atmosphere

MDPI

Article

Research on Prediction Methods of Anthropogenic Economic Pollutant Emission Index by Coupling of Ensemble Machine Learning and Time-Series Models Under Multiple Features

Mengzhen Li [1], Yang Cao [2,*] and Jianlei Lang [1]

[1] Key Laboratory of Beijing on Regional Air Pollution Control, College of Environmental Science and Engineering, Beijing University of Technology, Beijing 100124, China
[2] School of Computer Science and Artificial Intelligence, Beijing Wuzi University, Beijing 101149, China
* Correspondence: caoyang@bwu.edu.cn

Abstract

Predicting the anthropogenic economic pollutant emissions index helps balance economic growth and environmental protection. In this study, a set of constructed features was derived using provincial-level basic data from 1995 to 2023. By constructing a weighted ensemble strategy incorporating Extreme Gradient Boosting, light gradient boosting machine, Random Forest, and Multi-Layer Perceptron, and integrating it with the Autoregressive Integrated Moving Average model and Shapley Additive Explanations, an anthropogenic economic pollutant emission index (AS_GPI) forecasting model was finally established, with basic and constructed features employed as its inputs (AFEA-AG). The AS_GPI forecasting model (EA-AG) was developed without constructed features, with all other settings consistent with the AFEA-AG model for comparison. Results show that the proposed model achieves high forecasting accuracy for AS_GPI across four typical pollutants, with R^2 values exceeding 0.95 for the AS_GPI_NOx and AS_GPI_$PM_{2.5}$. The mean absolute percentage error was as low as 0.0744. The forecasting model with constructed features yielded lower errors and higher stability in its prediction results compared with the one without such features. Feature contribution analysis revealed differing key contributors, with AS_GPI lagged values and economic-related characteristics among underlying variables exhibiting strong predictive importance. The 2024 projection results indicated certain disparities in pollution control effectiveness between key and non-key regions. Further analysis of historical and predicted data revealed a nationwide decline in AS_GPI between 1995 and 2024. Beijing and Shanghai achieved notable environmental quality improvements through anthropogenic emission reductions. The coefficient of variation values of different AS_GPIs reveal spatial heterogeneity and differences. Nationwide efforts should prioritize the control of anthropogenic NMVOC and NOx emissions. This framework provides a prediction method that offers certain reference for the development of the economy and the environment.

Keywords: air pollution; emission intensity per GDP; anthropogenic source emissions; ensemble machine learning; feature engineering

Check for updates

Received: 18 March 2026
Revised: 1 May 2026
Accepted: 4 May 2026
Published: 8 May 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

Despite the significant progress China has made in combating air pollution and the continuous improvement in air quality [1,2], certain challenges remain in addressing surface $PM_{2.5}$ and O_3 pollution across mainland China [3,4]. NOx and NMVOCs are key precursors

of both $PM_{2.5}$ and O_3, while $PM_{2.5}$ and SO_2 affect $PM_{2.5}$ formation. Thus, focusing on all four pollutants is essential for improving $PM_{2.5}$ and O_3 air quality. Air pollution is closely linked to human health and wellbeing [5–7] and has attracted widespread attention [8,9]. Reconciling rapid development with improvements in environmental quality has become a crucial aspect of urban development [10,11]. Effective air pollution control fundamentally depends on regulating pollutant emissions [12,13]. Air pollution emissions are not only the core indicator for measuring pollution levels but also a key basis for formulating governance policies [13–15]. Therefore, an in-depth exploration of the relationship between air pollution emissions and economic growth [16] (such as pollutant emission intensity per unit of gross domestic product (GDP)) is necessary, providing policymakers with scientific support for targeted pollution control and helping formulate more targeted environmental protection strategies.

Researchers around the world have conducted extensive studies on the relationship between economic development and environmental pollution [17–20]. The existing research covers several popular areas: using economic and emissions data to predict either emissions or economic indicators; developing quantitative indices to characterize the relationship between economic activity and pollution; and exploring the feedback mechanisms and mutual influences between economic development and pollution, among others. Multiple linear regression and multiple polynomial regression models were employed to forecast Iran's CO_2 emissions for the year 2030 [21]. The gray model was used to forecast and comparatively analyze China's CO_2 emissions [22]. Several studies [23–25] have examined the relationship between urban economic development and environmental pollution from the perspective of the pollutant emission index per unit of GDP (GPI), which quantifies the amount of anthropogenic pollutants emitted per unit of GDP. Investigating the feedback mechanisms and mutual influences between economic development and pollution has also attracted considerable attention. One study [26] employed environmental Kuznets curves, regulatory effect models, and panel threshold models to analyze the respective impacts of economic growth and industrial upgrading on environmental pollution. The Spatial Durbin model [27] was employed to investigate the mechanisms by which urbanization influences atmospheric pollution. In summary, constructing a quantitative index for economic–pollution relationships (anthropogenic emissions per unit GDP) enables rapid analysis of the connection between economic growth and air pollution. Currently, few studies employ machine learning methods to rapidly forecast the pollutant emission intensity index per unit of GDP. Most research instead focuses on using machine learning [28] to predict pollution emissions or pollutant concentrations, or to investigate the feedback and mutual influence mechanisms between economic development and pollution. Furthermore, the majority of existing studies suffer from limitations such as relying on a single machine learning model or failing to achieve deep coupling among multiple models. This may lead to issues like potentially inadequate predictive accuracy and an inability to comprehensively capture complex relationships.

The machine learning models commonly employed in prediction studies related to air pollution and economic aspects do not only include traditional linear regression models [29] and spatial econometric models. They also include ensemble tree models (such as machine learning gradient boosting (XGBoost [30]) and lightweight gradient boosting machines (LightGBM)), decision tree models [31], support vector machines, and neural networks (such as Multi-Layer Perceptron (MLP) [32]). Recently, deep learning methods (e.g., CNN [33] and LSTM [34]) have been widely used in air pollution and economic forecasting. In addition, time-series models (such as Autoregressive Integrated Moving Average (ARIMA) [35], Vector Autoregressive Moving Average (VARMA) [36] and ensemble models [37], etc.) are also widely applied in some research regions. Linear regression models are

simple and interpretable but struggle to capture nonlinear relationships [38]; decision tree models can handle nonlinearity but face challenges in ultrahigh dimensions [39]; support vector machines involve higher computational complexity; and traditional neural network models suffer from poor interpretability [40]. Ensemble models [41–43] can combine the strengths of different types of models, demonstrating robust predictive capabilities. They can compensate for the limitations of individual models, ultimately achieving more stable and generalizable predictive outcomes across different scenarios.

The application of machine learning models to atmospheric forecasting relies critically on the quality of input features [44]. High-quality features encompass [45] not only fundamental characteristics but also multidimensional attributes, including proportional, interactive, statistical, and temporal dimensions. These effectively mitigate limitations inherent in traditional panel data, such as overly stringent linear assumptions, neglect of interaction effects, and underestimation of nonlinear feedback within complex systems. Consequently, developing constructed features [45] can translate the intricate interplay between physical laws, societal behaviors, and policy interventions into computable variables. Furthermore, when model changes amplify a feature's influence on outputs, current predictions may paradoxically diminish that feature's importance [46]. Therefore, attributing the causes behind predictive outcomes is crucial for enhancing model performance. The Shapley-based explainability (SHAP) [47] values represent a unified approach to feature attribution, drawing upon game theory [48] and local interpretability. Identifying the primary contributing features that quantify the impact on GDP anthropogenic pollutant emission index predictions lays a crucial foundation for enhancing forecasting accuracy and exploring the relationship between prediction outcomes and individual feature elements.

Overall, most existing studies focus on utilizing machine learning models to predict atmospheric pollution emissions or investigate the reciprocal effects between pollution and the economy. Few studies integrate multiple machines learning models, linking temporal modules and interpretable components while coupling constructed features to forecast economic and anthropogenic atmospheric pollution emission indices.

To examine the relationship between economic activity and anthropogenic atmospheric pollutant emissions, the Anthropogenic Source Emission Index for Gross Domestic Product (AS_GPI) has been introduced. Based on the basic and constructed features, this study employs a weighted ensemble of XGBoost [49], LightGBM, Random Forest (RF) [50,51] and MLP [52,53] machine models coupled with the SHAP model and combined with an ARIMA model for annual feature updates in the forecasting set (test set), to establish an AFEA-AG prediction model. The model successfully forecasted the 2024 AS_GPI for typical anthropogenic air pollutants (SO_2, NOx, $PM_{2.5}$, and NMVOC) across 31 provinces in China. To validate the necessity of establishing constructed features, an EA-AG prediction model was developed without constructed features (only including base features), jointly using the ensemble model and the time-series model. A comparative validation of the two predictive models on the test set was conducted. Furthermore, this study analyzed the CV values and relative change rates of the AS_GPI for these four pollutants in 31 provinces of China from 1995 to 2024. It achieved effective forecasting of the AS_GPI for the four conventional air pollutants while identifying key contributing factors influencing forecast outcomes. These findings provide methodological support for local governments in formulating differentiated air pollution control policies.

2. Data and Method

2.1. Data Setting

The anthropogenic emission data for the four pollutants were obtained from the Multi-scale Emission Inventory Model for China (MEICv2), which estimates anthropogenic

atmospheric pollutant emissions [54–56]. This dataset encompasses five sectors, power generation, industry, residential, transport, and agriculture (http://meicmodel.org.cn (accessed on 3 May 2026)), covering the years 1995 to 2023. Provincial rural total electricity consumption data were obtained primarily from official statistical yearbooks, covering the years 1995 to 2023. Night-time illumination data for each province are primarily derived from calibrated, annual China-wide DMSP-OLS-equivalent data obtained via the Defense Meteorological Satellite Program's Operational Line-Scanning System (DMSP-OLS) and the Suomi National Polar-orbiting Partnership-Visible Infrared Imaging Radiometer Suite (SNPP-VIIRS) data [57]. These data were subsequently organized by province, aggregating the calibrated night-time light grids for 1995–2023 according to China's provincial administrative boundaries, prefecture-level city boundaries, and county-level district boundaries. The primary aggregated metric is the annual total value. Population and population density data cover all 31 provinces in China. The time range covered by population density data is from 1995 to 2022 [58]. China's provincial GDP, GDP per capita, and the shares of primary, secondary, and tertiary industries in GDP were all sourced from the China Statistical Yearbook, covering the years 1995 to 2023. The detailed introduction of the basic data is presented in Table 1.

Table 1. Data sources and descriptions.

Name	Data Source	Detailed Explanation
Emissions Data	MEIC emissions inventory	Anthropogenic source emission inventory
Night-time Light Data	DMSP-OLS Data	Provincial-level night-time light data
Rural Electricity Consumption	Statistical Yearbooks	Provincial rural total electricity consumption of the whole society
Population Density	The dataset established by Liu et al. [58].	Provincial-level population density data
Economic Data	Statistical Yearbooks	GDP, gross domestic product per capita, and the proportion of primary/secondary/tertiary industries in GDP

2.2. Methodology

2.2.1. Definition of the AS_GPI

The GPI reflects the degree of coordination between economic growth and anthropogenic atmospheric environmental pressure [24]. It measures the quantity of pollutants emitted per unit of GDP within a given period for a nation or region. The expression of the GPI is shown in Formula (1). In this study, we define the AS_GPI, which represents the relationship between GDP and anthropogenic atmospheric pollutant emissions, calculated as the ratio of anthropogenic atmospheric pollutant emissions to GDP. Its unit is 10^4 tons per 100 million Chinese yuan.

$$\mathrm{GPI} = \frac{Q_{\mathrm{emissions}}}{\mathrm{GDP}} \quad (1)$$

$Q_{\mathrm{emissions}}$ denotes pollutant emissions, while GDP represents gross domestic product.

2.2.2. An Introduction to the AFEA-AG and EA-AG Models for AS_GPI Prediction

This study utilizes provincial GDP data from all provinces across China for the period 1995–2023, alongside annual emission data for four anthropogenic air pollutants (SO_2, NOx, $PM_{2.5}$, and NMVOC). Following the methodological approach outlined in Section 2.2.1, it calculated the distribution patterns of the AS_GPIs for these four pollutants (SO_2, NOx, $PM_{2.5}$, and NMVOC) across China's 31 provinces from 1995 to 2023.

The technical roadmap for AFEA-AG established in this paper, which integrates advanced features, ensemble modeling, and ARIMA coupling, is illustrated in Figure 1. Constructed features refer to comprehensively constructed variables that are recalculated based on basic variables. For example, they include the pairwise product terms of basic features, rather than only original single input variables. First, constructed data features are established through foundational data loading. Because AS_GPI values are relatively small, input units are initially recorded in tones per hundred million yuan. Subsequent validation and results are converted back to 10^4 tons per hundred million yuan. Preliminary base data processing is then conducted. Constructed features are then calculated from the basic features. As all base data points exceed zero and exhibit skewed distributions, logarithmic transformation is applied prior to feature extraction. This transformation enables tree-based models to better capture underlying patterns. These constructed features primarily encompass basic features, interaction features, and historical time-series features (such as long-term trend decomposition features). When constructing the temporal features of the target variable, only historical data prior to the current time point are utilized, thereby preventing data leakage and eliminating the risk of data exposure. Table 2 details the classification of specific constructed features. Constructed features ultimately entering model training undergo feature importance and stability screening, as shown in Table 2. Regarding the lagged features, each pollutant-specific AS_GPI only incorporates its own lagged features during the operational phase, whereas for other data features, all pollutant AS_GPIs require these features as inputs for prediction.

The data feature set is then partitioned chronologically into training, validation, and test sets. The test set comprises all provincial feature data in 2023. The validation set contains all provincial feature data in 2022. The training set encompasses provincial feature data from 1995 to 2021. Given the continuity of time-series data and the fact that the 2023 data have undergone practical validation, no separate test set was established for the 2024 forecast. After the accuracy of the 2023 test set was validated, the 1995–2022 training set was employed for forecasting 2024, with the 2023 data serving as the validation set. To ensure no overlap between datasets, standardization was performed. Standardization parameters were derived solely from the training set, with the validation and test sets fully simulating 'true unknown data' to eliminate data leakage risks.

In addition, the Autoregressive Integrated Moving Average (ARIMA) model was employed to forecast constructed features for the prediction (test) year, providing these constructed features to both individual tree models and ensemble tree models. Subsequently, the validated predictive features were merged with the original features (excluding those from the prediction/test set years). When forecasting the features for 2023, the input data covers the period from 1995 to 2022. For forecasting the features for 2024, if the actual data of 2023 are available, the input data shall include the period from 1995 to 2023; if the actual feature data of 2023 are unavailable, the input data shall remain as the dataset spanning from 1995 to 2022, and the Auto-Regressive Integrated Moving Average (ARIMA) model shall be adopted to perform a two-step forecast for the features without actual data for 2023 to acquire the 2024 feature.

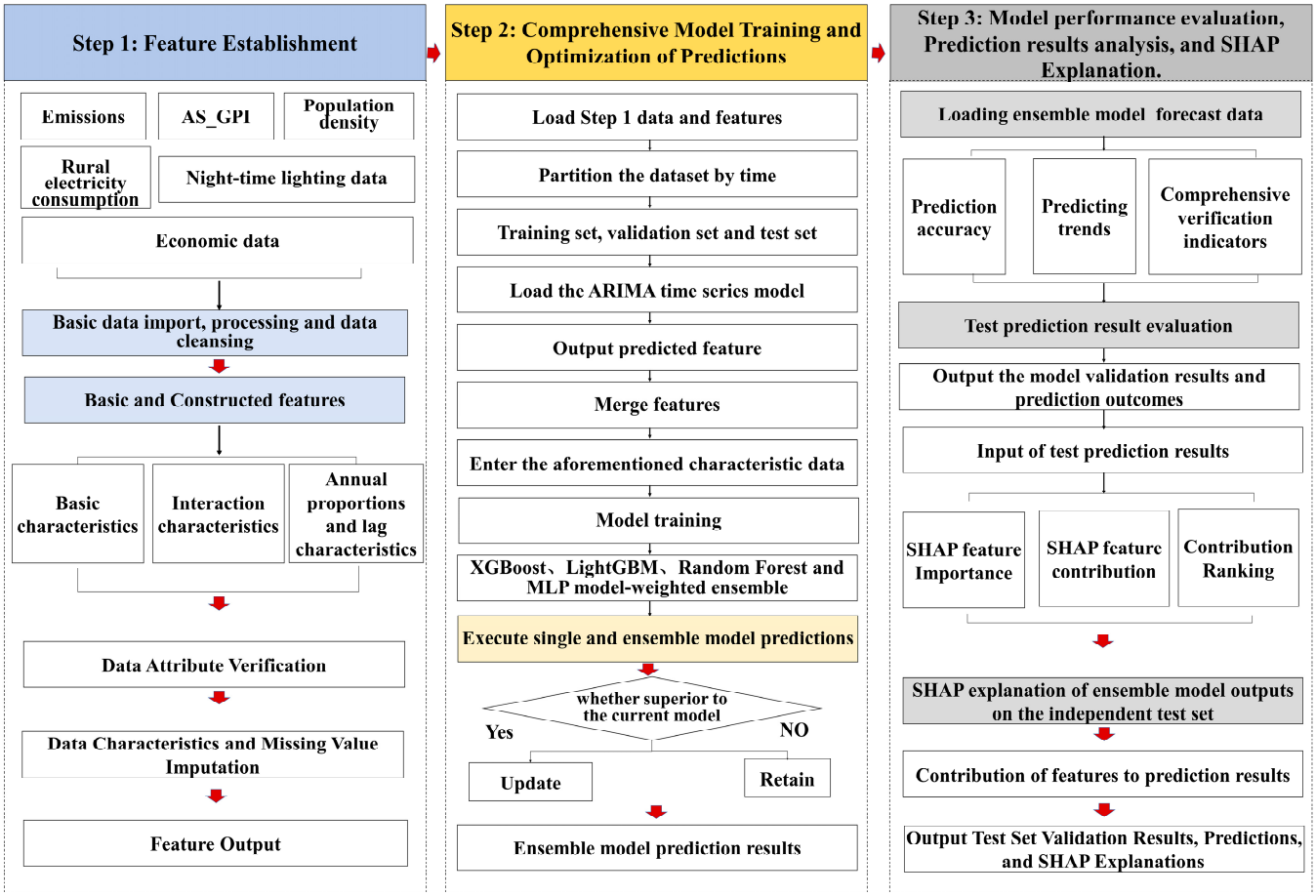


Figure 1. Technical roadmap of AFEA-AG prediction model.

These combined constructed features are then input into individual model training. The models are subsequently weighted and integrated using XGBoost [59], LightGBM, RF and MLP, with the integration weights determined by the validation set results. Dedicated weights were calculated for each sample in the ensemble model, such that models with smaller prediction errors are assigned higher weights for their corresponding samples. A maximum weight threshold is also set to prevent any single model from gaining excessive dominance. Specifically, this method dynamically calculates adaptive weights for each base learner (including XGBoost, LightGBM, RF, and MLP) for each validation sample based on the relative prediction error of each model on the validation set. Models with a lower prediction error are assigned higher weights, while constraints are imposed to limit the maximum weight of a single model and ensure the minimum weight of each model, avoiding excessive dependence on any single model and improving the robustness of the ensemble system. After calculation, the sample-specific weights and the global average ensemble weights are saved and can be directly loaded for subsequent prediction. On this basis, prediction uses the precomputed adaptive weights, enabling reliable prediction output. In addition, we observe that some provinces exhibit significantly lower GPI values compared with others. To alleviate the potential bias caused by regions with larger magnitudes, a targeted weighting strategy is applied to assign higher weights to these underrepresented provinces. Following this adjustment, model training was then performed, followed by predictions from both individual models and the ensemble model. The improvement of the current prediction models is evaluated, with the ensemble model's prediction results outputted. The target variable data are then converted back to their original scale. Subsequently, the prediction evaluation model was employed for

outcome analysis and validation, with specific verification results presented in Figure 2. Finally, utilizing the SHAP interpretability method, feature importance rankings were generated for the ensemble model's predictions, outputting feature contribution results. This identified key factors significantly influencing the AS_GPI across the four pollutant categories, providing a reference for atmospheric pollution control. The above introduces the AFEA-AG model for AS_GPI prediction.

Table 2. Constructed features.

Feature Category	Primary Construction		Explanation
Basic Data Characteristics	Emissions data, rural electricity consumption, night-time lighting data, population density economic data (as detailed in Table 1).		Data sourced from statistical yearbooks, DMSP-OLS data and MEIC inventories.
Interaction characteristics	Single pollution emission × basic data	Pollutant emissions (SO_2, NOx, NMVOC, $PM_{2.5}$) × population density, pollutant emissions × proportion of the secondary industry in GDP. Pollutant emissions × the proportion of the tertiary industry in GDP.	Multiplication of emissions and partial basic features.
	Population density × some basic characteristics	Population density × regional gross domestic product. Population density × night-time light data. Population density × gross domestic product per capita.	The product of population density and economic/environmental indicators.
	Regional gross domestic product × selected fundamental characteristics	Regional gross domestic product × night-time lighting data.	Regional gross domestic product multiplied by the night-time illumination.
	Night-time lighting data	Night-time lighting data × gross domestic product per capita.	Night-time lighting data multiplied by the gross domestic product per capita.
Annual ratio characteristics	Annual ratio characteristic of per capita regional GDP		Annual proportion characteristics of per capita GDP in provinces.
Temporal characteristics	Temporal decomposition characteristics of AS_GPI history for four pollutants	AS_GPI lagging one-year feature.	The temporal variation characteristics of AS_GPI time series.

The input data features for the EA-AG model comprise only fundamental data characteristics, excluding other constructed features. Apart from the aforementioned differences, all other settings remain identical to those of the AFEA-AG model.

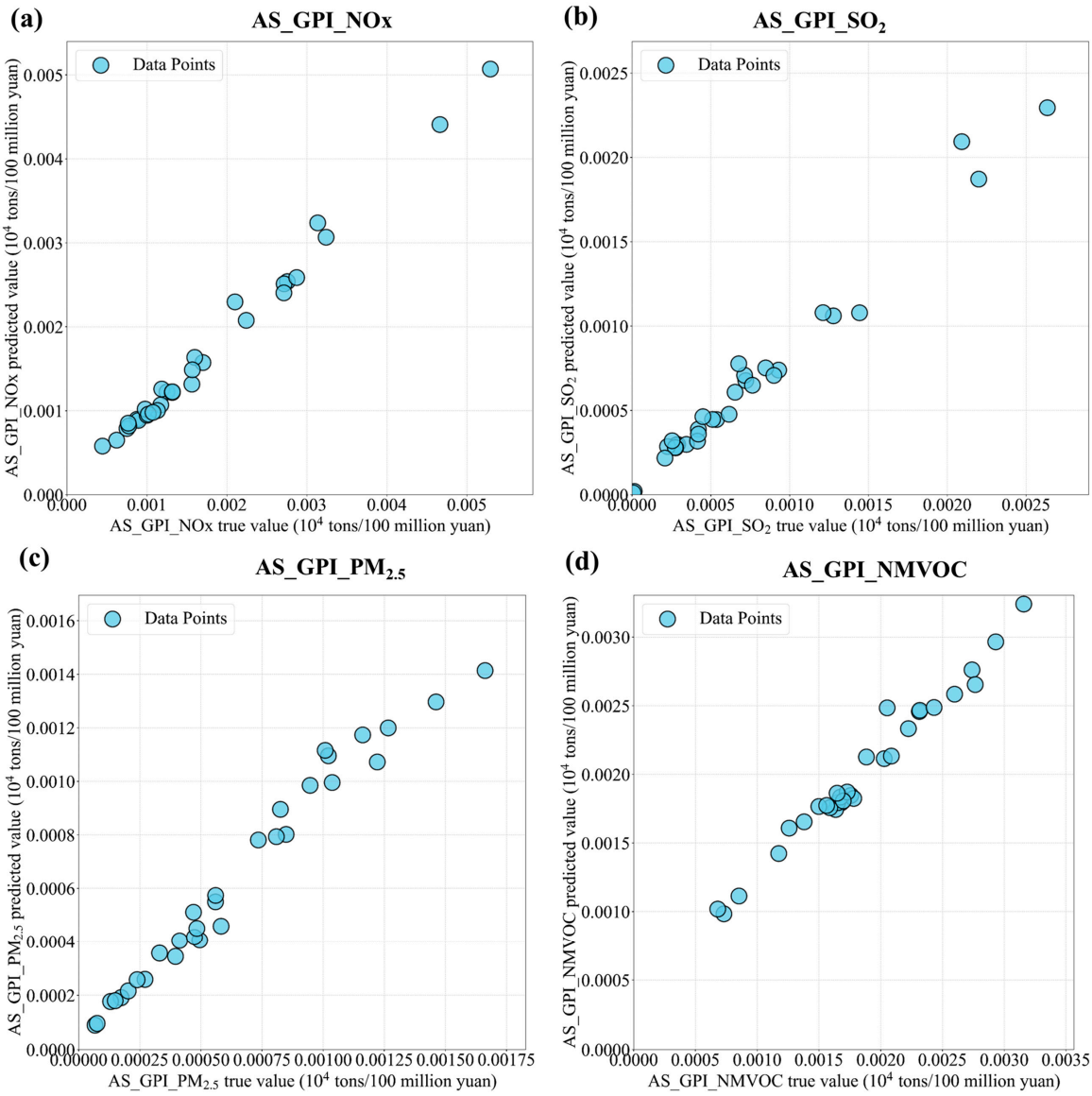


Figure 2. Model validation results for AS_GPI of four pollutants in the 2023 test dataset based on AFEA-AG model. (**a**) Model validation results of AS_GPI_NOx; (**b**) Model validation results of AS_GPI_SO_2; (**c**) Model validation results of AS_GPI_$PM_{2.5}$; (**d**) Model validation results of AS_GPI_NMVOC.

3. Results and Discussion

3.1. AS_GPI AFEA-AG Predictive Model Validation

To validate the efficacy of machine learning methods in predicting the AS_GPI for four pollutants, this study trained and tested an ensemble model—based on the weighted coupling of RF, XGBoost, LightGBM and MLP—alongside these four traditional models on the same dataset. A unified set of performance evaluation metrics was employed: coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). As illustrated in Figure 2, the comparison between the prediction results of the four machine learning models and the ensemble model against the actual values on the test set is presented. When combined with the model validation metrics in Table 3, the results for the AS_GPIs of the four pollutants indicate that the ensemble model demonstrates significantly superior predictive performance for the AS_GPI_NOx and AS_GPI_$PM_{2.5}$ compared to other pollutants on the test set. Specifically, the ensemble model achieved a high R^2, exceeding 0.95 on the test sets for these two pollutants,

alongside extremely low error metrics: RMSE values of 0.0001 and 0.0001, respectively, MAE values of 0.0001 and 0.0001, respectively, and MAPE values of 0.0744 and 0.1087. For the AS_GPI_NMVOC, the test set R^2 was 0.8984, with an RMSE of 0.0002 and MAE of 0.0002. Regarding the AS_GPI_SO_2, the ensemble model achieved an RMSE of 0.0001 and MAE of 0.0001. Validation results indicate that the AFEA-AG model performs robustly across multiple pollutants. The four individual models (XGBoost, LightGBM, RF and MLP) exhibited distinct performance across the four AS_GPI metrics, indicating that when employing a single model for multiple AS_GPI predictions, it lacks universal applicability. By comparing the prediction results of four individual models and the ensemble model, it was found that in the majority of AS_GPI predictions, the ensemble model demonstrated excellent R^2 fit and extremely low error rates. It effectively integrated the strengths of the four individual models, achieving relatively accurate prediction outcomes, and possesses a high degree of universality.

Table 3. Validation results for four pollutants using the AFEA-AG model in 2023 test set. The units of RMSE and MAE are the same as those of the target variable, while MAPE is dimensionless.

AS_GPI	Indicator	XGBoost	LightGBM	RF	MLP	Ensemble
AS_GPI_NOx	R^2	0.9662	0.9744	0.9823	0.9558	0.9837
	RMSE	0.0002	0.0002	0.0002	0.0002	0.0001
	MAE	0.0002	0.0001	0.0001	0.0002	0.0001
	MAPE	0.0991	0.0881	0.0803	0.1371	0.0744
AS_GPI_SO_2	R^2	0.9355	0.8963	0.9413	0.9051	0.9509
	RMSE	0.0002	0.0002	0.0002	0.0002	0.0001
	MAE	0.0001	0.0001	0.0001	0.0001	0.0001
	MAPE	0.1596	0.1919	0.1874	0.2234	0.1488
AS_GPI_$PM_{2.5}$	R^2	0.9454	0.9480	0.9535	0.9152	0.9668
	RMSE	0.0001	0.0001	0.0001	0.0001	0.0001
	MAE	0.0001	0.0001	0.0001	0.0001	0.0001
	MAPE	0.1153	0.1240	0.1247	0.2440	0.1087
AS_GPI_NMVOC	R^2	0.8659	0.8944	0.8903	0.7716	0.8984
	RMSE	0.0002	0.0002	0.0002	0.0003	0.0002
	MAE	0.0002	0.0002	0.0002	0.0002	0.0002
	MAPE	0.1309	0.1120	0.1239	0.1468	0.1168

Figure 2 presents model validation results for the AS_GPI of four pollutants in the 2023 test dataset based on the AFEA-AG model. Most data points cluster tightly around the diagonal of the graph, with only some points deviating slightly from it. This distribution indicates that the model has high prediction accuracy. It was found that for some provinces whose numerical values have some differences from those of most other provinces, the deviation from the true values is even greater. The results show that there is a certain similarity between the predicted values in the test set and the actual values.

3.2. Comparative Analysis of Two GPI Prediction Models

To validate the effectiveness of incorporating constructed features, Figure 3 compares the forecast results of the AFEA-AG model against the EA-AG model. Both models share identical conditions except for the inclusion of constructed features. Validation results for the four pollutants using the EA-AG model on the 2023 test set are available in Table S1 of the Supplementary Materials. As shown in Figure 3a,b, the prediction results of the two AS_GPI models exhibit certain similarities. Among the four pollutants' AS_GPI prediction results, the AS_GPI_SO_2 for China's 31 provinces demonstrated greater largest values in the mean relative difference between predicted and actual values compared

to the other three AS_GPIs. Furthermore, the median relative difference between the AS_GPI_NMVOC predicted and actual values was higher than that of the other three AS_GPIs. Regarding the predicted results of the AS_GPI_NMVOC, the outputs from both prediction models are generally slightly higher than the true values. For the other AS_GPIs, some predictions are relatively low while others are relatively high, but overall, they fall within a reasonably acceptable range. A comparison of the validation results on the test set for the two prediction models revealed that the AFEA-AG model exhibited a smaller discrepancy between the predicted values of the four AS_GPIs and the actual values than the EA-AG model.

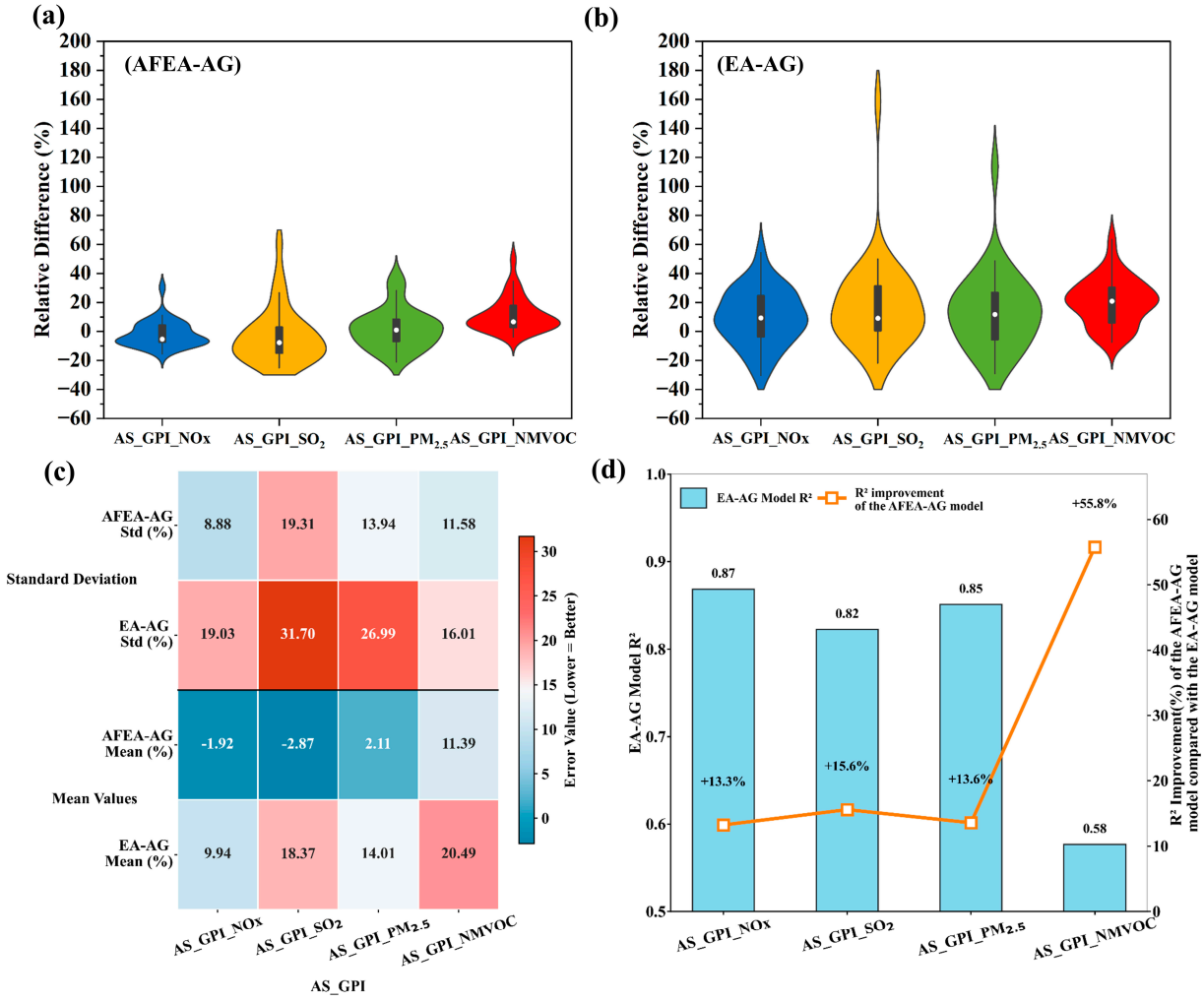


Figure 3. The comparation results of the prediction outcomes on the 2023 test set derived from the AFEA-AG model, compared to those from the EA-AG prediction model. (**a**) The relative difference between the predicted values based on the AFEA-AG model and true values. (**b**) The relative difference between the predicted values derived from the EA-AG model and true values. (**c**) Population standard deviation and average results of relative differences between predicted values and actual values from the AFEA-AG model and EA-AG prediction model. (**d**) The correlation results between the prediction and true values based on the EA-AG prediction model, and correlation improvement based on the AFEA-AG model compared to the EA-AG prediction model. The relative difference is defined as (predicted − true)/true × 100%.

Figure 3c illustrates the population standard deviation and mean of prediction relative difference for both models. Overall, the absolute mean of prediction relative difference of the AFEA-AG model was lower than that of the EA-AG model for the AS_GPI_SO_2 and AS_GPI_$PM_{2.5}$. Moreover, the standard deviation of prediction relative differences

for the AFEA-AG model is lower than that of the EA-AG model across all four AS_GPI types. As shown in Figure 3d, the R^2 between the predicted and true values is higher for the AFEA-AG model, which outperforms the EA-AG model across all four AS_GPI types, achieving an improvement rate of 55.8% for the AS_GPI_NMVOC.

A comprehensive comparison shows that the prediction results of the AFEA-AG model on the 2023 test set are more stable than those of the EA-AG model, exhibiting a closer trend to the true values. Overall, the average simulation validation results for the four pollutants demonstrate that the AFEA-AG prediction model has more reasonable outcomes than the EA-AG model.

In summary, compared to the other four independent machine learning models examined in this study (XGBoost, LightGBM, RF and MLP), the ensemble model and the constructed feature-coupled prediction model strike a balance between performance and stability. They mitigate single-point risks by achieving equilibrium among top-performing individual models rather than pursuing absolute maximum R^2 values, thereby balancing model characteristics. By circumventing the limitations inherent in individual models, it ultimately achieves a triple enhancement in accuracy, robustness, and result reliability.

3.3. Comprehensive Analysis of the AFEA-AG Model's 2024 Prediction Results and 1995–2023 Historical Data

Based on the AFEA-AG forecasting model proposed herein, the AS_GPIs for four pollutant categories across 31 provinces nationwide were projected for 2024. Figure 4a presents the 2024 AS_GPIs for these four pollutants in major pollution control cities within Beijing, Hebei, Tianjin, Shanxi, Shaanxi, Henan, Jiangsu, Anhui, Zhejiang, Shandong, Shanghai, and Guangdong. Hebei Province's AS_GPI_NMVOC and AS_GPI_NOx exhibit higher values compared to other key air pollution control cities, indicating that this region requires rationalized pollution control strategies for NOx and NMVOC, respectively. Among the air pollution control provinces, Shanxi Province's AS_GPI_SO_2 values exceed those of other provincial-level units, indicating that this region requires enhanced coordinated management of anthropogenic emissions of SO_2.

To better understand the AS_GPIs of pollutants across all provinces, the AS_GPIs for four pollutants in non-key air pollution control regions were also analyzed. As shown in Figure 4b, within the same pollutant categories, in 2024, Guizhou's AS_GPI_SO_2, Ningxia's AS_GPI_NOx, and Heilongjiang's AS_GPI_$PM_{2.5}$ were higher than those of the other provinces depicted. Tibet's AS_GPI_SO_2 was lower than the values in other non-priority pollution control zones. Compared to non-priority pollution control regions, the overall AS_GPI values across all four categories are lower in priority pollution control regions. This indicates that pollution control measures are more effective in priority pollution control provinces. Comparing Figure 4a,b reveals that, overall, the AS_GPI for NOx and NMVOCs in most non-priority pollution control provinces was slightly higher than the AS_GPI for the other two pollutants. This indicates that, while maintaining rapid economic development, it remains necessary to strengthen effective atmospheric pollution control for NOx and NMVOCs in non-priority pollution control provinces. Furthermore, significant disparities exist in the AS_GPI values for NOx across 19 provinces in 2024, whereas $PM_{2.5}$ AS_GPI values exhibit smaller inter-provincial variations. This underscores the necessity for differentiated atmospheric pollution control strategies targeting NOx across distinct provincial contexts.

Moreover, to analyze the 2024 AS_GPI projections for each province nationwide, Figure 4c illustrates the relative differences between the predicted AS_GPI values for four pollutants in each province for 2024 and the actual values recorded in 2023. Here, the relative difference is defined as (predicted value for 2024 − actual value for 2023)/actual value for 2023 × 100%. Comparisons reveal that the median relative differences between

the 2024 AS_GPI predictions and the 2023 actual values for all four pollutants are negative. This indicates that the 2024 AS_GPI predictions for these pollutants show a downward trend across most provinces, suggesting an anticipated improvement in environmental quality in 2024 compared to 2023. Additionally, the median relative difference for the AS_GPI_NMVOC was found to be higher than other AS_GPIs, indicating that the projected improvement in anthropogenic environmental quality for NMVOC is less pronounced compared to other pollutants.

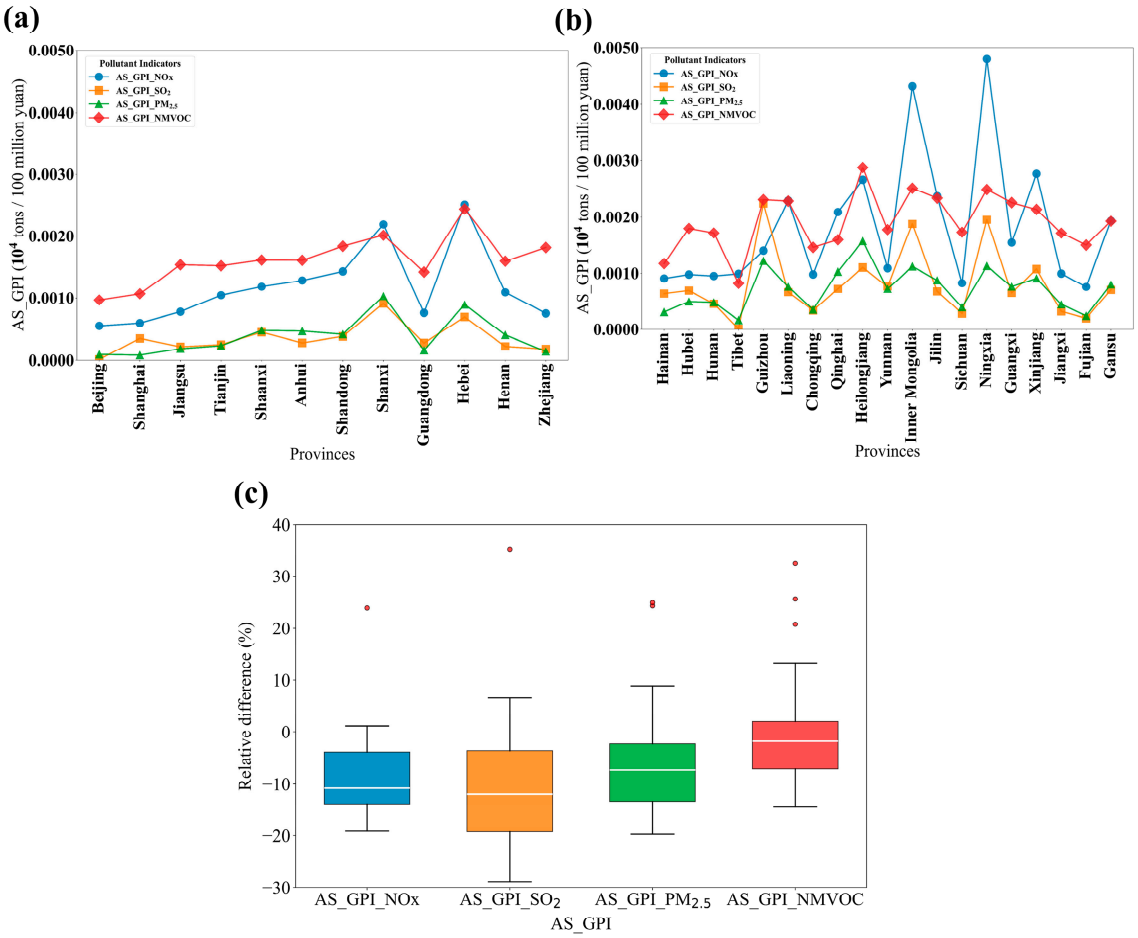


Figure 4. Distribution of AS_GPI values in 2024 and 2025 across nationally designated key pollution control cities and non-key cities. (**a**) AS_GPI in 2024 across nationally designated key pollution control cities. (**b**) The 2024 AS_GPI prediction value distribution across non-key national pollution control cities. (**c**) The relative differences between the AS_GPI prediction values of four pollutants in each province in 2024 and the actual values in 2023.

Regardless of whether they are key or non-key air pollution control cites, a comparison of the AS_GPIs for the four pollutants reveals that NOx and NMVOC exhibit higher AS_GPI values than $PM_{2.5}$ and SO_2 in most cities. This indicates that NMVOC and NOx will remain key pollutants requiring focused emission reductions in the future. Overall, the AS_GPI projections reveal significant disparities in pollution control effectiveness between key and non-key regions. Areas such as Shanxi, Hebei, Guizhou, and Inner Mongolia face considerable pollution control pressures, whereas Beijing, Shanghai, Tibet, and Fujian maintain relatively low levels. Future emission reduction efforts should emphasize regional differentiation and pollutant category differentiation, with particular focus on enhancing the management of NOx and NMVOC [60].

Figure 5 shows the relative change rate of the AS_GPI_NOx, AS_GPI_SO_2, AS_GPI_$PM_{2.5}$ and AS_GPI_NMVOC in 2024 compared to 1995. The 2024 data are the predicted results from the AFEA-AG model. As illustrated in Figure 5, the absolute sum of the AS_GPI changes for the four pollutants in Beijing is notably higher than that of other provincial-level administrative units. This indicates that Beijing has maintained relatively stringent control over anthropogenic pollution sources. Conversely, Heilongjiang Province exhibits the lowest absolute sum of relative AS_GPI change rates for the four pollutants, suggesting that the quality of anthropogenic pollution management in relation to economic development within this province requires further enhancement compared to other regions. Tibet has the largest absolute value of the change rate of the AS_GPI_NMVOC. This indicates that targeted pollution control strategies for different pollutants should be further implemented across all provinces in China.

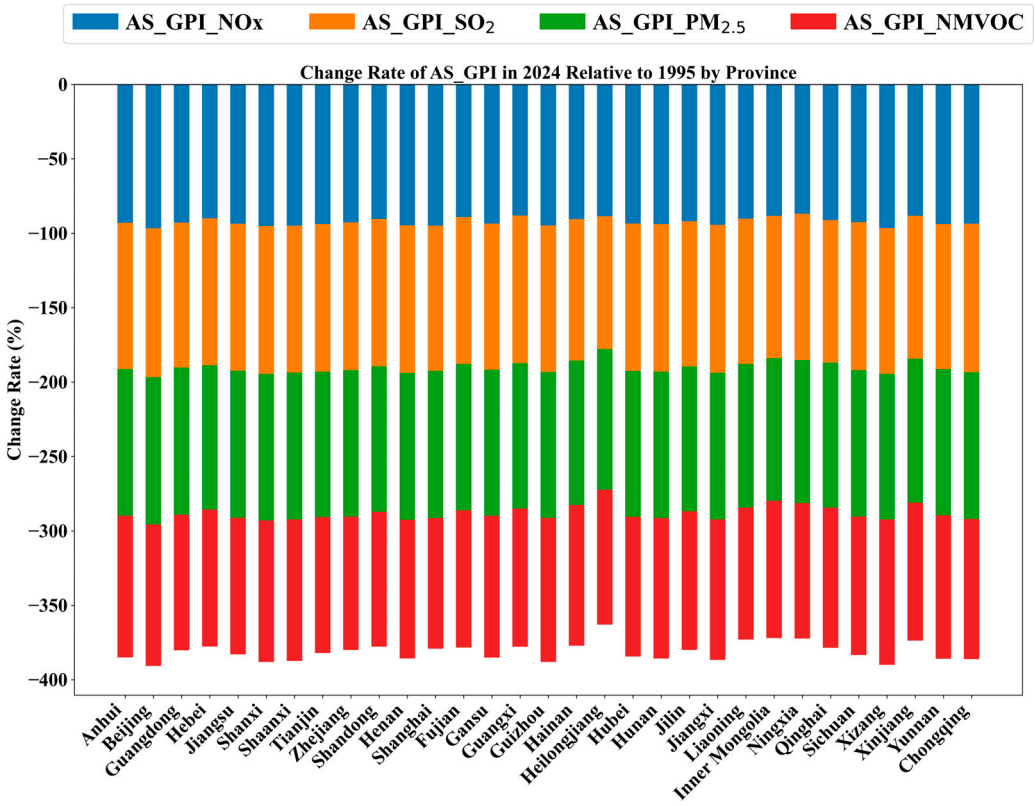


Figure 5. Relative change rate of AS_GPI_NOx, AS_GPI_SO_2, AS_GPI_$PM_{2.5}$ and AS_GPI_NMVOC in 2024 compared to 1995. Historical data from 1995 to 2023 are calculated values, while data for 2024 are predicted values.

Figure 6 shows the spatial distribution of the coefficient of variation (CV) of the AS_GPI for four pollutants across Chinese provinces from 1995 to 2024. The CV was calculated as the ratio of the population standard deviation to the mean, as the data cover the complete period 1995–2024.The 2024 data are the predicted results from the AFEA-AG model. The CV quantifies the temporal variability of AS_GPI values, with higher values indicating greater inter-annual fluctuations in the AS_GPI. From 1995 to 2024, the CV of the AS_GPI for the four atmospheric pollutants in China exhibited significant pollutant-specific differentiation and spatial heterogeneity: the AS_GPI_$PM_{2.5}$ recorded the highest overall CV, with values ranging from 1.00 to 1.24 in most provinces across the country. This indicates that the AS_GPI_$PM_{2.5}$ experienced the most dramatic inter-annual fluctuations during the study period, and was most significantly affected by policies including pollution control and energy structure adjustment. The AS_GPI_NOx showed the lowest overall

CV, with most provinces recording values between 0.56 and 0.85, and only a few provinces exhibiting high CV values. The CV values of the AS_GPI_SO_2 and AS_GPI_NMVOC fell at an intermediate level. The spatial differences in CV across the four AS_GPIs essentially reflect the implementation intensity of emission reduction policies for different pollutants: the AS_GPI_$PM_{2.5}$ and AS_GPI_SO_2, as core indicators of China's air pollution prevention and control, were subject to strong policy interventions, resulting in dramatic inter-annual fluctuations in the AS_GPI_$PM_{2.5}$. For all four pollutants, high-CV regions of the AS_GPI were concentrated in economically developed, densely populated, and industrially intensive eastern and central China, while low-CV regions were stably distributed in remote, economically underdeveloped western and northeastern provinces with low anthropogenic activity. This pattern directly reflects regional differences in anthropogenic pollution control policies. These findings highlight the need to formulate differentiated and targeted regional pollution control strategies tailored to the fluctuation characteristics of different pollutants in the future.

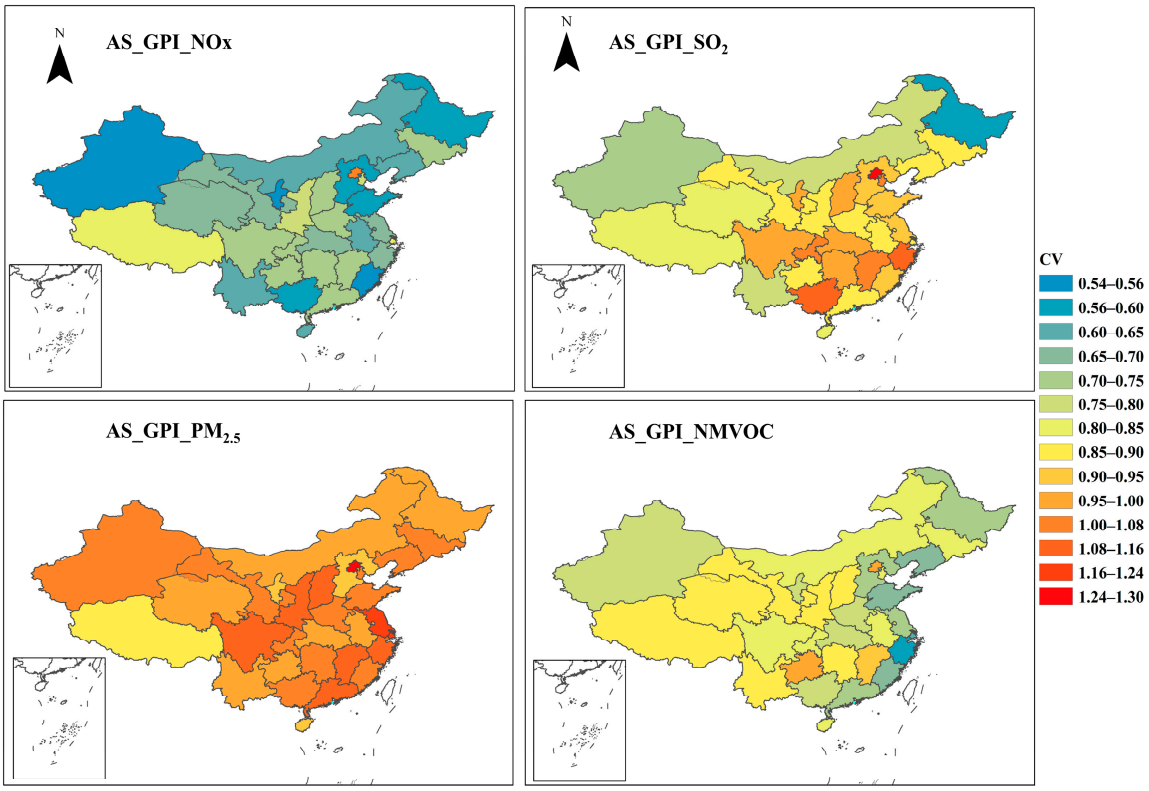


Figure 6. The coefficient of variation (CV) of the AS_GPI for four types of pollutants in each province of China from 1995 to 2024.

3.4. A Study on the AS_GPI Interpretable Predictor Variable Analysis for Four Pollutants

To better reveal the key features associated with AS_GPI prediction outcomes, Figure 7 displays the significant feature outputs and SHAP values derived from the AFEA-AG model predictions. As illustrated, SHAP values exhibit distribution on both positive and negative sides, indicating that the influence of these variables may vary across samples due to regional differences in economic scale and governance efficiency. Figure 7 demonstrates that the feature variables influencing prediction outcomes and their respective contribution levels vary across different pollutants. For AS_GPI_NOx predictions, the top five contributing features are: its own one-year lagged feature, its per capita regional GDP, NMVOC emissions, regional gross domestic product, and NOx emissions. The predictive accuracy of the AS_GPI_SO_2 relies more heavily on its own historical lagged features,

SO_2 emissions × the proportion of the secondary industry in GDP, SO_2 emissions, and the proportion of the tertiary industry in GDP. Additionally, the interaction between SO_2 emissions and the tertiary sector's share of GDP plays a significant role. For the AS_GPI_$PM_{2.5}$, predictive accuracy primarily hinges on its own historical lag characteristics, per capita regional GDP, $PM_{2.5}$ emissions, the interaction between $PM_{2.5}$ emissions and the proportion of the tertiary industry in GDP, and NMVOC emissions.

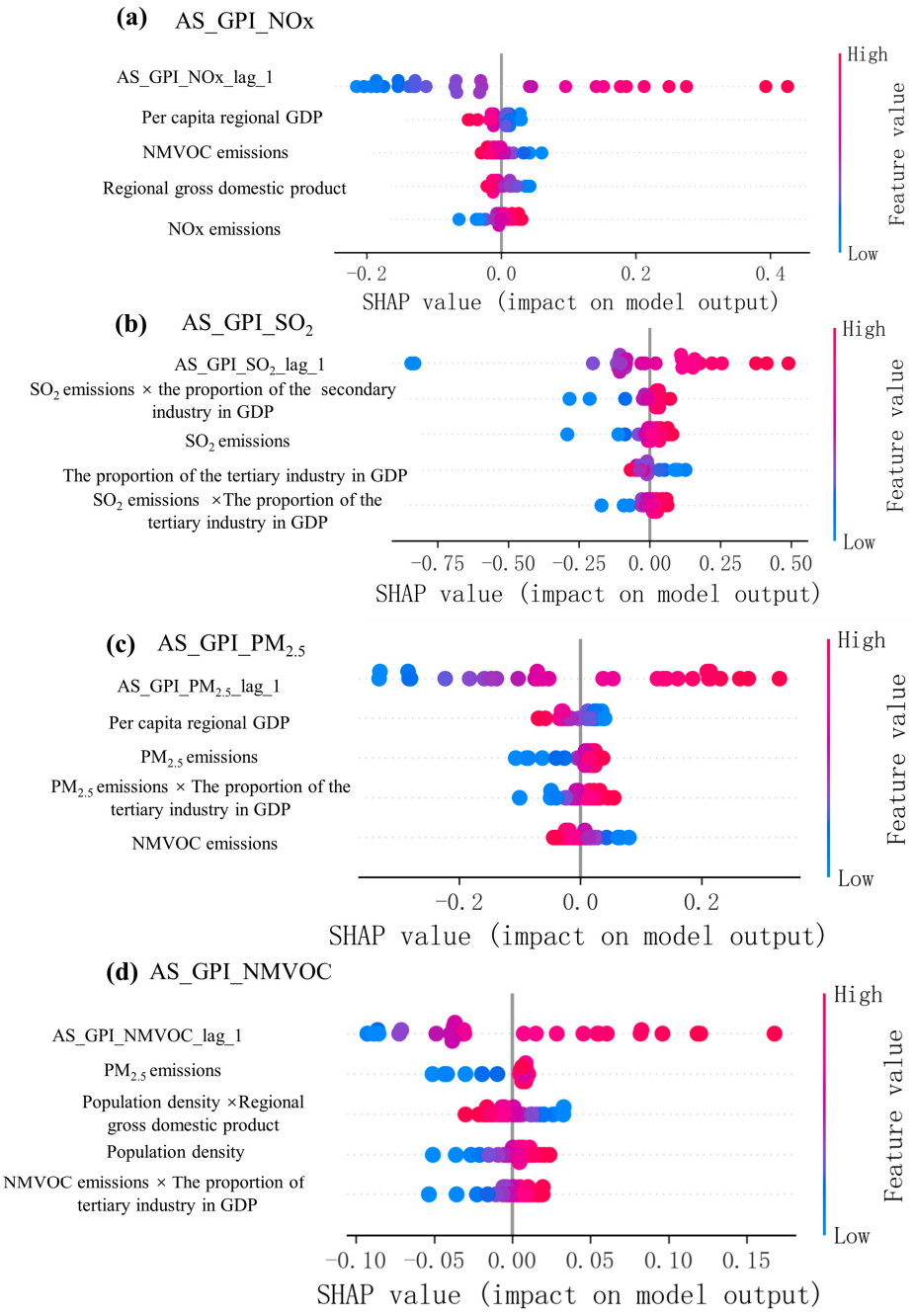


Figure 7. SHAP contribution explanatory variable distribution of AS_GPI for four pollutants in the 2023 test set predicted by the AFEA-AG model. (**a**) The SHAP values of AS_GPI_NOx. (**b**) The SHAP values of AS_GPI_SO_2. (**c**) The SHAP values of AS_GPI_$PM_{2.5}$. (**d**) The SHAP values of AS_GPI_NMVOC.

As for the AS_GPI_NMVOC, similar to the other three AS_GPI models. The historical lag characteristic of the AS_GPI_NMVOC itself contributes the most to the prediction accuracy of AS_GPI_NMVOC emissions. $PM_{2.5}$ emissions play crucial roles in predict-

ing AS_GPI_NMVOC outcomes, suggesting potential synergistic mechanisms between NMVOC and other pollutants. Additionally, population density × regional gross domestic product, population density, and NMVOC emissions × the proportion of the tertiary industry in GDP also exert significant influence. Comparing the four AS_GPI metrics reveals that different characteristics exhibit a wide range of contribution values to the AS_GPI_SO_2, with many characteristics showing significant negative contributions to the AS_GPI_SO_2.

A comprehensive comparison of the contribution characteristics of the four AS_GPIs reveals that the historical lagging nature of the AS_GPI itself consistently ranks foremost in its contribution to predictive performance. Overall, the economic factors influencing the predictions of the four AS_GPIs all exhibit similar contribution patterns. Economic scale indicators (such as regional GDP and per capita GDP), as well as the proportions of the three industrial structures, have always been important factors. There is usually a significant interaction effect between economic indicators, industrial structure proportions and pollutant emissions, jointly reflecting the stable and widespread influence of economic development and industrial structure on the intensity of anthropogenic emissions. Concurrently, socio-economic characteristics, such as the proportion of primary, secondary and tertiary industries alongside regional GDP and variations in population density and pollutant emissions, also play a significant role in forecasting. These elements collectively reflect the combined effect of regional economic development level, industrial layout, and human activity intensity, demonstrating the collaborative driving effect of economic scale, industrial structure, and population concentration on the spatiotemporal variation in pollutant emissions. They are the key socio-economic foundation for enhancing the explanatory power and predictive rationality of the model.

3.5. Uncertainty Analysis and Future Improvements

This study established an ensemble model based on four weighted algorithms—XGBoost, LightGBM, RF and MLP—combined with the ARIMA time-series model. By integrating the SHAP explainability model and constructed features, it performed numerical prediction and contribution identification for the AS_GPI values of four typical pollutants. Although the AS_GPI predictions and contribution attribution results for the four pollutants demonstrated reasonable interpretability, discrepancies persist between the predicted and actual AS_GPI values. Notably, greater prediction uncertainty is observed for provinces whose values differ substantially from the overall provincial average, indicating inherent uncertainty in the forecasts. Key uncertainties in this predictive study stem primarily from some sources: parameter selection within traditional machine learning models, parameter configuration of the time-series model, the volume of feature variable data, inherent uncertainties in the emission inventory, the integration methodology of the ensemble model, uncertainty in statistical data and atmospheric environmental governance policies. In particular, the emission inventory used in this study is a widely recognized estimated dataset rather than in situ observations, and its inherent uncertainty will be inherited by the model predictions. Given the potential impact of policy shifts and regulatory controls on future predictions, the model needs to conduct cautious extrapolation. The uncertainty mentioned above is also an area where this research needs improvement in the future.

To address these limitations, future research can make improvements in several aspects. Higher-resolution emission inventories and foundational data can be integrated to reduce data-related uncertainties. More cases at different scales can be studied to enhance the reliability of the model. When larger and higher-resolution datasets become available, more complex deep learning architectures could be explored to capture finer spatiotemporal patterns and potentially improve predictive performance. Furthermore, the proposed framework can be extended to predict the AS_GPI for more pollutants to enable a more

comprehensive assessment. Additionally, more refined spatial stratification strategies can be designed to enhance the stability of predictions for regions with strong heterogeneity. Explicitly incorporating technology-related features as input variables in machine learning models could further improve predictive performance by capturing the effects of pollution control measures. These improvements will help to enhance the accuracy, robustness, and generalization ability of the prediction framework in future applications.

4. Conclusions

To better investigate the relationship between economic development and the co-evolution of atmospheric pollution, this paper utilizes provincial-level data from China spanning from 1995 to 2023. Selecting NOx, SO_2, $PM_{2.5}$, and NMVOC as representative pollutants, it constructs a predictive model (AFEA-AG) for the AS_GPIs of these four pollutants. The AFEA-AG model employs weighted ensembles of XGBoost, LightGBM, RF and MLP, coupled with the SHAP model and integrated with an ARIMA time-series model. It combines these with constructed feature elements derived from fundamental indicators (economic output, industrial structure, energy consumption, night-time light emissions, etc.) to forecast the AS_GPIs for typical anthropogenic air pollutants across China's 31 provinces. Research indicates that the AFEA-AG model established herein can accurately predict the AS_GPI for four typical pollutants (SO_2, NOx, $PM_{2.5}$, and NMVOC), achieving R^2 values exceeding 0.95 in predicting the AS_GPI for NOx and $PM_{2.5}$. A comparison of AS_GPI prediction models with and without constructed features revealed that the prediction model incorporating constructed features coupled with ARIMA yields more reasonable outcomes than the model without such constructed features. The correlation coefficient R^2 for prediction errors in the AFEA-AG model outperforms the EA-AG model across all four AS_GPI types, achieving an improvement rate of 55.8% for the AS_GPI_NMVOC.

Feature contribution analysis revealed differing key contributors across pollutants, with historical decomposition trends, economic-related characteristics and interaction features playing prominent roles in prediction. From 1995 to 2024, the AS_GPI values for the four pollutants in the majority of provinces exhibited a downward trend. This outcome demonstrates that the air pollution prevention and control measures rolled out by the state have played a positive role in enhancing environmental pollution governance effectiveness. For these four AS_GPIs, the regions with a high CV of the AS_GPI are mainly concentrated in the eastern and central parts of China, which are economically developed, densely populated and industrially concentrated. However, the regions with a low CV are consistently distributed in the remote, economically underdeveloped and less human-occupied western and northeastern provinces. Moreover, there are certain differences in the CV among the four AS_GPIs. A comprehensive comparison of the contribution characteristics of the AS_GPI across four pollutants reveals that the historical lagging nature of the AS_GPI itself consistently ranks foremost in its contribution to predictive performance. Concurrently, socio-economic characteristics such as the proportion of primary, secondary and tertiary industries alongside regional GDP, as well as variations in population density ratios and pollutant emissions, also play a significant role in forecasting.

This study not only achieves effective prediction of the AS_GPI for four pollutants but also identifies key factors influencing the AS_GPI of these pollutants. The findings provide data support for local governments to assess regional green development quality and formulate differentiated environmental governance policies, while also offering a reference for advancing ecological civilization construction and achieving sustainable development goals in China.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/atmos17050480/s1, Table S1: Validation results for four pollutants using the EA-AG model in the 2023 test set.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available on request.

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

Abbreviations

The following abbreviations are used in this manuscript:

AS_GPI	Anthropogenic economic pollutant emissions index
SO_2	Sulfur dioxide
NOx	Nitrogen oxides
$PM_{2.5}$	Fine particulate matter
NMVOC	Non-methane volatile organic compounds
AS_GPI_NOx	Anthropogenic economic NOx pollutant emissions index
AS_GPI_$PM_{2.5}$	Anthropogenic economic $PM_{2.5}$ pollutant emissions index
AS_GPI_NMVOC	Anthropogenic economic NMVOC pollutant emissions index
AS_GPI_SO_2	Anthropogenic economic SO_2 pollutant emissions index
AFEA-AG	Anthropogenic economic pollutant emission index forecasting model integrating basic and constructed features, ensemble machine learning, and ARIMA coupling
EA-AG	Anthropogenic economic pollutant emissions index forecasting model that integrates basic, ensemble learning and ARIMA coupling
XGBoost	Extreme gradient boosting
LightGBM	Light gradient boosting machine
RF	Random forest
MLP	Multi-layer perceptron
SHAP	Shapley additive explanations
GDP	Gross domestic product
DMSP-OLS	Defense Meteorological Satellite Programmer Operational Line-Scanning System
LSTM	Long short-term memory
VARMA	Vector autoregressive moving average
MEIC	Multi-scale Emission Inventory Model for China
SNPP-VIIRS	Suomi National Polar-orbiting Partnership-Visible Infrared Imaging Radiometer Suite
GPI	Gross pollutant output index
R	Correlation coefficient
RMSE	Root mean square error
MAE	Mean absolute error
MAPE	Mean absolute percentage error
BTH	Beijing–Tianjin–Hebei
ARIMA	Autoregressive integrated moving average
CV	Coefficient of variation

References

1. Li, L.; He, M.; Liang, X.; Deng, H.; Yang, R. How Does the Impact of a Two-Stage Air Pollution Control Policy on Air Quality Different? Evidence from 258 Cities in China. *Atmos. Pollut. Res.* **2024**, *15*, 102148. [CrossRef]
2. Xue, T.; Wang, R.; Wang, M.; Wang, Y.; Tong, D.; Meng, X.; Huang, C.; Ai, S.; Li, F.; Cao, J.; et al. Health Benefits from the Rapid Reduction in Ambient Exposure to Air Pollutants after China's Clean Air Actions: Progress in Efficacy and Geographic Equality. *Natl. Sci. Rev.* **2024**, *11*, nwad263. [CrossRef] [PubMed]
3. Ma, M.; Liu, M.; Song, X.; Liu, M.; Fan, W.; Wang, Y.; Xing, H.; Meng, F.; Lv, Y. Spatiotemporal Patterns and Quantitative Analysis of Influencing Factors of PM2.5 and O_3 Pollution in the North China Plain. *Atmos. Pollut. Res.* **2024**, *15*, 101950. [CrossRef]
4. Shao, Z.; Zheng, X.; Zhao, J.; Liu, Y. Evaluating the Health Impact of Air Pollution Control Strategies and Synergies among PM2.5 and O_3 Pollution in Beijing-Tianjin-Hebei Region, China. *Environ. Res.* **2025**, *274*, 121348. [CrossRef]
5. Iram, S.; Qaisar, I.; Shabbir, R.; Pomee, M.S.; Schmidt, M.; Hertig, E. Impact of Air Pollution and Smog on Human Health in Pakistan: A Systematic Review. *Environments* **2025**, *12*, 46. [CrossRef]
6. Kebe, M.; Traore, A.; Sow, M.; Fall, S.; Tahri, M. Human Health Risk Evaluation of Particle Air Pollution (PM10 and PM2.5) and Heavy Metals in Dakar's Two Urban Areas. *Asian J. Atmos. Environ.* **2025**, *19*, 7. [CrossRef]
7. Chen, S.; Yu, W.; Shen, Y.; Lu, L.; Meng, X.; Liu, J. Unraveling the Mechanisms Underlying Air Pollution-Induced Dysfunction of the Oral-Gut-Brain Axis: Implications for Human Health and Well-Being. *Asian Biomed.* **2025**, *19*, 21–35. [CrossRef] [PubMed]
8. Zhang, Q.; Zheng, Y.; Tong, D.; Shao, M.; Wang, S.; Zhang, Y.; Xu, X.; Wang, J.; He, H.; Liu, W.; et al. Drivers of Improved PM2.5 Air Quality in China from 2013 to 2017. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 24463–24469. [CrossRef]
9. Li, Y.; Chen, B.; Yang, S.; Jiao, Z.; Zhang, M.; Yang, Y.; Gao, Y. Advances in Environmental Pollutant Detection Techniques: Enhancing Public Health Monitoring and Risk Assessment. *Environ. Int.* **2025**, *197*, 109365. [CrossRef]
10. Yu, B.; Fang, D. Decoupling Economic Growth from Energy-Related PM2.5 Emissions in China: A GDIM-Based Indicator Decomposition. *Ecol. Indic.* **2021**, *127*, 107795. [CrossRef]
11. Liu, H.; Li, X.; Zheng, S. Balancing Economic Growth and Environmental Conservation: Assessing Supportive Policies in Resources-Based Cities in China. *Systems* **2024**, *12*, 521. [CrossRef]
12. Ye, Q.; Li, J.; Tang, X.; Chen, X.; Kong, L.; Chen, H.; Yang, W.; Du, H.; Pan, X.; Wang, W.; et al. Uncertainties in the Simulated Intercontinental Transport of Air Pollutants in the Springtime from Emission and Meteorological Inputs. *Atmos. Environ.* **2023**, *293*, 119431. [CrossRef]
13. Chen, Y.; Xu, R.; Qu, G.; Li, Z.; Yang, Y.; Xiang, K.; Li, J. A Review of Research on Coordinated Air Pollution Control and Carbon Dioxide Reduction in the Silicon Industry: Coordinated Accounting and Emission/Control. *Silicon* **2025**, *17*, 935–951. [CrossRef]
14. Tong, Y.; Guo, X.; Guo, F.; Yue, T. Advances in Primary Air Pollutant Emissions of Industrial Boilers in China: Emission Characteristics, Emission Inventories, and Mitigation Potentials. *Sustainability* **2025**, *17*, 1987. [CrossRef]
15. Costa, M.A.M.; Albini, G.; Souza, A.J.D.; Oliveira, E.A.; Oliveira, L.F.; Sarti, A.; Dussan, K.J. Waste for Energy Production: Monitoring and Controlling Pollutant Emissions for a Sustainable Economy. *Waste Biomass Valoriz.* **2025**. [CrossRef]
16. He, J.; Wang, H. Economic Structure, Development Policy and Environmental Quality: An Empirical Analysis of Environmental Kuznets Curves with Chinese Municipal Data. *Ecol. Econ.* **2012**, *76*, 49–59. [CrossRef]
17. Wang, X.; Lu, C.; Cao, Y.; Chen, L.; Abedin, M.Z. Decomposition, Decoupling, and Future Trends of Environmental Effects in the Beijing-Tianjin-Hebei Region: A Regional Heterogeneity-Based Analysis. *J. Environ. Manag.* **2023**, *331*, 117124. [CrossRef]
18. Xiong, J.; Xu, D. Relationship between Energy Consumption, Economic Growth and Environmental Pollution in China. *Environ. Res.* **2021**, *194*, 110718. [CrossRef] [PubMed]
19. Song, L.; Kir, K.F.; Shu, P. Relationships among Energy Consumption, Air Pollution and Economic Growth in China. *E3S Web Conf.* **2020**, *194*, 4003. [CrossRef]
20. Osuntuyi, B.V.; Lean, H.H. Environmental Degradation, Economic Growth, and Energy Consumption: The Role of Education. *Sustain. Dev.* **2023**, *31*, 1166–1177. [CrossRef]
21. Hosseini, S.M.; Saifoddin, A.; Shirmohammadi, R.; Aslani, A. Forecasting of CO_2 Emissions in Iran Based on Time Series and Regression Analysis. *Energy Rep.* **2019**, *5*, 619–631. [CrossRef]
22. Ding, S.; Hu, J.; Lin, Q. Accurate Forecasts and Comparative Analysis of Chinese CO_2 Emissions Using a Superior Time-Delay Grey Model. *Energy Econ.* **2023**, *126*, 107013. [CrossRef]
23. Dong, X.; Yang, M. Relationship between Regional Economic Development and Environmental Pollution Based on GPI. *J. Chang. Univ. Nat. Sci. Ed.* **2013**, *33*, 106–110. (In Chinese)
24. Yang, M.; Dong, X. GPI Analysis of the Relationship between Urban Economic Development and Industrial Pollutant Emission. *Environ. Sci. Manag.* **2013**, *38*, 41–46. (In Chinese)
25. Zhao, P.; Wu, Y.; Dong, X. Research on the Relationship between Economic Growth and Environmental Pollution in Shaanxi Province—Based on the Gross Domestic Product Pollutant Emission Index GPI. *J. Xi'an Shiyou Univ. Soc. Sci. Ed.* **2015**, *24*, 19–23. (In Chinese)

26. Xi, B.; Zhai, P. Economic Growth, Industrial Structure Upgrading and Environmental Pollution: Evidence from China. *Kybernetes* **2023**, *52*, 518–553. [CrossRef]
27. Zhang, L.; You, S.; Zhang, M.; Zhang, S.; Yi, S.; Zhou, B. The Effects of Urbanization on Air Pollution Based on a Spatial Perspective: Evidence from China. *Front. Environ. Sci.* **2022**, *10*, 1058009. [CrossRef]
28. Men, Y.; Li, Y.; Luo, Z.; Jiang, K.; Yi, F.; Liu, X.; Xing, R.; Cheng, H.; Shen, G.; Tao, S. Interpreting Highly Variable Indoor PM2.5 in Rural North China Using Machine Learning. *Environ. Sci. Technol.* **2023**, *57*, 18183–18192. [CrossRef]
29. Kilian, S.; Meisenbacher, K.; Bischoff, M.S. Linear and Logistic Regression Analysis Interpret Correlations and Make Predictions. *Gefasschirurgie* **2023**, *28*, 462–472. [CrossRef]
30. Lu, B.; Meng, X.; Dong, S.; Zhang, Z.; Liu, C.; Jiang, J.; Herrmann, H.; Li, X.; Wang, M.; Wang, A.; et al. High-Resolution Mapping of Regional VOCs Using the Enhanced Space-Time Extreme Gradient Boosting Machine (XGBoost) in Shanghai. *Sci. Total Environ.* **2024**, *949*, 175176. [CrossRef]
31. Zagow, M.; Elbany, M.; Darwish, A.M. Identifying Urban, Transportation, and Socioeconomic Characteristics across US Zip Codes Affecting CO_2 Emissions: A Decision Tree Analysis. *Energy Built Environ.* **2025**, *6*, 484–494. [CrossRef]
32. Afzal, S.; Ziapour, B.M.; Shokri, A.; Shakibi, H.; Sobhani, B. Building Energy Consumption Prediction Using Multilayer Perceptron Neural Network-Assisted Models; Comparison of Different Optimization Algorithms. *Energy* **2023**, *282*, 128446. [CrossRef]
33. Chen, R.; Hu, J.; Song, Z.; Wang, Y.; Zhou, X.; Zhao, L.; Chen, B. The Spatiotemporal Distribution of NO2 in China Based on Refined 2DCNN-LSTM Model Retrieval and Factor Interpretability Analysis. *Remote Sens.* **2023**, *15*, 4261. [CrossRef]
34. Jiang, J.; Wan, X.; Zhu, F.; Xiang, D.; Hu, Z.; Mu, S. A Deep Learning Framework Integrating Transformer and LSTM Architectures for Pipeline Corrosion Rate Forecasting. *Comput. Chem. Eng.* **2026**, *204*, 109365. [CrossRef]
35. Hou, L.; Chen, H. The Prediction of Medium- and Long-Term Trends in Urban Carbon Emissions Based on an ARIMA-BPNN Combination Model. *Energies* **2024**, *17*, 1856. [CrossRef]
36. Düker, M.; Matteson, D.S.; Tsay, R.S.; Wilms, I. Vector AutoRegressive Moving Average Models: A Review. *WIREs Comput. Stat.* **2025**, *17*, e70009. [CrossRef]
37. Makar, P.A.; Cheung, P.; Hogrefe, C.; Akingunola, A.; Alyuz, U.; Bash, J.O.; Bell, M.D.; Bellasio, R.; Bianconi, R.; Butler, T.; et al. Critical Load Exceedances for North America and Europe Using an Ensemble of Models and an Investigation of Causes of Environmental Impact Estimate Variability: An AQMEII4 Study. *Atmos. Chem. Phys.* **2025**, *25*, 3049–3107. [CrossRef]
38. Tremblay, A.; Newman, A.J. Modeling Nonlinear Relationships in ERP Data Using Mixed-Effects Regression with R Examples. *Psychophysiology* **2015**, *52*, 124–139. [CrossRef] [PubMed]
39. Liu, W.; Tsang, I.W. Making Decision Trees Feasible in Ultrahigh Feature and Label Dimensions. *J. Mach. Learn. Res.* **2017**, *18*, 2814–2849.
40. Feng, S.; Wallace, E.; Grissom, A.; Iyyer, M.; Rodriguez, P.; Boyd-Graber, J. Pathologies of Neural Models Make Interpretations Difficult. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 31 October–4 November 2018*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 3719–3728. [CrossRef]
41. Mahmoud, M.F.; Arabi, M.; Pallickara, S. Harnessing Ensemble Machine Learning Models for Improved Salinity Prediction in Large River Basin Scales. *J. Hydrol.* **2025**, *652*, 132691. [CrossRef]
42. Mienye, I.D.; Sun, Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access* **2022**, *10*, 99129–99149. [CrossRef]
43. Cheng, Y.; Zeng, J.; Wang, Z.; Song, D. A Health State-Related Ensemble Deep Learning Method for Aircraft Engine Remaining Useful Life Prediction. *Appl. Soft Comput.* **2023**, *135*, 110041. [CrossRef]
44. Wu, Y.; Xi, X.; He, J. AFGSL: Automatic Feature Generation Based on Graph Structure Learning. *Knowl.-Based Syst.* **2022**, *238*, 107835. [CrossRef]
45. Zheng, A.; Casari, A. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*, 1st ed.; O'Reilly Media, Inc.: Newton, MA, USA, 2018; ISBN 978-1-4919-5324-2.
46. Zhou, Z.; Hooker, G. Unbiased Measurement of Feature Importance in Tree-Based Methods. *ACM Trans. Knowl. Discov. Data* **2021**, *15*, 1–21. [CrossRef]
47. Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions 2017. *arXiv* **2017**, arXiv:1705.07874.
48. Štrumbelj, E.; Kononenko, I. Explaining Prediction Models and Individual Predictions with Feature Contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647–665. [CrossRef]
49. Chen, H.; Deng, G.; Liu, Y. Monitoring the Influence of Industrialization and Urbanization on Spatiotemporal Variations of AQI and PM2.5 in Three Provinces, China. *Atmosphere* **2022**, *13*, 1377. [CrossRef]
50. Ma, W.; Feng, Z.; Cheng, Z.; Chen, S.; Wang, F. Identifying Forest Fire Driving Factors and Related Impacts in China Using Random Forest Algorithm. *Forests* **2020**, *11*, 507. [CrossRef]
51. Wang, T.; Xu, W.; Bao, A.; Yuan, Y.; Zheng, G.; Naibi, S.; Huang, X.; Wang, Z.; Zheng, X.; Bao, J.; et al. Mapping of Forest Structural Parameters in Tianshan Mountain Using Bayesian-Random Forest Model, Synthetic Aperture Radar Sentinel-1A, and Sentinel-2 Imagery. *Remote Sens.* **2024**, *16*, 1268. [CrossRef]

52. Qin, Y.; Li, C.; Shi, X.; Wang, W. MLP-Based Regression Prediction Model for Compound Bioactivity. *Front. Bioeng. Biotechnol.* **2022**, *10*, 946329. [CrossRef]
53. Duan, M.; Sun, Y.; Zhang, B.; Chen, C.; Tan, T.; Zhu, Y. PM2.5 Concentration Prediction in Six Major Chinese Urban Agglomerations: A Comparative Study of Various Machine Learning Methods Based on Meteorological Data. *Atmosphere* **2023**, *14*, 903. [CrossRef]
54. Zheng, B.; Tong, D.; Li, M.; Liu, F.; Hong, C.; Geng, G.; Li, H.; Li, X.; Peng, L.; Qi, J.; et al. Trends in China's Anthropogenic Emissions since 2010 as the Consequence of Clean Air Actions. *Atmos. Chem. Phys.* **2018**, *18*, 14095–14111. [CrossRef]
55. Li, M.; Liu, H.; Geng, G.; Hong, C.; Liu, F.; Song, Y.; Tong, D.; Zheng, B.; Cui, H.; Man, H.; et al. Anthropogenic Emission Inventories in China: A Review. *Natl. Sci. Rev.* **2017**, *4*, 834–866. [CrossRef]
56. Geng, G.; Liu, Y.; Liu, Y.; Liu, S.; Cheng, J.; Yan, L.; Wu, N.; Hu, H.; Tong, D.; Zheng, B.; et al. Efficacy of China's Clean Air Actions to Tackle PM2.5 Pollution between 2013 and 2020. *Nat. Geosci.* **2024**, *17*, 987–994. [CrossRef]
57. Wu, Y.; Shi, K.; Chen, Z.; Liu, S.; Chang, Z. Developing Improved Time-Series DMSP-OLS-like Data (1992–2019) in China by Integrating DMSP-OLS and SNPP-VIIRS. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4407714. [CrossRef]
58. Liu, L.; Cao, X.; Li, S.; Jie, N. A 31-Year (1990–2020) Global Gridded Population Dataset Generated by Cluster Analysis and Statistical Learning. *Sci. Data* **2024**, *11*, 124. [CrossRef]
59. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 785–794. [CrossRef]
60. Chen, T.; Chu, B.; Ma, J.; Ma, Q.; Liu, Q.; Wang, S.; He, K.; Zhao, J.; He, H. Ozone Pollution in China: Current Status and Control Strategies. *Engineering* 2025, *in press*. [CrossRef]

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