Article

Spatiotemporal Patterns of Air Pollutants over the Epidemic Course: A National Study in China

Kun Qin [1,2,†], Zhanpeng Wang [1,2,†], Shaoqing Dai [2,3,†], Yuchen Li [2,4,5], Manyao Li [1,2], Chen Li [1,2], Ge Qiu [1,2], Yuanyuan Shi [1,2], Chun Yin [1,2], Shujuan Yang [2,6] and Peng Jia [1,2,7,8,*]

1 School of Resource and Environmental Sciences, Wuhan University, Wuhan 430072, China
2 International Institute of Spatial Lifecourse Health (ISLE), Wuhan University, Wuhan 430072, China
3 Faculty of Geo-Information Science and Earth Observation, University of Twente, 7500 AE Enschede, The Netherlands
4 MRC Epidemiology Unit, University of Cambridge, Cambridge CB2 1TN, UK
5 Department of Geography, The Ohio State University, Columbus, OH 43210, USA
6 West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu 610041, China
7 Hubei Luojia Laboratory, Wuhan 430072, China
8 School of Public Health, Wuhan University, Wuhan 430071, China
* Correspondence: jiapengff@hotmail.com
† These authors contributed equally to this work.

Abstract: Air pollution has been standing as one of the most pressing global challenges. The changing patterns of air pollutants at different spatial and temporal scales have been substantially studied all over the world, which, however, were intricately disturbed by COVID-19 and subsequent containment measures. Understanding fine-scale changing patterns of air pollutants at different stages over the epidemic's course is necessary for better identifying region-specific drivers of air pollution and preparing for environmental decision making during future epidemics. Taking China as an example, this study developed a multi-output LightGBM approach to estimate monthly concentrations of the six major air pollutants (i.e., $PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and CO) in China and revealed distinct spatiotemporal patterns for each pollutant over the epidemic's course. The 5-year period of 2019–2023 was selected to observe changes in the concentrations of air pollutants from the pre-COVID-19 era to the lifting of all containment measures. The performance of our model, assessed by cross-validation R^2, demonstrated high accuracy with values of 0.92 for $PM_{2.5}$, 0.95 for PM_{10}, 0.95 for O_3, 0.90 for NO_2, 0.79 for SO_2, and 0.82 for CO. Notably, there was an improvement in the concentrations of particulate matter, particularly for $PM_{2.5}$, although PM_{10} exhibited a rebound in northern regions. The concentrations of SO_2 and CO consistently declined across the country over the epidemic's course ($p < 0.001$ and $p < 0.05$, respectively), while O_3 concentrations in southern regions experienced a notable increase. Concentrations of air pollutants in the Beijing–Tianjin–Hebei region were effectively controlled and mitigated. The findings of this study provide critical insights into changing trends of air quality during public health emergencies, help guide the development of targeted interventions, and inform policy making aimed at reducing disease burdens associated with air pollution.

Keywords: air pollutant; $PM_{2.5}$; PM_{10}; emerging hot spot analysis; multi-output LightGBM

1. Introduction

Air pollution has been standing as one of the most pressing global challenges, predominantly emanating from anthropogenic activities, such as chemical emissions from industries, exhaust emissions from vehicles, and the combustion of fossil fuels [1,2]. Air pollutants that are harmful to human health mainly include particulate matter with aerodynamic diameter <2.5 μm ($PM_{2.5}$) and <10 μm (PM_{10}), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), and carbon monoxide (CO) [3,4]. The formation and dispersion of air

pollutants carry profound repercussions that extend across diverse domains, including climate change, ecological well-being, and human health [5]. The changing patterns of air pollutants at different spatial (e.g., country, province/state, city) and temporal (e.g., yearly, seasonal, monthly) scales have been substantially studied all over the world, which, however, were intricately disturbed by the Coronavirus Disease 2019 (COVID-19) pandemic and subsequent containment measures [6–9]. It was observed that the concentrations of NO_2 and $PM_{2.5}$ decreased by about 60% and 31% in 34 countries during the lockdown period, with mixed trends for O_3 [10]. Another study reported significant declines in NO_2, SO_2, CO, $PM_{2.5}$, and PM_{10} levels in twenty major cities across six continents [11]. Understanding fine-scale changing patterns of air pollutants at different stages over the epidemic's course is necessary for better identifying region-specific drivers of air pollution and preparing for environmental decision making during future epidemics [12,13].

Measures against COVID-19 have been implemented to different degrees across provincial units and even cities during 2020–2022, for instance, in China [14,15]. Existing studies have only focused on short-term changing patterns of air pollutant concentrations during the implementation of COVID-19 containment measures, especially soon after the onset of COVID-19 [16,17]. For example, one previous study on the basis of ground monitoring data of air pollutants from 86% of the Chinese cities reported that the air quality index decreased on average by approximately 11.0% from January 2019 to July 2020 [18]. Other studies have been carried out in sparse areas. For example, one study conducted in the five northern provinces/municipalities reported decreased NO_2 and $PM_{2.5}$ and increased O_3 from January to March 2020 [19]; another study conducted in Shanghai reported a decline in daily concentrations of $PM_{2.5}$, PM_{10}, and NO_2 from March to June 2022 [20]. A full picture of spatiotemporal patterns of all major air pollutants across the country remains lacking. To devise comprehensive mitigation strategies for air pollution that can be applied at all levels (from national to local), it is essential to understand the collective dynamics of all major air pollutants.

To unveil spatiotemporal characteristics of air pollutants and evolutionary patterns of pollutant distribution, this study aimed to estimate monthly concentrations of the six major air pollutants (i.e., $PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and CO) in China and reveal distinct spatiotemporal patterns for each pollutant over the course of the epidemic. A 5-year period of 2019–2023 was selected to observe their changes from the pre–COVID-19 era to the lifting of all containment measures. The findings of this study provide critical insights into changing trends of air quality during public health emergencies, help guide the development of targeted interventions, and inform policy making aimed at reducing disease burdens associated with air pollution.

2. Methods

2.1. Datasets

The data used in this study included ground-based measurements of air pollutants, satellite-derived data, and other auxiliary data (Table 1).

2.1.1. Ground-Based Measurements

The hourly concentrations of $PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and CO during 2019–2023 were obtained from approximately 2020 national air quality monitoring stations administered by the China National Environmental Monitoring Center (Figure 1). They were then averaged over days and further over months to calculate monthly mean concentrations of air pollutants. The hourly in situ observations were conducted using either point analyzers or open path analyzers from ambient air quality continuous automated monitoring systems. $PM_{2.5}$ and PM_{10} were measured by the tapered element oscillating microbalance or the β-attenuation method with a precision of ± 1.5 or 0.1 $\mu g/m^3$, respectively. NO_2, SO_2, and O_3 were measured by a differential optical absorption spectroscopy (DOAS) method using open path analyzers or alternative methods (i.e., chemiluminescence, ultraviolet fluorescence, and UV spectrophotometry, respectively) using point analyzers. CO was

measured using the non-dispersive infrared absorption method or the gas filter correlation infrared absorption method with point analyzers. The concentrations of four gaseous pollutants (NO_2, SO_2, O_3, and CO) were measured with a mean relative error of less than 5%. All monitors underwent standard calibration once a week and precision tests every three months.

Table 1. Summary of the datasets.

Variable	Content	Unit	Spatial Resolution	Temporal Resolution	Data Source
Ground-based measurements					
$PM_{2.5}$	Ground-monitored $PM_{2.5}$ concentration	$\mu g/m^3$			
PM_{10}	Ground-monitored PM_{10} concentration	$\mu g/m^3$			
O_3	Ground-monitored O_3 concentration	$\mu g/m^3$	In situ	Hourly	China Environmental Monitoring Center
NO_2	Ground-monitored NO_2 concentration	$\mu g/m^3$			
SO_2	Ground-monitored SO_2 concentration	$\mu g/m^3$			
CO	Ground-monitored CO concentration	mg/m^3			
Satellite-derived data					
AOD	Aerosol optical depth	–	1 km × 1 km	Daily	Moderate-resolution Imaging Spectroradiometer (MODIS)
AAI	Absorbing aerosol index	–			
TROPOMI CO	CO column number density	mol/m^2			
TROPOMI NO_2	Tropospheric NO_2 column number density	mol/m^2	3.5 km × 7 km	Daily	Tropospheric Monitoring Instrument (TROPOMI)
TROPOMI O_3	O_3 column number density	mol/m^2			
Auxiliary data					
TEM	Temperature at 2 m	K			
DT	Dewpoint temperature at 2 m	K			
WU	U-component of wind at 10 m	m/s			
WV	V-component of wind at 10 m	m/s	0.1° × 0.1°		
SP	Surface pressure	hPa			
ET	Total evaporation	mm		Monthly	European Centre for Medium-Range Weather Forecasts Reanalysis version 5 (ERA5)
PRE	Total precipitation	mm			
BLH	Boundary layer height	m			
RH	Relative humidity	%			
UVB	Downward UV radiation at the surface	J/m^2	0.25° × 0.25°		
SSR	Surface net solar radiation	J/m^2			
STRD	Surface net thermal radiation	J/m^2			
EI	Emission inventory	$kt\ CO_2/cell$	0.1° × 0.1°	Annual	Global Infrastructure Emissions Detector (GID)
NTL	Nighttime light	$nW/sr/cm^2$	500 m × 500 m	Monthly	Visible Infrared Imaging Radiometer (VIIRS)
POP	Population counts	–	1 km × 1 km	Annual	LandScan Global Population Data
NDVI	Normalized difference vegetation index	–	1 km × 1 km	Monthly	Moderate-resolution Imaging Spectroradiometer (MODIS)
LUC	Land use cover	–	1 km × 1 km	Annual	
DEM	Surface elevation	m	90 m × 90 m	–	Shuttle Radar Topography Mission (SRTM)
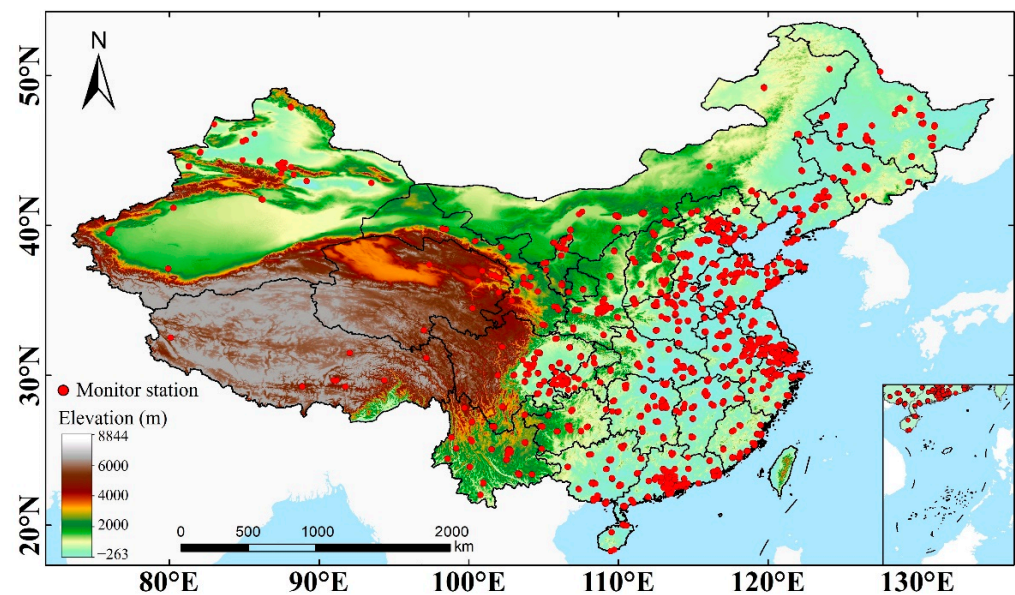
Figure 1. Spatial distribution of national automatic air quality monitoring stations in China.

Variations in the number of ground monitoring stations across cities may cause missing values in the original monitoring data of air quality, attributed to factors such as calibration of monitoring instruments, daily maintenance, and issues like communication failures or power outages. In this study, preprocessing involved the elimination of missing data and the evaluation of outliers using the Laida criterion, which excluded records falling outside the range of ($\mu - 3\sigma$, $\mu + 3\sigma$), where μ and σ denote means and standard deviation, respectively. Per the "China Ambient Air Quality Standard" (GB 3095-2012) [21], daily mean values were calculated from effective hourly data [22]. Data on a given day were considered invalid if recorded for less than 20 h. Similarly, data for calculating monthly mean values were considered invalid if recorded for less than 27 days in a month (or less than 25 days in February), resulting in 92,284 valid data records.

2.1.2. Satellite-Derived Data

The data products of the moderate resolution imaging spectroradiometer (MODIS), equipped on the National Aeronautics and Space Administration's (NASA) Terra/Aqua satellites, and of the tropospheric monitoring instrument (TROPOMI), equipped on the European Space Agency's (ESA) Sentinel-5P satellite, were used in this study. Specifically, the MODIS-derived multi-angle implementation of atmospheric correction (MAIAC) aerosol optical depth (AOD) data (product number MCD19A2) offers a spatial resolution of 1 km × 1 km and a temporal resolution of 1 day. The MAIAC uses time series analysis and image processing methods for atmospheric correction and aerosol inversion in regions with dark vegetation coverage and bright surfaces (e.g., deserts), thereby enhancing the effective observation range [23]. MAIAC offers quality assurance bands that signify retrieval quality, encompassing a cloud mask, a land/water/snow mask, and an adjacency mask indicating proximity to cloud or snow. MAIAC AOD at 550 nm was utilized, excluding pixels affected by cloud contamination or snow cover. Monthly AOD was derived by computing the average of all valid values for each image element over the month. The TROPOMI enables effective observation of trace gas components worldwide [24], where the offline Level 3 products included the absorbing aerosol index (AAI) and column number densities of O_3, CO, and NO_2 with a spatial resolution of approximately 3.5 km × 7 km. The raw data were aggregated to monthly averages and resampled to 10 km × 10 km grids, where observations were available for the largest number of pixels. For the few pixels without values, we employed a time linear interpolation method to interpolate based on the closest date before and after [25–27].

2.1.3. Auxiliary Data

The auxiliary data used in this study included meteorological, land surface, and socioeconomic data. The meteorological data were from the fifth generation European center for medium-range weather forecasts atmospheric reanalysis of the global climate (ERA5), produced by the Copernicus climate change service, which included 2m temperature (TEM), 2m dewpoint temperature (DT), 10m u-component of wind (WU), 10m v-component of wind (WV), surface pressure (SP), total evaporation (ET), total precipitation (PRE), boundary layer height (BLH), relative humidity (RH), downward UV radiation at the surface (UVB), surface net solar radiation (SSR), and surface net thermal radiation (STRD) [28,29]. Land surface data included the normalized difference vegetation index (NDVI) from the MOD13A3 product, the digital elevation model (DEM) from the shuttle radar topography mission (STRM) digital elevation dataset, and land use cover (LUC) data from the MCD12Q1 product. Socioeconomic data included population from the annual LandScan Population Data Global 1 km, emission inventory (EI) from the Global Infrastructure emissions Detector (GID) with a spatial resolution of $0.1° × 0.1°$ and covering 1990–2022, and nighttime light (NTL) data from the visible infrared imaging radiometer suite (VIIRS). All auxiliary data were converted into 10 km × 10 km grids using a bilinear interpolation method, to be consistent with the satellite-derived data, except for LUC data, which were resampled using a majority resampling method [30].

2.2. Extraction of Spatial and Temporal Features

Given the pronounced variability in the distribution of air pollutants over space and time, considerable fluctuations in their concentrations occurred. Incorporating latitude and longitude coordinates as spatial locations has been proved to be untenable for decision-making tree models, as it ostensibly encodes geographical information and thus predisposes the issue of threshold segmentation during feature fitting [31].

To address this issue, this study embraced a geocoding method to delineate relative spatial positions and capture regional variations. Specifically, we computed the distances from each grid to the centroid and four corner points of the rectangular grid (i.e., D1, D2, D3, D4, D5) [32]. We utilized a haversine method to transform latitudes and longitudes into spherical distances.

$$DIS = 2 \times r \times \text{asin}\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\gamma_2 - \gamma_1}{2}\right)}\right) \tag{1}$$

where r denotes Earth's mean radius ($\approx$ 6371 km) and γ and φ denote the longitude and latitude of a given point on the sphere, respectively.

To represent temporal information, the day of the year and month are commonly used metrics, which, however, do not adequately convey the ongoing progression and seasonal patterns inherent in temporal data. To address this issue, in this study, we converted months into cartesian coordinates (t_x and t_y) [33], which involved normalizing the time period to the range from 0 to 2π and transforming it into polar coordinates as follows:

$$\begin{bmatrix} t_x \\ t_y \end{bmatrix} = \begin{bmatrix} \cos\left(2\pi\frac{Month}{T}\right) \\ \sin\left(2\pi\frac{Month}{T}\right) \end{bmatrix} \tag{2}$$

where T is equal to 12.

2.3. Algorithm Description

In this study, we combined a multi-output regressor with a LightGBM framework to develop a multi-output LightGBM model, which was capable of estimating concentrations of the six air pollutants by employing ground-based measurements, satellite-derived data, auxiliary data, and spatial and temporal features. Subsequently, we analyzed predictive results from the model using an emerging hot spot analysis to evaluate the patterns of

hot and cold spots of diverse air pollutants in different regions, revealing their latest spatiotemporal patterns and trends of dynamic changes (Figure 2).
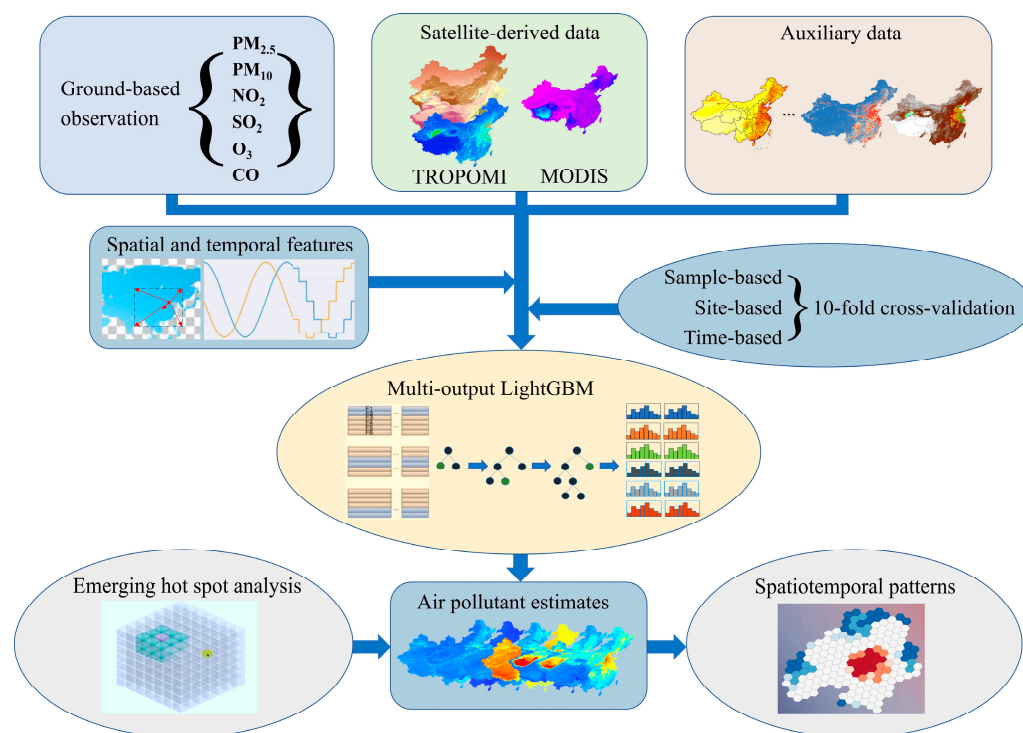
Figure 2. Flowchart of the modeling and analysis process for this study. LightGBM, light gradient boosting machine; MODIS, moderate resolution imaging spectroradiometer; TROPOMI, tropospheric monitoring instrument.

A multi-output regression algorithm was integrated with a LightGBM framework to develop a multi-output LightGBM model, which is capable of predicting concentrations of the six air pollutants that are highly correlated with multiple predictors. LightGBM is an open-source framework that implements the Gradient-Boosted Decision Tree (GBDT) algorithm [34]. Regarding runtime efficiency, it outperforms the well-known traditional machine learning models, such as Extreme Gradient Boosting (XGBoost) and Extremely Randomized Trees (ERTs). This is owing to several key optimizations, including gradient-based one-side sampling, exclusive feature bundling, the histogram-based algorithm, and the leaf-wise tree growth strategy with depth restriction. These optimizations together reduce the complexity of the model and the risk of overfitting, thereby enhancing the model's efficiency. Also, the efficient algorithmic structure of LightGBM could significantly mitigate the challenge of processing large volumes of environmental data efficiently [35,36].

The multi-output regression algorithm is a machine learning task approach aimed at predicting multiple output tasks for each input sample [37]. It assigns a set of specialized regressors to each target, thereby extending regressors that do not inherently support regression with multiple objectives. Each target can be accurately represented by a regressor and, therefore, accessed through its corresponding model, which provides a straightforward strategy for extending single-output regression models to support multiple objectives. The six air pollutants originate from certain shared sources and exhibit chemical or physical connections under identical meteorological conditions. Consequently, constructing multiple single-output models for individual pollutants involves using similar predictors and model structures, which leads to duplicated efforts. In contrast, employing a multi-output model to simultaneously estimate concentrations of these pollutants can leverage their correlations more effectively, thereby enhancing efficiency [38]. To compare the benefits of adopting the

multi-output regression algorithm versus a single-output regression approach, we trained a separate single-output LightGBM model for each pollutant using the same samples.

2.4. Model Evaluation

Three distinct 10-fold cross-validation (CV) methods, i.e., sample-based, site-based, and time-based, were used to comprehensively evaluate the performance of our multi-output LightGBM model [39,40]. The sample-based CV randomly partitions the target dataset into ten subsets and, alternately, uses nine for model training and one for model testing (i.e., repeated ten times to ensure that each subset was used for testing), which facilitates a comprehensive evaluation of generalization capabilities of the model across various subsets. The site-based CV considers monitoring sites while partitioning the target dataset and ensures that both training and test sets include data from different locations, which enhances the model's ability to predict across the study area. The time-based CV considers the time of data collection while partitioning the target dataset and ensures that both training and test sets include data collected in different periods, which improves the model's ability to generalize over time.

We employed several evaluation metrics to comprehensively assess the performance of our model, including the coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE). R^2 gauges the ability to explain the total variance, with a value closer to 1 indicating superior performance. The RMSE and MAE quantify prediction errors, with the RMSE exhibiting a greater sensitivity to large errors. The calculation methods of these metrics are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

$$MAE = \frac{1}{n} \times \sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

where y_i denotes the observed concentrations from monitoring stations; $\overline{y}_i$ denotes the mean concentration; $\hat{y}_i$ denotes the predicted concentration; and n denotes the number of samples.

In addition, during the model fitting, the importance of each feature, including satellite derived (AAI, TROPOMI CO, TROPOMI NO$_2$, TROPOMI O$_3$, AOD), auxiliary (TEM, RH, PRE, ET, BLH, DT, SP, WU, WV, UVB, SSR, STRD, EI, NTL, POP, NDVI, LUC, DEM), and other generated features (D1, D2, D3, D4, D5, t_x, t_y), was calculated and normalized to show relative contributions of predictors and enhance the interpretability of our model.

2.5. Trend Analysis

To assess spatial clustering and temporal variations of air pollutants over the 5-year period, an emerging hot spot analysis was used to identify spatiotemporal patterns of concentrations of air pollutants over different periods. It considers both spatial and temporal dimensions by forming a space–time cube from successive layers of data from different time cross-sections to evaluate whether observed clusters or outliers are statistically significant [41]. Specifically, a combination of two statistical methods was used: the Getis–Ord Gi* statistic identifies the location and size of spatial clusters of concentrations, and then the Mann–Kendall trend test detects temporal trends of concentrations at each location [42–44]. In the Getis–Ord Gi*, specific parameters were set for neighborhood distance and time steps to detect statistically significant spatial clusters, with both hot (high concentrations) and cold spots (low concentrations) identified and the corresponding z-scores and *p*-values generated. In this study, the neighborhood distance was set as 10 km (~0.1°) and the time step was set as 1 month. In the Mann–Kendall trend test, temporal trends of hot and cold spots were evaluated and all locations were classified into the seventeen spatiotemporal

patterns: eight (changing) patterns (i.e., new, continuous, intensifying, persistent, diminishing, dispersed, oscillating, and historical) of hot and cold spots separately, as well as the category "non-significance" [41]. To avoid complicated interpretation, the patterns existing in <1% of the study area were not shown in the results.

3. Results

3.1. Spatial and Temporal Distribution of Air Pollutants

Different spatiotemporal patterns of $PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and CO were observed during 2019–2023 (Figure 3). $PM_{2.5}$ and PM_{10} consistently exhibited higher levels, especially in the Taklimakan Desert, North China Plain, Beijing–Tianjin–Hebei (BTH) region, and Yangtze River Delta (YRD) in the east. In contrast to the widespread and high concentrations of particulate matter, NO_2, SO_2, and CO levels were higher in coastal areas, particularly around the BTH region. O_3 exhibited a different changing pattern, with notable increases observed in the west, north, northwest, and coastal regions, indicating that O_3 has a broader impact area compared to the other pollutants.
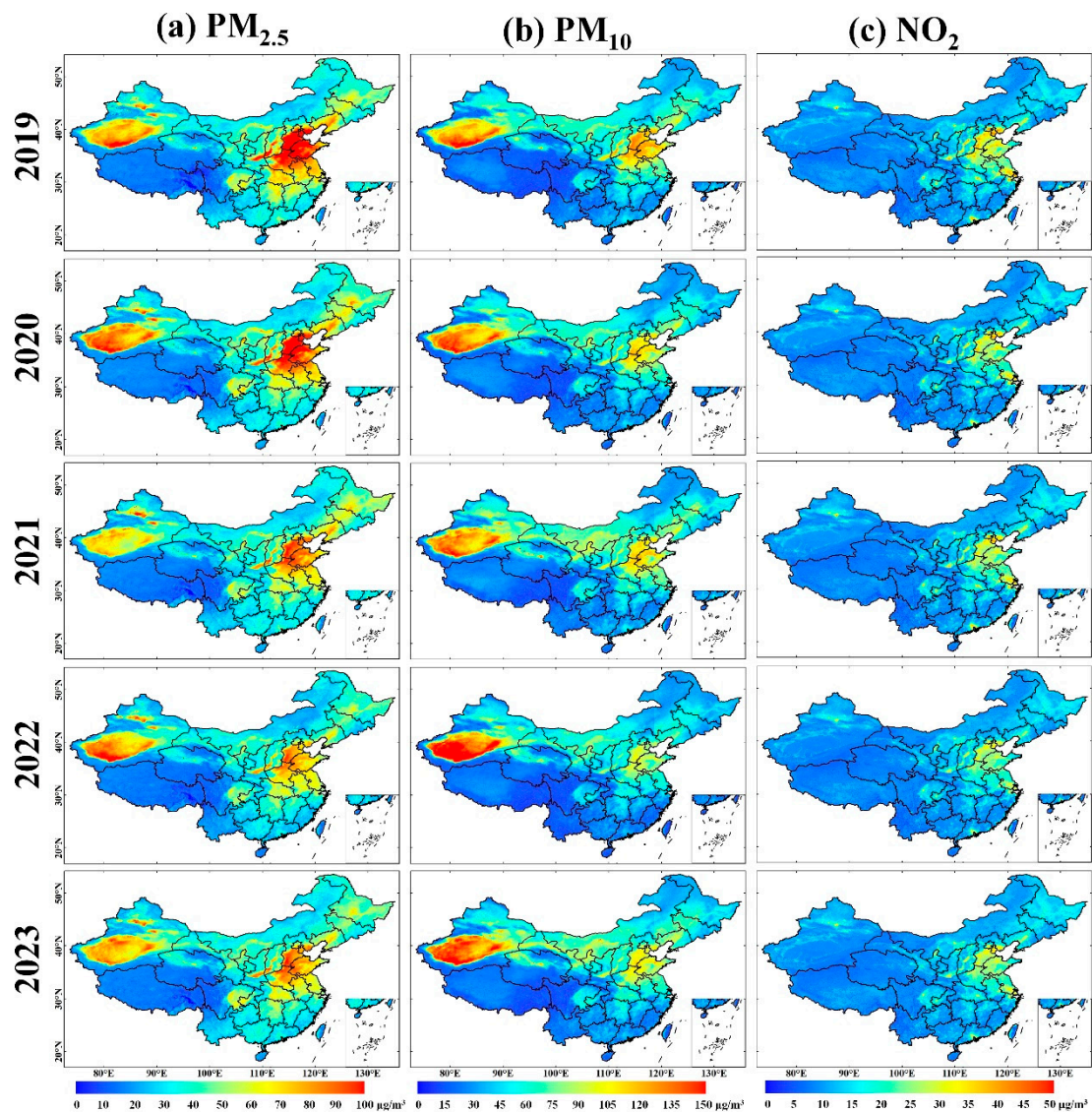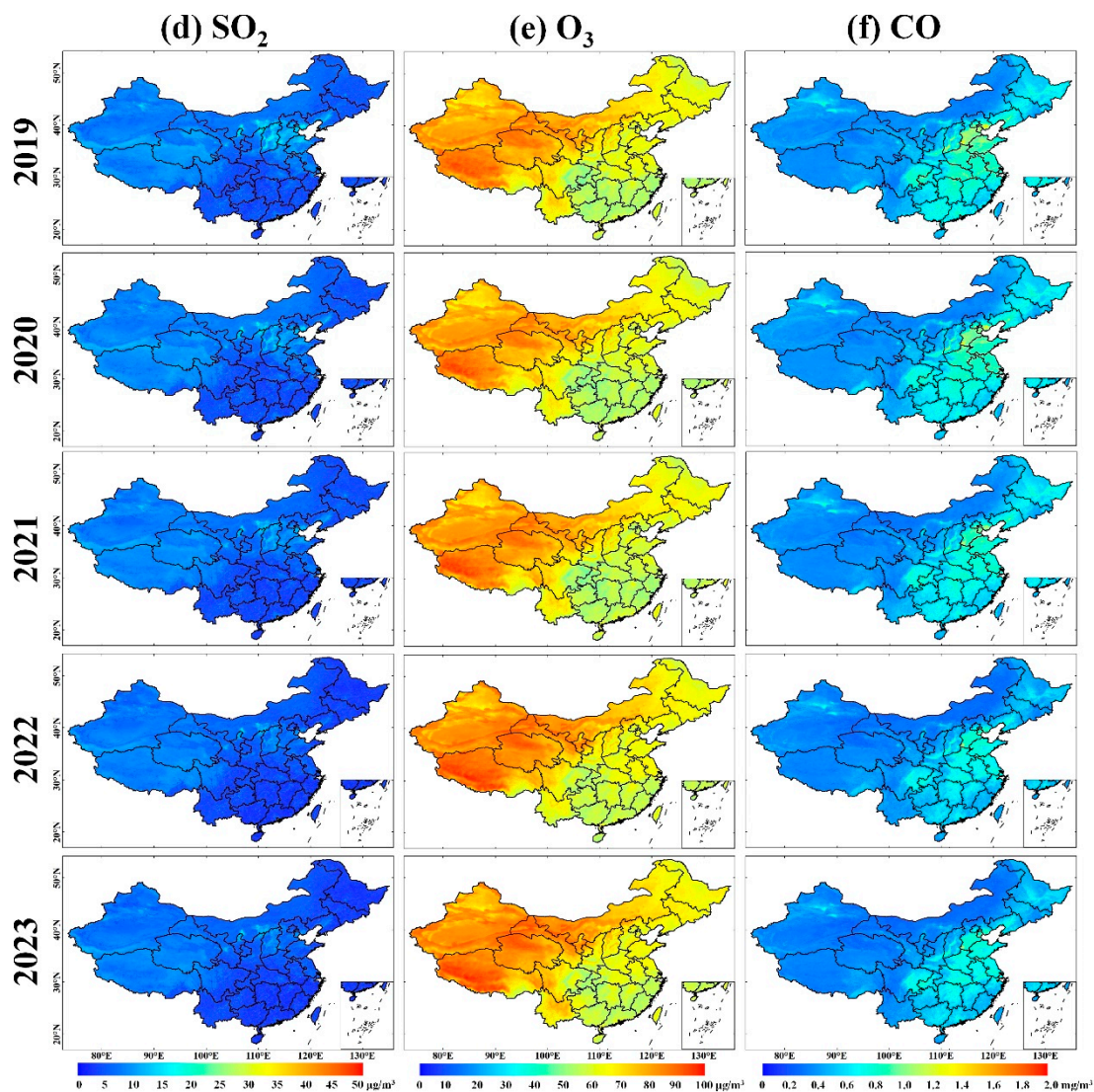

Figure 3. *Cont.*

Figure 3. Spatial distribution of annual mean concentration of $PM_{2.5}$ (**a**), PM_{10} (**b**), NO_2 (**c**), SO_2 (**d**), O_3 (**e**), and CO (**f**) in China from 2019 to 2023. The unit is mg/m^3 for CO and $\mu g/m^3$ for other air pollutants.

The trend of mean concentrations of pollutants over the past 60 months demonstrated a clear seasonal pattern (Figure 4). Monthly and annual data all over China were generated by directly averaging the estimated monthly concentrations at each grid. $PM_{2.5}$, NO_2, SO_2, and CO levels reached peaks in the spring and winter, likely due to increased emissions from heating-related coal and fossil fuel use. While PM_{10} is generally consistent with this trend, it is worth noting that it peaked twice each year, likely due to dust storms that are common in spring, leading to elevated levels post-winter. O_3 displays the opposite trend, with lower levels in winter and higher levels in summer. From 2019 to 2023, the annual mean concentrations of $PM_{2.5}$, PM_{10}, NO_2, and O_3 demonstrated a consistent pattern characterized by an initial decrease followed by a subsequent increase (Table 2). Both $PM_{2.5}$ and PM_{10} experienced a resurgence in 2022, while O_3 underwent a rapid increase following a modest decline from 2019 to 2020. The NO_2 level fluctuated continuously, whereas SO_2 and CO levels were either stable or declining.
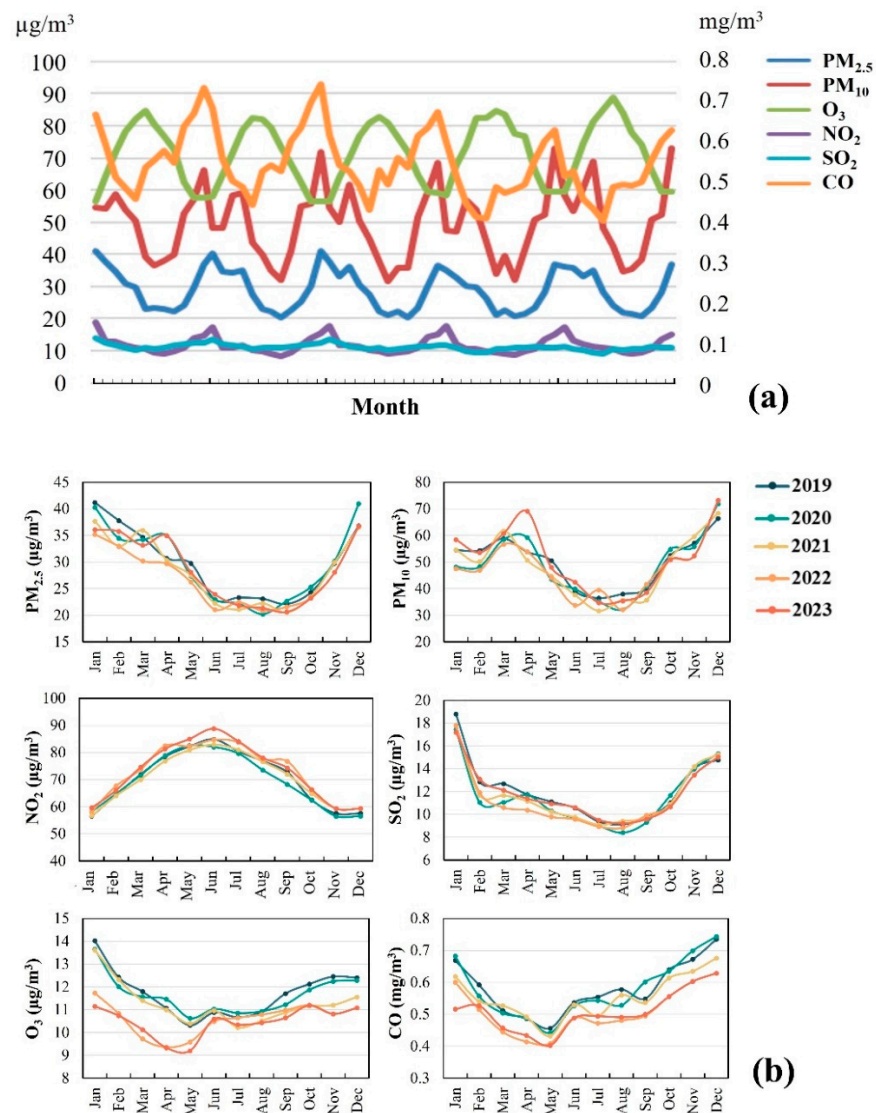
Figure 4. Monthly mean concentrations of the six major air pollutants (**a**) and their interannual differences (**b**) in China from 2019 to 2023. The unit is mg/m^3 for CO and µg/m^3 for other air pollutants.

Table 2. Annual mean concentrations of the six major air pollutants from 2019 to 2023. The unit is mg/m^3 for CO and µg/m^3 for other air pollutants.

Year	PM$_{2.5}$	PM$_{10}$	NO$_2$	SO$_2$	O$_3$	CO
2019	29.7 ± 11.4	50.1 ± 24.6	12.2 ± 5.6	11.7 ± 2.0	70.5 ± 10.6	0.58 ± 0.17
2020	29.7 ± 11.0	49.0 ± 23.8	11.6 ± 5.0	11.7 ± 1.9	69.6 ± 10.6	0.58 ± 0.15
2021	28.4 ± 9.4	48.5 ± 22.5	11.7 ± 5.1	11.3 ± 1.9	70.4 ± 11.2	0.55 ± 0.15
2022	27.4 ± 9.5	47.7 ± 26.3	11.4 ± 4.5	10.6 ± 1.9	72.7 ± 11.0	0.51 ± 0.13
2023	28.7 ± 9.4	51.5 ± 25.5	11.9 ± 4.7	10.5 ± 1.9	73.1 ± 10.9	0.51 ± 0.14

3.2. Spatiotemporal Patterns of Air Pollutants

Spatial clustering patterns of individual air pollutants in China from January 2019 to December 2023 were predominantly oscillating hot or cold spots, with some areas featuring persistent or intensifying spots and a few areas featuring diminishing spots (Figure 5). Among them, SO$_2$ and CO exhibited a statistically significant downward trend nationwide ($p < 0.05$ for SO$_2$, and $p < 0.001$ for CO), while the remaining four pollutants display increasing and/or decreasing trends only in specific areas.
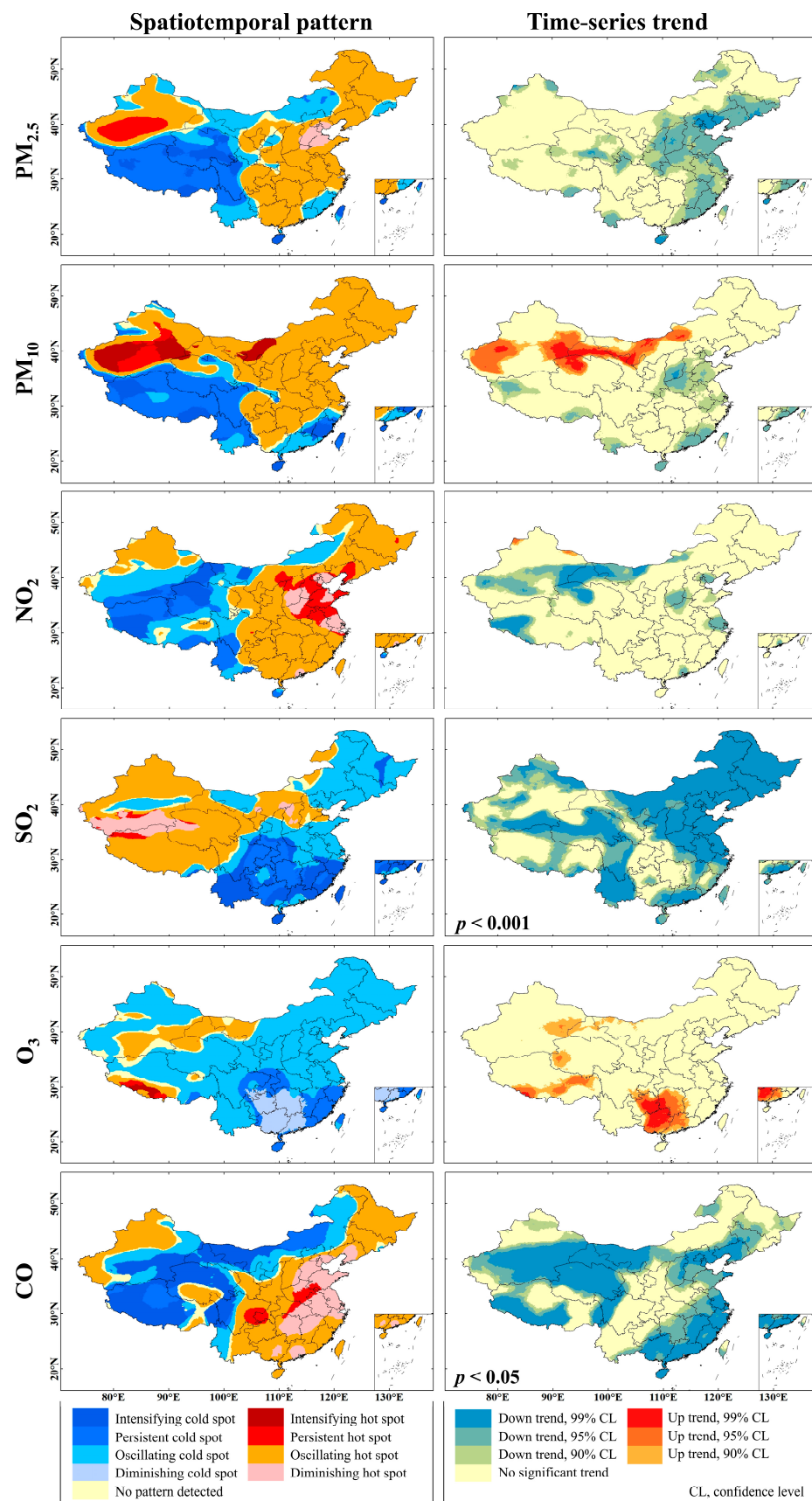
Figure 5. Spatial distribution patterns (**left**) and the corresponding temporal trends (**right**) of air pollutants in China from 2019 to 2023, with *p*-values of the significant trends marked.

While $PM_{2.5}$ and PM_{10} shared similar spatial patterns, with persistent and sometimes increasing cold spots in the western region (near the Tibetan Plateau) and persistent hot spots in the northwest, PM_{10} showed intensifying hot spots in the outer regions of the Taklimakan Desert and the North China Plain. This is likely attributed to the long-distance transport of sand and dust caused by the desert's monsoon climate. In contrast, $PM_{2.5}$ showed improvements with an oscillating cold spot in the North China Plain and a diminishing hot spot in the BTH region, indicating the effectiveness of control measures for air pollution. However, the temporal trends for $PM_{2.5}$ and PM_{10} diverged, with $PM_{2.5}$ trending downward, especially on the southeast coast, and PM_{10} trending upward in the north and northwest.

NO_2 and CO shared similar spatial patterns, with oscillating cold spots becoming more pronounced towards the northwest and persistent hot spots in the southeast. Some hot spots gradually vanished in the BTH region. In terms of temporal trends, NO_2 showed a downward trend in the northwest where the area of cold spots increased. CO exhibited a declining trend in most areas of the country ($p < 0.05$). SO_2 predominantly exhibited oscillating, persistent, and intensifying cold spots in the southeast, with oscillating, diminishing hot spots, and a few persistent hot spots in the northwest. Such spatial distribution, with increasing cold spots and disappearing hot spots, indicated a decrease in the concentrations of air pollutants, which is consistent with the significant downward trend nationwide ($p < 0.001$). The spatial distribution of O_3 was mainly oscillating cold spots, with a significant hot spot near the Himalayas on the southern edge of the Qinghai–Tibet Plateau, and a decreasingly persistent cold spot in southern China.

3.3. Model Performance

3.3.1. Feature Importance of Predictor Variables

All predictor variables contributed roughly equally to the six air pollutants (Figure 6). The majority of variables contributed more than 2%, with AAI being the most important feature (5.5%). Other key predictors were also vital, including TROPOMI NO_2, O_3, CO, and MAIAC AOD, with each accounting for around 4%. Meteorological factors exerted substantial influences and contributed over 40% in total. Among auxiliary data, population, nighttime light, and emissions inventory held relatively high importance, whereas the importance of LUC was minor, possibly because NDVI conveys part of land cover information already. Other generated features, denoting relative spatial positions, contributed about 18.5% to the total importance. Conversely, temporal information was less critical, potentially due to the limited values of months, making it less distinguishable when converted to polar coordinates. This phenomenon aligns with the highly consistent results between site-based and sample-based CV, whereas time-based CV yielded relatively inferior results.
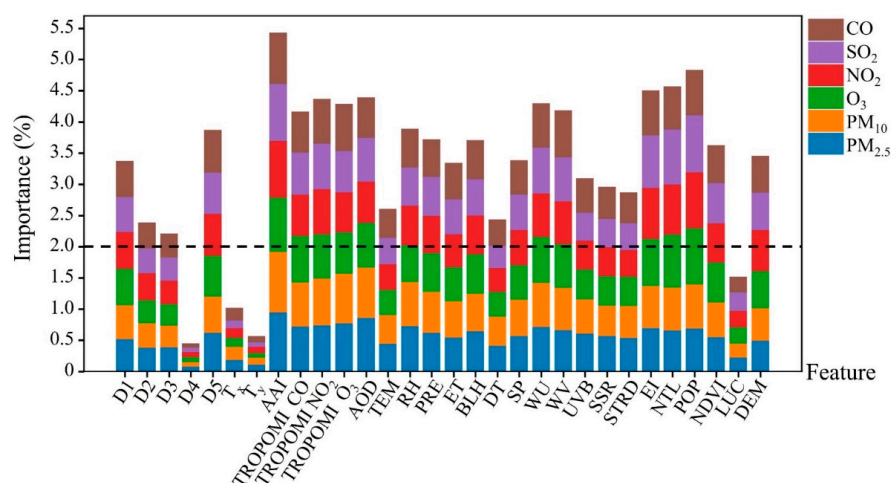

Figure 6. Importance (%) of each feature during model construction.

3.3.2. Predictive Accuracy of CV Results

In assessing model performance through three distinct 10-fold CV approaches, $PM_{2.5}$, PM_{10}, NO_2, and O_3 achieved a fairly good degree of fitting, and the results for SO_2 and CO were also commendable (Figure 7). Overall, our multi-output LightGBM demonstrated satisfactory results in training and fitting the concentrations of the six major air pollutants.

The R^2 from the sample-based CV was 0.92 for $PM_{2.5}$, 0.95 for PM_{10}, 0.95 for O_3, 0.90 for NO_2, 0.79 for SO_2, and 0.82 for CO. The RMSE was 6.1 $\mu g/m^3$ for $PM_{2.5}$, 9.0 $\mu g/m^3$ for PM_{10}, 3.8 $\mu g/m^3$ for NO_2, 2.7 $\mu g/m^3$ for SO_2, 5.3 $\mu g/m^3$ for O_3, and 0.11 mg/m^3 for CO (Table 3). These values suggest the model's ability to explain the variance in the observed data, reflecting its high predictive capability, reliability, and valuable contribution to the estimation of air pollutant concentrations.

Table 3. Accuracy of 10-fold cross-validation (CV) for different air pollutants across China. The units of the RMSE and MAE are mg/m^3 for CO and $\mu g/m^3$ for other air pollutants. CV, cross-validation; MAE, mean absolute error; RMSE, root-mean-square error.

|	Sample-Based CV			Site-Based CV			Time-Based CV		
	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE
$PM_{2.5}$	0.92	6.11	3.09	0.90	6.91	3.66	0.82	9.17	5.27
PM_{10}	0.95	9.01	5.54	0.91	11.52	6.61	0.77	18.69	10.53
NO_2	0.90	3.79	2.80	0.90	3.96	2.95	0.85	4.75	3.56
SO_2	0.79	2.67	1.65	0.77	2.83	1.78	0.70	3.23	2.01
O_3	0.95	5.27	3.84	0.93	5.85	4.35	0.89	7.34	5.65
CO	0.82	0.11	0.08	0.79	0.12	0.09	0.73	0.14	0.10

From the results of site-based CV, the predictive performance of the model for various pollutants remained consistent and robust spatially. The R^2, RMSE, and MAE values aligned closely with the results of the sample-based CV, showing good adaptability to the differences between monitoring stations and the overall accuracy of the model at the site level.

At the temporal level, the model maintains high predictive performance across various indicators over an extended period. Although the R^2 and RMSE indicators may be slightly lower than those of the other two CV methods, the predictive ability of the model remains comparatively robust at different years despite occasional deviations.
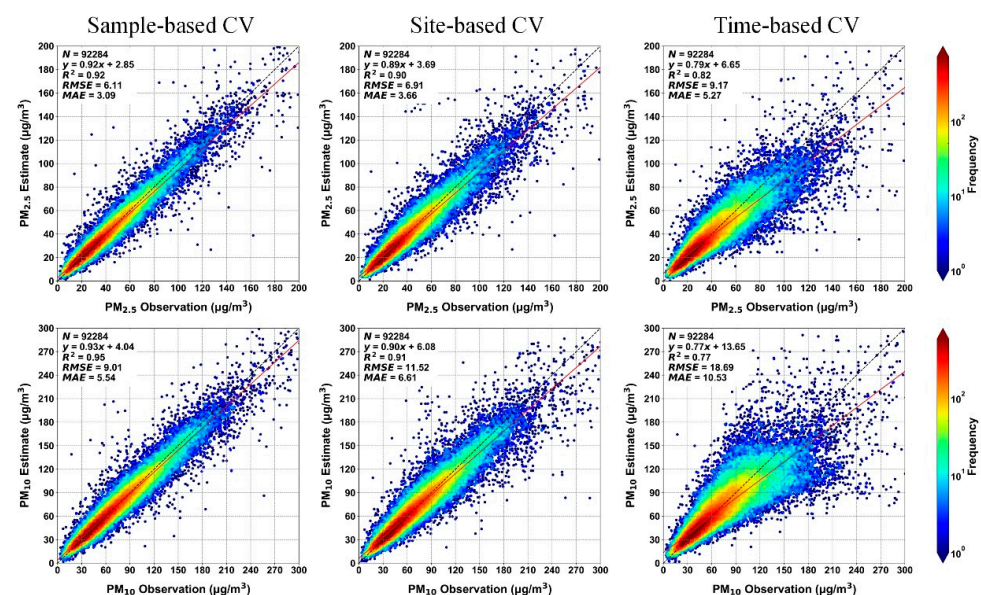

Figure 7. *Cont.*
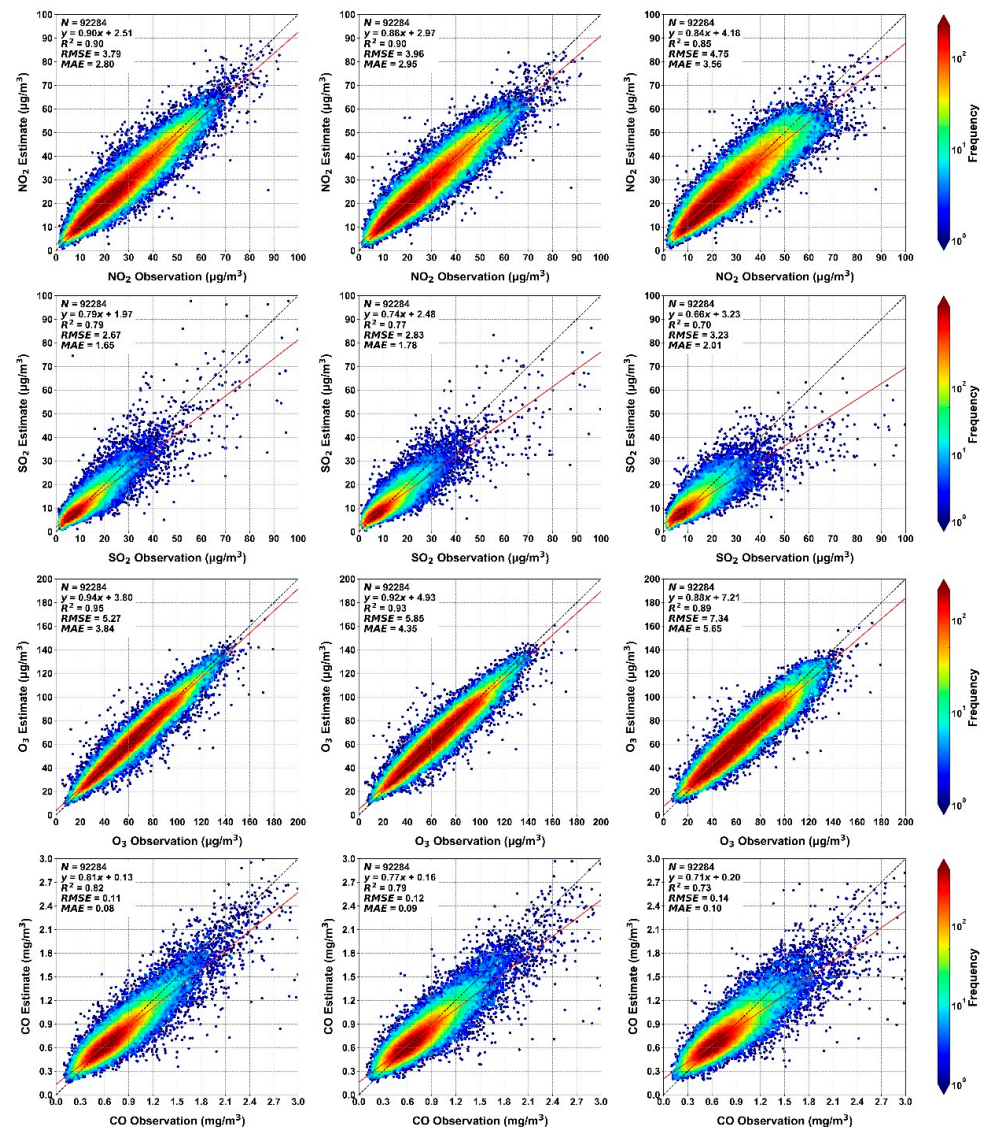
Figure 7. Density scatter plots of 10-fold cross-validation (CV) results of our multi-output LightGBM model. Solid lines denote the best-fit lines derived from linear regression, and dashed lines denote the 1:1 line. The provided information includes the sample size (N), coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE). The units of the RMSE and MAE are mg/m^3 for CO and $\mu g/m^3$ for other air pollutants.

3.3.3. Spatial and Temporal Robustness of the Results

The site-based CV accuracies of the model for the estimation of air pollutant concentrations at different locations of ground-based monitoring sites showed the robustness of the model across regions (Figure 8). Notably, stations in the southeast coastal region, including the BTH and YRD, demonstrated stable R^2 (around 0.90), coupled with commendable RMSE performance. In contrast, sites in the western and northern regions exhibited notably lower accuracy, reflecting the impact of sparse and uneven distributions of the monitoring network on model performance. For instance, in the Tibetan Plateau, the model predicted a modest R^2 value for particulate matters, possibly below 0.70. However, the inherently good air quality of the region could make the RMSE perform well, where small errors could lead to relatively large fluctuations. This pattern was supported by high concentrations of O_3 and CO with high precision in the Qinghai–Tibet Plateau. In the Taklimakan Desert region of northwest China, despite a substantial RMSE in $PM_{2.5}$ and PM_{10} estimates, R^2 remained relatively robust considering its extremely high concentrations of air pollutants.

The considerable disparities in the levels of air pollutants between these adjacent areas, influenced by topography and climate, explained the observed variations.
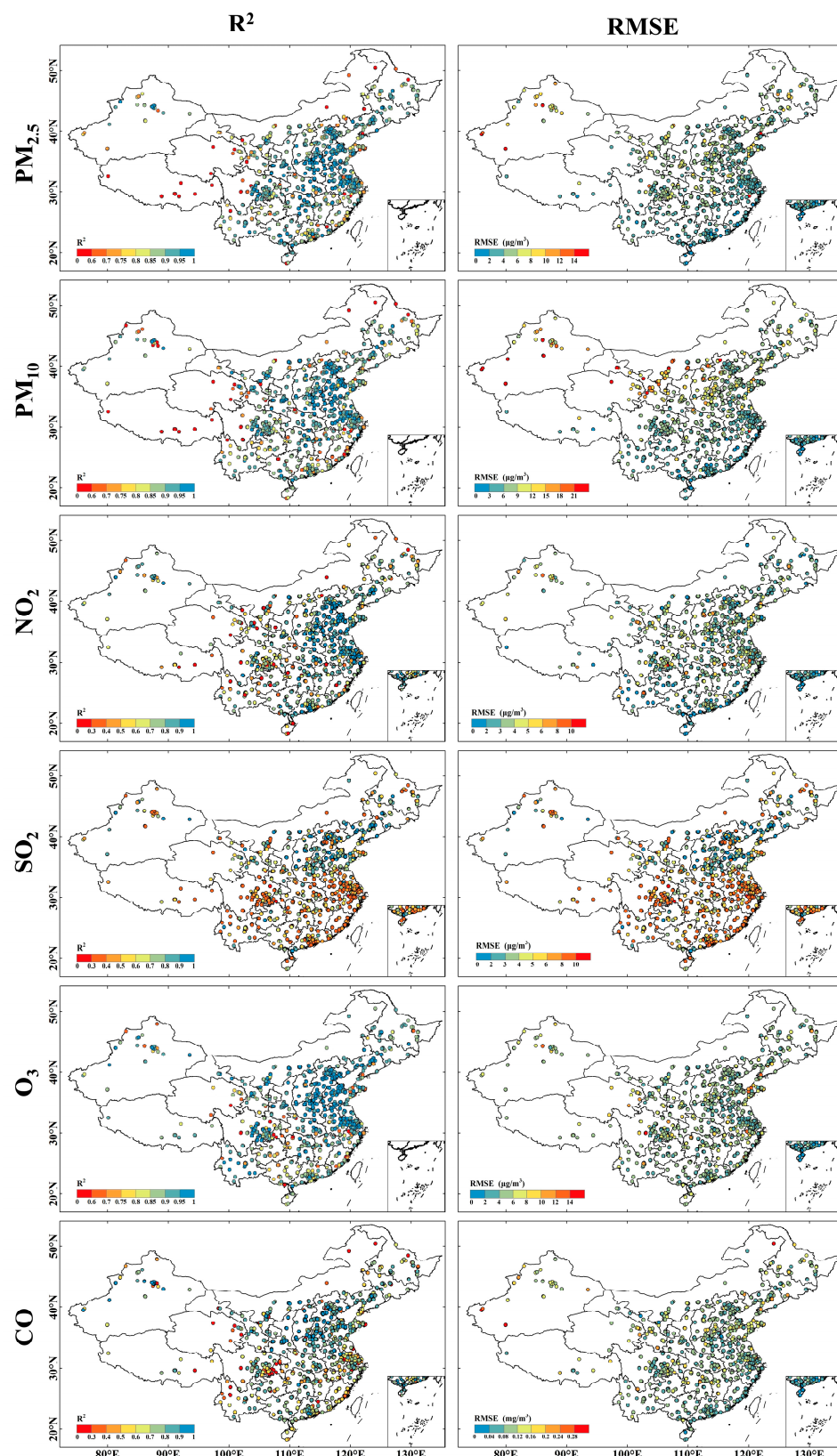
Figure 8. Spatial distributions of the site-based cross-validation results. RMSE, root-mean-square error. The units of the RMSE are mg/m^3 for CO and μg/m^3 for other air pollutants.

Additional insights into the model performance across different years were from the stability of the predictive capacity for the six air pollutants, indicated by R^2 and the RMSE, which remained relatively stable in each year, demonstrating the model's robust temporal flexibility (Figure 9). A slight decline in R^2 was observed from 0.84 in 2019 to 0.75 in 2023, while a continuous decrease in the RMSE from 3.19 to 2.03 $\mu g/m^3$ and a notable reduction in outliers and high-value points suggest a narrowing range due to a decline in the annual mean concentrations of SO_2. Notably, the model exhibited a specific instance of reduced accuracy in fitting $PM_{2.5}$ in 2021 ($R^2 = 0.81$, RMSE = 8.9 $\mu g/m^3$) compared to other years ($R^2 = 0.93$–0.96, RMSE = 4.0–5.2 $\mu g/m^3$). Such discrepancies may be attributed to a sudden decrease in the annual mean concentration of $PM_{2.5}$ for that year, indicating that the model is less resilient to such anomalous rapid changes.
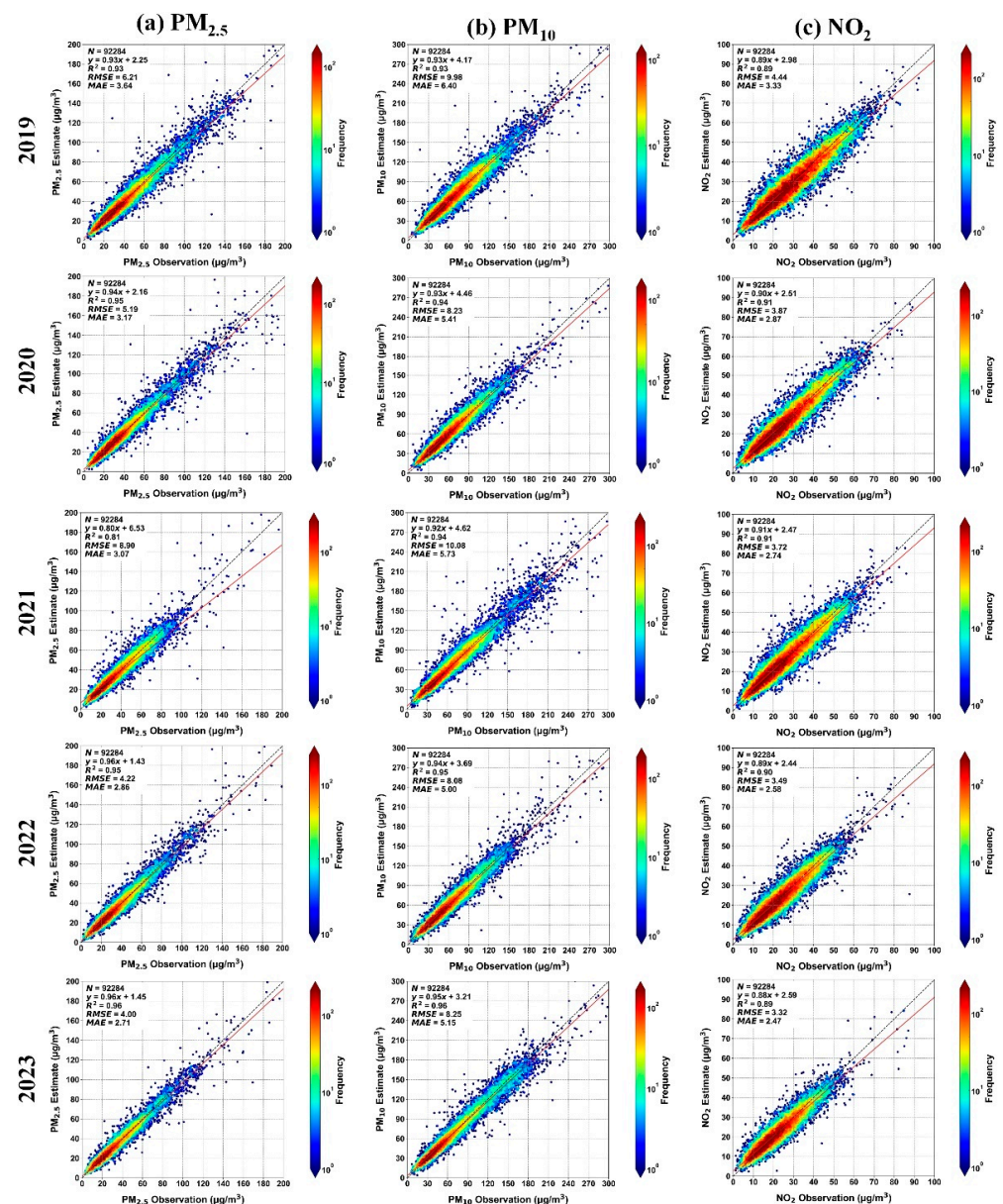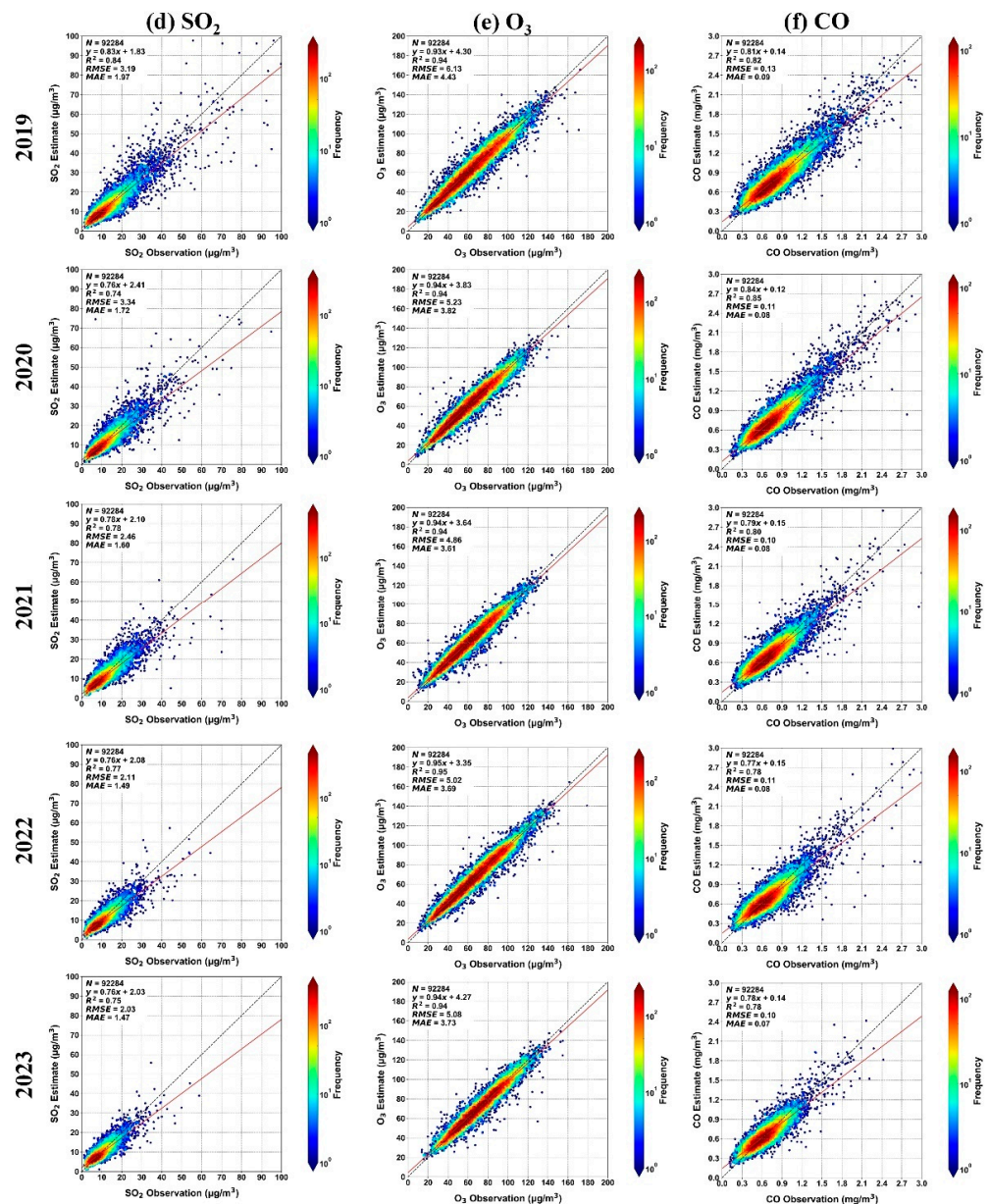

Figure 9. *Cont.*

Figure 9. Density scatter plots of yearly sample-based cross-validation (CV) results across China from 2019 to 2023. Solid lines denote the best-fit lines derived from linear regression, and dashed lines denote the 1:1 line. The provided information includes the sample size (N), coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE). The units of the RMSE and MAE are mg/m^3 for CO and μg/m^3 for other air pollutants. The pollutants from left to right are PM$_{2.5}$ (**a**), PM$_{10}$ (**b**), NO$_2$ (**c**), SO$_2$ (**d**), O$_3$ (**e**), and CO (**f**). Shown from top to bottom are the years 2019–2023 in order.

3.3.4. Comparisons of Multi-Output and Single-Output Models

The accuracy of multi-output and single-output LightGBM models was comparable (Table 4). However, in terms of efficiency, the multi-output LightGBM considerably outperformed the single-output LightGBM. Although LightGBM has been recognized for its efficient and fast computing, the total time spent on the single-output model, particularly in reading the predictors (satellite-derived data), was six times that of the multi-output model. Such discrepancy arose as the six single-output models redundantly read the same input data multiple times. Moreover, in practical applications, employing six single-output models entailed six rounds of optimal parameter searching and feature filtering, rendering

single-output models more time-consuming than multi-output models, especially in the context of 10-fold CV.

Table 4. Comparisons of the accuracy of 10-fold cross-validation (CV) for multi-output and single-output LightGBM.

		Sample-Based CV		Site-Based CV		Time-Based CV	
		R^2	RMSE	R^2	RMSE	R^2	RMSE
$PM_{2.5}$	Multi-output	0.92	6.11	0.90	6.91	0.82	9.17
	Single-output	0.92	6.05	0.90	6.85	0.77	10.24
PM_{10}	Multi-output	0.95	9.01	0.91	11.52	0.77	18.69
	Single-output	0.94	9.15	0.91	11.56	0.76	19.13
NO_2	Multi-output	0.90	3.79	0.90	3.96	0.85	4.75
	Single-output	0.90	3.94	0.89	3.97	0.85	4.80
SO_2	Multi-output	0.79	2.67	0.77	2.83	0.70	3.23
	Single-output	0.78	2.75	0.77	2.83	0.69	3.25
O_3	Multi-output	0.95	5.27	0.93	5.85	0.89	7.34
	Single-output	0.95	5.28	0.93	5.81	0.89	7.48
CO	Multi-output	0.82	0.11	0.79	0.12	0.73	0.14
	Single-output	0.81	0.11	0.79	0.12	0.74	0.14

4. Discussion

This study developed a multi-output LightGBM model to estimate the monthly mean concentrations of $PM_{2.5}$, PM_{10}, NO_2, SO_2, O_3, and CO simultaneously based on ground-based measurements of air pollutants, satellite-derived data, and other auxiliary data. During 2019–2023, the levels of SO_2 and CO decreased steadily, while $PM_{2.5}$, PM_{10}, NO_2, and O_3 experienced declines and then rebounds, with the annual mean concentrations of PM_{10} (mainly in the north and northwest) and O_3 (mainly in the southwest and south) in 2023 exceeding those in 2019. The levels of $PM_{2.5}$, NO_2, and CO in the BTH region significantly declined, with diminishing clusters of high concentrations and increasing clusters of low concentrations. This study advances environmental research and provides new approaches and important evidence for future air quality research and management.

In general, the 10-fold CV performance of our model demonstrates high accuracy and robustness for multiple air pollutants, offering reliable data support for spatiotemporal analysis of concentrations of air pollutants. Comparatively, the accuracy of the model is better for $PM_{2.5}$, PM_{10}, NO_2, and O_3 than SO_2 and CO, possibly due to their complex nature, lower concentrations, and smaller spatial variations. The fitted lines on the basis of the data points in 10-fold CV figures had a slope of <1 and a small positive intercept for all six air pollutants, which implies that the model tended to slightly underestimate the actual concentrations of air pollutants to a larger degree in regions with higher concentrations. Therefore, the predictive capability of the model in regions with a high level of air pollutants deserves more effort to improve [45,46].

Our study reveals a complex interplay between COVID-19 containment measures and air pollution levels. The initial declines in $PM_{2.5}$, PM_{10}, O_3, and NO_2 concentrations coincide with the onset of lockdown measures, possibly reflecting a direct impact of reduced economic activities and travel restrictions on air quality. The swift rebound of PM_{10} and O_3 levels after the lift of lockdown measures may be due to a rapid resumption of industrial activities and urban traffic [47]. The steady decreases in SO_2 and CO levels may possibly be due to the effectiveness of sustained control initiatives for air pollution that continued during the pandemic [48,49]. The significant improvement in air quality in the BTH region, with the decreased clusters of high concentrations and increased clusters of low concentrations for $PM_{2.5}$, NO_2, and CO, may suggest that pandemic control policies and shifts in industrial operations had a lasting positive effect on regional air quality. Some

nuanced spatial patterns, such as the increases of PM_{10} in the north and northwest and of O_3 in the southwest and south of China, raise questions about regional disparities in the environmental impact of the pandemic. These variations may reflect differences in local policy responses, the nature of industrial activities, and the persistence of pre-pandemic pollution sources [50].

It is crucial to consider the broader context of the pandemic. It acted as an unplanned experiment in global emission reductions, offering insights into the potential of concerted actions for air quality improvement. The clear seasonal variations in the levels of air pollutants, particularly the peaks in spring and winter, underscore the influences of both anthropogenic activities and natural events on air quality. Such understanding is vital for formulating strategies to maintain the improvement of air quality over the course of future public health emergencies [12].

This study has some limitations. First, spatiotemporal patterns observed in this study may vary by spatial and temporal scales at which the concentrations of air pollutants were modeled. Thus, modeling at different spatial scales may provide more comprehensive perspectives of examining spatiotemporal patterns of air pollutants [51,52]. Second, considering the precision of the sensors and relevant techniques, caution should be exercised in interpreting the findings, as they may be influenced by potential estimation biases arising from in situ sensors for air pollutant monitoring. Third, varying parameters selected during the modeling, such as spatial neighborhood size and time step in the space–time cube, might lead to variations in patterns of air pollutants. However, this is also considered an inherent challenge due to the modifiable areal unit problem [53–56]. Future studies could choose higher temporal resolutions (e.g., daily) to examine the impact of abrupt events on the concentrations of air pollutants to assess longer-term effects of different governance and emission control measures, conduct in-depth investigations on the associations between air pollutants and public health emergencies, and better understand the impact of human activities on air quality [57,58]. Last but not least, like many other previous studies, this study paid more attention to the accuracy rather than the degree of simplicity of the model. The following rule usually applies: the simpler the model, the better and the greater the universality of the model. Future efforts are warranted towards simplifying the model without compromising its accuracy.

5. Conclusions

This study developed a multi-output LightGBM model to estimate monthly mean concentrations of the six common air pollutants and examine their spatiotemporal patterns in China from 2019 to 2023. The findings indicate that our newly developed model could accurately estimate various air pollutants. Significant declines in SO_2 and CO were observed across the country, with the other four air pollutants showing increasing and/or decreasing trends in specific areas. Air pollutant levels in the BTH region were apparently mitigated, as shown by the diminishing clusters of high concentrations and increasing clusters of low concentrations. Our findings enhance the comprehension regarding spatiotemporal disparities in air pollutant changes and offer valuable insights into the dynamics and evolution of these variations over the course of public health emergencies.

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]
2. Zhou, X.; Zhang, X.; Wang, Y.; Chen, W.; Li, Q. Spatio-Temporal Variations and Socio-Economic Drivers of Air Pollution: Evidence from 332 Chinese Prefecture-Level Cities. *Atmos. Pollut. Res.* **2023**, *14*, 101782. [CrossRef]
3. Landrigan, P.J. Air Pollution and Health. *Lancet Public Health* **2017**, *2*, e4–e5. [CrossRef]
4. Brunekreef, B.; Holgate, S.T. Air Pollution and Health. *Lancet* **2002**, *360*, 1233–1242. [CrossRef]
5. Silva, R.A.; West, J.J.; Lamarque, J.-F.; Shindell, D.T.; Collins, W.J.; Faluvegi, G.; Folberth, G.A.; Horowitz, L.W.; Nagashima, T.; Naik, V. Future Global Mortality from Changes in Air Pollution Attributable to Climate Change. *Nat. Clim. Chang.* **2017**, *7*, 647–651. [CrossRef]
6. Cooper, M.J.; Martin, R.V.; Hammer, M.S.; Levelt, P.F.; Veefkind, P.; Lamsal, L.N.; Krotkov, N.A.; Brook, J.R.; McLinden, C.A. Global Fine-Scale Changes in Ambient NO_2 during COVID-19 Lockdowns. *Nature* **2022**, *601*, 380–387. [CrossRef]
7. Castells-Quintana, D.; Dienesch, E.; Krause, M. Air Pollution in an Urban World: A Global View on Density, Cities and Emissions. *Ecol. Econ.* **2021**, *189*, 107153. [CrossRef]
8. Sicard, P.; Agathokleous, E.; Anenberg, S.C.; De Marco, A.; Paoletti, E.; Calatayud, V. Trends in Urban Air Pollution over the Last Two Decades: A Global Perspective. *Sci. Total Environ.* **2023**, *858*, 160064. [CrossRef]
9. Rudke, A.P.; Martins, J.A.; Hallak, R.; Martins, L.D.; de Almeida, D.S.; Beal, A.; Freitas, E.D.; Andrade, M.F.; Koutrakis, P.; Albuquerque, T.T.A. Evaluating TROPOMI and MODIS Performance to Capture the Dynamic of Air Pollution in São Paulo State: A Case Study during the COVID-19 Outbreak. *Remote Sens. Environ.* **2023**, *289*, 113514. [CrossRef]
10. Venter, Z.S.; Aunan, K.; Chowdhury, S.; Lelieveld, J. COVID-19 Lockdowns Cause Global Air Pollution Declines. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 18984–18990. [CrossRef]
11. Fu, F.; Purvis-Roberts, K.L.; Williams, B. Impact of the COVID-19 Pandemic Lockdown on Air Pollution in 20 Major Cities around the World. *Atmosphere* **2020**, *11*, 1189. [CrossRef]
12. Huang, Y.; Yang, S.; Zou, Y.; Su, J.; Wu, C.; Zhong, B.; Jia, P. Spatiotemporal Epidemiology of COVID-19 from an Epidemic Course Perspective. *Geospat. Health* **2022**, *17*. [CrossRef]
13. Jia, P.; Yang, S. Are We Ready for a New Era of High-Impact and High-Frequency Epidemics? *Nature* **2020**, *580*, 321–322. [CrossRef]
14. Burki, T. Dynamic Zero COVID Policy in the Fight against COVID. *Lancet Respir. Med.* **2022**, *10*, e58–e59. [CrossRef]
15. Liu, Y.; Saltman, R.B. Policy Lessons from Early Reactions to the COVID-19 Virus in China. *Am. J. Public Health* **2020**, *110*, 1145–1148. [CrossRef]
16. Pei, Z.; Han, G.; Ma, X.; Su, H.; Gong, W. Response of Major Air Pollutants to COVID-19 Lockdowns in China. *Sci. Total Environ.* **2020**, *743*, 140879. [CrossRef]
17. Gao, C.; Zhang, F.; Fang, D.; Wang, Q.; Liu, M. Spatial Characteristics of Change Trends of Air Pollutants in Chinese Urban Areas during 2016–2020: The Impact of Air Pollution Controls and the COVID-19 Pandemic. *Atmos. Res.* **2023**, *283*, 106539. [CrossRef]
18. Zhang, Q.; Mao, X.; Wang, Z.; Tan, Y.; Zhang, Z.; Wu, Y.; Gao, Y. Impact of the Emergency Response to COVID-19 on Air Quality and Its Policy Implications: Evidence from 290 Cities in China. *Environ. Sci. Policy* **2023**, *145*, 50–59. [CrossRef]
19. Lv, Y.; Tian, H.; Luo, L.; Liu, S.; Bai, X.; Zhao, H.; Zhang, K.; Lin, S.; Zhao, S.; Guo, Z.; et al. Understanding and Revealing the Intrinsic Impacts of the COVID-19 Lockdown on Air Quality and Public Health in North China Using Machine Learning. *Sci. Total Environ.* **2023**, *857*, 159339. [CrossRef]
20. Ma, Q.; Wang, J.; Xiong, M.; Zhu, L. Air Quality Index (AQI) Did Not Improve during the COVID-19 Lockdown in Shanghai, China, in 2022, Based on Ground and TROPOMI Observations. *Remote Sens.* **2023**, *15*, 1295. [CrossRef]
21. Ministry of Ecology and Environment of the People's Republic of China. Ambien Air Quality Standards (GB 3095-2012). 2012. Available online: https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/dqhjbh/dqhjzlbz/201203/t20120302_224165.shtml (accessed on 1 April 2024). (In Chinese)
22. Wang, Z.; Tan, Y.; Guo, M.; Cheng, M.; Gu, Y.; Chen, S.; Wu, X.; Chai, F. Prospect of China's Ambient Air Quality Standards. *J. Environ. Sci.* **2023**, *123*, 255–269. [CrossRef]

23. Qin, W.; Fang, H.; Wang, L.; Wei, J.; Zhang, M.; Su, X.; Bilal, M.; Liang, X. MODIS High-Resolution MAIAC Aerosol Product: Global Validation and Analysis. *Atmos. Environ.* **2021**, *264*, 118684. [CrossRef]

24. Veefkind, J.P.; Aben, I.; McMullan, K.; Förster, H.; De Vries, J.; Otter, G.; Claas, J.; Eskes, H.J.; De Haan, J.F.; Kleipool, Q. TROPOMI on the ESA Sentinel-5 Precursor: A GMES Mission for Global Observations of the Atmospheric Composition for Climate, Air Quality and Ozone Layer Applications. *Remote Sens. Environ.* **2012**, *120*, 70–83. [CrossRef]

25. Wu, S.; Huang, B.; Wang, J.; He, L.; Wang, Z.; Yan, Z.; Lao, X.; Zhang, F.; Liu, R.; Du, Z. Spatiotemporal Mapping and Assessment of Daily Ground NO_2 Concentrations in China Using High-Resolution TROPOMI Retrievals. *Environ. Pollut.* **2021**, *273*, 116456. [CrossRef]

26. Kim, M.; Brunner, D.; Kuhlmann, G. Importance of Satellite Observations for High-Resolution Mapping of near-Surface NO_2 by Machine Learning. *Remote Sens. Environ.* **2021**, *264*, 112573. [CrossRef]

27. Goldberg, D.L.; Harkey, M.; de Foy, B.; Judd, L.; Johnson, J.; Yarwood, G.; Holloway, T. Evaluating NO_x Emissions and Their Effect on O_3 Production in Texas Using TROPOMI NO_2 and HCHO. *Atmos. Chem. Phys.* **2022**, *22*, 10875–10900. [CrossRef]

28. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [CrossRef]

29. Zuo, C.; Chen, J.; Zhang, Y.; Jiang, Y.; Liu, M.; Liu, H.; Zhao, W.; Yan, X. Evaluation of Four Meteorological Reanalysis Datasets for Satellite-Based $PM_{2.5}$ Retrieval over China. *Atmos. Environ.* **2023**, *305*, 119795. [CrossRef]

30. Mu, X.; Wang, S.; Jiang, P.; Wang, B.; Wu, Y.; Zhu, L. Full-Coverage Spatiotemporal Estimation of Surface Ozone over China Based on a High-Efficiency Deep Learning Model. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 103284. [CrossRef]

31. Jacquemin, B.; Lepeule, J.; Boudier, A.; Arnould, C.; Benmerad, M.; Chappaz, C.; Ferran, J.; Kauffmann, F.; Morelli, X.; Pin, I.; et al. Impact of Geocoding Methods on Associations between Long-Term Exposure to Urban Air Pollution and Lung Function. *Environ. Health Perspect.* **2013**, *121*, 1054–1060. [CrossRef]

32. Wei, J.; Li, Z.; Cribb, M.; Huang, W.; Xue, W.; Sun, L.; Guo, J.; Peng, Y.; Li, J.; Lyapustin, A.; et al. Improved 1 Km Resolution $PM_{2.5}$ Estimates across China Using Enhanced Space–Time Extremely Randomized Trees. *Atmos. Chem. Phys.* **2020**, *20*, 3273–3289. [CrossRef]

33. Yang, N.; Shi, H.; Tang, H.; Yang, X. Geographical and Temporal Encoding for Improving the Estimation of $PM_{2.5}$ Concentrations in China Using End-to-End Gradient Boosting. *Remote Sens. Environ.* **2022**, *269*, 112828. [CrossRef]

34. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.

35. Zhong, J.; Zhang, X.; Gui, K.; Wang, Y.; Che, H.; Shen, X.; Zhang, L.; Zhang, Y.; Sun, J.; Zhang, W. Robust Prediction of Hourly $PM_{2.5}$ from Meteorological Data Using LightGBM. *Natl. Sci. Rev.* **2021**, *8*, nwaa307. [CrossRef]

36. Ma, J.; Zhang, R.; Xu, J.; Yu, Z. MERRA-2 PM2. 5 Mass Concentration Reconstruction in China Mainland Based on LightGBM Machine Learning. *Sci. Total Environ.* **2022**, *827*, 154363. [CrossRef]

37. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

38. Yang, Q.; Kim, J.; Cho, Y.; Lee, W.-J.; Lee, D.-W.; Yuan, Q.; Wang, F.; Zhou, C.; Zhang, X.; Xiao, X.; et al. A Synchronized Estimation of Hourly Surface Concentrations of Six Criteria Air Pollutants with GEMS Data. *NPJ Clim. Atmos. Sci.* **2023**, *6*, 94. [CrossRef]

39. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. In Proceedings of the IJCAI, Montreal, QC, Canada, 20–25 August 1995; Volume 14, pp. 1137–1145.

40. Fushiki, T. Estimation of Prediction Error by Using K-Fold Cross-Validation. *Stat. Comput.* **2011**, *21*, 137–146. [CrossRef]

41. Esri. How Emerging Hot Spot Analysis Works. 2016. Available online: https://pro.arcgis.com/en/pro-app/3.1/tool-reference/space-time-pattern-mining/learnmoreemerging.htm (accessed on 1 April 2024).

42. Ord, J.K.; Getis, A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geogr. Anal.* **1995**, *27*, 286–306. [CrossRef]

43. Mann, H.B. Nonparametric Tests against Trend. *Econom. J. Econom. Soc.* **1945**, *13*, 245–259. [CrossRef]

44. Kendall, M.G. *Rank Correlation Methods*; Griffin: Watertown, WI, USA, 1948.

45. Thunis, P.; Clappier, A.; de Meij, A.; Pisoni, E.; Bessagnet, B.; Tarrason, L. Why Is the City's Responsibility for Its Air Pollution Often Underestimated? A Focus on $PM_{2.5}$. *Atmos. Chem. Phys.* **2021**, *21*, 18195–18212. [CrossRef]

46. Friberg, M.D.; Zhai, X.; Holmes, H.A.; Chang, H.H.; Strickland, M.J.; Sarnat, S.E.; Tolbert, P.E.; Russell, A.G.; Mulholland, J.A. Method for Fusing Observational Data and Chemical Transport Model Simulations to Estimate Spatiotemporally Resolved Ambient Air Pollution. *Environ. Sci. Technol.* **2016**, *50*, 3695–3705. [CrossRef]

47. Li, C.; Chen, Z.; Wang, X.; Wan, Y.; Zhao, Z. The Impact of COVID-19 on Economy, Air Pollution and Income: Evidence from China. *Stoch. Environ. Res. Risk Assess.* **2023**, *37*, 3343–3354. [CrossRef]

48. Silva, A.C.T.; Branco, P.T.; Ferrini Rodrigues, P.; Sousa, S.I. Sustainable Policies for Air Pollution Reduction after COVID-19 Pandemic: Lessons Learnt from the Impact of the Different Lockdown Periods on Air Quality. *Sustain. Dev.* **2023**, *31*, 959–975. [CrossRef]

49. Wang, R.; Peñuelas, J. Monitoring Compliance in Pandemic Management with Air Pollution Data: A Lesson from COVID-19. *Environ. Sci. Technol.* **2021**, *55*, 13571–13574. [CrossRef]

50. Dong, L.; Chen, B.; Huang, Y.; Song, Z.; Yang, T. Analysis on the Characteristics of Air Pollution in China during the COVID-19 Outbreak. *Atmosphere* **2021**, *12*, 205. [CrossRef]

51. Zhang, X.; Zhang, M.; Zhao, Z.; Huang, Z.; Deng, Q.; Li, Y.; Pan, A.; Li, C.; Chen, Z.; Zhou, M. Obesogenic Environmental Factors of Adult Obesity in China: A Nationally Representative Cross-Sectional Study. *Environ. Res. Lett.* **2020**, *15*, 044009. [CrossRef]

52. Yang, S.; Liang, X.; Dou, Q.; La, Y.; Cai, J.; Yang, J.; Laba, C.; Liu, Q.; Guo, B.; Yu, W. Ethnic Disparities in the Association between Ambient Air Pollution and Risk for Cardiometabolic Abnormalities in China. *Sci. Total Environ.* **2022**, *838*, 155940. [CrossRef]

53. Openshaw, S. The Modifiable Areal Unit Problem. In *Concepts and Techniques in Modern Geography*; Geo Books: Norwich, UK, 1984.

54. Jia, P.; Yu, C.; Remais, J.V.; Stein, A.; Liu, Y.; Brownson, R.C.; Lakerveld, J.; Wu, T.; Yang, L.; Smith, M. Spatial Lifecourse Epidemiology Reporting Standards (ISLE-ReSt) Statement. *Health Place* **2020**, *61*, 102243. [CrossRef]

55. Jia, P.; Stein, A. Using Remote Sensing Technology to Measure Environmental Determinants of Non-Communicable Diseases. *Int. J. Epidemiol.* **2017**, *46*, 1343–1344. [CrossRef] [PubMed]

56. Chen, X.; Ye, X.; Widener, M.J.; Delmelle, E.; Kwan, M.-P.; Shannon, J.; Racine, E.F.; Adams, A.; Liang, L.; Jia, P. A Systematic Review of the Modifiable Areal Unit Problem (MAUP) in Community Food Environmental Research. *Urban Inform.* **2022**, *1*, 22. [CrossRef]

57. Jia, P.; Stein, A.; James, P.; Brownson, R.C.; Wu, T.; Xiao, Q.; Wang, L.; Sabel, C.E.; Wang, Y. Earth Observation: Investigating Noncommunicable Diseases from Space. *Annu. Rev. Public Health* **2019**, *40*, 85–104. [CrossRef] [PubMed]

58. Jia, P.; Liu, S.; Yang, S. Innovations in Public Health Surveillance for Emerging Infections. *Annu. Rev. Public Health* **2023**, *44*, 55–74. [CrossRef] [PubMed]