

Article

Spatiotemporal Evolution and Spatial Spillover Effects of Urban Air Pollution Sources in the Yangtze River Economic Belt of China

Xue-Ying Wang^a, Kai-Rong Liang^a, Chen-Yu Fan^a, Andrea Appolloni^b, Hao Wang^a

^a *School of Management Science and Engineering, Nanjing University of Information Science & Technology, Nanjing, China.*

^b *Department of Management and Law, University of Rome Tor Vergata, Rome, Italy.*

Abstract

Rapid urban advancement has exacerbated air-pollution risks and calls for evidence-based refined urban environmental governance. Based on panel observational data from 33 cities in China's Yangtze River Economic Belt (YREB) spanning 2014-2022, this study establishes a three-tier sequential descriptive-evidence-chain framework combining XGBoost-SHAP interpretable machine-learning, Moran's I spatial-autocorrelation diagnosis, and SAR spatial-econometric estimation. Moving beyond simple technical combination of existing tools, this workflow enables systematic cross-comparison between variable predictive performance and spatial-statistical patterns. Analytical results show that: (1) PM_{2.5}, agricultural fertilizer application, electricity consumption, and NO₂ possess high predictive importance for AQI variations; (2) Urban pollution sources exhibit significant spatial dependence, particularly for PM_{2.5} and electricity consumption; (3) Spatial-effect decomposition identifies strong positive local statistical associations of PM_{2.5} and SO₂ concentrations with urban AQI, while their indirect spatial-statistical associations are counter-intuitively negative. Agricultural-fertilizer application presents offsetting negative-direct and positive-indirect spatial-statistical associations, whereas electricity consumption, NO₂, and natural gas show negligible overall spatial-statistical associations. This work offers empirical insights to support differentiated and coordinated air-pollution governance within the YREB.

Keywords: Air pollution; Machine learning; Extreme Gradient Boosting; Spatial econometrics; Yangtze River Economic Belt



1 Introduction

Rapid urbanization and industrialization in China have led to severe air pollution, posing significant risks to public health and sustainable development. The YREB, home to over 40% of China's population and accounting for more than 40% of its GDP, contributes a disproportionate share of national emissions. Recent data indicate that more than 67% of cities in the region exceed the national annual PM_{2.5} standard ($\leq 35 \mu\text{g}/\text{m}^3$), with frequent haze episodes in major urban centers such as Wuhan, Nanjing, and Changsha. A regional smog event in early 2021 triggered red alerts across the Yangtze River Delta, underscoring persistent challenges in source identification and transboundary governance.

To support effective mitigation strategies, there is an urgent need for analytical frameworks that accurately identify pollution sources, capture spatial heterogeneity, and inform coordinated policy responses. Machine learning models—particularly XGBoost and random forest—have demonstrated high accuracy in air quality prediction and variable importance ranking (Zhang et al., 2021; Satish et al., 2025). When combined with explainable AI (XAI) techniques like SHAP, these models can enhance interpretability, providing insights into feature contributions and interactions (He et al., 2026; Song et al., 2025). Such approaches offer advantages over conventional linear regression models in capturing complex nonlinear relationships and higher-order interactions in environmental data.

Spatial dependence and non-stationarity are critical yet frequently overlooked dimensions in pollution analysis. While geographically weighted regression (GWR) allows for localized parameter estimation (Brunsdon et al., 1996), it is not specifically designed to estimate endogenous spatial spillover effects across neighboring region. Moreover, many deep learning models remain “black-box” systems, hindering their adoption in policy design where transparency is essential (Rudin, 2019). Although recent studies have begun integrating XAI with spatial modeling (Yan and Yang, 2025; Sen et al., 2025), few studies systematically link model interpretability with spatial econometric analysis to examine whether variables with high predictive importance also exhibit significant direct and indirect spatial associations with air pollution.

Three key limitations persist in current research: (1) Limited interpretability of data-driven models restricts their utility for evidence-based decision-making; (2) Neglect of spatial heterogeneity reduces the regional relevance of identified drivers; and (3) inadequate assessment of spatial dependence and cross-regional spillovers limits our understanding of pollution diffusion mechanisms over time.

To address these gaps, this study proposes an integrated framework combining XGBoost-SHAP for interpretable feature selection and spatial autoregressive (SAR) models with k-nearest-neighbor (KNN) spatial weight matrices to quantify spatial spillovers. Using multi-year panel data from 33 cities in YREB (2014–2022), we identify key predictors of AQI and assess their predictive importance, with

particular emphasis on PM_{2.5}, energy use, and agricultural inputs. Our approach enables the joint assessment of local spatial associations and cross-regional spillover patterns, advancing beyond static or spatially isolated analyses.

The main contributions are threefold: First, we develop a three-stage sequential analytical framework that integrates XGBoost-SHAP predictive-importance screening, Moran's I spatial-clustering diagnosis, and SAR-based spatial-effect decomposition. Moving beyond a simple hybrid combination of separate tools, this framework enables systematic comparison between variable-level predictive importance and spatial-statistical patterns, thereby improving the interpretability and policy relevance of the empirical analysis. Second, we identify significant spatial dependence alongside counter-intuitive negative indirect spatial associations of PM_{2.5} and SO₂ with AQI, despite their strong positive direct associations. This empirical pattern highlights the potential importance of strengthened inter-regional coordination in air-pollution governance within the YREB, while the underlying mechanisms behind such counter-intuitive indirect effects remain open for further investigation. Third, we find that agricultural-fertilizer application exhibits considerable predictive importance for AQI from a machine-learning perspective, yet its direct and indirect spatial associations largely offset each other, resulting in a negligible overall spatial association with AQI. This empirical discrepancy between high predictive importance and a near-zero overall spatial association represents an interesting finding that merits further exploration in future research.

The structure of the remaining sections of this paper is organized as follows. Section 2 reviews the relevant literature and identifies research gaps in integrating explainable AI with spatial econometrics. Section 3 outlines the study area, data sources, pre-processing procedures and research methodology. Section 4 presents XGBoost-SHAP-derived variable predictive-importance results. Section 5 analyzes spatial-clustering patterns of key variables using global and local Moran's I statistics. Section 6 estimates the SAR panel model, decomposes spatial spillover effects, and implements a series of robustness checks. Section 7 concludes this study and proposes corresponding policy recommendations.

2 Literature Review

2.1 Research Methodologies in Air Pollution Studies

The methodological framework for air pollution research has evolved considerably, reflected not only in advances in analytical tools but also in the deepening and expansion of analytical approaches. Previous studies often relied separately on spatial statistics or machine learning. In the spatial analysis literature, researchers have employed models such as spatial autocorrelation analysis and spatial econometric models to examine regional coordination and spillover effects in environmental governance (Xu et al., 2024; Yang et al., 2022). In the machine learning literature, some studies have

utilized algorithms such as XGBoost and ensemble frameworks integrated with SHAP to predict air quality and enhance model interpretability (Gürtepe et al., 2025; Wang et al., 2022). Notably, Chang et al. (2024) combined spatial analysis with machine learning to classify urban pollution patterns across 285 cities, identifying distinct driving mechanisms—“urban carrying capacity driven,” “natural condition driven,” and “intermediate type”—and revealing persistent regional disparities despite an overall decline in pollutant levels.

In summary, research methodologies in air pollution studies have evolved from isolated applications of spatial statistics or machine learning toward integrated, multi-method frameworks. While these advances have improved modeling accuracy, predictive performance and interpretability, challenges remain in fully addressing spatial heterogeneity and the “black box” nature of machine learning models. Building on existing efforts, this paper proposes an integrated framework combining spatial statistics, XGBoost, and SHAP values—aiming to improve both the identification and interpretation of key predictors of air pollution, thereby providing more actionable evidence for urban air pollution control.

2.2 Analysis of the Causes of Air Pollution

Air pollution is influenced by a wide range of complex factors. Numerous studies have been conducted to examine the factors associated with emissions and concentrations of air pollutants—such as SO₂, NO₂, and PM_{2.5}—as well as changes in the AQI (Ren and Matsumoto, 2020). Regarding socioeconomic factors, population (Wang et al., 2016; Zhao et al., 2018), economic conditions (Wang et al., 2016), secondary industry activity (Zeng et al., 2019; Han et al., 2019), and urbanization (Han et al., 2014) have been widely identified as important determinant. Extending this line of research to the joint governance of air pollution and carbon emissions, Liu et al. (2024) further investigated the synergy between pollution and carbon reduction, identifying environmental regulation as a key driver in promoting this synergy. Annual precipitation, elevation, normalized difference vegetation index and the proportion of the secondary industry were also found to positively influence this synergistic effect. Recent high-resolution modeling of Beijing’s atmosphere indicates that fossil fuel combustion—particularly from residential heating and on-road diesel vehicles—serves as a major co-source of CO₂ and primary PM_{2.5}, with emissions strongly modulated by boundary layer dynamics and seasonal energy demand (Liang et al., 2025). Hence, the aforesaid literature demonstrates that air pollution is shaped by interacting socioeconomic, environmental, energy-related, and policy factors, highlighting the need for analytical approaches capable of capturing such multidimensional relationships.

Overall, research on the causes of air pollution has become increasingly systematic, covering multiple dimensions such as socioeconomic factors, natural conditions, and policy interventions. However, most existing studies focus on individual factors and lack in-depth exploration of their

interactive effects. This study integrates multi-source data and applies interpretable machine learning methods to further reveal the nonlinear association between multiple predictors and air quality and their relative predictive importance.

2.3 Environmental Policy Research

Environmental policy evaluation has advanced toward integrated and adaptive approaches. While environmental regulation may initially constrain firm performance, it can foster long-term competitiveness through technological innovation (Zhu & Xu, 2022). Policy effectiveness may also depend on the coordinated use of regulatory, market-based, and multi-level governance instruments. In large, heterogeneous regions like the Yangtze River Economic Belt, uniform assessment approaches may fail to adequately capture substantial regional disparities. For instance, Li et al. (2025) developed a differentiated green development assessment integrating interpretable machine learning, identifying rural income, ecological conservation, and industrial structure as key drivers. This data-driven framework supports targeted interventions. Moreover, strategic assessment mechanisms—moving beyond command-and-control toward cross-sectoral coordination and economic incentives—can unlock policy synergies. For instance, Hou et al. (2024) demonstrate significant pollution-carbon co-reduction potential in agriculture, underscoring the role of regional coordination and structural drivers.

In summary, while methodological advances have enhanced the scientific rigor of environmental policy assessment, most studies remain focused on single policy instruments or sectoral interventions, with limited attention to spatially mediated interactions and cross-domain synergies. To address this research gap, this study incorporates policy-related variables into an integrated analytical framework that jointly considers predictive importance, spatial heterogeneity, and cross-regional spatial association.

The aforesaid literature indicates that air pollution research has increasingly adopted integrated, data-driven methodologies. As summarized in Table 1, existing studies have incorporated different combinations of spatial heterogeneity, multi-source data, policy-related factors, machine learning, and model interpretability. However, few studies integrate these dimensions within a unified sequential framework. In particular, machine-learning and spatial-econometric analyses are often conducted separately, limiting systematic comparison between the predictive importance of individual variables and their spatial-statistical relationships. To address this gap, this study develops a three-stage sequential analytical framework comprising XGBoost–SHAP predictive-importance screening, global and local Moran’s *I* spatial-clustering diagnosis, and SAR-based spatial-effect decomposition. Rather than simply combining independent analytical tools, the proposed framework connects these three stages to enable systematic comparison between variable-level predictive importance and spatial-statistical patterns. In doing so, it enhances the interpretability and policy relevance of air quality

analysis and provides a more coherent empirical basis for coordinated urban environmental governance.

Table 1. Comparison between relevant literature and this paper

Literature	Spatial Analysis	Machine Learning	Policy Synergy Analysis	Multisource Data Integration	Model Interpretability
Liang et al.(2025)	✓			✓	
Xu et al. (2024)	✓		✓	✓	
Yang et al. (2022)	✓			✓	
Gürtepe et al. (2025)	✓	✓		✓	
Chang et al. (2024)		✓	✓	✓	
Ren et al. (2020)	✓		✓	✓	
Zhu et al. (2022)			✓	✓	
Liu et al. (2024)		✓	✓	✓	
Hou et al.(2024)	✓		✓	✓	✓
This Paper	✓	✓	✓	✓	✓

3. Research Area and Methodology

3.1 Research Area Overview

The Yangtze River Economic Belt (YREB) is one of China's major regional development initiatives, extending across the Yangtze River Basin and connecting the eastern, central, and western regions of the country. It consists of 11 provinces and municipalities, including Shanghai, Jiangsu, Zhejiang, and Anhui, and extends approximately from 99°E to 123°E and from 24°N to 33°N. Covering approximately 2.05 million km², the YREB accounts for about 21.4% of China's total land area. The region is characterized by a west-to-east decline in elevation and plays a strategically important role in China's socioeconomic development, ecological conservation, and regional innovation (Yang et al., 2024).

As the “golden corridor” of China's economic development, the YREB is home to more than 40% of the country's population and contributes approximately 46% of its GDP. Its diverse climatic conditions, complex topography, rapid urbanization, and intensive economic activity have contributed to substantial air pollution pressures across many cities. Against this background, this study examines 33 representative cities along the YREB, including Shanghai, Hangzhou, and Panzhihua, to identify key factors associated with air quality and examine their spatial patterns and cross-regional associations, thereby providing empirical evidence for coordinated regional environmental governance.

3.2 Data Sources

3.2.1 Air Quality Data

This study uses annual data on six major air pollutants (PM_{2.5}, PM₁₀, O₃, NO₂, SO₂, and CO) from 33 key cities in the YREB, including Shanghai, Hangzhou, and Nanjing, from 2014 to 2022. The pollutant concentration data are processed in accordance with the *Technical Regulation on Ambient Air Quality Index (on Trial)* (HJ 633-2012), and the corresponding air quality index (AQI) is calculated for each city. As a composite indicator of ambient air quality, AQI captures overall pollution conditions involving multiple pollutants and is therefore suitable for assessing air quality across cities in the YREB. In accordance with the *Ambient Air Quality Standards* (GB 3095-2012), AQI is evaluated using nationally standardized criteria, ensuring consistency and comparability in air quality assessments across cities in the study area.

3.2.2 Explanatory Variables

This research considers a broad set of factors associated with air quality, encompassing environmental conditions, pollution-related indicators, socioeconomic characteristics, resource consumption, and industrial activities. Environmental variables include temperature, precipitation, urban green coverage area, and industrial wastewater discharge, with data primarily obtained from provincial and municipal departments of ecology and environment. The data cover the period from 2014 to 2022, with temperature and precipitation measured as annual averages (Shanghai Municipal Ecology and Environment Bureau, 2018; Jiangsu Provincial Department of Ecology and Environment, 2019).

Social and livelihood factors include resident population and private car ownership. Indicators of economic development and resource use include the value added of the secondary industry, regional GDP, completed floor area of buildings, and agricultural fertilizer use. Energy- and industry-related indicators include electricity consumption and natural gas/liquefied petroleum gas consumption. Data for these variables are collected from provincial and municipal statistical yearbooks, national economic and social development statistical bulletins, and ecological and environmental reports (Shanghai Municipal Statistics Bureau, 2019; Jiangsu Provincial Bureau of Statistics, 2020).

3.3 Data preprocessing notes

For numeric variables, missing values were imputed using the median value of the corresponding variable. Potential outliers were identified using the interquartile range (IQR) rule and winsorized at the corresponding lower and upper thresholds to reduce the influence of extreme observations. Categorical variables were converted into dummy variables applying the `dummyVars` function in the `caret` package, and any duplicated column names generated during dummy encoding were renamed to ensure uniqueness. Numerical predictors were subsequently standardized using Z-score normalization.

For the machine-learning analysis, the dataset was divided into a training set (80%) and a test set (20%). To avoid information leakage, preprocessing parameters, including the medians used for missing-value imputation, outlier thresholds, and parameters for Z-score standardization, were estimated from the training set and then applied to the test set. Although tree-based XGBoost models are generally insensitive to feature scaling, Z-score standardization was retained to ensure a consistent preprocessing procedure across the different analytical stages and robustness checks.

3.4 Research Approach

This study adopts a structured sequential research framework consisting of five stages: **Data Acquisition** → **Machine Learning** → **Spatial Autocorrelation Analysis** → **Spatial Spillover Effect Analysis** → **Policy Recommendations**. These stages are systematically connected to provide a comprehensive assessment of air quality across 33 selected cities in the YREB. Specifically, machine learning is first employed to identify variables with high predictive importance for AQI, followed by spatial autocorrelation analysis to characterize their spatial clustering patterns. Spatial econometric analysis is then used to examine direct and indirect spatial associations, providing a basis for developing coordinated regional policy implications. The overall research framework is illustrated in Fig. 1.

4. Analysis of Predictive Factors Using XGBoost-SHAP

4.1 XGBoost Model Development and Evaluation

To characterize the potentially nonlinear relationships between AQI and its predictors, an XGBoost regression model was developed. The dataset was divided into a training set (80%) and a test set (20%). Model hyperparameters were tuned exclusively on the training set using grid search with five-fold cross-validation. The optimal hyperparameter combination was then used to train the final XGBoost model on the full training set. Early stopping was incorporated during model training to reduce the risk of overfitting.

The predictive performance of the final model was evaluated on the held-out test set using the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE). As reported in Table 2, the model achieved an R^2 of 0.754, an RMSE of 6.119, and an MAE of 4.541, indicating that it captured a substantial proportion of the variation in AQI and achieved satisfactory out-of-sample predictive performance. Additional diagnostic analyses showed no pronounced systematic patterns in the prediction residuals (Figs. 2–3), providing further evidence of stable model performance. Based on this predictive performance, the fitted XGBoost model was subsequently interpreted using SHAP to quantify the relative predictive importance of individual variables and characterize their contributions to model predictions.

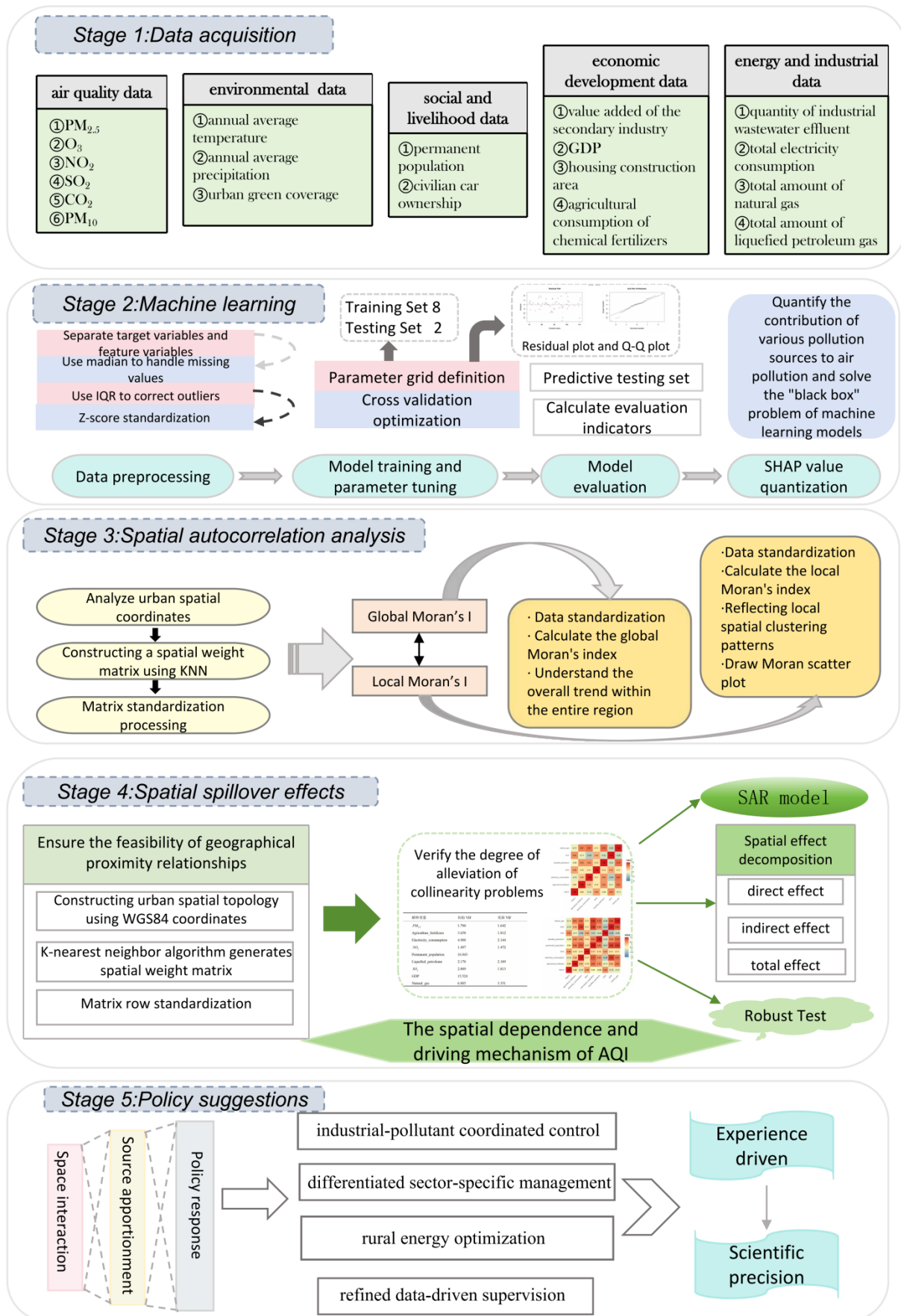


Fig. 1 Technical Roadmap for Urban Air Pollution Source Analysis

Table 2. 5-fold Cross-Validation Evaluation Metrics Results

R^2	RMSE	MAE
0.7540858	6.119329	4.541427

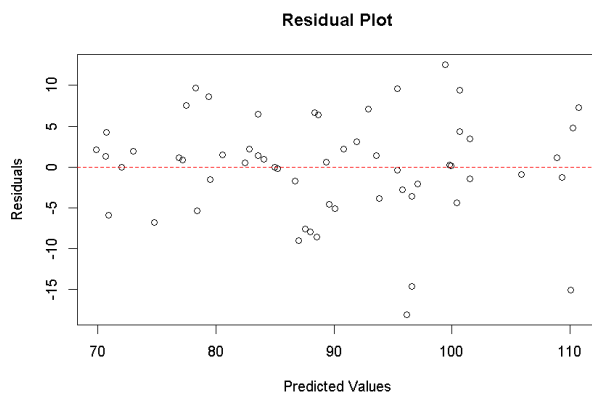


Fig. 2. Residual Plot

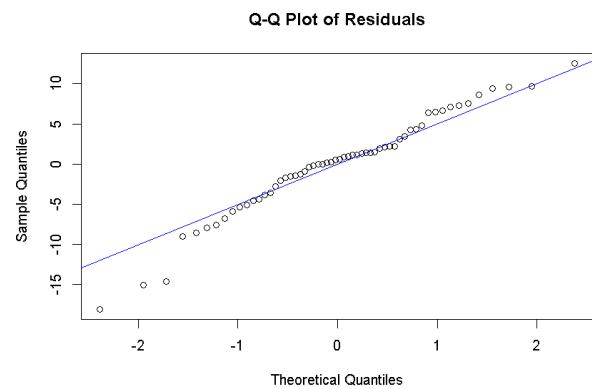


Fig. 3. Q-Q Plot

As shown in Fig. 2, the prediction residuals are generally distributed around zero without a pronounced systematic pattern across the range of predicted AQI values, although several relatively large residuals are observed. The Q–Q plot in Fig. 3 shows that most residuals follow the reference line reasonably closely, with some deviations in the tails. Overall, these diagnostic results suggest no substantial systematic prediction bias, complementing the out-of-sample performance metrics reported in Table 2.

4.2 SHAP-Based Analysis of Predictive Importance

To assess the relative contribution of each predictor to the XGBoost model, we calculate the mean absolute SHAP value for each variable across the 33 YREB cities. As shown in Fig. 4, a larger mean absolute SHAP value indicates a greater contribution of a variable to the model’s AQI predictions. Because absolute SHAP values capture the magnitude rather than the direction of a variable’s contribution, the ranking in Fig. 4 should be interpreted as predictive importance rather than as evidence of causal effects.

The results reveal substantial variation in predictive importance across variables. $PM_{2.5}$ is the most important predictor of AQI, with a mean absolute SHAP value of 5.14, substantially exceeding those of the other variables. This result is expected because $PM_{2.5}$ concentration is one of the pollutant components directly incorporated into AQI assessment. Agricultural fertilizer use ranks second, with a mean absolute SHAP value of 2.85. Although this finding does not by itself establish a causal relationship with AQI, existing research provides a plausible environmental mechanism: fertilizer application can contribute to ammonia (NH_3) emissions, which may facilitate the formation of secondary particulate matter. The relatively high predictive importance of fertilizer use therefore suggests that

agricultural activities may contain information relevant to variations in regional air quality and warrants further spatial investigation.

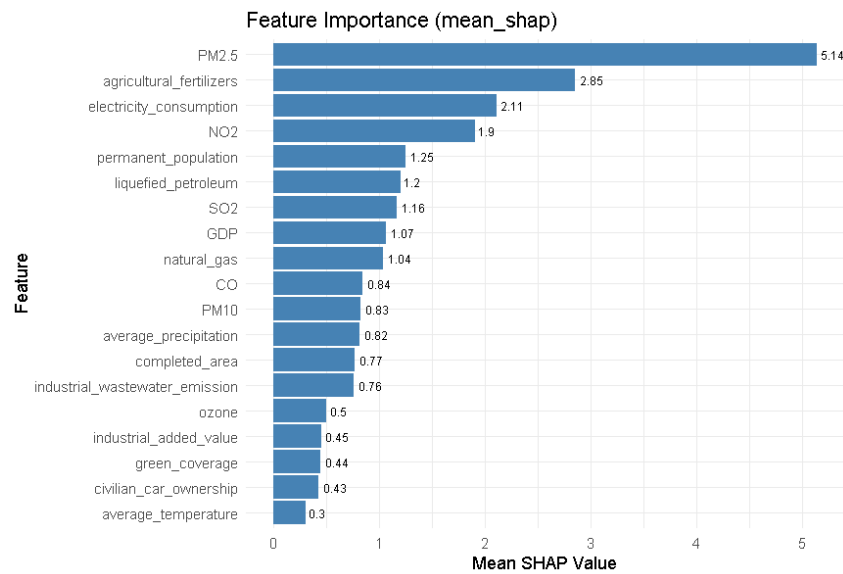


Fig. 4. Feature Importance Ranking Plot Based on SHAP Absolute Values

Electricity consumption is the third-ranked predictor, with a mean absolute SHAP value of 2.11. Its relatively high ranking may reflect its association with energy demand and industrial and economic activity, which can be accompanied by pollutant emissions depending on the regional energy mix and production structure. NO₂ ranks fourth, with a mean absolute SHAP value of 1.90. As an important air pollutant closely associated with combustion processes, road traffic, and industrial activities, NO₂ provides substantial predictive information for variations in AQI.

The remaining variables exhibit relatively lower predictive importance. Resident population, liquefied petroleum gas consumption, SO₂, GDP, and natural gas consumption have mean absolute SHAP values ranging from 1.04 to 1.25, indicating moderate contributions to the model's predictions. In contrast, variables such as O₃, secondary-industry value added, urban green coverage, private car ownership, and average temperature show comparatively low mean absolute SHAP values in the fitted model. Importantly, lower SHAP importance does not imply that these variables are environmentally unimportant; rather, it indicates that they contribute less to the predictive performance of the XGBoost model given the other variables included in the analysis. These differences between predictive importance and potential spatial relationships motivate the subsequent spatial autocorrelation and spatial econometric analyses.

4.3 Distributional Analysis of SHAP Values

To further examine the heterogeneity of predictor contributions across observations, Fig. 5 presents boxplots of the SHAP values for all variables included in the XGBoost model. Unlike the mean absolute SHAP values reported in Fig. 4, which summarize the overall predictive importance of each variable, the distributions in Fig. 5 provide additional information on the direction, magnitude, and variability of individual variables' contributions to model predictions across observations.

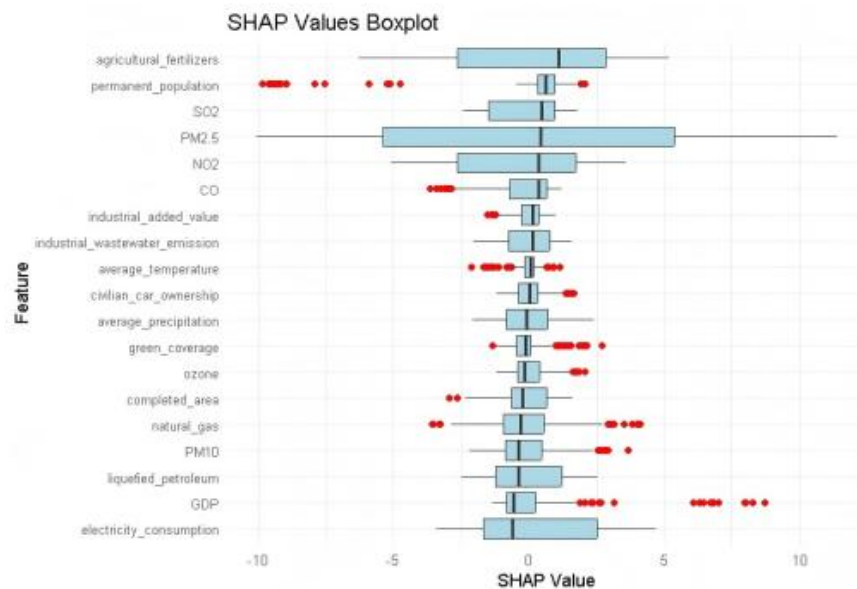


Fig. 5. Distribution of SHAP Values across Predictors

As shown in Fig. 5, PM_{2.5}, agricultural fertilizer use, and electricity consumption exhibit relatively wide SHAP-value distributions, consistent with their high predictive importance identified in Fig. 4. In particular, PM_{2.5} shows substantial variation in its SHAP contributions across observations, indicating considerable heterogeneity in its contribution to predicted AQI. Agricultural fertilizer use and electricity consumption also display relatively dispersed SHAP values, suggesting that their contributions to AQI predictions vary across cities and observations. By contrast, several variables, including temperature, precipitation, and urban green coverage, exhibit SHAP values concentrated more closely around zero, indicating comparatively smaller contributions to model predictions for most observations. Several predictors also display asymmetric distributions and extreme SHAP values, suggesting that their predictive contributions may vary substantially under different observed conditions.

It is important to emphasize that SHAP values quantify how individual predictors contribute to the predictions generated by the fitted XGBoost model; they should not be interpreted as causal effects or direct pollution-source contributions. Moreover, a positive or negative SHAP value indicates whether a particular observation of a predictor increases or decreases the model prediction relative to its baseline prediction, rather than establishing a universally positive or negative relationship between that predictor

and AQI. Accordingly, the SHAP results provide evidence on predictive importance and heterogeneous predictive contributions, but do not by themselves establish spatial dependence or cross-city spillover relationships.

Overall, the XGBoost–SHAP analysis identifies substantial differences in both the overall predictive importance and observation-level contributions of the variables associated with AQI. However, predictive importance does not necessarily correspond to spatial clustering or spatial-statistical associations. This distinction motivates the subsequent spatial analyses: Section 4 examines the spatial clustering patterns of key variables using global and local Moran’s I, while Section 5 further evaluates their direct and indirect spatial associations using the SAR model. The sequential analysis therefore allows the predictive patterns identified through machine learning to be systematically compared with the spatial patterns revealed by spatial statistical and econometric methods.

5 Spatial Autocorrelation Analysis

To examine the spatial dependence and clustering patterns of AQI and its key predictors, this study conducts global and local spatial autocorrelation analyses using Global Moran’s I and Local Indicators of Spatial Association (LISA), respectively. Global Moran’s I is used to assess whether the spatial distribution of each variable exhibits significant overall clustering, whereas LISA statistics identify localized clusters and spatial outliers. Statistical significance is evaluated using the corresponding standardized test statistics. These analyses provide evidence on the spatial structure and heterogeneity of air quality and related variables across YREB cities and establish an empirical basis for the subsequent spatial econometric analysis.

5.1 Selection of Spatial Weight Matrix

The specification of the spatial weight matrix is critical because it defines the spatial relationships among cities and therefore affects the measurement of spatial dependence. For the YREB, the spatial weight matrix should account for its large geographical extent and highly uneven distribution of cities. The region spans eastern, central, and western China, with relatively dense urban clusters in the east and more dispersed cities in the central and western areas. Such spatial heterogeneity makes a purely contiguity-based specification less suitable, particularly because nonadjacent cities may still belong to the same broader regional system.

Several alternative spatial weight matrices can be considered. A contiguity-based matrix defines spatial relationships according to shared geographical boundaries, but it may produce an uneven connectivity structure when the observational units are irregularly distributed. An inverse-distance matrix incorporates geographical proximity and distance decay but requires the specification of distance functions or thresholds and may assign very small weights to relatively distant observations (Anselin,

1995). By comparison, a KNN matrix defines the spatial neighborhood of each city according to its K geographically nearest cities, ensuring that every city is connected to a fixed number of neighboring observations.

The KNN specification is particularly suitable for the present study for two reasons. First, it accommodates the uneven spatial distribution of 33 cities in YREB while avoiding isolated observations in relatively sparse areas. Second, it provides a consistent neighborhood structure across cities, facilitating comparison of spatial dependence throughout the study area. Accordingly, this study employs a KNN spatial weight matrix as the baseline specification to characterize spatial relationships among the selected YREB cities. To assess whether the empirical results are sensitive to the specification of spatial connectivity, alternative spatial weight matrices are further considered in the robustness analysis.

5.2 Global Moran's I Index

The Global Moran's I index is used to assess the overall spatial autocorrelation of a variable across this study area, with values ranging from -1 to 1. Moran's $I > 0$ indicates that cities with similar values tend to be spatially clustered, whereas a negative value indicates spatial dispersion, with neighboring cities tending to exhibit dissimilar values. A Moran's I close to zero suggests weak or no detectable global spatial autocorrelation. Statistical significance is assessed to determine whether the observed spatial pattern differs significantly from that expected under the null hypothesis of spatial randomness. The Global Moran's I statistic is calculated as follows:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

where n denotes the number of spatial units, w_{ij} represents the elements of the spatial weight matrix, x_i denotes the observed value, $\bar{x}$ is the mean of the observed values, $S_0 = \sum_i \sum_j w_{ij}$ is the sum of all elements in the spatial weight matrix.

Based on Eq. (1), annual Global Moran's I statistics are calculated for PM_{2.5}, agricultural fertilizer use, NO₂, and electricity consumption from 2014 to 2022 using the KNN spatial weight matrix specified in Section 4.1. These variables are selected because they exhibit relatively high predictive importance in the preceding XGBoost–SHAP analysis. Statistical significance is evaluated at the 5% level. The resulting Moran's I statistics and corresponding p-values are shown in Table 3.

As shown in Table 3, the four variables exhibit markedly different spatial patterns over the study period. PM_{2.5} displays consistently positive Moran's I values and statistically significant spatial autocorrelation in most years, indicating persistent spatial clustering across YREB cities. Electricity consumption exhibits an even more stable pattern, with positive and statistically significant Moran's I values from 2014 to 2022, suggesting persistent clustering of cities with similar levels of electricity consumption. In contrast, agricultural fertilizer use shows small negative Moran's I values that are statistically insignificant in all years, providing little evidence of systematic global spatial clustering. NO₂ exhibits greater temporal variation: its Moran's I remains negative throughout the study period, but statistically significant negative spatial autocorrelation is observed only in selected years. Overall, these results reveal substantial heterogeneity in the spatial structures of the variables identified as important predictors by the XGBoost–SHAP model.

Table 3. Global Moran's I Index Values and P-values for selected variables

year	PM _{2.5}		agricultural_fertilizers		NO ₂		electricity_consumption	
	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value	Moran's I	p-value
2014	0.1838	0.035	-0.0760	0.284	-0.0241	0.398	0.2540	0.003
2015	0.2193	0.009	-0.0959	0.117	-0.0424	0.488	0.2409	0.001
2016	0.1735	0.016	-0.0801	0.254	-0.0798	0.318	0.2281	0.007
2017	0.1194	0.059	-0.0737	0.233	-0.1141	0.137	0.2143	0.008
2018	0.1838	0.020	-0.0698	0.246	-0.1329	0.080	0.2118	0.009
2019	0.2324	0.010	-0.0880	0.161	-0.1557	0.029	0.2053	0.009
2020	0.2601	0.006	-0.0886	0.141	-0.0937	0.214	0.1855	0.024
2021	0.2912	0.007	-0.0809	0.156	-0.0784	0.297	0.2185	0.009
2022	0.2705	0.003	-0.0793	0.144	0.0183	0.223	0.2121	0.013

Importantly, the spatial autocorrelation results do not necessarily correspond to the predictive-importance rankings obtained from SHAP. For example, agricultural fertilizer use ranks highly in predictive importance but exhibits no statistically significant global spatial autocorrelation during the study period, whereas electricity consumption displays both relatively high predictive importance and persistent positive spatial clustering. This contrast further illustrates that predictive importance and spatial dependence capture distinct dimensions of the data. The following subsections therefore examine the temporal and spatial characteristics of these key variables in greater detail.

5.2.1 Persistent Positive Spatial Autocorrelation of PM_{2.5}

PM_{2.5} shows positive Global Moran's I values from 2014 to 2022, with statistically significant spatial autocorrelation in most years (Table 3). The magnitude of Moran's I varies over time rather than following a strictly monotonic trend, but the consistently positive values indicate a persistent tendency for cities with similar PM_{2.5} levels to be spatially clustered. Such clustering may reflect a combination of shared regional emission structures, similar socioeconomic and energy-use patterns, and atmospheric transport across neighboring cities. These results highlight the regional nature of PM_{2.5} pollution and

suggest that city-level mitigation measures may benefit from coordination across neighboring jurisdictions.

5.2.2 Weak Global Spatial Autocorrelation of Agricultural Fertilizer Application

In contrast to PM_{2.5}, agricultural fertilizer use exhibits small negative Moran's I values throughout the study period, none of which are statistically significant at the 5% level. The results therefore provide little evidence of systematic global spatial autocorrelation in fertilizer use across the 33 YREB cities. This weak spatial structure may be related to substantial cross-city heterogeneity in agricultural intensity, cropping patterns, land-use conditions, and agricultural production practices. Notably, this result contrasts with the high predictive importance of fertilizer use identified by the XGBoost–SHAP analysis. The discrepancy suggests that a variable can contribute substantially to AQI prediction without necessarily exhibiting a strong global spatial clustering pattern, further demonstrating that predictive importance and spatial dependence capture different dimensions of the data.

5.2.3 Temporally Varying Negative Spatial Autocorrelation of NO₂

NO₂ shows negative Global Moran's I values throughout 2014–2022, although the spatial autocorrelation is statistically insignificant in most years. A statistically significant negative Moran's I is observed in 2019 (Moran's I = -0.156 , $p = 0.029$), indicating that neighboring cities with dissimilar NO₂ levels were more prevalent in that year than would be expected under spatial randomness. The predominance of statistically insignificant values in the remaining years suggests that the global spatial structure of NO₂ is relatively weak and temporally unstable. Such temporal variation may be associated with differences in traffic intensity, industrial activity, emission-control measures, and other city-specific conditions. However, the Global Moran's I results alone do not identify the mechanisms underlying the temporary negative spatial association, which requires further investigation using local spatial patterns and additional contextual evidence.

5.2.4 Persistent Positive Spatial Autocorrelation of Electricity Consumption

Electricity consumption displays consistently positive and statistically significant Global Moran's I values throughout 2014–2022, indicating persistent spatial clustering among YREB cities. Although the magnitude of Moran's I varies moderately over time, the overall pattern remains stable, suggesting that cities with similar levels of electricity consumption tend to be geographically clustered. This spatial structure may reflect regional similarities in economic development, industrial structure, population concentration, and energy demand. Together with its relatively high SHAP-based predictive importance, the persistent spatial clustering of electricity consumption indicates that energy-use patterns constitute an important dimension for understanding the spatial heterogeneity of air quality-related factors across the YREB.

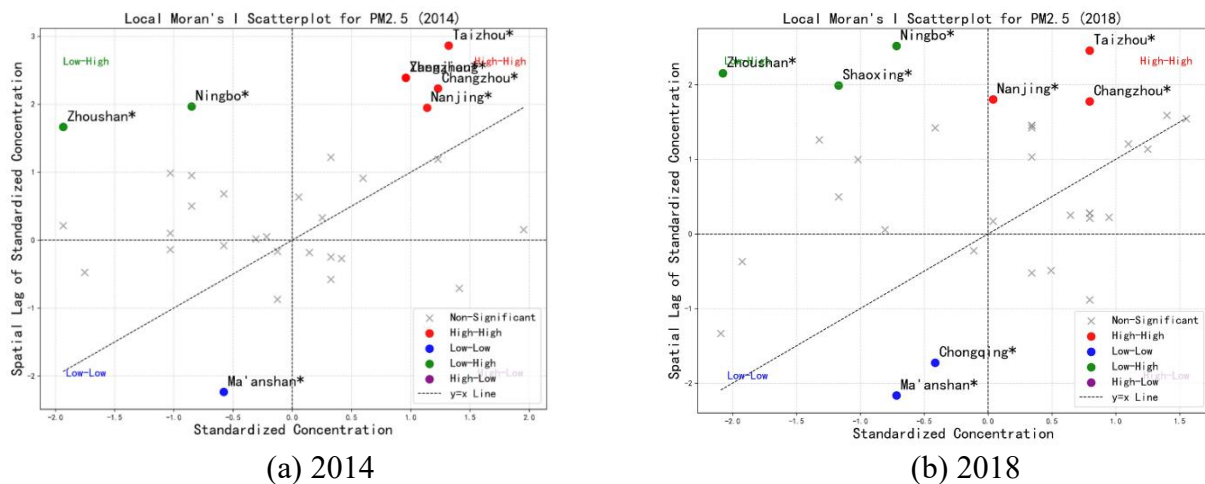
5.3 Local Moran's I Index

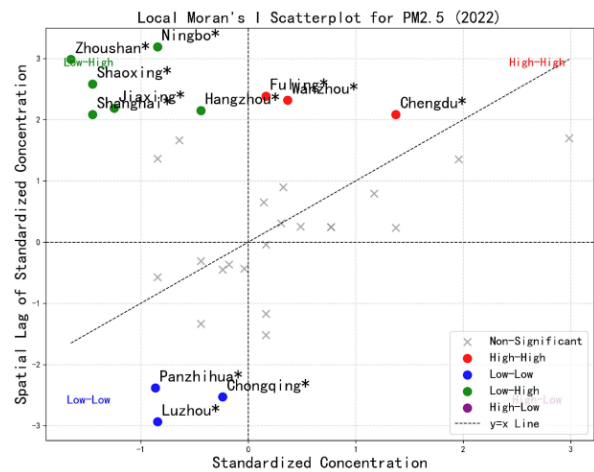
In order to identify the spatial clustering patterns or outliers of specific geographic units and their neighborhoods, we further calculate the Local Moran's I index to detect the local spatial autocorrelation characteristics of the spatial data. The formula for calculating the Local Moran's I index is as follows:

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_{j=1}^n w_{ij} (x_j - \bar{x}) \quad (2)$$

In Eq. (2), x_i represents the observation value of a geographic unit, $\bar{x}$ represents the mean, s^2 represents the variance, and w_{ij} represents the elements of the spatial weight matrix. The results can be divided into four categories: High-High, Low-Low, High-Low, and Low-High. The LISA clustering map visualizes spatial heterogeneity. In pollution analysis, High-High clusters may indicate the core areas of pollution spread, while High-Low outliers reflect isolated pollution sources. This index complements the global Moran's I and is suitable for formulating refined spatial governance strategies.

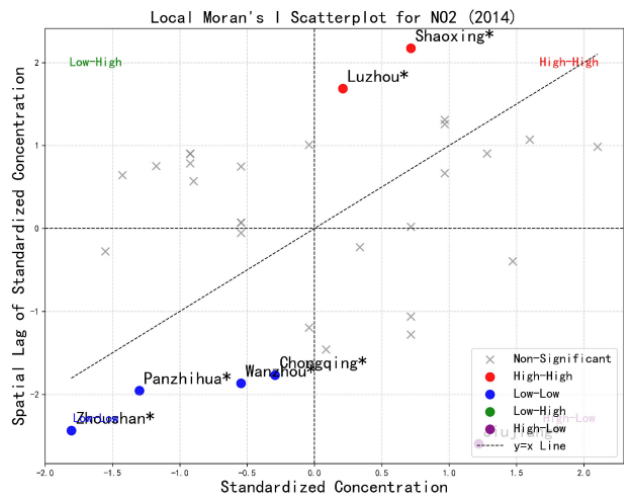
Based on Eq. (2), using the KNN adjacency matrix, 2014, 2018, and 2022 were selected as change points at four-year intervals. Local Moran scatter plots for PM_{2.5}, NO₂, agricultural fertilizer application, and social electricity consumption were created as shown in Figs. 6 to 9.



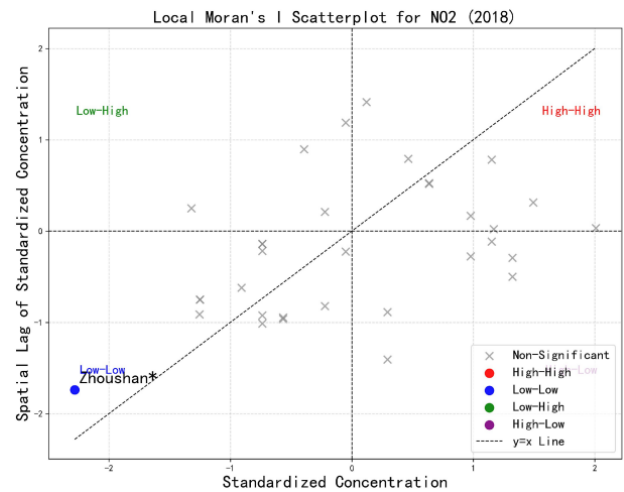


(c) 2022

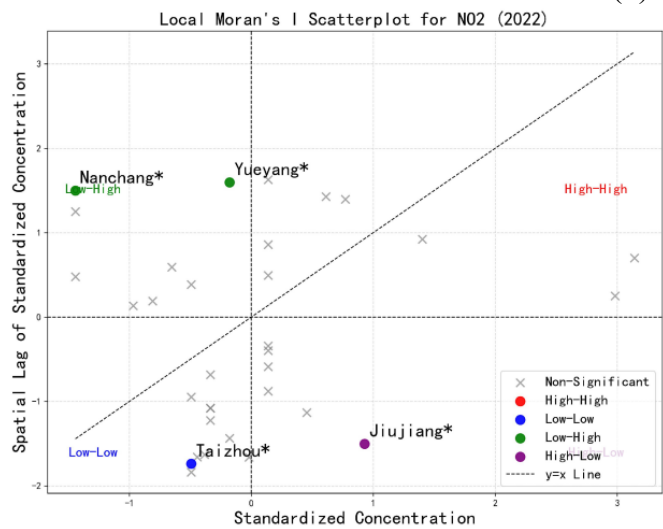
Fig. 6. PM_{2.5} Levels in 2014, 2018 and 2022



(a) 2014



(b) 2018



(c) 2022

Fig. 7. NO₂ Levels in 2014, 2018 and 2022

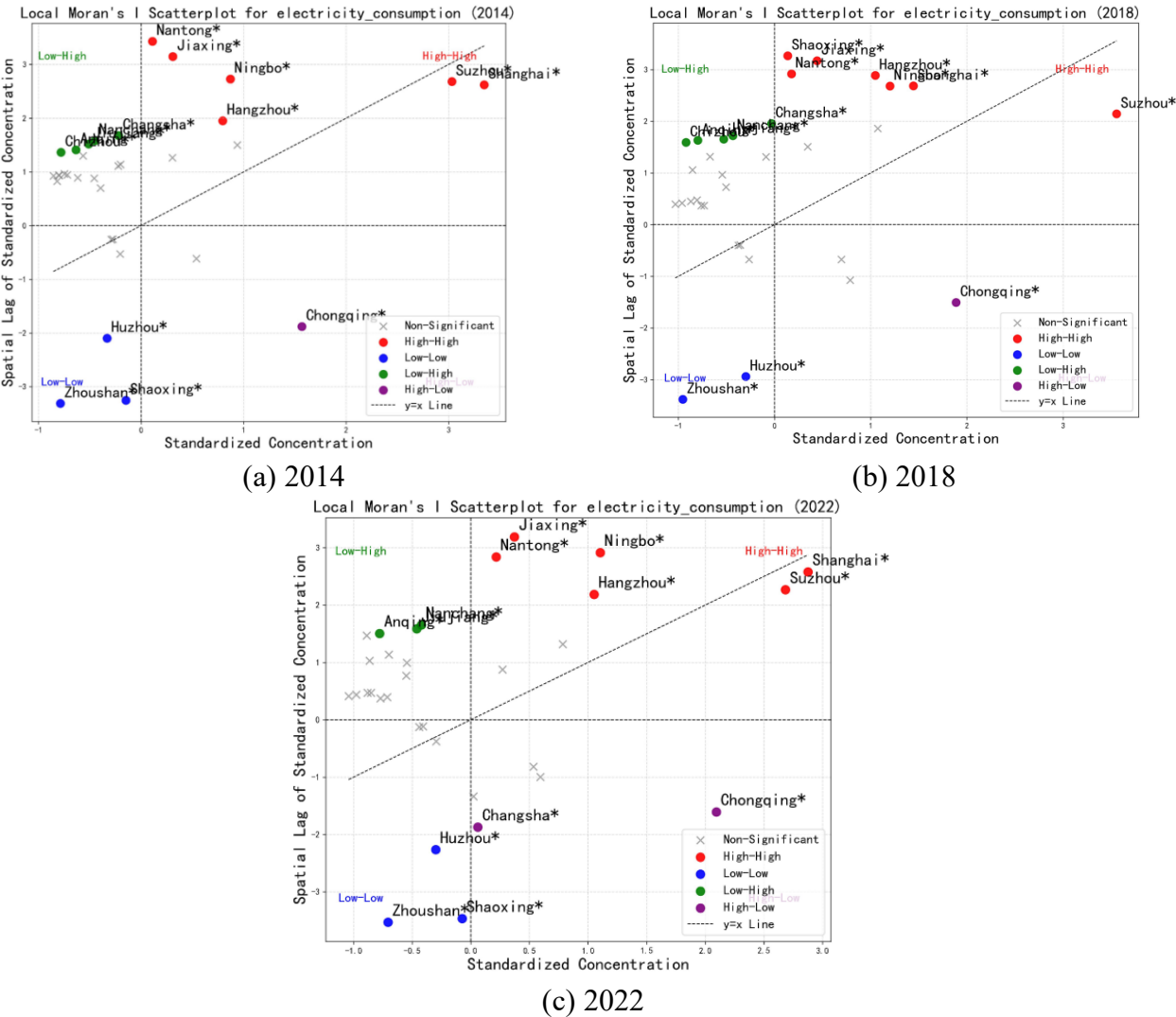
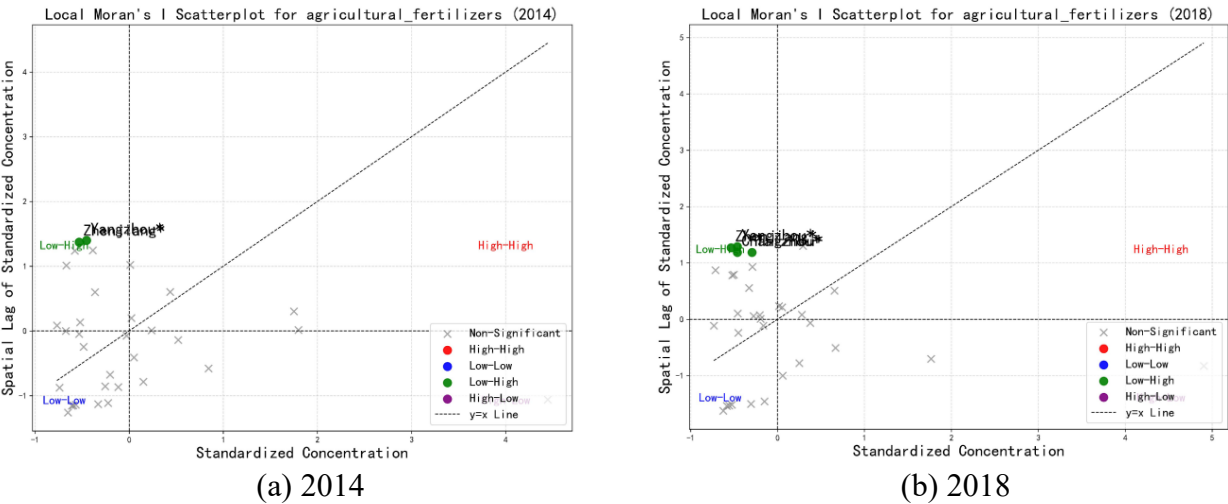
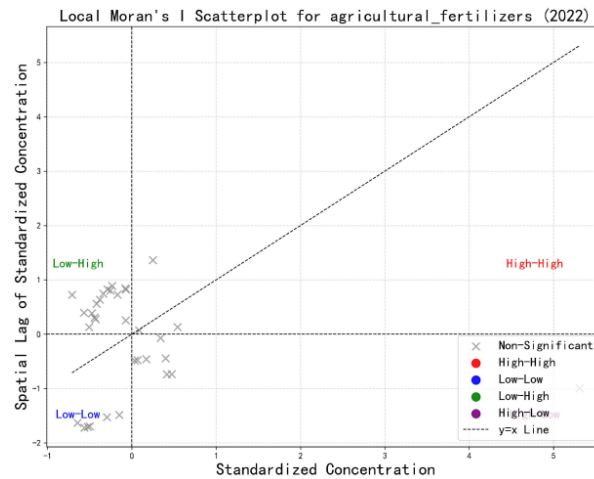


Fig. 8. electricity consumption Levels in 2014, 2018 and 2022





(c) 2022

Fig. 9. agricultural fertilizer application Levels in 2014, 2018 and 2022

From Figs. 6 to 9, we observe that $PM_{2.5}$ and electricity consumption (electricity_consumption) exhibit more significant spatial aggregation characteristics at the three points in time (2014, 2018, and 2022) compared to agricultural fertilizer application (agricultural_fertilizers) and NO_2 . Moreover, the Moran's I index at a 5% significance level indicates that these two variables exhibit stable positive spatial correlation with AQI. Subsequent analysis will focus on the spatiotemporal evolution of $PM_{2.5}$ and electricity consumption.

As shown in Table 4, upstream cities—Chongqing, Panzhihua, and Luzhou—formed a persistent “low-low” (LL) cluster for $PM_{2.5}$, reflecting sustained low pollution and improving air quality. In contrast, midstream cities (Zhenjiang, Yangzhou, Changzhou, Nanjing) were “high-high” (HH) clusters in 2014, indicating severe pollution and positive spatial dependence. By 2022, most exited this cluster, demonstrating the effectiveness of pollution control. Downstream cities—including Ningbo, Zhoushan, Shaoxing, Jiaxing, Shanghai, Huzhou, and Hangzhou—consistently exhibited “low-high” (LH) clustering, indicating good local air quality but exposure to surrounding high-pollution areas, highlighting cross-regional influence.

Table 4. Changes in Local Moran’s I Index of PM_{2.5}

year	LH	LL	HH	HL
2014	Ningbo and Zhoushan formed a Low-High cluster.	Ma'anshan constituted a Low-Low cluster for intensity.	Taizhou, Zhenjiang, Yangzhou, Changzhou, and Nanjing formed High-High clusters for PM _{2.5} intensity.	none
2018	Shaoxing also became a member of this Low-High cluster.	Chongqing also joined this cluster.	Zhenjiang and Yangzhou exited the High-High cluster zone.	none
2022	Jiaxing, Shanghai, Huzhou, and Hangzhou joined the cluster.	Panzhihua and Luzhou entered the Low-Low cluster, while Ma'anshan exited this category.	Taizhou, Nanjing, and Changzhou exited the High-High cluster, while Jiangjin, Chengdu, Wanzhou, and Fuling became new members of this zone.	none

In Fig. 8, we observe that the local Moran’s I index of electricity consumption exhibits spatial clustering characteristics from 2014 to 2022. To further capture its dynamic evolution, Table 5 is created, which show persistent spatial clustering in electricity consumption across the YREB from 2014 to 2022. Chongqing consistently exhibited a “high-low” (HL) pattern, indicating high local consumption relative to lower-intensity neighbors. In central regions, Changsha, Nanchang, Jiujiang, and Anqing formed LH clusters in 2014, but by 2022, Changsha shifted to HL, reflecting a reversal in spatial dependence. Downstream cities—Shanghai, Suzhou, Nantong, Hangzhou, Ningbo, and Jiaxing—formed a strong HH cluster in 2014, indicating intensive and spatially correlated consumption. However, by 2022, Shaoxing and Wuxi exited this cluster, and Shaoxing returned to LL, signaling a dispersion of consumption patterns, likely due to energy-saving policies. These dynamics reveal evolving regional disparities and the impact of socioeconomic factors on energy use. The shifting spatial autocorrelation underscores the need to account for neighborhood effects in energy and environmental planning.

Table 5. Changes in Local Moran’s I Index of Electricity Consumption

Year	LH	LL	HH	HL
2014	Changsha, Nanchang, Jiujiang, and Anqing formed Low-High clusters for electricity consumption intensity.	Huzhou, Shaoxing, and Zhoushan formed a Low-Low cluster.	Shanghai, Suzhou, Nantong, Hangzhou, Ningbo, and Jiaxing formed a High-High cluster.	Chongqing constituted a High-Low cluster.
2018	Chizhou joined the Low-High cluster.	Shaoxing exited the Low-Low cluster.	Shaoxing and Wuxi became part of the High-High cluster.	Chongqing formed a High-Low agglomeration area.
2022	Changsha and Chizhou exited the Low-High cluster.	Shaoxing has returned to the Low-Low cluster.	Shaoxing and Wuxi exited the High-High cluster.	Changsha joined the High-Low cluster.

This section systematically investigates the spatial agglomeration characteristics and dynamic evolution patterns of key variables, including PM_{2.5}, NO₂, agricultural fertilizer application, and electricity consumption, in cities along the YREB using Moran's I index and LISA analysis. The empirical results demonstrate statistically significant spatial clustering features for PM_{2.5} and electricity consumption, confirming their non-random spatial distributions and highlighting substantial regional synergistic effects in air quality patterns. These findings fundamentally necessitate incorporating spatial-dependence considerations into the analysis of AQI determinants. Neglecting such spatial dependence may yield biased parameter estimates, weaken model interpretability, and impair the plausibility of subsequent policy-related inferences. It should be noted that the preceding SHAP-XGBoost workflow serves only for predictive-feature screening, and subsequent SAR outputs capture conditional statistical associations rather than strict causal effects.

6. Spatial Spillover Effect Analysis

Building upon the feature-importance ranking derived from the XGBoost-SHAP workflow in Section 4 and the spatial-autocorrelation diagnosis outcomes from Section 5, this section constructs a spatial autoregressive (SAR) framework to quantitatively assess spatial-spillover mechanisms for AQI. With key predictors pre-screened via data-driven machine-learning outputs, the SAR model enables us to evaluate both local-city impacts and cross-neighbourhood spillover relationships, complementing the purely predictive perspective from SHAP and revealing whether variables with high predictive importance also generate measurable spatial-statistical associations across cities within the YREB. This study selects SAR as the baseline spatial panel specification, whose core theoretical motivation lies in capturing spatial dependence of the dependent variable AQI — namely air-quality spillover effects across neighbouring cities, the key mechanism represented by the spatial-lag term of the outcome variable within SAR. We recognize that the Spatial Durbin Model (SDM) can additionally accommodate spatially-lagged explanatory variables to identify spillover effects originating from predictors. Nevertheless, our sample is limited to 33 cities across 9 years; adopting SDM would markedly expand the number of estimated parameters and consume substantial degrees of freedom, potentially impairing statistical power for spatial-parameter inference. Therefore, the SAR model is adopted as the baseline specification in this study.

6.1 Spatial Autoregressive Model

The air pollution intensity in the YREB is influenced by multiple temporal and spatial factors, exhibiting spatial effects. Spatial lag models, spatial error models, and spatial Durbin models are the main methods for spatial regression modeling. In this section, we construct a spatial autoregressive model (SAR) based on spatial panel data to systematically analyze the spatial dependence of AQI and its driving mechanisms. The research uses the WGS84 coordinate system to construct the urban spatial

topology, and the k-nearest neighbor algorithm is used to generate the spatial weight matrix. This matrix is then standardized to ensure comparability of geographic proximity relationships. The effectiveness of multicollinearity mitigation was verified through variable correlation heatmaps (Figs. 10 and 11) and Variance Inflation Factor (VIF) tests.

As Dormann et al. (2013) stated, multicollinearity is a common issue in regression analysis, requiring multiple methods for evaluation and mitigation. The initial model included nine explanatory variables (PM_{2.5}, agricultural fertilizer application, and electricity consumption). After removing redundant variables, the maximum VIF dropped to 3.351 (Table 6), indicating effective mitigation. Fig. 10 shows high intercorrelations in the original set, leading to inflated VIFs. In contrast, Fig. 11 reveals reduced correlations after optimization, consistent with VIF results. This visual and statistical convergence confirms improved variable selection and strengthens the robustness of the subsequent SAR model.

Table 6. Multicollinearity Analysis of Explanatory Variables

Explanatory variables	Initial VIF	Updated VIF
PM _{2.5}	1.790	1.642
Agriculture_fertilizers	3.658	1.812
Electricity_consumption	4.988	2.144
NO ₂	1.497	1.472
Permanent_population	16.843	-
Liquefied_petroleum	2.178	2.109
SO ₂	2.009	1.813
GDP	15.524	-
Natural_gas	6.885	3.351

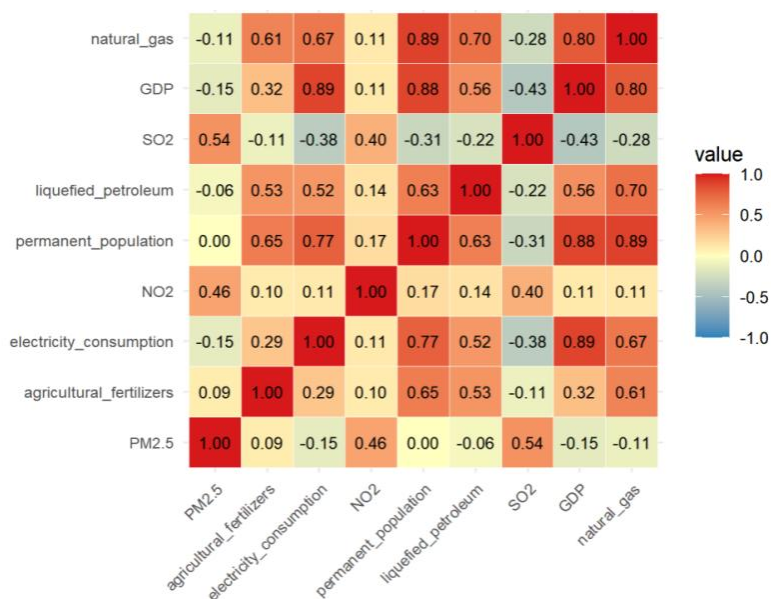


Fig. 10. Correlation Matrix of Independent Variables in the Initial Model

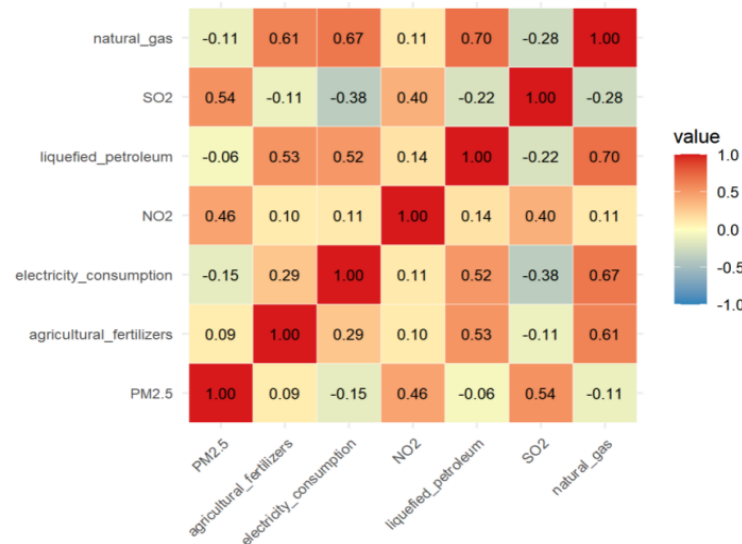


Fig. 11. Correlation Matrix of Independent Variables in the Updated Model

To examine conditional statistical associations between explanatory variables (e.g., PM_{2.5}, agricultural fertilizer application, and social electricity consumption) and AQI, and to quantify the transmission mechanism of spatial spillover effects on urban pollution, we have constructed a regression framework based on the SAR model (as shown in Eq. (3)) to systematically evaluate the spatial dependence and effect of each factor.

$$(1 - \rho W) AQI_{it} = \beta_i X_{it} + \varepsilon \quad (3)$$

The baseline SAR panel model employs city-level individual fixed-effects only and excludes time-fixed-effects. City-individual fixed-effects are introduced to absorb time-invariant unobserved city-specific heterogeneity across our 33 sample cities, such as persistent geographic conditions, topographic features, and inherent urban endowments. These time-invariant unobserved characteristics constitute important sources of omitted-variable bias in air-pollution research; individual fixed-effects help mitigate such confounding bias. Time-fixed-effects are not included in the baseline specification. Time-fixed-effects would absorb common year-specific nationwide shocks, such as unified national environmental-policy adjustments and macroeconomic fluctuations. However, our dataset consists of only 33 cross-sectional cities and 9 time-periods (2014-2022). Adopting two-way (city-and-time) fixed-effects in the baseline would consume considerable degrees of freedom and potentially reduce statistical power for estimating spatial-spillover parameters. For this reason, the two-way fixed-effect SAR is reserved as a robustness test in Section 6.2.2 instead of being used as our baseline setting.

In this model, the target parameter is the air quality composite index of city i in year t ; W represents the spatial weight matrix; X_{it} includes the main explanatory variables, such as PM_{2.5}, agricultural fertilizers, electricity consumption, NO₂, liquefied petroleum, SO₂, and natural gas; ρ is the spatial autoregressive coefficient; represents the regression coefficients of the independent variables; and ε is

the random error term (typically assumed to be independently and identically distributed).

Table 7. Regression Results of the Model

Explanatory variables	Coefficient	P-value
PM _{2.5}	0.3594	<0.001
agriculture_fertilizers	-0.1195	0.048
electricity_consumption	<0.0001	0.870
NO ₂	-0.0788	0.223
liquefied_petroleum	<0.0001	0.023
SO ₂	0.5266	<0.001
natural_gas	<0.0001	0.734
$R^2 = 0.9090$, SAR parameter estimates: 0.3134 (P < 0.001)		

According to the regression results shown in Table 7, the spatial autoregressive coefficient ($\lambda = 0.3134$) in the SAR model is significant, indicating that there is a significant spatial dependence and spillover effect between different regions for AQI. Improving air quality in a specific region not only benefits that region but also improves the air quality of neighboring regions by reducing the diffusion of pollutants. Specifically, the regression coefficients for PM_{2.5} and SO₂ are significantly positive ($p < 0.001$), marking them key variables showing strong positive statistical associations with local AQI. For every 1-unit increase in PM_{2.5}, the target area's impact increases by 0.36 units, likely due to atmospheric diffusion leading to transboundary transmission of pollutants. The strong positive correlation with SO₂ (coefficient=0.527) further supports the Inter-regional influence characteristic of air pollutants. Although the coefficient for liquefied petroleum is relatively small, its significance ($p = 0.023$) suggests that long-term accumulation may generate observable spatial externalities. The significant negative correlation with agricultural fertilizers may reflect the purifying effect of vegetation cover in agricultural areas, reducing pollution levels in neighboring regions. Power consumption, NO₂, and natural gas do not show significant spatial effects ($p > 0.05$), indicating that their statistical associations with AQI are spatially confined.

Through spatial effect decomposition, the differential spatial mechanisms of environmental variables on air quality can be revealed. The direct effect reflects the immediate impact of the variables on the local region, while the indirect effect represents the cross-domain effect on neighboring regions through spatial interaction. The total effect is the sum of the direct and indirect effects, representing the comprehensive impact of the variable on the overall study area. This study, based on SAR and spatial multiplier decomposition method, constructs a $k=6$ nearest neighbor spatial weight matrix (W) and calculates the spatial multiplier matrix $S = S(1 - \rho W)^{-1}$, decomposing the regression coefficients into direct $\beta \times \text{diag}(S)$, indirect $\beta \times (S - \text{diag}(S))$, and total effects $\beta \times S$. The spatial autoregressive coefficient ρ is estimated through Eq. (3), and the construction of the spatial weight matrix ensures the

topological consistency of geographical proximity (using the WGS84 coordinate system and KNN).

Table 8. Decomposition Results of Spatial Effects

Variable	Direct Effect	Indirect Effect	Total Effect
PM _{2.5}	0.366	-0.350	0.016
agriculture_fertilizers	-0.122	0.116	-0.0053
electricity_consumption	>-0.0001	<0.0001	>-0.0001
NO ₂	-0.0801	0.0766	-0.0035
liquefied_petroleum	<0.0001	>-0.0001	<0.0001
SO ₂	0.536	-0.513	0.0234
natural_gas	<0.0001	>-0.0001	<0.0001

As shown in Table 8, spatial decomposition reveals: (1) PM_{2.5} and SO₂ exhibit asymmetric spatial effects. PM_{2.5} has a significant positive direct effect (+0.366), but a negative indirect effect (−0.350), leading to a near-neutral total effect (+0.016), possibly due to regional control spillovers or deposition. Similarly, SO₂ shows a direct effect of +0.536 and an indirect effect of −0.513, reflecting local pollution lock-in and inter-regional coordination. (2) Agricultural fertilizer application produces a negative direct statistical coefficient (−0.122) yet a positive indirect statistical coefficient (+0.116). The mechanisms giving rise to this contrasting statistical outcome cannot be resolved using current city-level panel data and require further targeted research. (3) Liquefied petroleum gas shows weak opposing direct and indirect effects, suggesting local pollution displacement with minimal net spillover. (4) Power consumption, NO₂, and natural gas exhibit no significant spatial effects, indicating localized impacts. Those findings imply that PM_{2.5} and SO₂ require strengthened regional joint control, while agricultural fertilizer application calls for watershed-level coordination to address transboundary non-point source pollution.

6.2 Robustness Test

6.2.1 Alternative Dependent Variable Test (PM_{2.5} substitution)

To alleviate potential mechanical correlation concerns arising from the composite nature of AQI, this study conducts a robustness test by replacing the dependent variable with single-pollutant PM_{2.5} concentration. All data preprocessing procedures, hyperparameter grids, and cross-validation strategies remain identical to the baseline model to ensure fair comparison.

The alternative PM_{2.5}-based model achieves satisfactory predictive performance with a test R² of 0.8708, confirming the reliability of the XGBoost framework. As shown in Figure 12, differences in SHAP feature importance rankings are observed between the AQI-based and PM_{2.5}-based specifications. Such divergence is methodologically reasonable: AQI is a comprehensive air quality index integrating multiple pollutant components, whereas PM_{2.5} is a single atmospheric pollutant with distinct formation mechanisms and meteorological sensitivities.

Most importantly, the core factors focused on in this study, including agricultural fertilizer application and electricity consumption, exhibit non-negligible predictive importance for air quality. Although their relative positions decline compared with the AQI baseline, they do not become trivial low-contribution predictors.

This verification demonstrates that the identified predictive relationships between influencing factors and air quality are robust and not distorted by the mechanical correlation inherent in the composite AQI indicator.

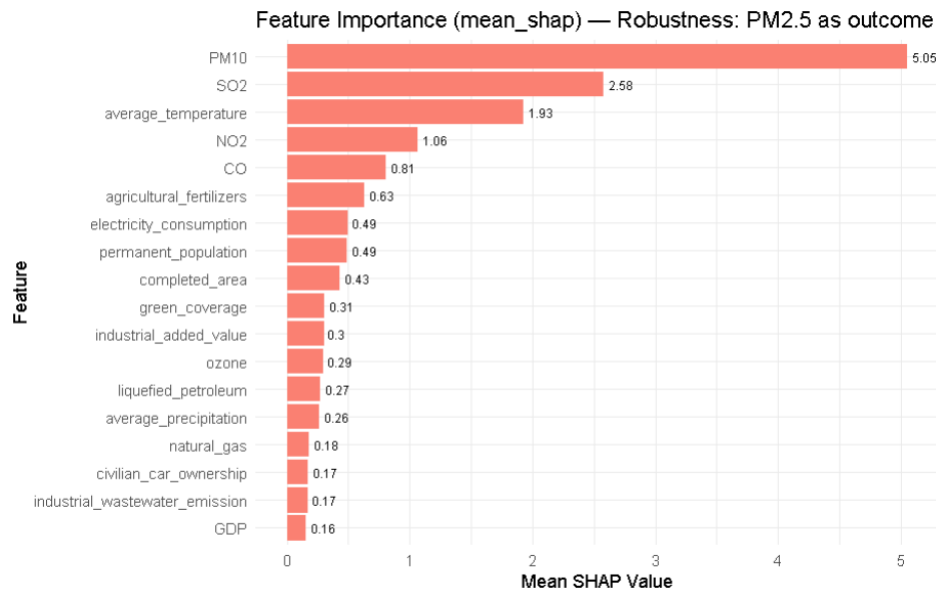


Fig. 12. SHAP-Based Feature Importance Ranking for PM_{2.5} Robustness Test

6.2.2 Robustness with alternative specifications

To verify the robustness of the model results, three different robustness tests are conducted in this study. First, the spatial weight matrix is replaced with an inverse distance weight matrix instead of the K-nearest neighbor matrix, and the model is re-estimated. The results show that the spatial autoregressive coefficient significantly increases from 0.3134 to 0.5545, and the model fit (R^2) improves from 0.9098 to 0.9153, indicating that the model has good adaptability to the choice of the weight matrix. Second, we estimate the two-way (city-and-time) fixed-effect SAR model. By adding time-fixed-effects on top of the baseline city-individual fixed-effects, this specification controls for common nationwide time-varying shocks and global temporal trends, serving as a complementary robustness check for our baseline model. The model fit significantly improves ($R^2 = 0.9219$), and the Akaike Information Criterion and Bayesian Information Criterion values decrease to 1685.116 and 1714.666, respectively, further suggesting that the model can effectively capture the interaction between time and spatial effects. Lastly, a subsample analysis is conducted by excluding special regions such as Shanghai. It is found that the coefficients of the core explanatory variables in the full sample and the subsample are almost identical, with a 0% difference, indicating that the model results are not sensitive to the presence of special regions.

In summary, all three robustness tests support the reliability of the core conclusions of this study and confirm the robustness and universality of the model results.

6.2.3 Sensitivity Analysis of the K-Nearest-Neighbor Spatial Weight Matrix

To mitigate the arbitrariness introduced by the choice of nearest-neighbor number in the KNN spatial-weight matrix, we re-estimate the SAR model with alternative nearest-neighbor values of $k=4,5,6,7$. Table 9 reports the estimation outputs. The spatial autoregressive coefficient λ fluctuates within a narrow range from 0.3115 to 0.3198 and remains close to the baseline SAR spatial-autoregressive coefficient 0.3134 reported in Table 7. Model-fit statistics including AIC, BIC and R^2 exhibit only minor numerical changes. Although the model with $k=4$ yields marginally lower AIC value, the overall difference across specifications is trivial. The spatial autoregressive coefficient remains statistically positive in all cases, and the sign and statistical significance of core explanatory variables do not reverse. Considering the uneven geographical distribution of cities within the YREB, $k=6$ is retained as our baseline setting to guarantee sufficient neighboring observations for sparsely-distributed western cities. The stable qualitative pattern across different k values validates the robustness of our baseline SAR regression outcomes.

Table 9. Sensitivity test for different K-nearest-neighbor specifications

k	λ	Log-likelihood	AIC	BIC	R^2
4	0.3157	-854.8	1726	1755	0.9114
5	0.3129	-856.2	1728	1758	0.9102
6	0.3115	-856.7	1729	1759	0.9097
7	0.3198	-857.0	1730	1759	0.9094

Note: Slight numerical divergence of λ between Table 7 (0.3134) and Table 9 (0.3115 for $k=6$) originates from numerical tolerance in maximum-likelihood iterative optimization. Sample, spatial-weight matrix and model specification remain identical across estimations. All specifications adopt individual city-fixed-effect SAR with row-standardized KNN weight matrices.

7 Conclusion and Policy Recommendations

7.1 Conclusion

This paper investigates spatial-statistical patterns and spillover characteristics of urban air-pollution-related variables to support spatial-governance optimization within the YREB. An innovative analytical framework is proposed to integrate XAI and spatial econometric models. Through the SHAP-XGBoost model, we obtain a quantitative ranking of features according to their predictive importance for AQI, while the KNN spatial weight matrix and Moran's I index reveal the spatial clustering characteristics of

pollutants. Furthermore, the SAR and effect decomposition methods are employed to analyze spatial spillover pathways and cross-regional transmission mechanisms of high-SHAP-value factors. This study overcomes the limitations of traditional pollution analysis methods in terms of interpretability and spatial interaction modeling and provides a decision-support system for the coordinated governance of air pollution in the YREB, which is grounded in both theoretical foundations and policy-oriented insights.

The main research findings are as follows: (1) Variables such as $PM_{2.5}$ and total electricity consumption exhibit high predictive feature importance for variations in the AQI. Among these, $PM_{2.5}$ exhibits the highest mean absolute SHAP value (5.14), marking it the top-performing predictive feature for air-quality variation. (2) In terms of spatial effects, $PM_{2.5}$ and SO_2 exhibit significant positive local direct effects (0.366 and 0.536, respectively), consistent with the expectation that local stationary emission sources such as coal-fired power plants and heavy industries may exhibit strong statistical covariation with local air-quality outcomes. However, both variables yield negative indirect statistical coefficients in the SAR effect-decomposition outputs. (3) Although agricultural fertilizer application produces a negative direct statistical coefficient for AQI, its overall statistical association with air quality is negligible. Our aggregate city-level dataset cannot pinpoint the underlying drivers for this statistical pattern, calling for further research. (4) The use of liquefied petroleum gas demonstrates marginal emission reduction benefits, while natural gas shows no significant spatial effect. This highlights the uneven progress in urban-rural energy substitution, indicating the need for stronger policy guidance to promote clean energy transitions in rural areas.

7.2 Policy Recommendations

Based on the empirical results of XGBoost-SHAP feature-importance identification, spatial autocorrelation tests, and SAR estimation, this section elaborates policy implications for collaborative air pollution governance in the YREB. The recommendations proceed from industrial-pollutant coordinated control, differentiated sector-specific management, rural energy optimization, and refined data-driven supervision, aiming to support targeted, differentiated, and coordinated air quality management across YREB cities.

(1) Build an integrated governance system combining local source regulation and cross-regional joint-prevention based on pollutant predictive importance and spatial spillover patterns. The XGBoost-SHAP results indicate that $PM_{2.5}$ and SO_2 exhibit dominant predictive importance for regional air quality, while SAR estimation confirms their significant local effects and negative spatial spillover characteristics. In this study, the negative spatial spillover is interpreted as a statistically-observed spatial association feature among neighboring cities rather than direct evidence of pollution transfer. Accordingly, YREB urban agglomerations should prioritize targeted source regulation for

industries with high PM_{2.5} and SO₂ emissions. Ultra-low-emission transformation of high-energy-consuming sectors such as steel and cement should be continuously promoted to reduce local pollutant accumulation. Given the spatially-correlated emission characteristics, isolated city-only regulation is insufficient for air-pollution control. Spatial-autocorrelation results further verify the obvious spatial aggregation of air pollutants, highlighting the necessity of regional joint prevention and control. Key urban clusters should strengthen standardized joint supervision of pollutant-intensive industries, unify regional environmental-access standards, and build a long-term coordinated-governance mechanism adapted to the spatial distribution of air pollution within the YREB.

(2) Adopt localized and adaptive agricultural management strategies considering the weak and spatially heterogeneous performance of fertilizer-related variables Empirical results show that agricultural fertilizer application presents non-negligible predictive importance in machine-learning identification but exhibits offsetting direct and indirect spatial effects in SAR models. This inconsistent empirical pattern implies that agricultural non-point-source pollution across the YREB is highly spatially heterogeneous and cannot be regarded as a uniform pollution source for the whole region. We therefore propose localized agricultural optimization: major grain-producing cities should promote soil-testing-based fertilization and organic-fertilizer substitution to mitigate potential agricultural pollution accumulation, while non-agriculture-dominant cities should focus on industrial and residential pollution supervision. Differentiated agricultural ecological management, rather than unified regional compensation mechanisms, is more consistent with the empirical spatial-heterogeneity findings.

(3) Optimize rural clean-energy substitution structure based on the identified imbalance in energy-related predictive signals Feature-importance results reveal distinct differences in the predictive relationship between different clean-energy sources and air quality. Liquefied petroleum gas shows a favorable predictive association with improved air-quality outcomes, while the predictive contribution of natural gas is weaker, reflecting an unbalanced performance of current rural energy-substitution practices. Counties and rural areas within the YREB should further expand clean-coal replacement and standardized clean-fuel promotion. Targeted energy-structure optimization can effectively alleviate the structural imbalance of rural energy-substitution measures and improve grassroots air-pollution-control efficiency.

(4) Strengthen multi-source environmental-data monitoring to support refined governance Although urban electricity consumption does not yield statistically-significant regression coefficients in the SAR model, it presents significant spatial correlation and non-negligible predictive importance in the machine-learning outputs. This inconsistency across different identification approaches suggests that socioeconomic energy-consumption indicators may exhibit correlation with air-quality conditions within certain latent spatial patterns. On the basis of this empirical discrepancy, this study proposes

strengthening multi-source environmental-data integration, including datasets on energy consumption, industrial emissions and urban geographic information, to support refined and dynamic environmental supervision.

Funding Statement

This research was partially supported by the Natural Science Foundation of Jiangsu Province Youth Fund Project (BK20250736), Canada-China Scholars' Exchange Program (202509040060).

Conflict of Interest Statement:

The authors declare no conflict of interest.

References

- [1] Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis*, 27(2), 93–115.
- [2] Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4), 281–298.
- [3] Chang, Y., Li, G., Zhang, P., et al. (2024). Relationships among six urban air pollutants and identification of pollution types: A case study of Chinese cities above prefecture level. *Atmospheric Pollution Research*, 15(7), 102160.
- [4] Dormann, C. F., Elith, J., Bacher, S., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46.
- [5] Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), 189–206.
- [6] Gürtepe, İ. Ç., Şenkal, İ. T., Ünal, A., Güllü, G., Aslanoğlu, Y., & Marshall, J. D. (2025). Machine learning-driven regional prediction of PM_{2.5} concentrations in the eastern Mediterranean: Bridging spatial data gaps in air quality monitoring. *Environmental Modelling & Software*, 192, 106586.
- [7] Han, L., Zhou, W., Li, W., et al. (2014). Impact of urbanization level on urban air quality: A case of fine particles (PM_{2.5}) in Chinese cities. *Environmental Pollution*, 194, 163–170.
- [8] Han, X., Li, H., Liu, Q., et al. (2019). Analysis of influential factors on air quality from global and local perspectives in China. *Environmental Pollution*, 248, 965–979.
- [9] He, H., Zhao, K., Yuan, Z., Shen, J., Lin, Y., Zhang, S., Wang, M., Wang, A., & Lian, P. (2026). A novel approach to identify the spatial characteristics of ozone-precursor sensitivity based on interpretable machine learning. *Journal of Environmental Sciences*, 159, 54–63.

- [10] Hou, M., Cui, X., Xie, Y., Lu, W., & Xi, Z. (2024). Synergistic emission reduction effect of pollution and carbon in China's agricultural sector: Regional differences, dominant factors, and their spatial-temporal heterogeneity. *Environmental Impact Assessment Review*, 107, 107543.
- [11] Jiangsu Municipal Statistics Bureau. (2020). Jiangsu statistical communique on the 2019 national economic and social development. <https://jszd.stats.gov.cn/gjtjjjsdczd/sjxxfb/content/9227f7b4-4845-4f45-a5f1-4b679dc5ac4a.html>.
- [12] Jiangsu Provincial Department of Ecology and Environment. (2020). 2019 Jiangsu environmental statistical annual report. https://sthjt.jiangsu.gov.cn/art/2020/11/19/art_83555_10044762.html
- [13] Li, L., Deng, C., Zhu, F., Song, X., Wang, E., & Xie, M. (2025). Beyond one-size-fits-all: Differentiated green development assessment integrating a hybrid approach in China's Yangtze River Economic Belt. *Environmental Impact Assessment Review*, 115, 108076.
- [14] Liang, Z., Cai, Q., Zeng, N., Tang, W., Han, P., Zhang, Y., Quan, W., Yao, B., Wang, P., & Liu, Z. (2025). Modeling the atmospheric CO₂ concentration in the Beijing region and assessing the impacts of fossil fuel emissions. *Environments*, 12(5), 156.
- [15] Liu, K., Ren, G., Dong, S., et al. (2024). The synergy between pollution reduction and carbon reduction in Chinese cities and its influencing factors. *Sustainable Cities and Society*, 106, 105348.
- [16] Ren, L., & Matsumoto, K. (2020). Effects of socioeconomic and natural factors on air pollution in China: A Spatial Panel Data Analysis. *Science of the Total Environment*, 740, 140155.
- [17] Rudin, C. (2019). Stop explaining black box machine learning models for high stake decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215.
- [18] Satish, M., Biswas, K. S., & Purkayastha, B. (2025). Predictive purity: Advancements in air pollution forecasting through machine learning. *Optical Memory and Neural Networks*, 34(2), 256–271.
- [19] Sen, Y., Jia, H., Fan, F., Zhao, N., Kuang, Y., Tang, X., Huang, Q., & Wang, X. (2025). Assessing ozone formation impact through SHAP interaction redistribution analysis: A novel framework for evaluating VOC photochemical loss and source interactions. *Environmental Impact Assessment Review*, 115, 108017.
- [20] Shanghai Municipal Ecology and Environment Bureau. (2018). Shanghai Ecological Environment Status Bulletin 2017. <https://sthj.sh.gov.cn/assets/html/110939-01.pdf>
- [21] Shanghai Municipal Statistics Bureau. (2020). Shanghai statistical yearbook 2019. China Statistics Press. <https://tjj.sh.gov.cn/tjnj/tjnj2019.htm>.
- [22] Song, X., Sun, Y., Wang, H., Huang, X., Han, Z., Shu, Y., Wu, J., Zhang, Z., Zhong, Q., Li, R., & Fan, Z. (2025). Uncovering soil heavy metal pollution hotspots and influencing mechanisms through machine learning and spatial analysis. *Environmental Pollution*, 370, 125901.

- [23] Tang, T., Lin, X., & Lu, M. (2024). Analysis of the dynamic mechanism of trans-boundary pollution collaborative control in the Yangtze River Economic Belt. *Environmental Research*, 257, 119327.
- [24] Wang, D., Thunéll, S., Lindberg, U., et al. (2022). Towards better process management in wastewater treatment plants: Process analytics based on SHAP values for tree-based machine learning methods. *Journal of Environmental Management*, 301, 113941.
- [25] Wang, Y., Han, R., & Kubota, J. (2016). Is there an environmental Kuznets curve for SO₂ emissions? A semi-parametric panel data analysis for China. *Renewable and Sustainable Energy Reviews*, 54, 1182–1188.
- [26] Xu, Y., Liu, Z., Walker, T. R., et al. (2024). Spatio-temporal patterns and spillover effects of synergy on carbon dioxide emission and pollution reductions in the Yangtze river delta region in China. *Sustainable Cities and Society*, 105, 105419.
- [27] Yan, Y., & Yang, Y. (2025). Revealing the synergistic spatial effects in soil heavy metal pollution with explainable machine learning models. *Journal of Hazardous Materials*, 482, 136578.
- [28] Yang, T., Zhou, K., & Ding, T. (2022). Air pollution impacts on public health: Evidence from 110 cities in Yangtze river economic belt of China. *Science of the Total Environment*, 851, 158125.
- [29] Zeng, J., Liu, T., Feiock, R., et al. (2019). The impacts of China's provincial energy policies on major air pollutants: A Spatial Econometric Analysis. *Energy Policy*, 132, 392–403.
- [30] Zhang, W., He, P., & Huang, J. (2018). The effect of agricultural eco-compensation on reducing non-point source pollution in China: Evidence from a field experiment. *Journal of Environmental Management*, 218, 952–961.
- [31] Zhang, Y., Li, X., & Wang, J. (2021). Application of spatial statistics and machine learning in urban air pollution source apportionment: A Review. *Environmental Pollution*, 288, 117834.
- [32] Zhao, S., Liu, S., Hou, X., et al. (2018). Temporal dynamics of SO₂ and NO_x pollution and contributions of driving forces in urban areas in China. *Environmental Pollution*, 242, 239–248.
- [33] Zhu, J., & Xu, J. (2022). Air pollution control and enterprise competitiveness: A re-examination based on China's clean air action. *Journal of Environmental Management*, 312, 114968.