*applied sciences*

MDPI

*Article*

# The Spatiotemporal Evolution Patterns and Spatial Differentiation Mechanisms of $PM_{2.5}$ Concentrations in East China Based on Multi-Source Fused Data

**Yuwei Lei [1,2], Tiange You [1,2,*], Jiangying Chen [1,2], Senyuan Lu [1,2] and Yihan Zhang [1,2]**

[1] College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China; 12411053023@fafu.edu.cn (Y.L.)
[2] Fujian Statistical Information Research Center, Fuzhou 350002, China
* Correspondence: 000q101026@fafu.edu.cn; Tel.: +86-133-2820-6536

**Abstract**

To reveal the spatiotemporal evolution of $PM_{2.5}$ concentrations in East China from 2000 to 2023 and quantify the multiscale driving effects of natural and socioeconomic factors, we integrated satellite-derived and ground-based monitoring data to construct a multi-source dataset. XGBoost-SHAP screened key factors; Theil–Sen trend, Hurst index, and Moran's I characterized spatiotemporal patterns; and the geographic detector and MGWR quantified driving mechanisms. $PM_{2.5}$ declined significantly (Sen slope: $-1.29$ $\mu g \cdot m^{-3} \cdot a^{-1}$ to $-0.03$ $\mu g \cdot m^{-3} \cdot a^{-1}$), with accelerated decrease after 2013. The Hurst index indicated sustainable improvement in the Yangtze River Delta core but reversal risk in southwestern Zhejiang and northern Fujian. Spatially, the north–south gradient intensified: high concentrations persisted in industrial regions (northern Jiangsu, northern Anhui, western Shandong), whereas low concentrations remained in ecological zones (southwestern Zhejiang, northern Fujian). Natural factors dominated spatial variation; temperature (q = 0.804) and precipitation (q = 0.724) showed the strongest explanatory power. MGWR further revealed the stronger negative effects of temperature and precipitation in the north than in the south, and a continuous spatial gradient of per capita GDP from coastal industrial clusters to inland ecological zones. These findings underscore the need for region-specific emission reduction strategies that account for climatic heterogeneity.

**Keywords:** $PM_{2.5}$ concentration; XGBoost-SHAP; geodetector; MGWR model; East China region; spatiotemporal heterogeneity

Check for updates

Academic Editors: Luciana Maria Baptista Ventura and Márcio de Almeida D'Agosto

Received: 5 June 2026
Revised: 6 July 2026
Accepted: 8 July 2026
Published: 13 July 2026

**Copyright:** © 2026 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.

## 1. Introduction

Air quality is globally recognized as a key factor for maintaining ecosystem balance and protecting public health. With the rapid advancement of industrialization and urbanization in China, complex air pollution problems—particularly those involving fine particulate matter ($PM_{2.5}$)—have become increasingly prominent, which poses a serious challenge to sustainable economic and social development [1]. Due to its physical and chemical properties—including small particle size, large specific surface area, ability to accumulate toxic and harmful substances, and capacity for long-range transport—$PM_{2.5}$ poses a direct and far-reaching threat to human health. It can penetrate deep into the alveoli via the respiratory tract and enter the bloodstream, triggering or exacerbating respiratory diseases, cardiovascular diseases, and lung cancer, thereby increasing morbidity and mortality rates and reducing life expectancy, imposing a has a key component of aerosols, it can

also affect a heavy disease burden on society [2,3]. At the same time, high concentrations of $PM_{2.5}$ can reduce visibility and compromise transportation safety; the energy balance of the Earth-atmosphere system through the scattering and absorption of solar radiation, thereby exerting complex feedback effects on regional and even global climate [4]. Consequently, the effective management of $PM_{2.5}$ has transcended the realm of a mere environmental issue, becoming a major public health and strategic concern that involves social well-being, economic costs, and national image. In the context of China's green development initiatives, accurately identifying the spatiotemporal patterns of $PM_{2.5}$ pollution and the underlying driving mechanisms is a prerequisite and key to implementing scientific [5], precise, and efficient coordinated governance. This is of critical theoretical and practical significance for formulating differentiated regional prevention and control strategies [6], evaluating the effectiveness of environmental policies, and ultimately achieving the long-term goals of the "Blue Sky Defense Campaign" [7].

Currently, scholars both domestically and internationally have conducted extensive research on the spatiotemporal heterogeneity of $PM_{2.5}$ concentrations and their driving mechanisms. Yang Qi et al. used the MGTWR model to reveal that $PM_{2.5}$ concentrations in the Yangtze River Economic Belt exhibited an inverted U-shaped decline from 2012 to 2022, with significant spatial variations; industrial structure, vegetation, and economic factors were identified as the primary influencing factors [8]. Chen Youliang et al. based on data from 2017 to 2022, found that $PM_{2.5}$ concentrations in the Yangtze River Delta region declined overall, exhibiting a spatial pattern of "higher in the west and lower in the east, higher in the north and lower in the south," with higher levels in winter and lower levels in summer. Public transportation and wind speed were the primary driving factors, while social factors showed significant interprovincial differences during the pandemic [9]. Wang Lili et al. using 2018 data, found that $PM_{2.5}$ concentrations in the Yangtze River Economic Belt exhibited low levels in summer and high levels in winter, with a spatially positive correlation and a spatial autocorrelation threshold of approximately 870 km. Altitude, topographic relief, and population density were key influencing factors, and the interaction among these factors significantly enhanced the explanatory power for pollution, with the interaction between industrial structure and altitude being particularly prominent [10]. X Zhang et al. found that $PM_{2.5}$ concentrations in North China increased between 2000 and 2017, with high levels concentrated in southeastern Hebei, northern Henan, and western Shandong. Urbanization and road network density exacerbated pollution, while the share of the tertiary sector had a mitigating effect [11].

Y Zhang et al. demonstrated that the drivers of $PM_{2.5}$ in China's coastal ports exhibit spatiotemporal heterogeneity; cargo throughput in ports around the Bohai Sea was positively correlated with $PM_{2.5}$, and this effect continued to strengthen in Beibu Gulf ports between 2013 and 2018 [12]. M. Hu et al. indicated that concentrations of most air pollutants in the Pearl River Delta region declined significantly between 2006 and 2019, with only ozone showing an upward trend. Spatially, concentrations exhibited a pattern of higher levels inland and lower levels along the coast. Pollutant concentrations were significantly correlated with meteorological factors and exhibited seasonal fluctuations, with winter presenting the most unfavorable dispersion conditions [13].

In terms of international research, studies on the driving mechanisms of $PM_{2.5}$ in developed regions such as Europe and North America began relatively early, providing an important reference for understanding the relative contributions of natural and socioeconomic factors. Multi-year observational data from the European Monitoring Program for Environmental Protection (EMEP) indicate that in industrial clusters in northwestern Europe, meteorological conditions—particularly the frequency of winter temperature inversions and boundary layer height—have a decisive influence on the seasonal fluctuations in

$PM_{2.5}$ concentrations. Although the explanatory power of temperature factors is lower than in the East Asian monsoon region, the precipitation-induced removal effect is more stable in an oceanic climate context [14]. Monitoring studies by the U.S. EPA based on the CASTNET network reveal that in the industrial belt of the U.S. Midwest, the spatial variation of $PM_{2.5}$ is strongly modulated by the secondary formation of sulfates and nitrates. The interaction between meteorological and socioeconomic factors exhibits seasonal characteristics distinct from those in East Asia: photochemical oxidation dominates secondary pollution formation in summer, while in winter, it is jointly driven by heating emissions and the cumulative effect of low temperatures [15]. A related study in India's Indo-Gangetic Plain region shows that $PM_{2.5}$ concentrations in this area are influenced by the combined effects of winter biomass burning, boundary layer height compression, and topographic blocking. The mechanism of temperature influence is primarily characterized by "emissions-meteorology" coupling, with winter $PM_{2.5}$ concentrations reaching more than 3.6 times those of summer, and anthropogenic emissions contributing 90–100% to winter $PM_{2.5}$ [16]. This represents a fundamental difference from the mechanism identified in this study for the East China region, which centers on dispersion conditions and chemical transformation.

From a global comparative perspective, an MGWR study based on global data from 2000 to 2019 systematically compared the $PM_{2.5}$ driving mechanisms in countries at different stages of development. It found that $PM_{2.5}$ concentrations in developed countries (the United States, the European Union, and Japan) are influenced more evenly by various factors, with relatively balanced regression coefficients, indicating that these countries have entered a mature governance stage characterized by "multi-factor synergistic regulation"; In contrast, $PM_{2.5}$ concentrations in developing countries (India, the Middle East, and Africa) are more significantly influenced by human activities and remain in the "emission-driven" stage [17]. The study also confirmed that temperature is the primary driver of $PM_{2.5}$ concentration changes worldwide, and that the negative impact of NDVI is universally observed across the globe.

As one of China's most developed regions, East China exhibits meteorological characteristics more similar to those of mature industrial regions in Europe and the United States; however, the spatial heterogeneity of socioeconomic factors remains significantly greater than in developed countries. Consequently, the region's air quality—particularly the issue of $PM_{2.5}$ pollution—is of significant research value. A systematic investigation of the spatiotemporal evolution patterns, formation mechanisms, and health and climate impacts of $PM_{2.5}$ in this region not only provides a scientific basis for the precise prevention and control of regional air pollution and the implementation of the "dual carbon" strategy but also holds exemplary significance for exploring green development pathways for high-density urban clusters. He Qianshan et al. based on MODIS AOD inversions, found that $PM_{2.5}$ concentrations in East China exhibit seasonal patterns of high levels in autumn and winter and low levels in spring and summer, as well as an interannual trend of first increasing and then decreasing [18]. These variations are modulated by seasonal differences in large-scale circulation factors such as the polar vortex, the subtropical high, and teleconnections, revealing the significant influence of the climatic background on regional pollution. However, most of the aforementioned studies employ global regression or models at a single spatial scale, making it difficult to distinguish differences in the spatial scales at which various factors operate. Furthermore, they have failed to systematically compare the explanatory power of natural versus socioeconomic factors. The integration of machine learning and spatial remote sensing remains an unexplored area in these regions.

Although previous studies have addressed air pollution in East China, there remain gaps in the analysis of the spatiotemporal evolution patterns and driving mechanisms of pollution, and these studies have not effectively integrated spatiotemporal heterogeneity

research with interpretable machine learning methods. The XGBoost algorithm, which incorporates a second-order Taylor expansion to optimize the loss function and combines regularization with weighted quantile techniques, has been widely applied in relevant research [19]. Meanwhile, the SHAP method, based on Shapley value theory from cooperative game theory, quantifies the marginal contribution of each feature to the model's output to construct an interpretability evaluation system that satisfies the axioms of additivity, consistency, and symmetry, thereby providing an effective solution to the "black box" problem in machine learning. Zheng Qiuping et al. revealed the meteorological driving mechanisms behind $PM_{2.5}$–$O_3$ co-pollution in Fujian Province [20]; Yao Xiong et al. used an XGBoost-SHAP model to quantify the driving effects on the evolution of ecological and environmental quality in the Min River Delta urban agglomeration [21]; Feng et al. used this model to analyze the environmental causes of spatial heterogeneity in land-use functions in arid regions [22]; Wu et al. expanded the methodological application of machine learning in predicting complex systems in Earth sciences [23]. Combining XGBoost with SHAP enables both high-precision prediction and in-depth analysis of feature mechanisms, forming a complete closed-loop of "precise prediction—interpretable analysis" that significantly enhances the theoretical value and application potential of the model in research on environmental issues such as $PM_{2.5}$. Regarding research methods for spatiotemporal heterogeneity, the academic community has gradually shifted from traditional statistical analysis to spatially explicit techniques, such as spatial autocorrelation analysis and geographically weighted regression. Chen Yimin et al. developed a geographically weighted random forest model, which effectively improved the spatial prediction accuracy of $PM_{2.5}$ in the Yangtze River Delta [24]; Zhang et al. quantified the multiscale effects of ecosystem service spillovers using a spatial optimization framework [25]; and Kan Z et al. empirically analyzed the scale effects and driving mechanisms of water pollutant emissions in the Yangtze River Delta urban agglomeration [26]. The Optimal Parameter Geodetector (OPGD), as an emerging spatial analysis tool, can effectively identify and quantify the explanatory power of different factors on the spatial variation of $PM_{2.5}$. Through interactive detection, it reveals synergistic or attenuating effects among multiple factors, thereby deepening our understanding of pollution formation mechanisms. Gao Feng et al. used this method to identify the dominant factors driving the spatiotemporal variability of $PM_{2.5}$ in Shanxi Province [27]; Dong Dong et al. revealed the mechanisms underlying the differentiation of traditional villages under the influence of spatial scale effects [28]; and Zhang Jiping et al. analyzed the spatial heterogeneity of key drivers of ecosystem services in the Chaobai River Basin [29]. However, the impact of different driving factors on habitat quality exhibits spatial heterogeneity, and multiscale geographically weighted regression models can overcome the limitations of traditional models by incorporating spatial location into the regression analysis.

This study utilized $PM_{2.5}$ data from East China for the period 2000–2023 and employed the XGBoost-SHAP method to quantify the effects of socioeconomic, meteorological, and topographic factors on $PM_{2.5}$. After screening key variables, the Geographic Detector and MGWR were applied to reveal regional differences and spatiotemporal heterogeneity in the driving factors. The study's innovation lies in the construction of a three-tiered framework—"feature selection—spatial detection—local modeling"—which reveals the unique driving mechanisms of $PM_{2.5}$ in East China and provides a Chinese case study for understanding the phased differences in $PM_{2.5}$ management across global industrial regions.

## 2. Materials and Methods

### 2.1. Overview of the Study Area

The East China region is located on the eastern coast of China and, according to national administrative divisions, encompasses Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, and Shandong provinces. While the region as a whole enjoys a high level of economic development, there is significant spatial heterogeneity in air quality. Among these, the Yangtze River Delta urban agglomeration (Jiangsu, Zhejiang, Shanghai, and Anhui)—one of the nation's most economically dynamic regions with the highest concentration of energy consumption—has faced prominent $PM_{2.5}$ pollution issues in recent years. Annual average concentrations generally exceed 50 $\mu g/m^3$ [30], with particularly severe fine particulate matter pollution in inland industrial clusters and urban core areas. Shandong Province, as a key component of the North China Plain, also experiences relatively high $PM_{2.5}$ concentrations due to its industrial structure and winter heating practices, with annual averages in most cities ranging between 45 and 60 $\mu g/m^3$ [31]. While pollution levels in Jiangxi Province are generally slightly lower, the combined effects of regional transport and local emissions result in annual concentrations in northern cities reaching 40–55 $\mu g/m^3$ [32]. In contrast, Fujian Province benefits from a terrain dominated by mountains and hills, the highest forest coverage rate in the nation, and the dispersal advantages of a subtropical maritime monsoon climate. Consequently, $PM_{2.5}$ concentrations are generally lower, with annual averages maintaining around 35 $\mu g/m^3$, and air quality consistently outperforms that of most provinces in East China [33,34].

The East China region exhibits significant multidimensional spatial variations: the northern part of the region has a temperate monsoon climate with high winter heating demand, while the southern part features a subtropical oceanic monsoon climate with excellent dispersion conditions; industrially, the core area of the Yangtze River Delta is dominated by high-tech industries, whereas western Shandong and northern Anhui still concentrate on heavy chemical industries. This continuous spatial variation within a single administrative region provides an ideal research environment, facilitating the standardization of control mechanisms while allowing for the separation of the independent and interactive effects of meteorological, economic, and industrial factors. It offers a research sample with comparative methodological advantages for multiscale spatial models such as MGWR, thereby aiding in the development of regionally differentiated collaborative governance strategies.

### 2.2. Data Sources and Preprocessing

This study uses a balanced panel of 77 prefecture-level cities in East China (2000–2023). Annual $PM_{2.5}$ concentrations were derived from a multi-source fusion: ground-based monitoring data (2013–2023) and satellite AOD (ACAG v5.GL.02) (2000–2012), calibrated via linear slope correction over the 2013–2015 overlap [35,36]. Raster data were zonal-averaged to city boundaries; yearbook indicators were used at the prefecture level. Explanatory variables covered meteorology, energy, transport, economic structure, vegetation, and social factors. Missing values were linearly interpolated, outliers were screened via the 3σ criterion, and all variables were Z-score standardized before estimation. Table 1 summarizes $PM_{2.5}$ concentrations by province. In selecting influencing factors, we comprehensively considered multidimensional elements such as meteorological, economic, industrial, transportation, vegetation, and social factors. Specifically, meteorological data were sourced from the https://www.ecmwf.int (ECMWF), vegetation cover data from the National Earth System Science Data Center (http://www.geodata.cn/), and industrial emissions data were compiled from the "Bulletins on Ecological and Environmental Conditions" published by each province. The remaining socioeconomic data were obtained from the "Statistical

Yearbooks" of each province for the period 2000–2023 and the "Urban Statistical Yearbooks." All explanatory variables were standardized to the same spatiotemporal scope to ensure data consistency for model construction.

**Table 1.** $PM_{2.5}$ Concentration Data for Selected Provinces

| Province/Municipality | Mean (μg/m$^3$) | Standard Deviation | Minimum | Median | Maximum | Coefficient of Variation (%) |
|---|---|---|---|---|---|---|
| Shanghai | 50.32 | 8.45 | 34.84 | 52.37 | 59.71 | 16.8 |
| Jiangsu | 52.86 | 12.34 | 23.10 | 53.20 | 84.41 | 23.3 |
| Zhejiang | 38.45 | 10.67 | 20.74 | 38.25 | 66.18 | 27.7 |
| Anhui | 49.28 | 13.56 | 21.38 | 48.56 | 90.07 | 27.5 |
| Fujian | 32.16 | 8.92 | 18.15 | 31.72 | 55.35 | 27.7 |
| Jiangxi | 42.83 | 10.45 | 22.15 | 42.66 | 68.93 | 24.4 |
| Shandong | 61.47 | 14.78 | 34.42 | 61.02 | 100.33 | 24.0 |
| Overall | 47.14 | 15.96 | 14.42 | 45.87 | 100.33 | 33.9 |

In selecting influencing factors, we adhered to the principles of mechanistic relevance, literature-based reference, and data availability and spatiotemporal consistency. Specifically, we prioritized variables that characterize the $PM_{2.5}$ formation process in terms of pollution emissions, secondary transformation, dispersion and deposition, and ecological buffering; we comprehensively referenced existing research findings on $PM_{2.5}$ driving mechanisms in the Yangtze River Delta, the Yangtze River Economic Belt, and eastern China; the variables must be continuously available and uniformly processed at the prefecture-level city scale for the period 2000–2023. Based on these criteria, an indicator system was ultimately constructed encompassing dimensions such as meteorology, economy, industry/energy, transportation, vegetation, and social scale. Specifically, meteorological data were sourced from the https://www.ecmwf.int (ECMWF), vegetation cover data from the National Earth System Science Data Center (http://www.geodata.cn/), and data on industrial emissions were compiled from the *Bulletins on the State of the Ecological Environment* published by each province. All other socioeconomic data were drawn from the *Statistical Yearbooks* of each province for the period 2000–2023 and the "Urban Statistical Yearbooks." All explanatory variables were standardized to the same spatiotemporal scope to ensure data consistency for model construction. See Table 2 for specific influencing factors.

**Table 2.** Table of influencing factors.

| Primary Indicator | Secondary Indicator | Unit |
|---|---|---|
| Meteorology (X1) | Annual wind speed (X11) | m/s |
| | Accumulated precipitation year by year (X12) | m |
| | Annual average temperature (X13) | °C |
| Economy (X2) | Per capita gross domestic product(X21) | yuan/person |
| | Regional GDP growth rate (X22) | / |
| | Proportion of the secondary industry (X23) | / |
| Industry (X3) | Industrial electricity (X31) | 10,000 kWh |
| | Total gas or natural gas supply (X32) | 10,000 m$^3$ |
| | Sewage treatment Volume (X33) | 10,000 m$^3$ |
| Transportation (X4) | Highway freight volume (X41) | 10,000 tons |
| | Highway passenger volume(X42) | 10,000 people |
| | Highway mileage (X43) | kilometers |
| Vegetation (X5) | NDVI (X51) | / |
| | Green coverage rate in built-up areas (X52) | / |
| Society (X6) | Year by year population (X61) | people |

### 2.3. Research Methods

This study established a three-tiered progressive analytical framework comprising "feature selection—spatial exploration—local modeling." The overall technical roadmap is illustrated in Figure 1. The functional roles and logical relationships among these methods are as follows:

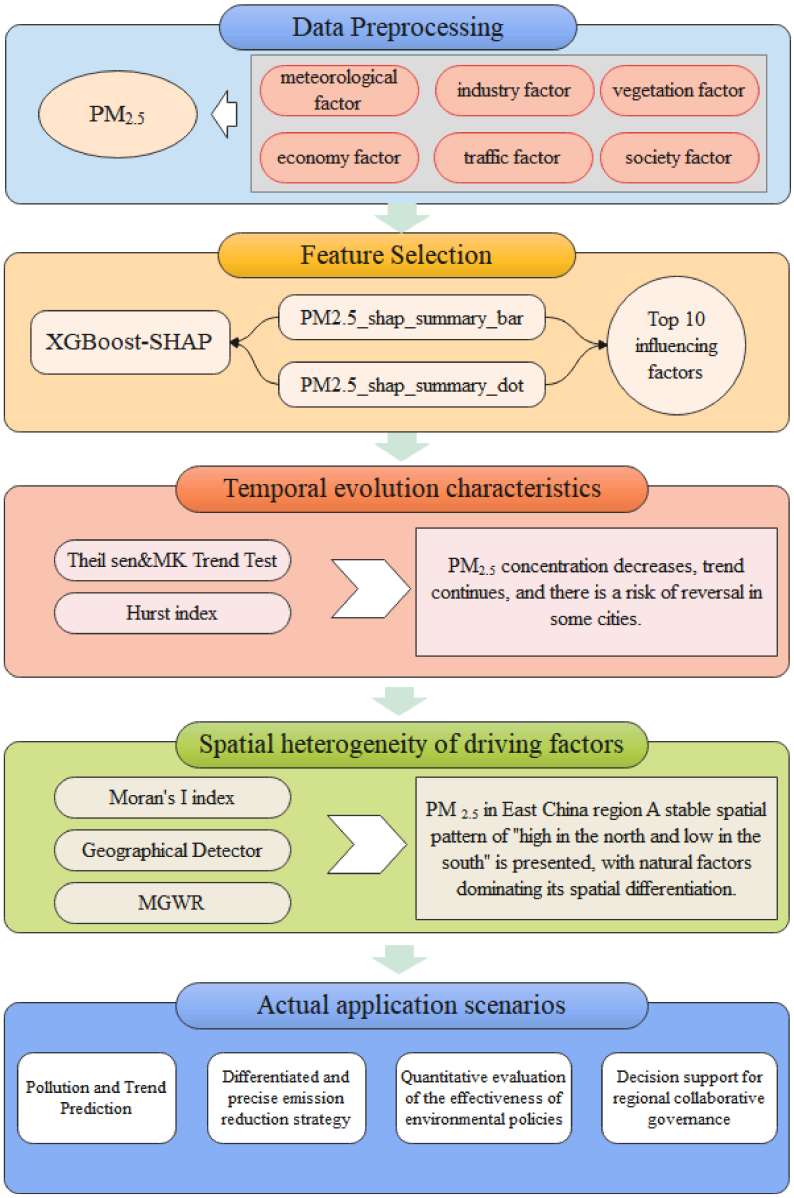



**Figure 1.** Technical flowchart.

First Layer: Feature Selection, which aims to address the question of "which factors are important." By leveraging XGBoost's nonlinear fitting capabilities to capture complex relationships among variables and quantifying the marginal contribution of each factor to $PM_{2.5}$ concentrations using SHAP values, this layer selects the top 10 key driving factors from 15 candidate variables, providing a variable foundation for subsequent spatial analysis. This step does not presuppose a specific functional form, thereby avoiding the reliance on linear assumptions inherent in traditional correlation analysis.

Second Layer: Spatial Detection, aimed at addressing the question of "what drives spatial heterogeneity." Building on the 10 factors selected by XGBoost-SHAP, the geographic detector uses the q-statistic to quantify each factor's explanatory power for $PM_{2.5}$ spatial heterogeneity and identifies interactive effects that either enhance or weaken the relationships between factors. The advantage of this method lies in its independence from linear assumptions and its ability to handle categorical variables, forming a complementary "nonlinear prediction—spatial heterogeneity detection" approach with XGBoost.

Third Layer: Heterogeneity analysis using OLS, GWR, and MGWR, aimed at addressing the question of "how influence mechanisms vary spatially." Geographic detectors reveal global explanatory power but cannot characterize the spatial non-stationarity of influence direction and intensity. Therefore, this study employs a progressive comparison of OLS (global baseline), GWR (single-scale local), and MGWR (multi-scale local) to gradually release spatial location parameters, ultimately using MGWR to reveal the spatial heterogeneity of the scales at which different factors operate.

ArcMap 10.8 (Esri, Redlands, CA, USA) was used for spatial data processing and basic spatial analysis. Machine learning modeling and statistical testing were performed in a Python 3.12.2 (Python Software Foundation, Wilmington, DE, USA) environment, using core libraries including XGBoost 2.1.3 (XGBoost Contributors, open-source community), SHAP 0.45.1 (Microsoft Research, Cambridge, MA, USA), and MGWR 2.2.1 (PySAL/Oshan et al., open-source community). GeoDetector analysis was performed using the GeoDetector Software in Excel (Wang Jinfeng, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China).

#### 2.3.1. XGBoost

XGBoost is an efficient machine learning algorithm based on the gradient boosting framework that achieves high-performance predictions by ensemble learning multiple weak learners [37,38]. The objective function of XGBoost consists of a loss function and a regularization term:

$$L(\theta) = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \quad (1)$$

In the formula, $y_i$ represents the true value of the $i$-th sample, and $\hat{y}_i$ represents the predicted value of the $i$-th sample. Assuming there are $k$ trees in the model, the expression for $\hat{y}_i$ is

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i) \quad (2)$$

Here, $k$ represents the number of weak learners, and each tree is generated from the results of the previous tree, that is

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3)$$

The objective function for XGBoost can be expressed as

$$L(\theta) = \sum_{i=1}^{n} [l(\hat{y}_i^{(t-1)}, y_i + f_t(x_i))] + \sum_{k=1}^{K} \Omega(f_k) + \text{const} \quad (4)$$

where const is a constant term; to optimize the objective function, we perform a second-order Taylor expansion of the objective function:

$$L(\theta) = \sum_{i=1}^{n}\left[l(\hat{y}_i^{(t-1)}, y_i) + g_i f_t(x_i) + \frac{1}{2}h_i f_t^2(x_i)\right] + \sum_{k=1}^{K}\Omega(f_k) + \text{const} \quad (5)$$

Here, $g_i$ is the first-order partial derivative of the loss function with respect to $x$ at point $i$, and $h_i$ is the second-order partial derivative of the loss function with respect to $x$ at point $i$. The value of each tree can be expressed as a mapping function:

$$f_k(x) = \omega_{q(x)}, \omega \in R^T, q : R^d \to \{1, 2, \ldots, K\} \quad (6)$$

Here, $\omega$ represents the leaf weights of the decision tree, and $q$ represents the mapping.

XGBoost adds two regularization terms to the error function of gradient-boosted decision trees to prevent overfitting. Let $T$ represent the number of leaf nodes and $\omega$ represent the score of the leaf nodes; the penalty term is expressed as

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T}\omega_j^2 \quad (7)$$

After factoring out the constant terms, optimizing the loss function terms, and expanding the regularization term, the objective function can be simplified to

$$L(\theta) = \sum_{j=1}^{T}\left[G_j\omega_j + \frac{1}{2}(H_j + \lambda)\omega_j^2\right] + \gamma T \quad (8)$$

Here, $G_j = \sum g_i$ represents the sum of the first-order partial derivatives of the samples contained in leaf node $j$, and $H_j = \sum h_i$ represents the sum of the second-order partial derivatives of the samples contained in leaf node $j$; both are constant real numbers.

The XGBoost model was trained using 5-fold cross-validation. Hyperparameters were determined via grid search, resulting in a learning rate of 0.05, a tree depth of 6, and a subsampling rate of 0.8. Fitting accuracy was evaluated using $R^2$, RMSE, MAE, and MAPE to assess the model's generalization ability.

#### 2.3.2. SHAP

SHAP is a game-theory-based model interpretation method used to quantify the contribution of each feature to the model's predictions [39–41]. Its core concept is derived from the Shapley value, which fairly allocates the influence of the prediction results to individual features.

SHAP interprets the model's output as the sum of the Shapley values for each feature, expressed as

$$\hat{y}_i = f_0 + \sum f_{i,j} = f_0 + f_{i,1} + f_{i,2} + \ldots + f_{i,m} \quad (9)$$

Here, $\hat{y}_i$ is the predicted value for the $i$-th sample, and $f_{i,m}$ is the Shapley value of the $m$-th feature for the $i$-th sample; that is, a Shapley value matrix containing $i \times m$ elements. The core idea behind Shapley values is to calculate the average marginal contribution of a feature to the model, expressed as

$$\phi_i(f, x) = \sum_{z' \subseteq x'}\frac{|z'|!(M-|z'|-1)!}{M!}[f_x(z') - f_x(z'\backslash\{\mathrm{i}\})] \quad (10)$$

When calculating the Shapley value $\phi_i(f, x)$ of feature $X$, $z'$ is a subset of all samples containing $x$, and $f_x(z') - f_x(z'/\mathrm{i})$ represents the marginal contribution under the subset $z'$ after adding feature $x$—that is, the difference between the mean of the output values after adding $x$ and the mean of the output values before adding $x$.

$$\omega(z') = \frac{|z'|!(M-|z'|-1)!}{M!} \tag{11}$$

The above expression represents the weighted marginal contribution, i.e., the contribution weights adjusted based on the probability of feature occurrence in the $z'$ subset.

2.3.3. Theil-Sen and Mann Kendall Trend Analysis

The Theil-Sen estimator and the Mann–Kendall trend test are robust, nonparametric trend analysis methods that are insensitive to outliers; they are widely used for trend detection in long-term time series in fields such as environmental science and hydrometeorology [42–44].

The Theil-Sen estimator estimates the magnitude of the trend by calculating the median of the slopes of the data at all adjacent time points; its results are not affected by outliers and are highly robust. The formula for the trend slope estimator $\beta$ is

$$\beta = \mathrm{median}\left(\frac{x_j - x_i}{j - i}\right) \forall i < j \tag{12}$$

Here, $x_i$ and $x_j$ represent the sequence values at time $i$ and $j$, respectively. The final trend slope $\beta$ is the median of all pairwise slopes. If $\beta > 0$, the sequence is in an upward trend; if $\beta < 0$, it is in a downward trend.

The Mann–Kendall test is a nonparametric test for trend significance that does not require the data to follow a specific distribution. The null hypothesis $H_0$ is that the series has no trend, and the alternative hypothesis $H_1$ is that the series exhibits a monotonic trend. At a given significance level $\alpha$, if $|Z| > Z_{1-\alpha/2}$, the null hypothesis is rejected, and it is concluded that the series exhibits a significant trend.

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}}, \text{if } S > 0 \\ 0, \text{if } S = 0 \\ \frac{S+1}{\sqrt{Var(S)}}, \text{if } S < 0 \end{cases} \tag{13}$$

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathrm{sgn}(x_j - x_i) \tag{14}$$

$$\mathrm{sgn}(x_j - x_i) = \begin{cases} 1, \text{if } x_j - x_i > 0 \\ 0, \text{if } x_j - x_i = 0 \\ -1, \text{if } x_j - x_i < 0 \end{cases} \tag{15}$$

For a sample size $n \geq 10$, the statistic $S$ approximately follows a normal distribution with the following variance, where $m$ is the number of groups with the same value, and $t_k$ is the width of the $k$-th node.

$$\mathrm{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{k=1}^{m} t_k(t_k-1)(2t_k+5)}{18} \tag{16}$$

#### 2.3.4. Hurst Index

The Hurst exponent is a nonparametric statistical method used to quantify the long-range dependence or persistence of a time series. The Hurst exponent typically ranges from 0 to 1. When H = 0.5, it indicates that the time series follows a random walk. Past increments are uncorrelated with future increments; the sequence has no memory and conforms to standard Brownian motion under the efficient market hypothesis. 0.5 < H < 1: indicates that the sequence exhibits persistence or trend amplification. That is, an upward trend in a past period tends to continue into the future, and the same applies to downward trends. The sequence exhibits long-term memory, and its fluctuations are smoother than those of a random sequence. 0 < H < 0.5: Indicates that the sequence exhibits anti-persistence or mean reversion. In contrast to persistence, upward trends from the past are more likely to be reversed by downward trends in subsequent periods, and vice versa; the sequence's fluctuations are more intense and frequent [45,46].

#### 2.3.5. Spatial Heterogeneity of $PM_{2.5}$

Moran's I is one of the core indicators used in spatial statistics to measure spatial autocorrelation [47,48]. Spatial autocorrelation describes the spatial distribution pattern of a geographic phenomenon or variable, specifically whether observations in one spatial unit are correlated with those in neighboring spatial units. This study calculated both global and local Moran's I to assess the spatial autocorrelation of $PM_{2.5}$ distribution in East China and to investigate its spatial clustering characteristics.

$$I_G = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{W}_{ij}(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{n} \sum_{j=1}^{n} \boldsymbol{W}_{ij} \sum_{i=1}^{n} (x_i - \overline{x})^2} \tag{17}$$

$$I_L = \frac{(x_i - \overline{x}) \sum_{j=1}^{n} \boldsymbol{W}_{ij}(x_j - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \tag{18}$$

In the equation, $I_G$ and $I_L$ represent the global and local Moran's indices, respectively; n is the total number of samples; $\boldsymbol{W}_{ij}$ is the spatial weight matrix; $x_i$ is the $PM_{2.5}$ value for the $i$-th region; and $\overline{x}$ is the average $PM_{2.5}$ value for the study area. The global Moran's index ranges from $-1$ to 1, and local spatial autocorrelation can be classified into five types: insignificant, high–high clustering, low–low clustering, high-low clustering, and low-high clustering.

#### 2.3.6. Geographic Detector

Geographic detectors are primarily used to reveal the spatial heterogeneity of geographic phenomena and to identify the underlying drivers [49–51]. The core principle is that if a certain independent variable ($X$) has a decisive influence on the spatial distribution of a dependent variable ($Y$), then the spatial distributions of the independent and dependent variables should exhibit similarity. This method does not rely on linear assumptions, imposes no strict requirements on data types, can effectively handle categorical variables, and detect interactions among factors; consequently, it is widely applied in driving force analysis across both natural and socio-economic domains.

The factor detector uses the q-statistic to measure the explanatory power of the independent variable $X$ on the dependent variable $Y$. The formula is as follows:

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} = 1 - \frac{SSW}{SST} \tag{19}$$

Here, $h$ denotes the stratum or category into which the independent variable $X$ has been stratified or classified; $N_h$ and $N$ represent the number of units in stratum $h$ and the total number of units in the entire population, respectively; $\sigma_h^2$ and $\sigma^2$ denote the variance of $Y$ values within stratum $h$ and the variance of $Y$ values across the entire population, respectively; and SSW and $SST$ represent the within-stratum variance and the total variance across the entire population, respectively.

The interaction detector identifies interactions by comparing the individual q-values of two independent variables, $X_1$ and $X_2$, with their combined q-value. The evaluation method is shown in Table 3 below:

**Table 3.** Interaction assessment methods.

| Criteria for Determining Interactions | Description of Interaction Type |
| --- | --- |
| $q(X_1 \cap X_2) < \min(q(X_1), q(X_2))$ | Nonlinear Attenuation |
| $\min(q(X_1), q(X_2)) < q(X_1 \cap X_2) < \max(q(X_1), q(X_2))$ | Single-Factor Nonlinear Attenuation |
| $q(X_1 \cap X_2) > \max(q(X_1), q(X_2))$ | Two-Factor Enhancement |
| $q(X_1 \cap X_2) = q(X_1) + q(X_2)$ | Independent |
| $q(X_1 \cap X_2) > q(X_1) + q(X_2)$ | Nonlinear Enhancement |

The q-statistic of the Geographical Detector is sensitive to discretization methods. In this study, Spearman's rank correlation test was used to assess ranking consistency, and the core factors were iterated over a range of 5–10 categories; a coefficient of variation (CV) for the q-value of less than 15% indicates that the results are robust.

#### 2.3.7. Ordinary Least Squares (OLS)

Ordinary Least Squares (OLS) [52,53] is the foundational and benchmark analytical method for spatial modeling. As a global regression model, both its constant term and the coefficients of the explanatory variables are fixed values, and it cannot account for spatial variations between regions. However, before constructing a Geographically Weighted Regression (GWR) model capable of capturing local effects, it is typically necessary to first use the OLS method to perform effective variable diagnostics and selection. Its basic form is as follows:

$$Y_i = \beta_0 + \beta_i X + \varepsilon_i \tag{20}$$

In this equation, $\beta_0$ is the constant term, $\beta_i$ is the regression coefficient, and $\varepsilon_i$ is the random error term.

#### 2.3.8. Multiscale Geographically Weighted Regression Model (MGWR)

Geographically weighted regression (GWR) [54,55] incorporates spatial location parameters into the classical linear regression framework; by performing differentiated fitting of variables within a local area, it effectively mitigates model bias caused by spatial non-stationarity. Multiscale Geographically Weighted Regression (MGWR) [56], as a significant extension of GWR, achieves key improvements in the bandwidth optimization mechanism. Bandwidth selection employs the AICc minimization criterion, and the kernel function is an adaptive Gaussian kernel. It allows different explanatory variables to have distinct

scales of influence, with the optimal bandwidth for each variable determined via the golden section search algorithm. This approach more accurately reveals the spatial heterogeneity in multifactorial driving processes, significantly enhancing the model's fitting accuracy and explanatory power. The basic mathematical expression of the MGWR model is as follows:

$$Y_i = \sum_{j=1}^{k} \beta_{bwj}(u_i, v_i) x_{ij} + \beta_0(u_i, v_i) + \varepsilon_i \quad (21)$$

Here, $Y_i$ represents the dependent variable $PM_{2.5}$, $x_{ij}$ represents the independent variables, $k$ represents the total number of spatial units included in the analysis, $(u_i, v_i)$ represents the spatial coordinates of the sample points, $\beta_0(u_i, v_i)$ represents the intercept at location $i$, and $\beta_{bwj}(u_i, v_i)$ represents the local regression coefficients for the $j$th variable at location $i$. A larger regression coefficient indicates a stronger effect of that variable on $PM_{2.5}$.

## 3. Results and Analysis

### *3.1. Multi-Source Data Fusion Validation*

To ensure the temporal continuity of multi-source data fusion and reduce systematic biases resulting from the integration of different data sources, this study selected the period from 2013 to 2015 as the overlap period between satellite inversion and ground-based monitoring. Using the annual average $PM_{2.5}$ concentration from ground-based monitoring as a reference, systematic validation was conducted at three levels, and linear regression correction was applied to the ACAG inversion data. The effectiveness of the correction was evaluated using the coefficient of determination ($R^2$), root mean square error (RMSE), mean absolute error (MAE), and mean bias (MBE). The results for the mean values are shown in Figure 2. The annual average $PM_{2.5}$ concentration across the entire region did not exhibit any abnormal jumps during the 2012–2013 transition period, and the magnitude of variation remained within the range of intra-satellite interannual fluctuations;

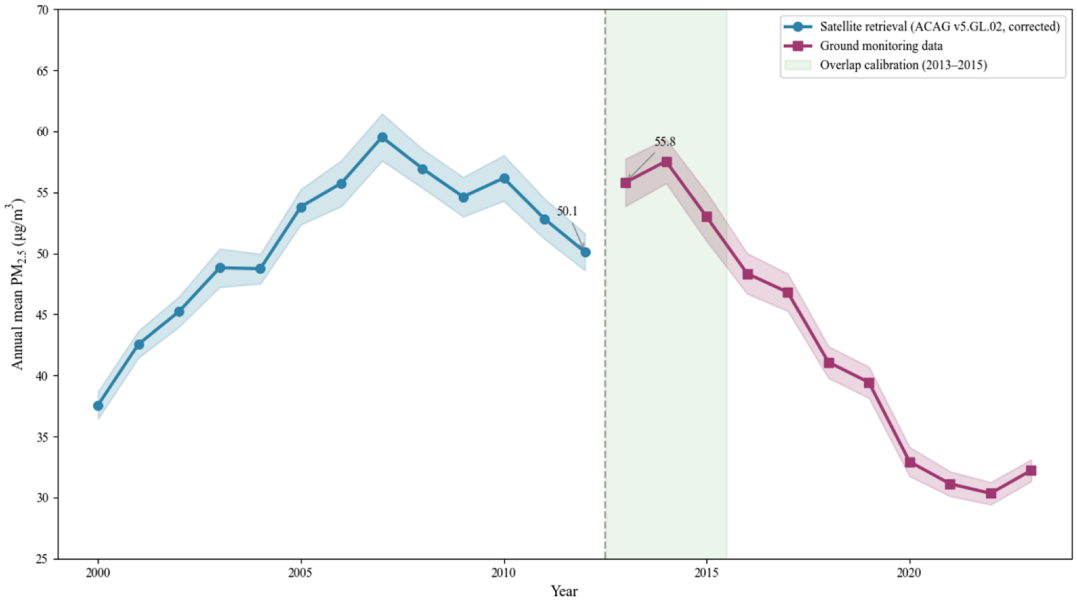



**Figure 2.** Time series of annual average $PM_{2.5}$ concentrations in 77 prefecture-level cities in East China.

The calibration accuracy results are shown in Figure 3. The linear calibration model for the overlap period is $y = 2.5 + 0.92x$, with an $R^2$ of 0.87, a slope close to 1, and a small intercept. The accuracy is comparable to that of similar products such as ACAG v5.GL.02 and the ChinaHighPM$_{2.5}$ dataset; This level of accuracy is consistent with validation results for similar satellite-derived PM$_{2.5}$ products: van Donkelaar et al. [57] validated the ACAG v5.GL.02 dataset using a global ground-based observation network, with the annual average concentration bias controlled within ±15%; Xiao et al. [58] used an ensemble machine learning method to estimate PM$_{2.5}$ in China, achieving a cross-validation $R^2$ of 0.79 and an RMSE of 21 μg/m$^3$; Ma, Z et al. [59] estimated PM2.5 in China using a GWR model, achieving a cross-validation $R^2$ of 0.64 and an RMSE of 32.98 μg/m$^3$. In this study, both the corrected $R^2$(0.87) and RMSE (5 μg/m$^3$) outperformed those of the aforementioned studies, indicating that the linear correction method achieves a better fit during the overlap period.

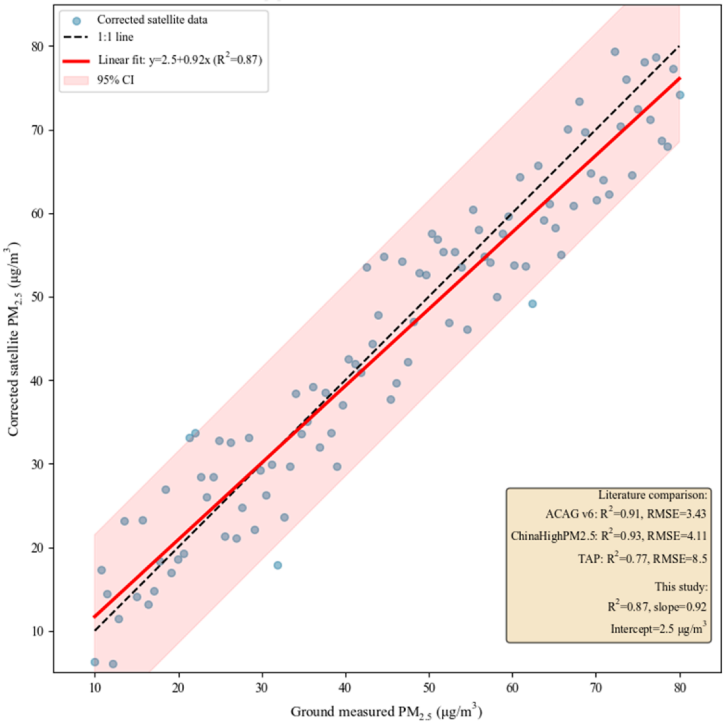



**Figure 3.** Validation of calibration results for the overlap period (2013–2015).

The distribution results are shown in Figures 4 and 5. The kernel density estimation curves for the three time periods exhibit a smooth transition at the boundaries. Neither the KS test ($p$ = 0.246) nor the Mann–Whitney U test ($p$ = 0.159) could reject the null hypothesis that the distributions before and after the transition period are consistent.

The above validation and Table 4 show that, prior to correction, $R^2$ was 0.78 and RMSE was 6.8 μg/m$^3$; after correction, $R^2$ increased to 0.87, RMSE decreased to 5 μg/m$^3$, and MAE was 3.9 μg/m$^3$, and the MBE was −1.6 μg/m$^3$. The ACAG satellite data, after linear correction, are statistically compatible with ground-based monitoring data, and the fused time series for 2000–2023 can be used for subsequent long-term trend analysis. This validation process drew upon the satellite PM$_{2.5}$ validation framework proposed by Li et al. [60], which employs a multidimensional evaluation system combining sample cross-validation, station cross-validation, and temporal cross-validation to ensure the reliability of the inverted data across both spatial and temporal dimensions.

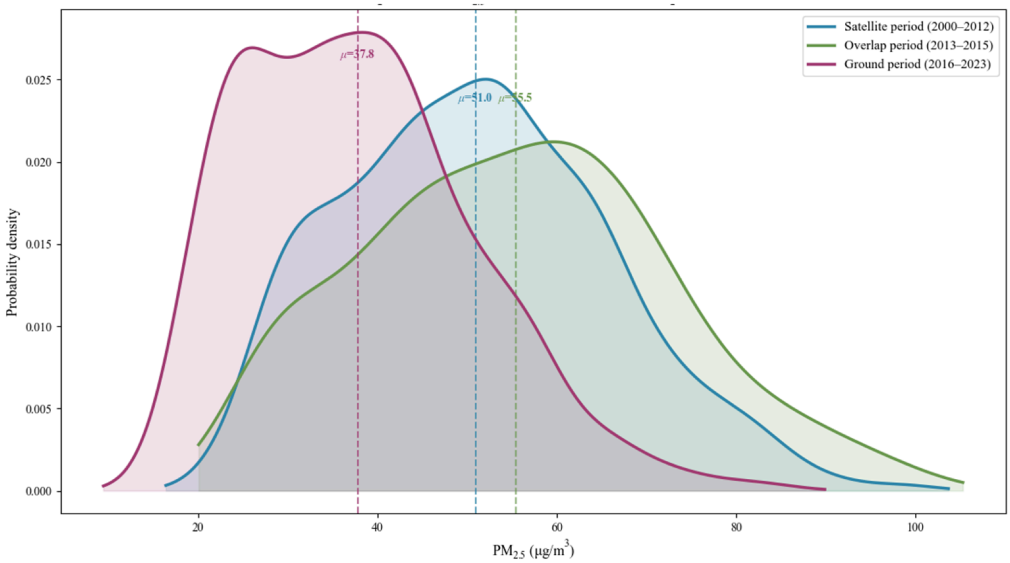



**Figure 4.** Three-time-period nuclear density estimation curve.

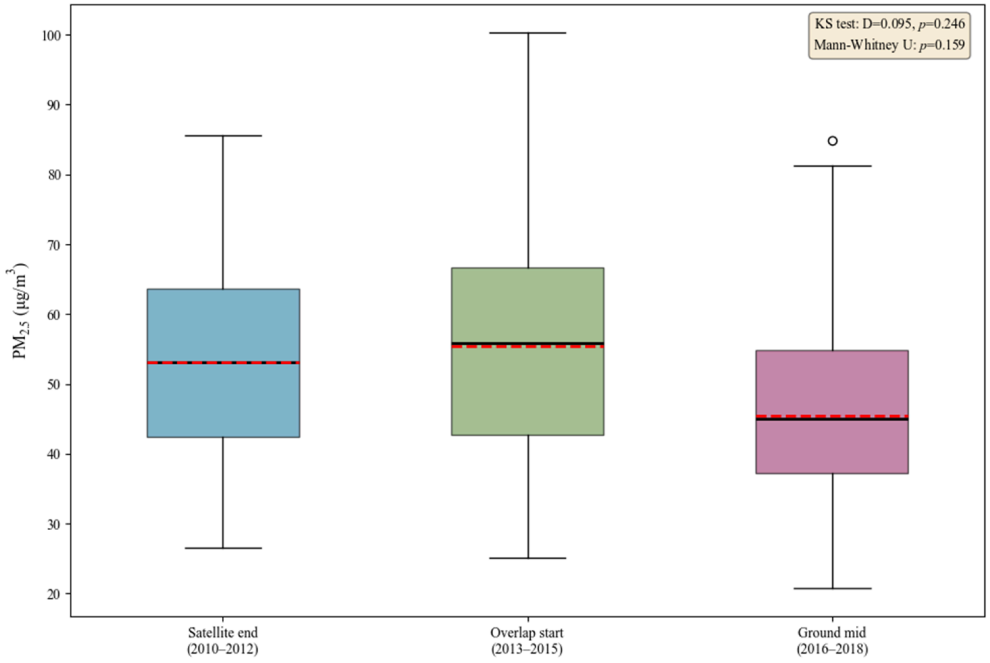



**Figure 5.** The KS test and the Mann–Whitney U test.

**Table 4.** Results before and after correction.

| Period/Indicator | $R^2$ | RMSE | MAE | MBE |
|---|---|---|---|---|
| Before Calibration | 0.78 | 6.8 | 5.4 | −5.2 |
| After Calibration | 0.87 | 5 | 3.9 | −1.6 |

### *3.2. Analysis of XGBoost-SHAP Results*

Based on multi-source data from 2000 to 2023, this study used the XGBoost machine learning model to model the factors influencing $PM_{2.5}$ concentrations and conducted an interpretability analysis of the model results using the SHAP framework. qThe results of the 5-fold cross-validation for XGBoost were as follows: $R^2$ = 0.89, RMSE = 4.82 μg/m$^3$, MAE = 3.56 μg/m$^3$, and MAPE = 8.7%, indicating that the model did not overfit and that the predictions were stable. Figure 6 presents a bar chart summarizing the SHAP

values for each influencing factor, reflecting the average absolute contribution of each variable to the model output. To strike a balance between model interpretability and complexity, this study employed the elbow analysis method to identify inflection points in the ranking of feature importance. When the number of factors increased from 9 to 10, the cumulative contribution rate rose from 91.2% to 94.7%, an increase of 3.5 percentage points; while increasing from 10 to 11 resulted in an increase of only 0.9 percentage points. Therefore, the top 10 factors were selected as variables for subsequent analysis, specifically: annual average temperature ($X_{13}$), regional GDP growth rate ($X_{22}$), share of the secondary industry ($X_{23}$), per capita regional GDP ($X_{21}$), total road mileage ($X_{43}$), annual cumulative precipitation ($X_{12}$), road passenger traffic ($X_{42}$), annual population ($X_{61}$), NDVI ($X_{51}$), and total coal/natural gas supply ($X_{32}$).

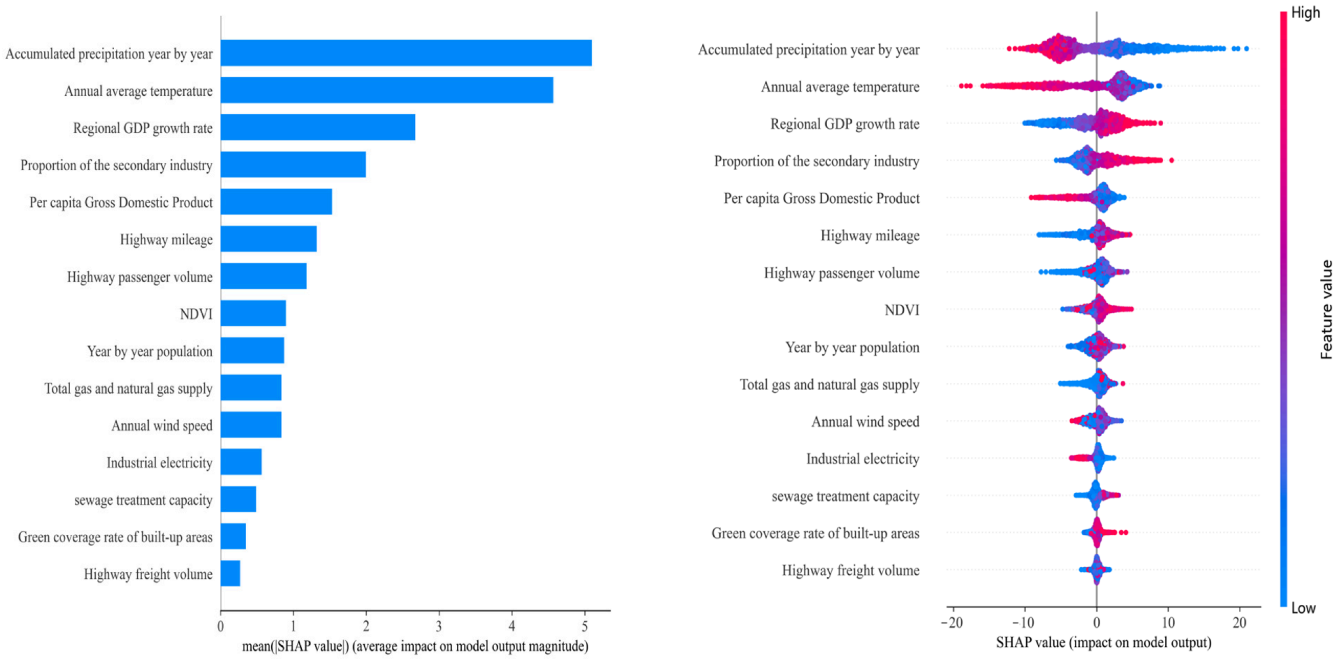



**Figure 6.** Summary plot of SHAP values for $PM_{2.5}$ influencing factors: (**a**) summary bar chart of SHAP values, and (**b**) SHAP value distribution plot.

Further analysis of the SHAP value summary map revealed that the effects of different factors on $PM_{2.5}$ concentrations exhibit significant directionality and heterogeneity. Natural factors such as precipitation, NDVI, and air temperature generally exerted a negative influence, indicating that increases in these factors help suppress $PM_{2.5}$ concentrations; In contrast, socioeconomic factors—such as the share of the secondary industry, total coal and natural gas supply, total road mileage, and road passenger traffic—generally exhibited a positive driving effect, reflecting the exacerbating role of energy consumption on pollution. The direction of influence differed markedly between per capita regional GDP and regional GDP growth rate, suggesting that the relationship between economic development and air quality might be nonlinear or exhibit regional heterogeneity.

### *3.3. Analysis of Temporal Variations in $PM_{2.5}$ Concentrations in East China*

#### 3.3.1. Theil-Sen and Mann Kendall Trend Analysis

Figure 7 shows the results of the Theil-Sen trend analysis and the Mann–Kendall test for $PM_{2.5}$ concentrations in the East China region. The region as a whole exhibited a significant trend toward improved air quality during the observation period, with all pixel Sen slopes

being negative, ranging from $-1.29\ \mu g \cdot m^{-3} \cdot a^{-1}$ to $-0.03\ \mu g \cdot m^{-3} \cdot a^{-1}$, confirming the widespread decline in $PM_{2.5}$ concentrations. This improvement process exhibited marked spatial heterogeneity, which is reflected not only in variations in the gradient of the decline rate but also in spatial differences in statistical significance, revealing underlying disparities among different sub-regions in terms of industrial structure, emission reduction efforts, and environmental carrying capacity. Specifically, in the core urban clusters of the Yangtze River Delta—such as Shanghai, southern Jiangsu, and northern Zhejiang—the Sen slopes ranged from $-1.29\ \mu g \cdot m^{-3} \cdot a^{-1}$ to $-1.08\ \mu g \cdot m^{-3} \cdot a^{-1}$, indicating the most rapid decline in concentrations. These trends all passed significance tests, highlighting the combined effectiveness of high-intensity industrial transformation and strict end-of-pipe treatment; In contrast, in traditional industrial and energy bases such as northern Jiangsu, northern Anhui, and central-Western Shandong, the Sen slopes range from $-0.81\ \mu g \cdot m^{-3} \cdot a^{-1}$ to $-0.57\ \mu g \cdot m^{-3} \cdot a^{-1}$. Although moderate improvement is evident, the rate of decline is relatively slower, indicating the lag in industrial restructuring and the persistent impact of complex local emission sources; meanwhile, in peripheral regions such as the southeastern coastal areas of Shandong and the southwestern mountainous areas of Zhejiang, the downward trend is the weakest.

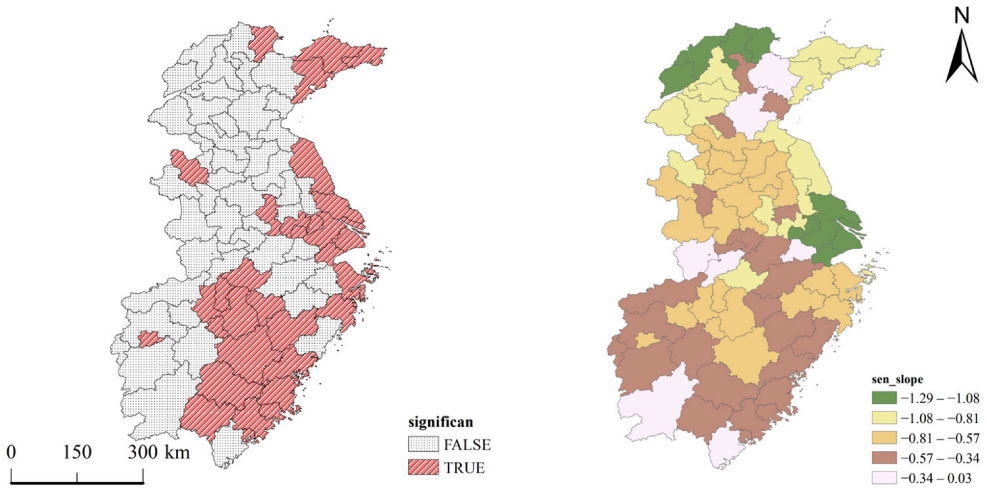



**Figure 7.** Theil-Sen and Mann–Kendall trend analysis of $PM_{2.5}$ concentration in East China, 2000–2023.

### 3.3.2. Hurst Index

A spatial analysis of the Hurst index based on $PM_{2.5}$ concentrations in East China from 2000 to 2023, as shown in Figure 8, revealed that the variation in $PM_{2.5}$ concentrations in this region exhibited complex long-term memory characteristics and significant spatial heterogeneity. Overall, the trend is dominated by persistence, but there are distinct differences in the underlying dynamic mechanisms within the region. In the core area of the Yangtze River Delta and the Shandong Peninsula urban agglomeration, the Hurst index generally exceeds 0.65, indicating strong persistence. This suggests that the downward trend in $PM_{2.5}$ concentrations in these regions exhibits significant temporal inertia, and the currently observed improvement in air quality is highly likely to continue in the future. This reflects the systematic and stable nature of their industrial restructuring and emission reduction policies; In regions undergoing industrial transformation and energy structure adjustment, such as northern Jiangsu, northern Anhui, and central-western Shandong, the Hurst index is primarily distributed within the moderate persistence range of 0.55–0.65. This suggests that future trends will generally align with current patterns; however, the strength of these trends may be subject to some uncertainty due to factors such as the pace of industrial restructuring, fluctuations in industrial activity, and regional pollution trans-

port; In ecologically sensitive and economically underdeveloped regions such as the hilly and mountainous areas of southwestern Zhejiang, the Dabie Mountains in western Anhui, and northern Fujian, anti-persistence signals with Hurst values below 0.45 have emerged. This indicates that $PM_{2.5}$ concentration dynamics in these areas are more susceptible to combined disturbances from local emission changes, interprovincial pollution transport, and interannual climate fluctuations, suggesting a potential for the current observed trend to reverse in the future.

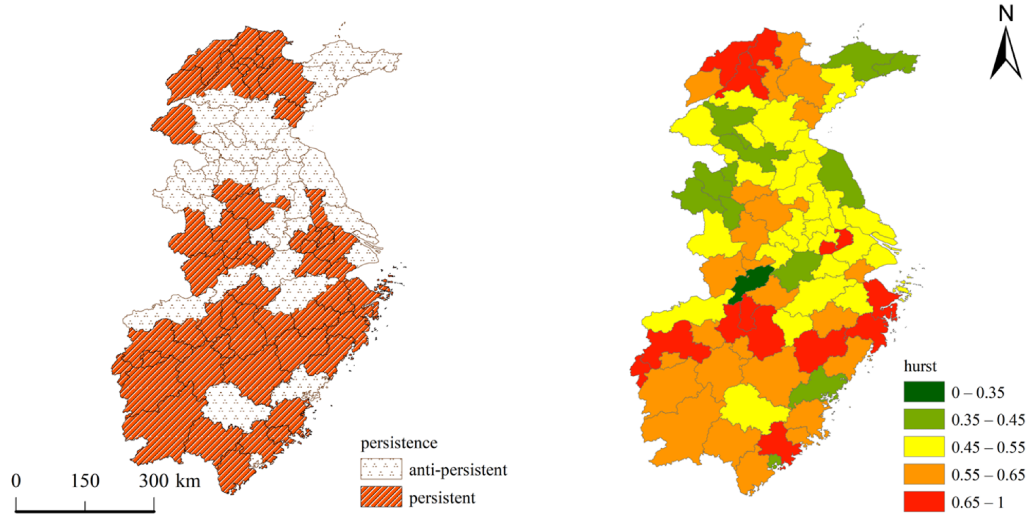



**Figure 8.** Spatial analysis of Hurst Index for $PM_{2.5}$ concentration in East China (2000–2023).

### *3.4. Spatial Variation in $PM_{2.5}$ Concentrations in East China*

The spatial autocorrelation analysis of $PM_{2.5}$ concentrations in East China from 2000 to 2023 is shown in Figure 9. The global Moran's I index remained consistently and significantly positive throughout the study period, confirming that the distribution of $PM_{2.5}$ concentrations exhibited significant spatial positive correlation and displays distinct spatial clustering characteristics. Regarding local spatial correlation patterns, high-concentration clusters were primarily distributed in industrial hubs such as Xuzhou, Suqian, and Lianyungang in Jiangsu Province; Huaibei, Suzhou, and Fuyang in Anhui Province; and Jinan, Zibo, and Dezhou in Shandong Province. These regions, characterized by industrial structures dominated by heavy industry and energy sectors, have formed stable pollution clusters; Conversely, areas with consistently low concentrations are found in ecological functional zones such as Lishui and Quzhou in Zhejiang Province, Nanping and Ningde in Fujian Province, as well as Huangshan in Anhui Province and Shangrao in Jiangxi Province. These regions feature high forest coverage and low industrial development intensity, constituting stable low-concentration clusters.

Overall, the spatial distribution of $PM_{2.5}$ concentrations in East China exhibited both significant spatial dependence and complex spatial heterogeneity. Although the overall concentration gradient has weakened somewhat, the basic spatial pattern of "high in the north and low in the south" remains firmly established in East China. This spatial distribution pattern clearly reflected the ongoing influence of regional differences in industrial structure and ecological background conditions on the spatial differentiation of pollution.

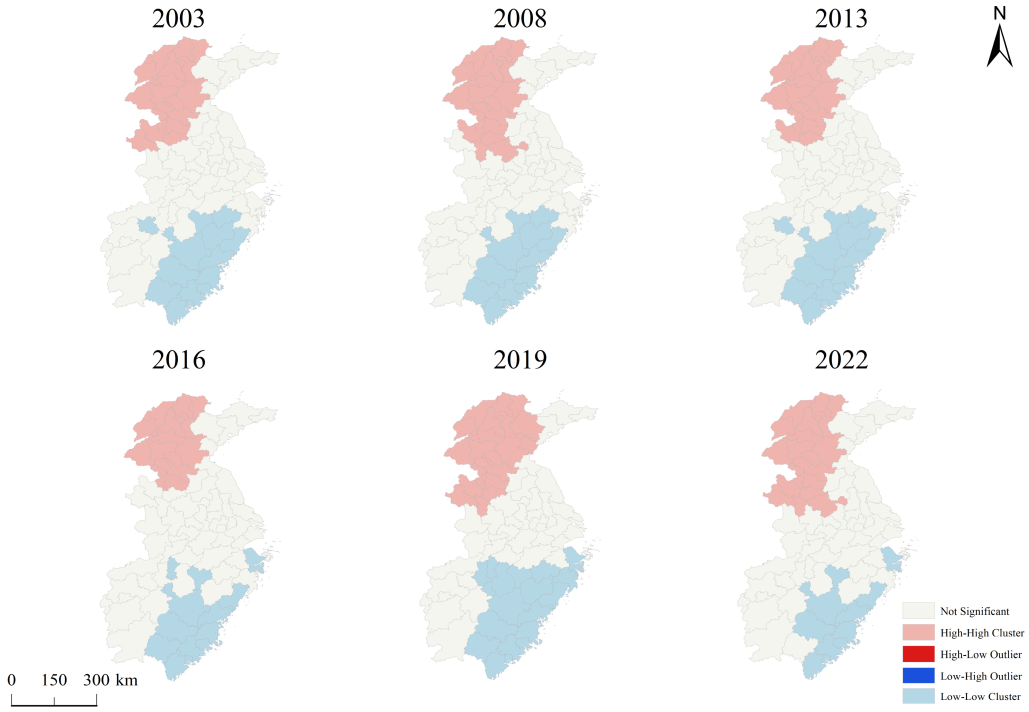



**Figure 9.** LISA clustering map of $PM_{2.5}$ concentration in East China.

### *3.5. Analysis of Factors Driving $PM_{2.5}$ Concentrations in East China*

#### 3.5.1. Selection of Dominant Factors

This study systematically evaluates the explanatory power and dominant factors of 10 key drivers identified using the XGBoost-SHAP algorithm regarding the spatial variation in $PM_{2.5}$ concentrations in East China. Using ArcGIS 10.8 and PyCharm 2024.3.4, we employed the statistical analysis tools to obtain the mean or sum of each factor within the grid cells. We then classified the independent variables using the natural breakpoint method and finally utilized the Geodetector to conduct factor exploration and interaction analysis, thereby revealing the key mechanisms influencing the spatial distribution of $PM_{2.5}$. Spearman's ρ for the ranked factor q-values was greater than 0.92, the *p*-values were less than 0.001, and the CV values were all less than 15%, indicating that the results are robust.

Figure 10 shows the results of the single-factor explanatory power analysis conducted by the Geographic Detector. The spatial variation in $PM_{2.5}$ concentrations in East China is driven by a combination of multidimensional factors. Among these, "annual average temperature" (q = 0.804) had the strongest explanatory power for spatial variation, significantly higher than other influencing factors, revealing the key controlling role of temperature in the regional pollution pattern; The secondary driving factor is "annual cumulative precipitation" (q = 0.724), indicating that meteorological factors have a very strong explanatory power for $PM_{2.5}$ concentrations in this region. The fact that both of these factors are far higher than socioeconomic factors demonstrates the fundamental shaping role of natural background conditions on the spatial pattern of $PM_{2.5}$. The remaining factors, such as "regional GDP growth rate" (q = 0.253) and "total road mileage" (q = 0.236), constitute the third tier of influence, reflecting to some extent that the region's economic and industrial development levels have a certain impact on $PM_{2.5}$ concentrations.

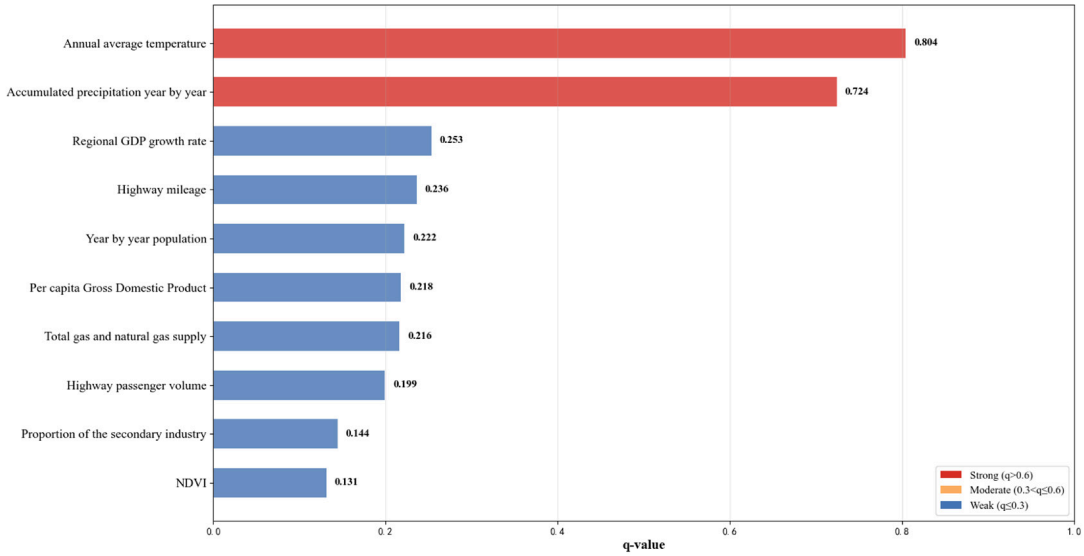



**Figure 10.** Ranking of single-factor explained variance for $PM_{2.5}$ concentration.

### 3.5.2. Interaction Analysis

Based on the results of the interactive detection analysis using the Geodetector, as shown in Figure 11, the formation mechanism of spatial variation in $PM_{2.5}$ concentrations in East China exhibits typical characteristics of synergistic enhancement driven by multiple factors. The Q-values obtained from pairwise interactions between the main factors are all greater than the explanatory power of any single factor, and all exhibited a two-factor enhancement relationship. This indicated that the regional pollution pattern was the result of interactions among multiple systems, including energy consumption, transportation activities, economic factors, and the natural environment. The interaction effects among meteorological factors are most pronounced. The interaction explanatory power between "annual cumulative precipitation" and "annual average temperature" reached 0.950, highlighting the fundamental regulatory role of the climate system in pollution dispersion. This implies that under adverse meteorological conditions—such as low temperatures and low rainfall—stricter measures for staggered industrial emissions and mobile source control should be implemented to offset the synergistic amplification effect of meteorological conditions on pollution accumulation. Concurrently, the interaction Q-value between "annual average temperature" and "share of the secondary industry" reached 0.617, indicating that high-temperature environments intensify the pollution effects of industrial emissions. Within the human activity system, the interaction Q-value between "road mileage" and "regional GDP growth rate" was 0.575, confirming that the synergistic effects of transportation network expansion and economic growth play a significant role in shaping the spatial patterns of pollution.

Overall, the intensity of interactions between natural factors and human activities is generally higher than that of interactions within purely human-activity systems, indicating that natural environmental conditions play a significant modulatory role in the response processes to human activities during the formation of pollution. This complex network of interactions suggests that efforts to control $PM_{2.5}$ pollution in East China require the establishment of a human emissions regulation system that takes climate conditions into account, with particular emphasis on the synergistic mechanisms involving industrial structure, transportation networks, and meteorological factors.

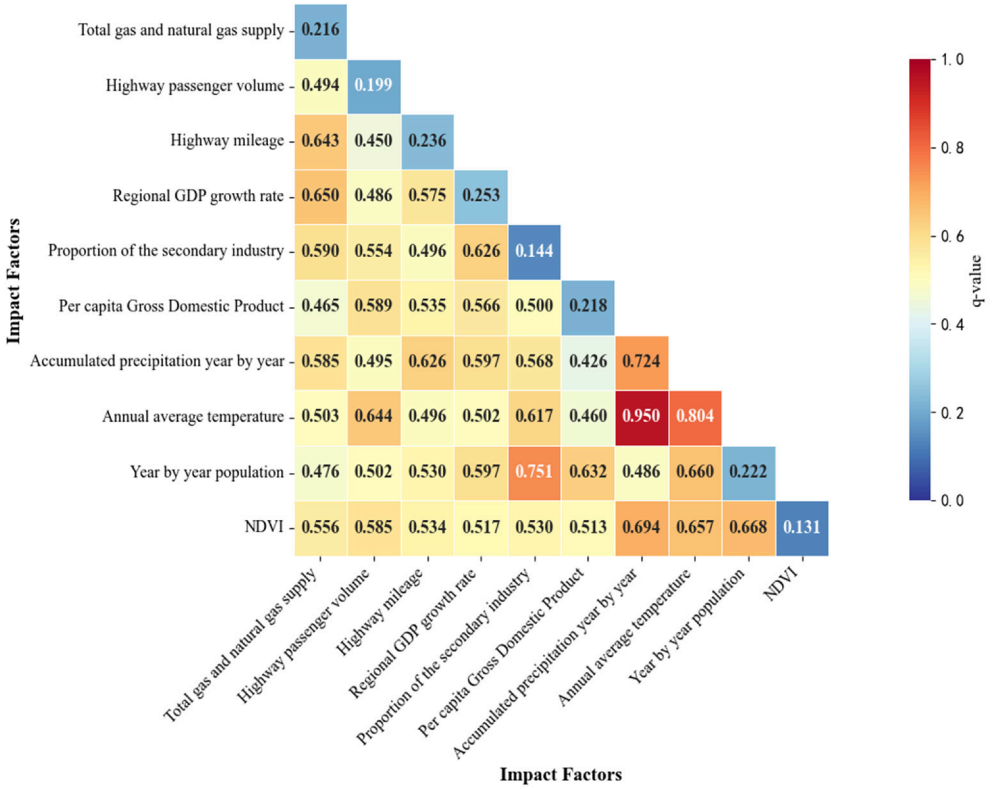



**Figure 11.** Interactive detection results of $PM_{2.5}$ influencing factors.

### *3.6. Analysis of Spatial Variability in $PM_{2.5}$ Drivers*

#### 3.6.1. VIF Test

Multiscale Geographically Weighted Regression (MGWR) is relatively sensitive to multicollinearity among explanatory variables; multicollinearity may lead to biased regression coefficient estimates and inflated standard errors. To ensure the robustness of the model estimates, a Variance Inflation Factor (VIF) diagnostic was performed on all candidate drivers prior to formal modeling. The results are shown in Figure 12 and Table 5. The VIF values for all 10 variables ranged from 1.36 to 3.17, all well below the empirical threshold of 5.0. The highest value was observed for population (VIF = 3.17), followed by total coal and natural gas supply (VIF = 2.29) and NDVI (VIF = 2.20). while the VIF values for the remaining variables were all below 2.2. This confirms that there is no severe multicollinearity among the drivers, and the variable structure meets the basic requirements for the regression model, allowing all variables to be included simultaneously in the subsequent MGWR analysis.

**Table 5.** VIF test results for 10 key factors.

| Variable | VIF |
|---|---|
| gas_supply | 2.29 |
| passenger_volume | 1.35 |
| road_mileage | 1.80 |
| gdp_growth_rate | 1.87 |
| secondary_industry_share | 1.59 |
| per_capita_gdp | 2.11 |
| annual_precipitation | 2.14 |
| annual_temperature | 1.82 |
| population | 3.16 |
| NDVI | 2.19 |

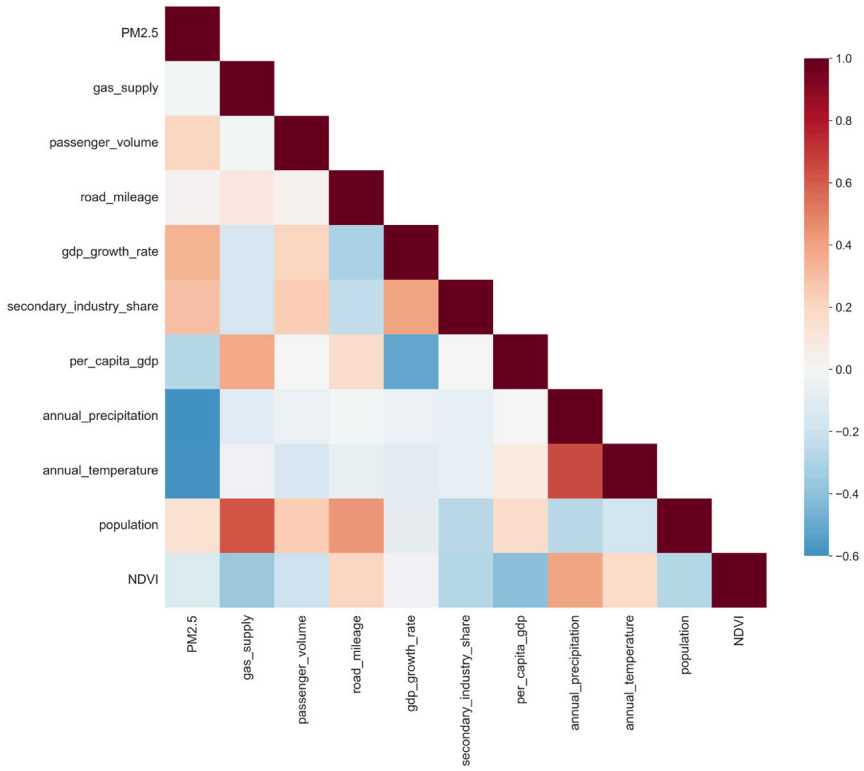



**Figure 12.** Heatmap of 10 key factors.

Between 2000 and 2023, the global Moran's I for $PM_{2.5}$ concentrations in the East China region was consistently greater than 0, and all values passed the significance test ($p < 0.001$), indicating that $PM_{2.5}$ concentrations in the region exhibit significant spatial autocorrelation and satisfy the basic assumptions of spatial regression analysis.

#### 3.6.2. Spatial Variability Analysis

This study selected three time points—2003, 2013, and 2023—to systematically characterize the evolution of $PM_{2.5}$ pollution in East China during three key phases: the early stages of rapid industrialization, the implementation of the Air Pollution Prevention and Control Action Plan, and the in-depth advancement of the "Dual Carbon" strategy. The selection of these cross-sectional years balances the need to cover major environmental policy milestones with data availability, providing a robust research foundation for using multiscale geographically weighted regression (MGWR) to reveal the spatiotemporal patterns of spatial heterogeneity in the underlying mechanisms.

The temporal trends and persistence characteristics during the study period have been thoroughly revealed by the Theil-Sen and Hurst indices, as well as spatial autocorrelation analysis. This study developed three models, Ordinary Least Squares (OLS), Geographically Weighted Regression (GWR), and Multiscale Geographically Weighted Regression (MGWR),to systematically investigate the spatial mechanisms by which various driving factors influence $PM_{2.5}$ concentrations. The results of the model comparison (Table 6) indicate that, given the MGWR model's sensitivity to sample size and the number of variables, this study used OLS and GWR as baseline models for comparison with MGWR. The results indicate that the MGWR model outperforms both OLS and GWR in terms of adjusted $R^2$ and AICc, suggesting that this model can more effectively capture the spatial non-stationarity of $PM_{2.5}$ drivers in the East China region. Meanwhile, further spatial autocorrelation tests on the model residuals revealed that the global Moran's I values for the OLS, GWR, and MGWR residuals were 0.33 ($p = 0$), 0.21 ($p = 0$), and $-0.01$ ($p = 0.002$),

respectively. The global Moran's I for the MGWR model residuals decreased to −0.01 ($p$ = 0.002), with the absolute value converging significantly compared to OLS (0.33) and GWR (0.21). Spatial dependence has been greatly mitigated, and the model adequately captured spatial structure, effectively reducing the spatial dependence of the residuals. The residuals tend toward a random distribution, indicating a good model fit. Therefore, at the current sample scale of 77 cities, the MGWR model demonstrated the best performance in terms of simulation accuracy and spatial adaptability, explaining 82.9% of the spatial variation in $PM_{2.5}$ concentrations in East China.

**Table 6.** Performance comparison of OLS, GWR, and MGWR models.

| Model | $R^2$ | Adjusted $R^2$ | AICc |
|---|---|---|---|
| OLS | 0.651 | 0.602 | 173.861 |
| GWR | 0.660 | 0.609 | 164.228 |
| MGWR | 0.883 | 0.829 | 128.947 |

The regression coefficients of the MGWR model reveal the spatial response of various driving factors to changes in $PM_{2.5}$. The regression coefficients were visualized using the natural breakpoint method on the ArcGIS 10.8 platform. The spatial analysis based on the MGWR model's regression coefficients is shown in Figures 13–15. The influence of each driving factor on changes in$PM_{2.5}$ exhibits significant geographic heterogeneity, and this heterogeneity shows distinct phased evolution across the three key stages.

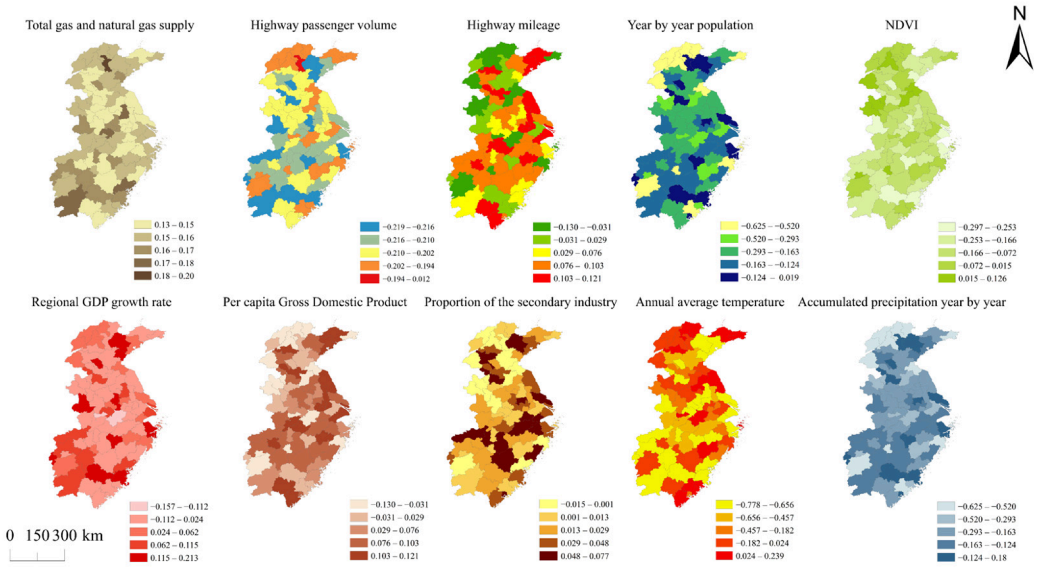



**Figure 13.** Spatial heterogeneity of factors affecting $PM_{2.5}$ in East China in 2003.

Natural factors have consistently played a dominant role, although the strength of their influence has varied over time. The absolute values of the regression coefficients for annual cumulative precipitation and annual average temperature were significantly higher than those of other socioeconomic factors in all three periods. Northern cities consistently exhibited a strong negative correlation, while the negative correlation in southern regions weakened somewhat, forming a clear spatial gradient pattern. In 2023, the absolute values of the regression coefficients for natural factors increased compared to 2003, indicating that the impact of climate change on $PM_{2.5}$ is intensifying.

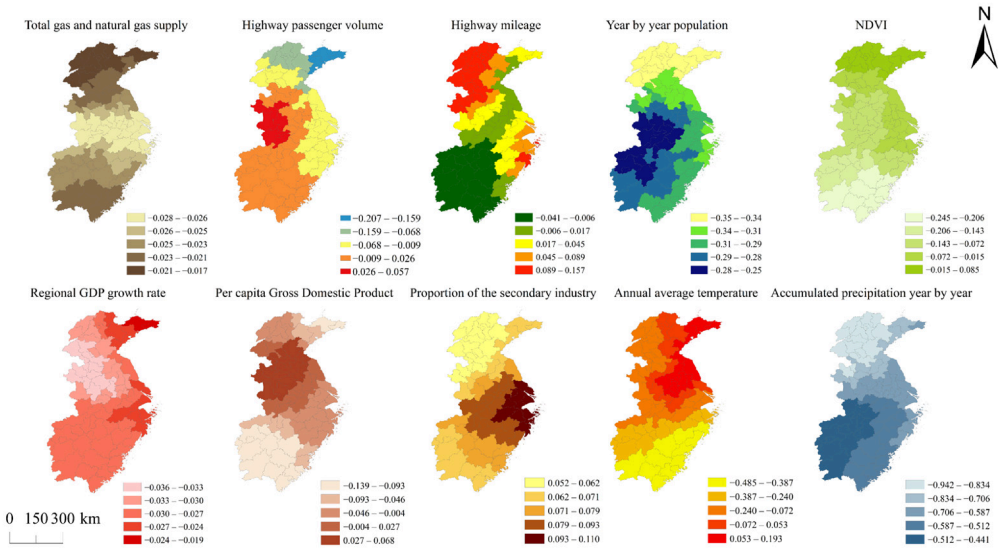



**Figure 14.** Spatial heterogeneity of factors affecting $PM_{2.5}$ in East China in 2013.

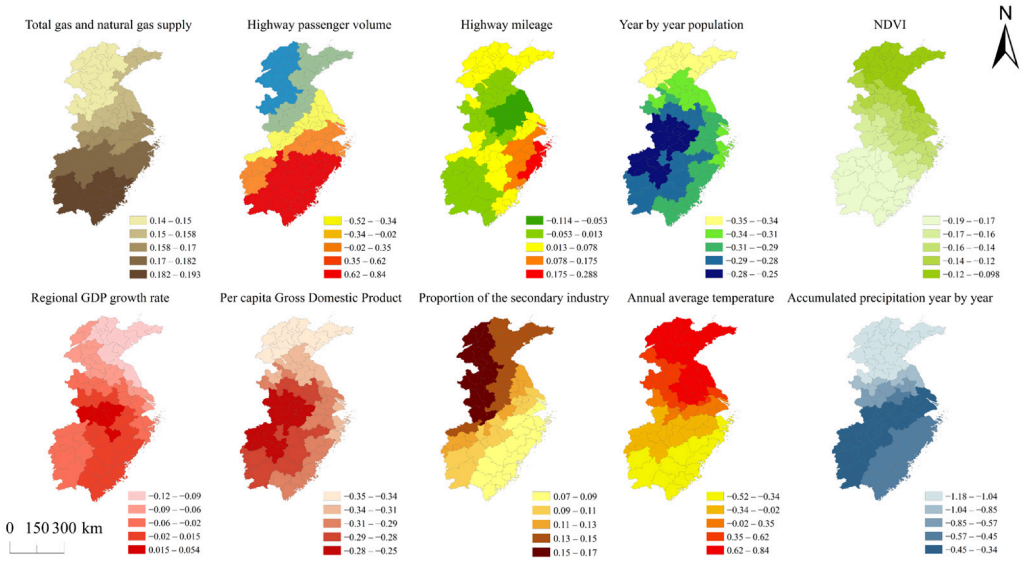



**Figure 15.** Spatial heterogeneity of factors affecting PM2.5 in East China in 2023.

The socioeconomic dimension exhibits distinct characteristics of phased evolution. The direction of influence of GDP per capita gradually evolved from a relatively homogeneous, weak positive correlation in 2003 to a coastal-inland divergence pattern in 2013, eventually forming a continuous spatial gradient by 2023 that transitions from coastal industrial clusters to inland ecological functional zones: the core area of the Yangtze River Delta and the Shandong Peninsula serve as positive drivers, while inland mountainous regions and ecological functional zones—such as southwestern Zhejiang, northern Fujian, and western Anhui—shifted to negative drivers. This pattern differs from a simple east–west dichotomy and better reflects the synergistic interaction between the stage of economic development and the ecological baseline. Between 2003 and 2013, the restraining effect of the secondary sector's share began to emerge in cities with higher levels of industrialization, and this characteristic became even more pronounced by 2023; the influence of infrastructure such as highways evolved from a weak spatial pattern in 2003 to a concentric diffusion pattern centered on transportation hubs by 2013, with this concentric structure becoming even clearer by 2023.

The NDVI index showed a negative correlation in most regions across the three periods, but the intensity of this effect exhibited significant regional variations and temporal evolution. In 2003, the negative impact of NDVI was relatively uniform; after 2013, differences among various ecological functional zones gradually widened; and by 2023, the negative impact in ecological conservation zones became more pronounced.

These characteristics of spatial heterogeneity revealed fundamental differences in the mechanisms of impact across different regions, providing an important basis for formulating targeted regional development policies. Coastal economic hubs (Shanghai, Southern Jiangsu, and Northern Zhejiang) should prioritize the optimization of industrial structure and climate adaptation; inland industrial transition zones (Northern Jiangsu, Northern Anhui, and Western Shandong) need to focus on balancing economic growth with infrastructure planning; while the Ecological Conservation Zones (Southwest Zhejiang, Northern Fujian, and Western Anhui) should strengthen the coordinated advancement of ecological protection and green development.

#### 3.6.3. Stability Test

Given that the "Total Gas/Natural Gas Supply" indicator is reported as a consolidated item in the Urban Statistical Yearbook—and thus cannot be broken down into separate supply volumes for coal-based gas and natural gas—and that the time-series changes in this indicator reflect the city's transition toward cleaner energy, its coefficient reflects the combined effect of the expansion in total energy consumption and the shift toward a cleaner energy structure, rather than the emissions contribution of any single highly polluting energy source. To verify the robustness of the MGWR model results, this study tested the model's explanatory power after excluding this variable from the model.

The results show (Table 7) that after excluding this variable, the model's adjusted $R^2$ decreased slightly from 0.829 to 0.820, and the AICc increased from 128.947 to 132.5; the model's overall explanatory power remained at a high level. Neither the direction nor the significance of the coefficients for the core explanatory variables underwent any substantial changes: air temperature and precipitation continued to exhibit a stable negative driving effect, with the strength of the negative correlation being significantly stronger in northern regions than in southern regions; the influence of per capita GDP continued to show a continuous spatial gradient transitioning from coastal industrial agglomeration zones to inland ecological functional zones; and the positive driving effects of socioeconomic factors, such as the share of the secondary industry and highway mileage, also remained stable. These results indicate that the MGWR model demonstrates good robustness in revealing the driving mechanisms of $PM_{2.5}$.

**Table 7.** Comparison of MGWR robustness test results.

| Indicator | Full Model | Model with Variables Excluded |
|---|---|---|
| $R^2$ | 0.883 | 0.875 |
| Adjusted $R^2$ | 0.829 | 0.820 |
| AICc | 128.947 | 130.961 |
| Number of variables | 10 | 9 |

## 4. Discussion

### *4.1. The Research in This Paper*

The analysis of $PM_{2.5}$ concentrations based on multi-source integrated data in this study is well-suited to the East China region. During the study period, $PM_{2.5}$ concentrations in East China showed a significant downward trend, and this improvement is closely linked to China's ongoing policies for air pollution control. In particular, since the implementation

of the 2013 Action Plan for Air Pollution Prevention and Control [61], measures such as industrial restructuring and the widespread adoption of end-of-pipe treatment facilities have effectively reduced regional particulate matter emissions. However, the extent of improvement varies significantly across different regions. The decline was most pronounced in the core area of the Yangtze River Delta, while improvements in traditional industrial zones such as northern Jiangsu and northern Anhui were relatively limited, reflecting the significant impact of regional differences in industrial structure on the effectiveness of pollution control efforts.

From a spatial perspective, $PM_{2.5}$ concentrations exhibit distinct spatial variations. Areas with high concentrations are primarily located in industrial clusters such as northern Jiangsu, northern Anhui, and western Shandong, where heavy chemical and energy industries dominate and emission intensities are high; conversely, areas with low concentrations are consistently found in ecological functional zones such as southwestern Zhejiang and northern Fujian, where high forest coverage and low levels of industrial development form natural ecological barriers. Hurst index analysis reveals that changes in $PM_{2.5}$ concentrations in ecologically sensitive areas—such as southwestern Zhejiang and northern Fujian—exhibit anti-persistence characteristics, indicating that the trend toward improved air quality in these regions remains uncertain. The causes can be summarized as follows: cross-regional pollution transport, the fact that these ecological functional zones are located downwind of the Yangtze River Delta industrial cluster, where changes in upstream emissions significantly disrupt local concentrations through interprovincial transport; when upstream emission reductions plateau or encounter unfavorable meteorological conditions, increased transport flux can offset the effects of local mitigation efforts [62]; and the lag effect of industrial relocation, as strict environmental policies in the core area of the Yangtze River Delta may drive energy-intensive industries to relocate to surrounding regions, placing some receiving areas under pressure from increased industrial emissions. Air quality improvement trends in these regions are susceptible to interference from local emission fluctuations and interprovincial transport, resulting in uncertainty; therefore, a continuous monitoring mechanism must be established. To preliminarily identify the potential impacts of transboundary transport, this study draws on the analytical framework of the HYSPLIT trajectory model and incorporates findings from previous research to support its conclusions. In winter, influenced by the Mongolian High, polluted air masses from the North China Plain can be transported to the Yangtze River Delta region via cold front events. Kang et al. simulated a cold front passage in January 2015 using the WRF-CMAQ model and found that the contribution of $PM_{2.5}$ from the North China Plain to the Yangtze River Delta could reach 29% during the cold front event [63]. Strong northwesterly winds during the cold front transported high-concentration particulate matter from upstream to downstream regions, after which subsidence under high-pressure conditions caused pollutants to linger within the boundary layer. Huang et al. further revealed the amplifying effect of aerosol-boundary layer interactions on transboundary transport, pointing out that this feedback mechanism can transport haze pollution from the North China Plain long distances to the Yangtze River Delta, with transport distances exceeding 1000 km [64]. Furthermore, an analysis by Yu et al. using the WRF-CHEM model of a severe smog event in Zhejiang in December 2013 showed that transboundary transport contributed as much as 62% to the increase in $PM_{2.5}$ concentrations in Zhejiang Province, with northerly airflows serving as the primary transport pathway [65]. These spatial differences in air mass origins partly explain why regions near the North China Plain, such as northern Jiangsu and northern Anhui, are prone to forming stable high-concentration zones, while regions influenced by clean air currents from the south, such as southwestern Zhejiang and northern Fujian, maintain a low-concentration pattern.

An analysis of the driving mechanisms indicates that natural factors play a dominant role in the spatial variation of $PM_{2.5}$. Annual average temperature and annual cumulative precipitation have the strongest explanatory power, reflecting the fundamental regulatory role of meteorological conditions on pollutant dispersion and wet deposition. The explanatory power of annual average temperature for the spatial variation in $PM_{2.5}$ concentrations reaches 0.804, significantly higher than that of other factors. This may be due to the fact that, under diffusion-controlled conditions, low temperatures compress the boundary layer thickness, suppressing vertical diffusion and leading to the accumulation of pollutants near the ground. The temperature gradient in East China—"cold in the north, warm in the south"—directly shapes the spatial pattern of "high pollution in the north, low pollution in the south"; The annual average temperature in East China ranges from 13 to 20 °C, which falls precisely within the active temperature window for the conversion of gaseous precursors into sulfates and nitrates [66]; spatial variations in temperature are amplified by chemical processes. Under stable atmospheric conditions, low-temperature areas are more prone to the formation of temperature inversions [67], which, when combined with stable weather, exacerbates pollution accumulation. Temperature comprehensively regulated physical diffusion, chemical transformation, and meteorological conditions; its high explanatory power reflects the fundamental role that climatic conditions play in shaping regional pollution patterns. The impact of the secondary industry's share on $PM_{2.5}$ exhibits significant spatial heterogeneity; while it generally exerts a positive driving effect, in some highly industrialized cities with advanced pollution control measures, the local regression coefficients are negative, indicating that the environmental effects of industrial structure variables are jointly moderated by development stage, technological level, and the intensity of pollution control measures. The interaction between natural factors and human activities significantly enhances the explanatory power of individual factors; for example, the interactive explanatory power of precipitation and temperature reaches 0.950, highlighting the synergistic impact mechanism between the climate system and anthropogenic emissions.

The results of the multiscale geographically weighted regression revealed spatial heterogeneity in the driving factors. Prior to applying the MGWR method, a multicollinearity test was conducted; the VIF diagnostic results showed that the VIF for all variables was below 3.2, indicating that multicollinearity among the variables was within a manageable range and would not distort the estimation of regression coefficients. Second, regarding sample size, the 77 prefecture-level cities in East China are sufficient for a global regression model (OLS), but present certain constraints for a local regression model such as MGWR. The core mechanism of MGWR is to select the optimal spatial bandwidth for each explanatory variable and perform weighted regression using neighboring sample points within that bandwidth. In this study, the adjusted $R^2$ of the MGWR model (0.883) improved by approximately 38% compared to OLS (0.602). The spatial distribution patterns of the MGWR regression coefficients provide a preliminary quantification and exploration of the spatial non-stationarity of $PM_{2.5}$ driving mechanisms under existing data conditions. These results offer valuable insights for comparing regional differences. Natural factors exhibited a stronger negative correlation in northern regions, forming a distinct spatial gradient; meanwhile, the direction of influence of per capita regional GDP varied across regions. Coastal developed urban agglomerations had entered a consumption-driven emissions phase, with rising energy consumption in transportation, construction, and the service sector. This finding suggests that differentiated strategies are needed for $PM_{2.5}$ pollution control in East China: economic core areas should focus on industrial upgrading and energy structure optimization; adjacent industrial zones need to balance environmental infrastructure development with economic growth; and ecological function zones should prioritize the coordinated advancement of ecological conservation and green development.

### 4.2. Comparison with Previous Known Findings

Compared with previous findings, this study found that $PM_{2.5}$ concentrations in East China exhibit a stable pattern of "high in the north and low in the south," which is generally consistent with the results of existing studies on the Yangtze River Delta and the eastern coastal regions of China. High-concentration areas are primarily concentrated in industrial clusters such as northern Jiangsu, northern Anhui, and western Shandong, indicating that a heavy industrial structure, high energy consumption intensity, and regional transport collectively shape the core areas of high pollution; conversely, low-concentration areas are mostly located in mountainous ecological functional zones such as southwestern Zhejiang and northern Fujian, suggesting that topographic barriers, vegetation cover, and lower industrial intensity act as buffers against $PM_{2.5}$.

Regarding the driving mechanisms, this study found that air temperature and precipitation have significantly greater explanatory power than socioeconomic factors. This is consistent with the conclusion by Hu et al. that meteorological conditions play a fundamental moderating role in air pollution in the Pearl River Delta [13], but differs from some studies that place greater emphasis on the direct impact of industrial emissions [10]. This discrepancy may stem from the fact that this study utilized city-scale annual average $PM_{2.5}$ concentrations and regional long-term time series, which more fully reflect the comprehensive modulating effect of climatic conditions on pollution dispersion and removal, while short-term fluctuations in industrial emissions are partially smoothed out on an annual scale.

By integrating multiple spatial analysis methods, this study systematically revealed the spatiotemporal evolution patterns and spatial distribution of $PM_{2.5}$ concentrations in East China. However, the study still has certain limitations, particularly regarding the insufficient quantification of the impact of interregional transport. Future research could incorporate atmospheric chemical transport models to further clarify the contributions of local emissions and regional transport, thereby providing more precise support for joint prevention and control efforts in the Yangtze River Delta region. Furthermore, as the "Dual Carbon" strategy is further advanced, exploring pathways for synergistic pollution and carbon reduction will be a key focus of future research.

Furthermore, the MGWR results indicate that the direction of the effects of per capita GDP and the share of the secondary sector exhibits significant spatial heterogeneity, which is consistent with the widely held view—both domestically and internationally—that "the relationship between economic development stages and pollution responses is not strictly linear." Coastal core cities may mitigate the pollution pressure resulting from the expansion of traditional industries through industrial upgrading and strengthened governance, whereas inland industrial transition zones continue to exhibit a pattern where economic growth and pollution accumulation coexist.

### 4.3. Future Research and Limitations

This study established an analytical framework of "feature screening—spatial detection—local modeling" and systematically identified the spatiotemporal evolution and driving mechanisms of $PM_{2.5}$ in East China; however, further research could be conducted in the following areas.

Refined analysis of transregional transport. This study primarily employed statistical attribution methods and integrated existing atmospheric chemical transport results to reveal the combined effects of meteorological and anthropogenic factors. In the future, atmospheric chemical transport models such as WRF-Chem and CMAQ could be utilized to conduct numerical simulations of typical pollution episodes, quantitatively separating the contributions of local emissions and regional transport, thereby providing a more

refined basis for source identification in the Yangtze River Delta's joint prevention and control efforts.

Expansion of research scope. This study focuses on $PM_{2.5}$ as a single pollutant; future research could be expanded to analyze the synergistic pollution mechanisms of $PM_{2.5}$ and $O_3$, and conduct studies on pathways for synergistic pollution and carbon reduction in line with the "Dual Carbon" goals. Additionally, incorporating health impact assessments to construct an integrated "pollution–health–climate" analytical framework would provide more comprehensive decision-making references for targeted regional governance. Furthermore, as the "Dual Carbon" strategy continues to advance, exploring pathways for synergistic pollution and carbon reduction will be a key focus of future research.

## 5. Conclusions

(1) Between 2000 and 2023, $PM_{2.5}$ concentrations in East China showed a significant downward trend. Based on Theil-Sen trend analysis, the Sen slopes for all pixels across the region were negative, ranging from $-1.29\ \mu g \cdot m^{-3} \cdot a^{-1}$ to $-0.03\ \mu g \cdot m^{-3} \cdot a^{-1}$, confirming the environmental reality of overall air quality improvement. This improvement process exhibited distinct phased characteristics: the decline was relatively slow prior to the implementation of the 2013 Action Plan for Air Pollution Prevention and Control, after which a period of rapid improvement began. Hurst index analysis indicated that changes in $PM_{2.5}$ concentrations in certain regions, such as southwestern Zhejiang and western Anhui, exhibited anti-persistence characteristics, suggesting a potential risk of reversal in the improvement trend in these areas. It is recommended to establish a cross-provincial joint prevention and control early warning mechanism. These regions are located downwind of the Yangtze River Delta industrial cluster, and their air quality is susceptible to interference from upstream emission changes and interannual climate fluctuations, posing a risk of reversal in the current improvement trend. Policy measures should strengthen the function of ecological barriers, strictly control local emission increases resulting from the gradual relocation of energy-intensive industries, and establish a mechanism for staggered industrial production that is dynamically adjusted based on meteorological conditions.

(2) The spatial distribution of $PM_{2.5}$ concentrations exhibited a distinct pattern of spatial differentiation. Spatial autocorrelation analysis indicated that high-concentration clusters persisted in industrially concentrated regions such as northern Jiangsu, northern Anhui, and western Shandong, including cities like Xuzhou, Suqian, Huaibei, and Suzhou; conversely, low-concentration clusters were consistently found in ecological functional zones such as southwestern Zhejiang and northern Fujian, encompassing areas like Lishui, Quzhou, and Nanping. This spatial pattern profoundly reflected the differences in regional industrial structures and ecological conditions. The heavy industrial structure and energy consumption characteristics of industrial clusters stand in stark contrast to the advantages of forest cover in ecological functional zones.

(3) Analysis of the driving mechanisms indicated that natural factors played a dominant role in the spatial variation of $PM_{2.5}$ concentrations. Results from the geographic detector showed that annual average temperature (q = 0.804) and annual cumulative precipitation (q = 0.724) had significantly higher explanatory power than other factors. The MGWR model revealed the spatial heterogeneity of driving factors: natural factors exhibit a stronger negative correlation in northern East China (e.g., northern Jiangsu, northern Anhui, and western Shandong), forming a distinct north–south spatial gradient; among socioeconomic factors, the share of the secondary industry generally exerted a positive driving effect, while the direction of the impact of per

capita regional GDP varied by region. It acted as a promoting factor in coastal urban agglomerations such as the core area of the Yangtze River Delta and the Shandong Peninsula, whereas it shifted to an inhibitory effect in inland mountainous and ecological conservation areas such as southwestern Zhejiang, northern Fujian, and western Anhui. This demonstrated a continuous spatial gradient transitioning from coastal to inland areas and from industrial agglomeration to ecological conservation, rather than a simple east–west dichotomy. This finding provided a scientific basis for formulating differentiated regional pollution control strategies.

**Author Contributions:** Conceptualization, methodology, and writing—original draft, Y.L.; investigation and formal analysis, Y.L. and J.C.; validation, S.L. and Y.Z.; resources, S.L. and Y.L.; data curation and visualization, Y.L. and J.C.; software, Y.L. and Y.Z.; writing—review and editing, and supervision, T.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Philosophy and Social Sciences Research Project of Education System in Fujian Province, China (No. JAS24029) and Fujian Provincial Social Science Planning Project (No. FJ2018B063).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** We thank the reviewers and editors for their suggestions on this work.

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

## References

1. Wan, L.; Bambrick, H.; Tong, M. Future health burden and economic costs associated with ambient $PM_{2.5}$ pollution in China amid net-zero transition. *Environ. Res.* **2025**, *283*, 122130. [CrossRef] [PubMed]
2. Chang, H.Y.; Tang, D.J.; Zhao, C.; Sun, C.Y.; Chen, F.G.; Guan, M.Y.; Zhang, Y. Source Apportionment and Ecological Risk Assessment of Metal Pollution in Atmospheric $PM_{2.5}$ in a City of North China. *Chin. J. Dis. Control Prev.* **2021**, *25*, 1251–1256+1263. [CrossRef]
3. Sang, C.; He, Y.Q.; Jia, J.; Chen, H.; Wang, M.Z. Effects of short-term exposure to $PM_{2.5}$ and its constituents on fasting plasms glucose and blood lipids. *Chin. Environ. Sci.* **2025**, *45*, 1118–1126. [CrossRef]
4. Sha, D.D.; Wang, H.L.; Zhu, B.; Lin, X.; Guo, T.; Shi, S.S.; Jiang, L.; Li, Y.E. Pollution Characteristics of Carbonaceous Aerosols in Winter $PM_{2.5}$: Analysis of a Regional Heavy Pollution Process in the Yangtze River Delta. *Chin. Environ. Sci.* **2017**, *37*, 3611–3622.
5. Kong, S.J.; He, M.Y. Impact and Mechanism Analysis of the "Three-Year Action Plan for Defending the Blue Sky" Policy Implementation on $CO_2$-$PM_{2.5}$-$O_3$ Coordinated Reduction. *Res. Environ. Sci.* **2025**, *38*, 1689–1700. [CrossRef]
6. Xiang, X.H. Clouds Disperse, Mist Fades: A Study on Air Pollution Control under Dual Synergy of "Region-Target". *Technol. Econ. Manag. Res.* **2025**, 8–14. [CrossRef]
7. Geng, G.; Xiao, Q.; Zheng, Y.; Tong, D.; Zhang, Y.; Zhang, X.; Zhang, Q.; He, K.; Liu, Y. Impact of China's Air Pollution Prevention and Control Action Plan on $PM_{2.5}$ chemical composition over eastern China. *Sci. China Earth Sci.* **2019**, *62*, 1872–1884. [CrossRef]
8. Yang, Q.; Wang, B.; Chen, Y.J. Spatio-Temporal Evolution Characteristics and Influencing Factors of $PM_{2.5}$ and $O_3$ Pollution in Zhejiang Province. *China Environ. Monit.* **2026**, *42*, 144–156.
9. Chen, Y.L.; Li, Q.Q.; Fan, Q.; Wang, Z.R. Spatial-Temporal Heterogeneity and Driving Factors of $PM_{2.5}$ in the Yangtze River Delta Urban Agglomeration. *J. Saf. Environ.* **2025**, *25*, 3252–3264. [CrossRef]
10. Wang, L.L.; Liu, X.J.; Li, D.; Sun, Y.Q. Geographical Detection of Spatial Heterogeneity and Drivers of $PM_{2.5}$ in the Yangtze River Economic Belt. *Environ. Sci.* **2022**, *43*, 1190–1200. [CrossRef] [PubMed]
11. Zhang, X.; Gu, X.; Cheng, C.; Yang, D. Spatiotemporal heterogeneity of $PM_{2.5}$ and its relationship with urbanization in North China from 2000 to 2017. *Sci. Total Environ.* **2020**, *744*, 140925. [CrossRef] [PubMed]
12. Zhang, Y.; Yang, Y.; Chen, J.; Shi, M.Y. Spatiotemporal heterogeneity of the relationships between $PM_{2.5}$ concentrations and their drivers in China's coastal ports. *J. Environ. Manag.* **2023**, *345*, 118698. [CrossRef] [PubMed]
13. Hu, M.; Wang, Y.; Wang, S.; Jiao, M.Y.; Huang, G.H.; Xia, B.C. Spatial-temporal heterogeneity of air pollution and its relationship with meteorological factors in the Pearl River Delta, China. *Atmos. Environ.* **2021**, *254*, 118415. [CrossRef]

14. Aas, W.; Mortier, A.; Bowersox, V.; Cherian, R.; Faluvegi, G.; Fagerli, H.; Hand, J.L.; Klimont, Z.; Galy-Lacaux, C.; Lehmann, C.M.B.; et al. Global and regional trends of atmospheric sulfur and nitrogen deposition. *Atmos. Chem. Phys.* **2019**, *19*, 15235–15254. [CrossRef]
15. Hand, J.L.; Schichtel, B.A.; Malm, W.C.; Chow, J.C. Particulate sulfate and nitrate trends in the United States from the IMPROVE network: 2001–2016. *Atmos. Environ.* **2019**, *214*, 116873. [CrossRef]
16. Pawar, H.; Garg, S.; Kumar, V.; Sachan, H.; Arya, R.; Sarkar, C.; Chandra, B.P.; Sinha, B. Quantifying the contribution of long-range transport to particulate matter (PM) mass loadings at a suburban site in the north-western Indo-Gangetic Plain (IGP). *Atmos. Chem. Phys.* **2015**, *15*, 9531–9546. [CrossRef]
17. Xu, X.; Shi, K.; Huang, Z.; Shen, J. What Factors Dominate the Change of $PM_{2.5}$ in the World from 2000 to 2019? A Study from Multi-Source Data. *Int. J. Environ. Res. Public Health* **2023**, *20*, 2282. [CrossRef] [PubMed]
18. He, Q.S.; Geng, F.H.; Wang, Y.Y.; Zhou, G.Q.; Gao, W.; Peng, Y.P. An Analysis of $PM_{2.5}$ Distribution, Variation and Formation in the Eastern China for 2000–2015. *Adv. Meteorol. Sci. Technol.* **2017**, *7*, 134–142+149. [CrossRef]
19. Zhao, X.Y.; Zhang, Q.; Yang, F.S.; Zheng, L.L.; Luo, J.X.; Shi, Z.H.; Wen, S.L. Analysis of Influencing Factors of $PM_{2.5}$ in Shaanxi Province Based on XGBoost-SHAP Method. *Res. Environ. Sci.* **2025**, *38*, 990–999. [CrossRef]
20. Zheng, Q.P.; Li, F.; Zhao, R.; Jiang, D.S.; Wang, H. Analysis of the Characteristics of $PM_{2.5}$-$O_3$ Compound Pollution and the Impact of Synoptic Weather Patterns in Fujian Province. *Ecol. Environ.* **2023**, *32*, 1440–1448. [CrossRef]
21. Yao, X.; Chen, X.; Wang, X.W. Spatio-Temporal Evolution and Driving Factors of Eco-Environmental Quality in the Min-Triangle Urban Agglomeration Based on XGBoost-SHAP Model. *Environ. Sci.* **2026**, *47*, 3758–3769. [CrossRef] [PubMed]
22. Feng, H.Y.; Zhang, X.B.; Shi, P.J.; Shi, J.; Wang, Z.Y. Environmental Interpretation of Spatial Heterogeneity in the Trade-Offs and Synergies of Land Use Functions: A Study Based on the XGBoost-SHAP Model. *J. Arid Land* **2025**, *17*, 1378–1401. [CrossRef]
23. Wu, T.; Tian, F.; Fu, Y.; Tang, B.; Li, X.N.; Chen, S.H.; Liu, G.D.; Wei, S.C. A Method for Predicting Lithium Associated with Aluminum-Bearing Rock Series in Wuzhengdao Area of Northern Guizhou Based on Machine Learning. *J. Earth Sci.* **2026**, *37*, 1–20. [CrossRef]
24. Chen, Y.M.; Su, Z.W.; Chen, Y.P.; Lin, Z.P. Modeling of $PM_{2.5}$ in the Yangtze River Delta Based on Geographically Weighted Random Forest. *Chin. Environ. Sci.* **2024**, *44*, 4240–4248. [CrossRef]
25. Zhang, K.L.; Fang, B.; Zhang, Z.C.; Xia, C.H.; Liu, Q.Q.; Liu, K. Spatial Optimisation Based on Ecosystem Service Spillover Effect and Cross-Scale Knowledge Integration: A Case Study of the Yellow River Basin. *J. Geogr. Sci.* **2025**, *35*, 1080–1122. [CrossRef]
26. Zhou, K.; Yin, Y.; Chen, Y.F. Drivers of Water Pollutant Discharge in Urban Agglomerations and Their Scale Effects: Empirical Analysis of 305 Counties in the Yangtze River Delta. *J. Geogr. Sci.* **2023**, *33*, 195–214. [CrossRef]
27. Gao, F.; Li, Y.X.; Gou, J.R.; Wang, Z.H.; Huang, R. Spatiotemporal Variation of $PM_{2.5}$ and Identification of Influencing Factors in Shanxi Province Using the Geodetector. *Res. Environ. Sci.* **2026**, *39*, 36–49. [CrossRef]
28. Dong, D.; Wen, X.; Li, Q.; Wang, Y.; Zhang, H. Spatial Differentiation Characteristics and Influencing Factors of Traditional Villages in Anhui Province Based on Spatial Scale Effect Using Geodetector. *J. Anhui Agric. Univ.* **2025**, *52*, 915–928. [CrossRef]
29. Zhang, J.P.; Qiao, Q.; Liu, X.N.; Zhu, L.H.; Liu, C.L.; Peng, Y. Identification of Key Driving Forces of Ecosystem Services and Analysis of Spatial Heterogeneity of Their Driving Effects in the Chaobai River Basin. *Res. Environ. Sci.* **2026**, *39*, 393–404. [CrossRef]
30. Shi, T.; Wang, S.; Yang, L.; Chen, W.Q.; Wang, Y.; Gao, J.J. Spatiotemporal Variation of $PM_{2.5}$ Concentration and Its Association with Landscape Patterns in East China. *Remote Sens. Technol. Appl.* **2024**, *39*, 435–447. [CrossRef]
31. Zang, X.H.; Lu, Y.T.; Yao, H.; Li, F.D.; Zhang, S.C. Spatiotemporal Distribution Characteristics of Major Air Pollutants in China. *Ecol. Environ. Sci.* **2015**, *24*, 1322–1329. [CrossRef]
32. Fan, J.; Huang, L.; Li, X. Spatiotemporal Variation Characteristics and Meteorological Causes of $PM_{2.5}$ Pollution in Jiangxi Province from 2015 to 2018. *Ecol. Environ. Sci.* **2020**, *29*, 1165–1172. [CrossRef]
33. Zhang, H.; Wu, Y.Z.; Mei, Y.Z.; Cheng, J.H.; Zheng, J.H. Spatio-Temporal Changes of Air Quality and Their Urban Agglomeration Distribution Characteristics in China: A Discussion on the Effectiveness of Air Pollution Control since the 13th Five-Year Plan Period. *J. Environ. Econ. Res.* **2023**, *8*, 46–65. [CrossRef]
34. Luo, Z.; Li, H.R. The Impact of the "Air Ten Articles" Policy Implementation on Air Quality. *China Ind. Econ.* **2018**, 136–154. [CrossRef]
35. van Donkelaar, A.; Martin, R.V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P.J. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, 847–855. [CrossRef] [PubMed]
36. Hammer, M.S.; van Donkelaar, A.; Li, C.; Lyapustin, A.; Sayer, A.M.; Hsu, N.C.; Levy, R.C.; Garay, M.J.; Kalashnikova, O.V.; Kahn, R.A. Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998–2018). *Environ. Sci. Technol.* **2020**, *54*, 7879–7890. [CrossRef] [PubMed]
37. Xu, Q.; Kang, S.; He, X.; Xu, M. Evaluating the affecting factors of glacier mass balance in Tanggula Mountains using explainable machine learning and the open global glacier model. *J. Mt. Sci.* **2025**, *22*, 466–488. [CrossRef]

38. Xin, Z.; Zhang, J.; Peng, K.; Zhang, J.; Zhang, C.; Wu, J.; Zhang, B.; Liu, Q. Explainable machine learning model for predicting molten steel temperature in the LF refining process. *Int. J. Miner. Metall. Mater.* **2024**, *31*, 2657–2669. [CrossRef]
39. Wang, M.; Tang, J.; Chu, M.; Shi, Q.; Zhang, Z. Prediction and optimization of flue pressure in sintering process based on SHAP. *Int. J. Miner. Metall. Mater.* **2025**, *32*, 346–359. [CrossRef]
40. Doan, Q.C.; Vu, H.K.; Trinh, T.K.T.; Bui, T.C.N. Examining the nonlinear and threshold effects of the 5Ds built environment to land values using interpretable machine learning models. *J. Geogr. Sci.* **2024**, *34*, 2509–2533. [CrossRef]
41. Liu, N.; Hao, Z.; Zhao, P. Spatiotemporal change of $PM_{2.5}$ concentration in Beijing-Tianjin-Hebei and its prediction based on machine learning. *Urban Clim.* **2024**, *58*, 102167. [CrossRef]
42. Gao, J.; Bian, H.; Zhu, C.; Tang, S. The response of key ecosystem services to land use and climate change in Chongqing: Time, space, and altitude. *J. Geogr. Sci.* **2022**, *32*, 317–339. [CrossRef]
43. Huang, Y.; Bai, F. Study on the variation law of drainage in salinized irrigation area based on Mann-Kendall test: A case study of Hetao Irrigation Area. *J. Irrig. Drain.* **2025**, *44*, 92–99. [CrossRef]
44. Zhao, Z.H.; Luo, Z.J.; Huang, L.Y.; Xing, L.T.; Sun, H.J.; Chen, H.L. Analysis of rainfall and groundwater changes in Jinan based on STL and Mann-Kendall method. *J. China Hydrol.* **2022**, *42*, 73–77. [CrossRef]
45. Xue, L.Y.; Yang, J.; Tang, Z.S.; Zhou, J.; Yue, L.; Meng, T.L. Analysis of phenological characteristics and climatic driving factors of natural grassland in China under climate change. *Acta Ecol. Sin.* **2026**, *46*, 1208–1220. [CrossRef]
46. Song, H.P.; Tao, H.; Dang, W.T.; Li, M.M.; Wang, Y.B. Study on the response of vegetation NDVI to climate change and human activities in Longdong area. *J. Chang. River Sci. Res. Inst.* **2026**, *43*, 97–104. [CrossRef]
47. Li, Q.; Lin, H.; Wang, P.; Dong, Q.T.; Wang, K.; Shuai, J.W.; Ye, F.F. Analysis of spatiotemporal dynamic patterns of gene expression during mouse embryonic development based on Moran's I and spatial transcriptomics. *Chin. Phys. B* **2025**, *34*, 08703. [CrossRef]
48. Feng, Y.; Chen, L.; Chen, X. The impact of spatial scale on local Moran's I clustering of annual fishing effort for Dosidicus gigas offshore Peru. *J. Oceanol. Limnol.* **2019**, *37*, 330–343. [CrossRef]
49. Chen, G.P.; Zhang, D.D.; Li, K.; Zhao, J.S.; Yang, H.B. Trade-offs/Synergies of multi-functional land use and their driving mechanisms in Kunming City. *Trans. Chin. Soc. Agric. Mach.* **2025**, *56*, 648–661. [CrossRef]
50. Ran, L.J.; Guan, D.J.; Zhou, L.L.; Chen, S.; Lu, Y.; Zhu, X.S. Identification of decoupling types between urban sprawl and ecosystem service flows in the Yangtze River Basin and improvement paths. *Environ. Sci.* **2025**. *online first*. [CrossRef]
51. Liu, H.; Zhang, J. Spatiotemporal patterns and influencing factors of synergistic development of carbon reduction, pollution reduction, and green expansion in the Yellow River Basin. *Environ. Sci.* **2025**. [CrossRef]
52. Peng, J.; Zhang, Z.; Fu, S.; Tian, J. Spatio-temporal effects of environmental factors on the innovation-oriented urban construction based on the MO-OLS method: A case study of Hubei Province. *Resour. Environ. Yangtze Basin* **2024**, *33*, 2562–2576. [CrossRef] [PubMed]
53. Yu, L.L.; Jiang, X.L.; Ba, W.L.; Chen, S.Y.; Yan, X.T.; Yue, Y.F. Medium- and long-term prediction of system equivalent inertia based on PCA-OLS model. *Proc. CSU-EPSA* **2024**, *36*, 101–109. [CrossRef]
54. Shen, Q.X.; Zhu, Y.J.; Han, Z.R.; Deng, Y.F.; Weng, R.J. Characteristics and impact mechanisms of spatiotemporal patterns of carbon emissions in China's urban agglomerations. *Chin. Geogr. Sci.* **2025**, *35*, 1328–1341. [CrossRef]
55. Zhang, M.; Zhang, Z.; Yu, F. Spatio-temporal pattern and attribution analysis of the mass elevation effect in the Tianshan Mountains in China. *J. Geogr. Sci.* **2023**, *33*, 2031–2051. [CrossRef]
56. Song, X.; Mi, N.; Mi, W.; Li, L.T. Spatial non-stationary characteristics between temperate grasslands based on a mixed geographically weighted regression model. *J. Geogr. Sci.* **2022**, *32*, 1076–1102. [CrossRef]
57. van Donkelaar, A.; Hammer, M.S.; Bindle, L.; Brauer, M.; Brook, J.R.; Garay, M.J.; Hsu, N.C.; Kalashnikova, O.V.; Kahn, R.A.; Lee, C.; et al. Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty. *Environ. Sci. Technol.* **2021**, *55*, 15287–15300. [CrossRef] [PubMed]
58. Xiao, Q.; Chang, H.H.; Geng, G.; Liu, Y. An Ensemble Machine-Learning Model To Predict Historical $PM_{2.5}$ Concentrations in China from Satellite Data. *Environ. Sci. Technol.* **2018**, *52*, 13260–13269. [CrossRef] [PubMed]
59. Ma, Z.; Hu, X.; Huang, L.; Bi, J.; Liu, Y. Estimating Ground-Level $PM_{2.5}$ in China Using Satellite Remote Sensing. *Environ. Sci. Technol.* **2014**, *48*, 7436–7444. [CrossRef] [PubMed]
60. Li, T.; Shen, H.; Yuan, Q.; Zhang, L. Validation Approaches for Satellite-Based $PM_{2.5}$ Estimation: Assessment and a New Approach. *Remote Sens. Environ.* **2019**, *225*, 219–228. [CrossRef]
61. Liu, H.; Shao, M.; Guo, L. China's air pollution control in the new era: Historical review, effectiveness evaluation, and path outlook. *J. Bus. Econ.* **2023**, 66–79. [CrossRef]
62. Yan, F.F.; Su, H.; Cheng, Y.F.; Huang, R.J.; Liao, H.; Yang, T.; Zhu, Y.Y.; Zhang, S.Q.; Sheng, L.F.; Kou, W.B.; et al. Frequent Haze Events Associated with Transport and Stagnation over the Corridor between North China Plain and Yangtze River Delta. *Atmos. Chem. Phys. Discuss.* **2024**, *24*, 2365–2376. [CrossRef]
63. Kang, H.; Zhu, B.; Gao, J.; He, Y.; Wang, H.; Su, J.; Pan, C.; Zhu, T.; Yu, B. Potential impacts of cold frontal passage on air quality over the Yangtze River Delta, China. *Atmos. Chem. Phys.* **2019**, *19*, 3673–3685. [CrossRef]

64. Huang, X.; Ding, A.; Wang, Z.; Ding, K.; Gao, J.; Chai, F.; Fu, C. Amplified transboundary transport of haze by aerosol–boundary layer interaction in China. *Nat. Geosci.* **2020**, *13*, 428–434. [CrossRef]
65. Yu, Y.; Xu, H.; Jiang, Y.; Chen, F.; Liu, D. A modeling study of $PM_{2.5}$ transboundary transport during a winter severe haze episode in southern Yangtze River Delta, China. *Atmos. Res.* **2021**, *253*, 105454. [CrossRef]
66. Hong, Y.; Sun, J.; Ma, Y.; Wang, Y.; Zhang, Y.; Liu, N.; Zhou, D. Formation and evolution of secondary particulate matter during heavy haze pollution episodes in winter in a severe cold climate region of Northeast China. *Environ. Sci. Pollut. Res.* **2022**, *29*, 67821–67836. [CrossRef] [PubMed]
67. Meng, C.; Cheng, T.; Gu, X.; Shi, S.; Bao, F. Contribution of meteorological factors to particulate pollution during winters in Beijing. *Sci. Total Environ.* **2019**, *656*, 977–985. [CrossRef] [PubMed]

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