Additionally, through the calculation results of evaluation indicators, we found that the prediction accuracy of APIIU-AQI was higher than APIIU-PM_{2.5}. In terms of RMSE and MAPE, the difference between the APIIU-AQI and APIIU-PM_{2.5} was also basically inconspicuous. However, the values of R^2 were all higher than 0.89, showing that the SC-RS-XGBoost model had a good fitting performance on APIIU-AQI and APIIU-PM_{2.5}. By verifying the effectiveness and feasibility of the proposed model, the SC-RS-XGBoost model accurately predicted the APIIU-AQI and APIIU-PM_{2.5} and provided helpful insights for the government's environmental protection and air pollution governance. This could make regional governments pay more attention to air pollution when developing industrialization and urbanization.

5. Conclusions

Most previous research works have analyzed AQI and PM_{2.5} spatiotemporal distribution using statistical methods and regression analyses to evaluate the air quality. The goal of the current research was to analyze and determine the influence of the development of industrialization and urbanization on AQI and PM_{2.5}. In this period of artificial intelligence, a SC-RS-XGBoost model for air pollution assessment could effectively improve the assessment accuracy and provide a new reference source for future air pollution management. Based on previous relevant research works, this research constructed a regional air pollution assessment indicators system to evaluate APIIU-AQI and AQIIU-PM_{2.5} using data from Henan Province, Hubei Province and Hunan Province. The indicators system of this research adopted three aspects, including industrialization indicators, urbanization indicators and meteorological indicators. The following conclusions were drawn from our findings:

(1) The principal indicators were screened using SC, and the weights of each indicator were determined according to the correlation coefficients. From the correlation analysis, there were differences in the impact of different indicators on AQI and $PM_{2.5}$. The indicators, including regional GDP, proportion of secondary industry, coal consumption, density of exhaust emissions, total city population, proportion of population, density of population and several relevant meteorological indicators, had a high influence on AQI. However, the proportion of secondary industry, coal consumption, exhaust emissions, density of exhaust emissions, population of city jurisdiction, total city population, administrative land areas and other related indicators had a high influence on $PM_{2.5}$.

(2) Using the weights equation to calculate the APIIU-AQI and APIIU-PM_{2.5} of each region from 2017 to 2019, we identified that 17 cities from Hubei Province, which has the largest regional GDP and the largest proportion of the population, had the largest APIIU-AQI in the past three years. Moreover, APIIU-AQI increased sharply compared with 2017 in the three provinces. Fourteen cities from Hunan Province, which has the largest

proportion of secondary industry and the largest administrative land area, had the largest APIIU-PM_{2.5} in the past three years. APIIU-PM_{2.5} has gradually increased compared with 2017 in the three provinces.

(3) This research verified that the SC-RS-XGBoost could be used as a method of air pollution assessment. Analyzing the influence of industrialization and urbanization on AQI and $PM_{2.5}$ is of great significance for future regional sustainable development. It is necessary for carrying out more advanced research and constructing a more accurate indicators system and prediction models for the various different regions to deliver useful information for the government officials and policymakers.

The present research contributed to the limited knowledge in the regions regarding the influence of industrialization and urbanization on spatiotemporal variations of AQI and $PM_{2.5}$ in three Chinese provinces, and it is important to conduct additional research in other regions on this topic. AQI and the concentration of $PM_{2.5}$ in different regions are affected by wind speed and direction at different altitudes to some extent. It is worth noting that this research was focused only on static and mild wind conditions, and the annual average wind speed was only observed at 70 m–80 m altitude. The result was limited, since the data selected for discussion were monitored under specific climate conditions. Hence, long-term continuous monitoring under various weather conditions is needed in future research. Due to differences in industrial and human activity emissions in different seasons, in order to explore the seasonal changes, it is necessary to consider the influence of industrialization and urbanization indicators on AQI and $PM_{2.5}$ in different seasons in future research.

Supplementary Materials: The following supporting information can be downloaded at: https: //www.mdpi.com/article/10.3390/atmos13091377/s1, Table S1: The statistical analysis of PM_{2.5} in 2017 in three provinces; Table S2: The statistical analysis of PM_{2.5} in 2018 in three provinces; Table S3: The statistical analysis of PM_{2.5} in 2019 in three provinces; Figure S1: Comparison of R^2 of SC-RS-SVR model, SC-SVR model, RS-SVR model, SVR model and ELM model; Figure S2: Comparison of RMSE of SC-RS-SVR model, SC-SVR model, RS-SVR model, SVR model and ELM model; Figure S3: Comparison of MAPE of SC-RS-SVR model, SC-SVR model, RS-SVR model, RS-SVR model and ELM model.

Author Contributions: Conceptualization, H.C.; methodology, H.C.; validation, G.D. and Y.L.; formal analysis, H.C.; investigation, H.C.; resources, H.C.; data curation, H.C.; Writing—Original draft preparation, H.C.; Writing—Review and editing, H.C.; supervision, G.D.; project administration, H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was greatly supported by the National Social Science Fund Key Project (15AGL013); Henan Provincial Department of Science and Technology Risk Management Innovation and Public Policy Soft Science Research Base, Henan Social Science Planning Project (2019BJJ030). Research on the construction of disaster prevention and reduction support system in large and medium-sized cities in Henan Province (22240041001); Henan Provincial Colleges and Universities Philosophy and Social Science Basic Research Major Project "Evaluation Research on Comprehensive Disaster Resilience Capacity of Chinese Communities" (2021JCZD04).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

(b) Real Value Real Value SC-RS-XGB SC-XGBoost 2.5 2.5 2.0 2 APIIU-AQI APIIU-AQI 1.5 1.5 1.0 0.5 75 imbe 75 umber 100 125 150 100 125 150 (c) (d) Real Value Real Value 2.5 RS-XGBoos 2.5 XGBoost 2.0 2.0 APIIU-AQI IOA PIIU-/ 1.5 1.5 1.0 0.5 0.5 125 150 100 125 150 Nu

Figure A1. Comparison of real values and predicted values of APIIU-AQI from (**a**) SC-RS-GBoost model, (**b**) SC-XGBoost model, (**c**) RS-XGBoost model and (**d**) XGBoost model.



Figure A2. Comparison of real values and predicted values of APIIU-PM_{2.5} from (**a**) SC-RS-XGBoost model, (**b**) SC-XGBoost model, (**c**) RS-XGBoost model and (**d**) XGBoost model.

References

- 1. Lelieveld, J.; Evans, J.S.; Fnais, M.; Giannadaki, D.; Pozzer, A. The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* **2015**, 525, 367–371. [CrossRef]
- Huang, F.F.; Pan, B.; Wu, J.; Chen, E.G.; Chen, L.Y. Relationship between exposure to PM_{2.5} and lung cancer incidence and mortality: A meta-analysis. *Oncotarget* 2017, *8*, 43322–43331. [CrossRef] [PubMed]
- 3. Bilal, M.; Nichol, J.; Nazeer, M.; Shi, Y.; Wang, L.C.; Kumar, K.; Ho, H.; Mazhar, U.; Bleiweiss, M.; Qiu, Z.F.; et al. Characteristics of fine particulate matter (PM_{2.5}) over urban, suburban and rural areas of Hong Kong. *Atmosphere* **2019**, *10*, 496. [CrossRef]

Appendix A

- Georgieva, E.; Syrakov, D.; Prodanova, M.; Etropolska, I.; Slavov, K. Evaluating the performance of WRF-CMAQ air quality modelling system in Bulgaria by means of the DELTA tool. *Int. J. Environ. Pollut.* 2015, 57, 272–284. [CrossRef]
- Robichaud, A. Surface data assimilation of chemical compounds over North America and its impact on air quality and Air Quality Health Index (AQHI) forecasts. *Air. Qual. Atmos. Health* 2017, *10*, 955–970. [CrossRef]
- 6. Xu, Y.N.; Liu, H.; Duan, Z. A novel hybrid model for multi-step daily AQI forecasting driven by air pollution big data. *Air. Qual. Atmos. Health* **2020**, *13*, 197–207. [CrossRef]
- 7. Loomis, D.; Grosse, Y.; Lauby-Secretan, B.; El Ghissassi, F.; Bouvard, V.; Benbrahim-Tallaa, L.; Guha, N.; Baan, R.; Mattock, H.; Straif, K.; et al. The carcinogenicity of outdoor air pollution. *Lancet. Oncol.* **2013**, *14*, 1262–1263. [CrossRef]
- 8. Balmes, J.R. Household air pollution from domestic combustion of solid fuels and health. *J. Allergy. Clin. Immul.* **2019**, 143, 1979–1987. [CrossRef]
- 9. Du, W.; Zhuo, S.J.; Wang, J.Z.; Luo, Z.H.; Chen, Y.C.; Wang, Z.L.; Lin, N.; Cheng, H.F.; Shen, G.F.; Tao, S. Substantial leakage into indoor air from on-site solid fuel combustion in chimney stoves. *Environ. Pollut.* **2021**, *291*, 118138. [CrossRef]
- 10. European Environment Agency. Air Quality in Europe—2017 Report, EEA Report No 13/2017. Available online: https://www.eea.europa.eu/publications/air-quality-in-europe-2017 (accessed on 8 May 2022).
- 11. European Environment Agency. Air Quality in Europe—2021 Report, EEA Report No 8/2021. Available online: https://www.eea.europa.eu//publications/air-quality-in-europe-2021 (accessed on 8 May 2022).
- Chen, R.J.; Kan, H.D.; Chen, B.H.; Huang, W.; Bai, Z.P.; Song, G.X.; Pan, G.W. Association of particulate air pollution with daily mortality: The China air pollution and health effects study. *Epidemiology* 2012, 175, 1173–1181. [CrossRef]
- 13. Ding, Y.H.; Liu, Y.J. Analysis of long-term variations of fog and haze in China in recent 50 years and their relations with atmospheric humidity. *Sci. China-Earth Sci.* 2014, 57, 36–46. [CrossRef]
- 14. Zheng, B.; Tong, D.; Li, M.; Liu, F.; Hong, C.P.; Geng, G.N.; Li, H.Y.; Li, X.; Peng, L.Q.; Qi, J.; et al. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. *Atmos. Chem. Phys.* 2018, *18*, 14095–14111. [CrossRef]
- 15. Xu, P.; Chen, Y.F.; Ye, X.J. Haze, Air Pollution, and Health in China. Lancet 2013, 382, 2067. [CrossRef]
- Li, X.; Zhang, Q.; Zhang, Y.; Zheng, B.; Wang, K.; Chen, Y.; Wallington, T.J.; Han, W.; Shen, W.; Zhang, X.Y.; et al. Source contributions of urban PM_{2.5} in the Beijing–Tianjin–Hebei Region: Changes between 2006 and 2013 and relative impacts of emissions and meteorology. *Atmos. Environ.* 2015, 123, 229–239. [CrossRef]
- GBD 2016 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 84 behavioral, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 2017, 390, 1345–1422. [CrossRef]
- Zhang, X.Y.; Wang, Y.Q.; Niu, T.; Zhang, X.C.; Gong, S.L.; Zhang, Y.M.; Sun, J.Y. Atmospheric aerosol compositions in China: Spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols. *Atmos. Chem. Phys.* 2012, *12*, 779–799. [CrossRef]
- 19. Wang, X.X.; Wang, L.Q.; Liu, Y.Y.; Hu, S.G.; Liu, X.Z.; Dong, Z.Z. A data-driven air quality assessment method based on unsupervised machine learning and median statistical analysis: The case of China. *J. Clean. Prod.* **2021**, *328*, 129531. [CrossRef]
- Du, X.H.; Chen, R.J.; Meng, X.; Liu, C.; Niu, Y.; Wang, W.D.; Li, S.Q.; Kan, H.D.; Zhou, M.G. The establishment of national air quality health index in China. *Environ. Int.* 2020, 138, 105594. [CrossRef]
- Feng, L.; Liao, W.J. Legislation, plans, and policies for prevention and control of air pollution in China: Achievements, challenges and improvements. J. Clean. Prod. 2016, 112, 1549–1558. [CrossRef]
- 22. Bo, M.; Salizzoni, P.; Clerico, M.; Buccolieri, R. Assessment of indoor-outdoor particulate matter air pollution: A Review. *Atmosphere* **2017**, *8*, 136. [CrossRef]
- 23. Kayes, I.; Shahriar, S.A.; Hasan, K.; Akhter, M.; Kabir, M.M.; Salam, M.A. The relationships between meteorological parameters and air pollutants in an urban environment. *Glob. J. Environ. Sci. Manag.* **2019**, *5*, 265–278.
- 24. Islam, M.; Sharmin, M.; Ahmed, F. Predicting air quality of Dhaka and Sylhet divisions in Bangladesh: A time series modeling approach. *Air. Qual. Atmos. Health* **2020**, *13*, 607–615. [CrossRef]
- Su, Y.; Yu, Y.Q. Dynamic early warning of regional atmospheric environmental carrying capacity. *Sci. Total Environ.* 2020, 714, 136684. [CrossRef] [PubMed]
- Bao, J.M.; Li, H.; Wu, Z.H.; Zhang, X.; Zhang, H.; Li, Y.F.; Qian, J.; Chen, J.H.; Deng, L.Q. Atmospheric carbonyls in a heavy ozone pollution episode at a metropolis in Southwest China: Characteristics, health risk assessment, sources analysis. *J. Environ. Sci.* 2022, 113, 40–54. [CrossRef]
- Wang, S.P.; Li, K.Q.; Liang, S.K.; Zhang, P.; Lin, G.H.; Wang, X.L. An integrated method for the control factor identification of resources and environmental carrying capacity in coastal zones: A case study in Qingdao, China. Ocean. Coast. Manag. 2017, 142, 90–97. [CrossRef]
- 28. Su, Y.; Yu, Y.Q. Spatial association effect of regional pollution control. J. Clean. Prod. 2019, 213, 540–552. [CrossRef]
- Zhai, B.X.; Chen, J.G. Development of a stacked ensemble model for forecasting and analyzing daily average PM_{2.5} concentrations in Beijing, China. *Sci. Total Environ.* 2018, 635, 644–658. [CrossRef]
- 30. Liu, H.; Jin, K.R.; Duan, Z. Air PM_{2.5} concentration multi-step forecasting using a new hybrid modeling method: Comparing cases for four cities in China. *Atmos. Pollut. Res.* **2019**, *10*, 1588–1600. [CrossRef]
- Zafra, C.; Angel, Y.; Torres, E. ARIMA analysis of the effect of land surface coverage on PM₁₀ concentrations in a high-altitude megacity. *Atmos. Pollut. Res.* 2017, *8*, 660–668. [CrossRef]

- 32. Liu, P.G. Simulation of the daily average PM₁₀ concentrations at Ta-Liao with Box–Jenkins time series models and multivariate analysis. *Atmos. Environ.* **2009**, *43*, 2104–2113. [CrossRef]
- Zhang, Y.; Bocquet, M.; Mallet, V.; Seigneur, C.; Baklanov, A. Real-time air quality forecasting, part I: History, techniques and current status. *Atmos. Environ.* 2012, 60, 632–655. [CrossRef]
- 34. Munawar, S.; Hamid, M.; Khan, M.S.; Ahmed, A.; Hameed, N. Health monitoring considering air quality index prediction using Neuro-Fuzzy Inference Model: A case study of Lahore, Pakistan. J. Basic Appl. 2017, 13, 123–132. [CrossRef]
- 35. Wang, D.Y.; Wei, S.; Luo, H.Y.; Yue, C.Q.; Grunder, O. A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine. *Sci. Total Environ.* **2017**, *580*, 719–733. [CrossRef]
- Li, H.M.; Wang, J.Z.; Yang, H.F. A novel dynamic ensemble air quality index forecasting system. *Atmos. Pollut. Res.* 2020, 11, 1258–1270. [CrossRef]
- 37. Phruksahiran, N. Improvement of air quality index prediction using geographically weighted predictor methodology. *Urban. Clim.* **2021**, *38*, 100890. [CrossRef]
- Du, Z.J.; Heng, J.N.; Niu, M.F.; Sun, S.L. An innovative ensemble learning air pollution early-warning system for China based on incremental extreme learning machine. *Atmos. Pollut. Res.* 2021, 12, 101153. [CrossRef]
- Li, Y.Y.; Huang, S.; Yin, C.X.; Sun, G.H.; Ge, C. Construction and countermeasure discussion on government performance evaluation model of air pollution control: A case study from Beijing-Tianjin-Hebei region. J. Clean. Prod. 2020, 254, 120072. [CrossRef]
- 40. Guo, J.M.; Zhao, M.J.; Xue, P.; Liang, X.; Fan, G.T.; Ding, B.H.; Liu, J.J.; Liu, J.P. New indicators for air quality and distribution characteristics of pollutants in China. *Build. Environ.* **2020**, *172*, 106723. [CrossRef]
- 41. Olstrup, H.; Johansson, C.; Forsberg, B.; Tornevi, A.; Ekebom, A.; Meister, K. A multi-pollutant air quality health index (AQHI) based on short-term respiratory effects in Stockholm, Sweden. *Int. J. Environ. Res. Publ. Health* **2019**, *16*, 105. [CrossRef]
- 42. Pozzer, A.; Bacer, S.; Sappadina, S.D.Z.; Predicatori, F.; Caleffi, A. Long-term concentrations of fine particulate matter and impact on human health in Verona. *Atmos. Pollut. Res.* **2019**, *10*, 731–738. [CrossRef]
- Feng, R.; Zheng, H.J.; Gao, H.; Zhang, A.R.; Huang, C.; Zhang, J.X.; Luo, K.; Fan, J.R. Recurrent neural network and random forest for analysis and accurate forecast of atmospheric pollutants: A case study in Hangzhou, China. *J. Clean. Prod.* 2019, 231, 1005–1015. [CrossRef]
- Franceschi, F.; Cobo, M.; Figueredo, M. Discovering relationships and forecasting PM₁₀ and PM_{2.5} concentrations in Bogotá, Colombia, using artificial neural networks, principal component analysis, and k-means clustering. *Atmos. Pollut. Res.* 2018, 9, 912–922. [CrossRef]
- Gao, M.; Yin, L.T.; Ning, J.C. Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. *Atmos. Environ.* 2018, 184, 129–139. [CrossRef]
- Li, X.; Peng, L.; Yao, X.J.; Cui, S.L.; Hu, Y.; You, C.Z.; Chi, T.H. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ. Pollut.* 2017, 231, 997–1004. [CrossRef] [PubMed]
- Liu, H.; Chen, C. Prediction of outdoor PM_{2.5} concentrations based on a three-stage hybrid neural network model. *Atmos. Pollut. Res.* 2020, *11*, 469–481. [CrossRef]
- 48. Pak, U.; Kim, C.; Ryu, U.; Sok, K.; Pak, S. A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction. *Air. Qual. Atmos. Health* **2018**, *11*, 883–895. [CrossRef]
- 49. Zhou, Y.T.; Zhao, X.Y.; Lin, K.P.; Wang, C.H.; Li, L.L. A Gaussian process mixture model-based hard-cut iterative learning algorithm for air quality prediction. *Appl. Soft. Comput.* **2019**, *85*, 105789. [CrossRef]
- 50. Yu, H.Y.; Gao, M.M.; Zhang, H.F.; Chen, Y. Dynamic modeling for SO₂-NO_x emission concentration of circulating fluidized bed units based on quantum genetic algorithm-Extreme learning machine. *J. Clean. Prod.* **2021**, *324*, 129170. [CrossRef]
- Middya, A.I.; Roy, S. Pollutant specific optimal deep learning and statistical model building for air quality forecasting. *Environ. Pollut.* 2022, 301, 118972. [CrossRef]
- 52. Sun, W.; Li, Z.Q. Hourly PM_{2.5} concentration forecasting based on feature extraction and stacking-driven ensemble model for the winter of the Beijing-Tianjin-Hebei area. *Atmos. Pollut. Res.* **2020**, *11*, 110–121. [CrossRef]
- 53. Ribeiro, V.M. Sulfur dioxide emissions in Portugal: Prediction, estimation and air quality regulation using machine learning. *J. Clean. Prod.* 2021, 317, 128358. [CrossRef]
- Han, G.X.; Shi, Y.J.; Lu, Y.L.; Liu, C.F.; Cui, H.T.; Zhang, M. Coupling relation between urbanization and ecological risk of PAHs on coastal terrestrial ecosystem around the Bohai and Yellow Sea. *Environ. Pollut.* 2021, 268, 115680. [CrossRef] [PubMed]
- 55. Yan, D.; Ren, X.H.; Zhang, W.L.; Li, Y.Y.; Miao, Y. Exploring the real contribution of socioeconomic variation to urban PM_{2.5} pollution: New evidence from spatial heteroscedasticity. *Sci. Total Environ.* **2022**, *806*, 150929. [CrossRef] [PubMed]
- Bilal, M.; Mhawish, A.; Nichol, J.E.; Qiu, Z.F.; Nazeer, M.; Ali, M.A.; Leeuw, G.D.; Levy, R.C.; Wang, Y.; Chen, Y.; et al. Air pollution scenario over Pakistan: Characterization and ranking of extremely polluted cities using long-term concentrations of aerosols and trace gases. *Remote. Sens. Environ.* 2021, 264, 112617. [CrossRef]
- 57. Liu, M.M.; Huang, Y.N.; Jin, Z.; Ma, Z.W.; Liu, X.Y.; Zhang, B.; Liu, Y.; Yu, Y.; Wang, J.N.; Bi, J.; et al. The nexus between urbanization and PM_{2.5} related mortality in China. *Environ. Pollut.* **2017**, 227, 15–23. [CrossRef] [PubMed]
- Shi, Y.; Bilal, M.; Ho, H.C.; Omar, A. Urbanization and regional air pollution across South Asian developing countries—A nationwide land use regression for ambient PM_{2.5} assessment in Pakistan. *Environ. Pollut.* 2020, 266, 115145. [CrossRef]

- 59. Chen, H.X. Fix-and-optimize and variable neighborhood search approaches for multi-level capacitated lot sizing problems. *Omega* **2015**, *56*, 25–36. [CrossRef]
- Gansterer, M.; Födermayr, P.; Hartl, R.F. The capacitated multi-level lot-sizing problem with distributed agents. *Int. J. Prod. Econ.* 2021, 235, 108090. [CrossRef]
- 61. Dunke, F.; Nickel, S. A general modeling approach to online optimization with lookahead. Omega 2016, 63, 134–153. [CrossRef]
- 62. Hu, L.Y.; Wang, C.; Ye, Z.R.; Wang, S. Estimating gaseous pollutants from bus emissions: A hybrid model based on GRU and XGBoost. *Sci. Total Environ.* **2021**, *783*, 146870. [CrossRef]
- 63. Ma, M.H.; Zhao, G.; He, B.H.; Li, Q.; Dong, H.Y.; Wang, S.G.; Wang, Z.L. XGBoost-based method for flash flood risk assessment. *J. Hydrol.* **2021**, 598, 126382. [CrossRef]
- 64. Dong, W.; Huang, Y.M.; Lehane, B.; Ma, G.W. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. *Automat. Constr.* **2020**, *114*, 103155. [CrossRef]
- 65. Ma, J.; Cheng, J.C.P.; Xu, Z.R.; Chen, K.Y.; Lin, C.Q.; Jiang, F.F. Identification of the most influential areas for air pollution control using XGBoost and Grid Importance Rank. *J. Clean. Prod.* **2020**, *274*, 122835. [CrossRef]
- Wang, Z.S.; Lv, J.G.; Tan, Y.F.; Guo, M.; Gu, Y.Y.; Xu, S.; Zhou, Y.H. Temporospatial variations and Spearman correlation analysis of ozone concentrations to nitrogen dioxide, sulfur dioxide, particulate matters and carbon monoxide in ambient air, China. *Atmos. Pollut. Res.* 2019, 10, 1203–1210. [CrossRef]
- 67. Zhou, Y.; Li, S.J. BP neural network modeling with sensitivity analysis on monotonicity based on Spearman coefficient. *Chem. Intell. Lab.* **2020**, 200, 103977. [CrossRef]
- Zhi, Y.J.; Jin, Z.H.; Lu, L.; Yang, T.; Zhou, D.Y.; Pei, Z.B.; Wu, D.Q.; Fu, D.M.; Zhang, D.W.; Li, X.G. Improving atmospheric corrosion prediction through key environmental factor identification by random forest-based model. *Corros. Sci.* 2021, 178, 109084. [CrossRef]
- 69. Rupakheti, D.; Yin, X.F.; Rupakheti, M.; Zhang, Q.G.; Li, P.; Rai, M.; Kang, S.C. Spatio-temporal characteristics of air pollutants over Xinjiang, northwestern China. *Environ. Pollut.* **2021**, *268*, 115907. [CrossRef]
- Kim, Y.; Lee, I.; Farquhar, J.; Kang, J.; Villa, I.M.; Kim, H. Multi isotope systematics of precipitation to trace the sources of air pollutants in Seoul, Korea. *Environ. Pollut.* 2021, 286, 117548. [CrossRef]
- Sanchez, G.M.; Nejadhashemi, A.P.; Zhang, Z.; Marquart-Pyatt, S.; Habron, G.; Shortridge, A. Linking watershed-scale stream health and socioeconomic indicators with spatial clustering and structural equation modeling. *Environ. Modell. Softw.* 2015, 70, 113–127. [CrossRef]
- Angelini, M.E.; Heuvelink, G.B.M. Including spatial correlation in structural equation modelling of soil properties. *Spat. Stat.* 2018, 25, 35–51. [CrossRef]