Quantifying the driving factors of particulate matter variabilities in the Beijing-Tianjin-Hebei and Yangtze River Delta regions from 2015 to 2022 by machine learning approach

Zhongfeng Pan[1,2], Hao Yin[3, *], Zhenda Sun[2], Chongyang Li[2], Youwen Sun[1,2,4 *], Cheng Liu[5,6,*]

[1] Institutes of Physical Science and Information Technology, Anhui University, Hefei 230601, China;

[2] Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, HFIPS, Chinese Academy of Sciences, Hefei 230031, China

[3] School of Energy and Environment, City University of Hong Kong, Hong Kong SAR, China

[4] School of Environmental Science and Optoelectronic Technology, University of Science and Technology of China, Hefei 230026, China

[5] Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230026, China

[6] Key Laboratory of Precision Scientific Instrumentation of Anhui Higher Education Institutes, University of Science and Technology of China, Hefei 230026, China

Corresponding author: Hao Yin (haoyin@cityu.edu.hk); Youwen Sun (ywsun@aiofm.ac.cn); Cheng Liu (chliu81@ustc.edu.cn)

Abstract. Accurately quantifying the relative roles of anthropogenic emissions and meteorological conditions is essential for understanding long-term particulate-matter (PM) variability. Using ground observations from 40 cities, GEOS-FP meteorology, CEDS emissions, and monthly LightGBM models, this study assesses the drivers of $PM_{2.5}$ and PM_{10} across the Beijing–Tianjin–Hebei (BTH) and Yangtze River Delta (YRD) regions during 2015–2022. The models demonstrate strong predictive skill ($R/R^2 = 0.82/0.67$ for $PM_{2.5}$ and $0.81/0.65$ for PM_{10}), with consistently high performance across cities. Both pollutants exhibit significant decreasing trends over the study period. Counterfactual experiments show that emission reductions overwhelmingly dominate these improvements: $PM_{2.5}$ emission-driven changes intensify from -9.1 μg m^{-3} in 2016 to -31.4 μg m^{-3} in 2022, while PM_{10} reductions strengthen from -9.8 to -42.9 μg m^{-3}. Meteorological contributions remain positive (approximately $+2$–4 μg m^{-3} for $PM_{2.5}$ and $+0.5$–3 μg m^{-3} for PM_{10}), indicating that air-quality gains were achieved despite meteorological conditions unfavorable to pollutant dispersion. SHAP attribution highlights 2-m air temperature (T2M), humidity (QV2M), and key

precursors (SO_2_total, BC_total, NOx_total) as dominant predictors, with interaction patterns showing that meteorological conditions modulate the effectiveness of precursor emissions and associated chemical formation. These results provide a comprehensive, data-driven assessment of the factors shaping PM variability in two major urban clusters of China.

1 Introduction

Particulate matter (PM) is a significant air pollutant and is also a critical research topic in environmental science due to its diverse sources, complex chemical composition, and profound impacts on human health (Zhang et al., 2022a). Classified by aerodynamic diameter, $PM_{2.5}$ (fine particles, ≤2.5μm) and PM_{10} (inhalable particles, ≤10μm) exert differential impacts on ecosystems and human health owing to their distinct physicochemical properties and environmental behaviors (WHO, 2021). Fine particles ($PM_{2.5}$) penetrate deep into the lungs and cross the alveolar–blood barrier into systemic circulation, while coarser particles (PM_{10}) deposit predominantly in the upper respiratory tract (Fu et al., 2024). Chronic exposure to $PM_{2.5}$ is linked to respiratory/cardiovascular diseases, declines in lung function, and impairment of the immune system (Franklin et al., 2008; Kioumourtzoglou et al., 2016), Whereas PM_{10} aggravates asthma , chronic obstructive pulmonary disease (COPD), and other respiratory conditions (Seaton et al., 1995). As two pivotal economic engines of China, the Beijing-Tianjin-Hebei (BTH) and Yangtze River Delta (YRD) regions, characterized by dense industrial clusters and populations, generate substantial industrial and transportation emissions, with high-intensity production and daily activities resulting in long-standing composite air pollution dominated by $PM_{2.5}$, PM_{10}, and ozone (Dai et al., 2021, 2023), posing persistent threats to human health and urban livability. Furthermore, PM pollution acidifies aquatic environments, disrupts ecosystem balance, degrades soils, and contributes to acid rain and terrestrial biosphere damage (Dominici et al., 2014; Jerrett, 2015).

To address severe air pollution problem, Chinese government implemented the Air Pollution Prevention and Control Action Plan (State Council of the People's Republic of China, 2013) and the Three-Year Action Plan for Winning the Blue Sky Defense Battle (State Council of the People's Republic of China, 2018).These initiatives led to substantial reductions in PM concentrations nationwide (Song et al., 2023). However, China's current Ambient Air Quality Standards (GB 3095-

2012) stipulate Grade II annual mean limits of 35 μg m^{-3} for $PM_{2.5}$ and 70 μg m^{-3} for PM_{10}, which significantly exceed the updated WHO guidelines (AQG 2021).

The dynamics of PM are shaped by anthropogenic precursor emissions—sulfur dioxide (SO_2), nitrogen oxides (NO_x), and ammonia (NH_3)—together with meteorological factors such as temperature, humidity, precipitation, pressure, and wind (Xiao et al., 2021). In addition to these inorganic precursors, volatile organic compounds (VOCs) also play an important role in secondary aerosol formation, particularly through pathways leading to secondary organic aerosols, as recognized in numerous atmospheric chemistry studies. $PM_{2.5}$ originates predominantly from traffic and industrial emissions, combustion processes (e.g., cooking, biomass burning), and secondary formation via atmospheric oxidation to sulfate, nitrate, and organic aerosols (Zhang et al., 2015). PM_{10} also includes coarse particles from fugitive dust (construction, agriculture) and secondary coarse-mode particulates (Wu and Huang, 2021). The SO_2, NO_x, and NH_3 in the atmosphere can be converted into secondary inorganic aerosols, which significantly regulate PM concentrations (Ding et al., 2019; Feng et al., 2021). Meteorological parameters, such as temperature, relative humidity, precipitation, pressure, and wind, critically influence PM generation, dispersion, and removal (Leung et al., 2018; Zhao et al., 2013). For instance, elevated temperatures accelerate SO_x/NO_x oxidation rates and fine PM formation (Chen et al., 2022). High humidity promotes particle hygroscopic growth, gas-to-particle conversion (e.g., secondary organic aerosols), and wet deposition, thereby altering PM size distribution and lifetime. These PM-meteorology interactions exhibit region- and year-specific nonlinear characteristics (Shen et al., 2017), challenging conventional linear modeling approaches(Zhao et al., 2018)

Machine learning (ML), with its capacity to capture complex, nonlinear relationships, has emerged as a powerful tool for atmospheric pollution research (Yin et al., 2022b). ML enhances source apportionment accuracy through multi-source data integration (meteorological, emission, socioeconomic), high-dimensional pattern recognition, and real-time adaptive analysis, enabling identification of complex pollutant interactions (Peng et al., 2024). For $PM_{2.5}$ and PM_{10} studies, ML facilitates quantitative disentanglement of meteorological and emission contributions, elucidates source-receptor relationships, and informs targeted mitigation strategies.

This study employs the LightGBM framework to quantify the drivers of $PM_{2.5}$ and PM_{10}

variability across the BTH and YRD regions during 2015–2022. By leveraging LightGBM's capability to model nonlinear emission–meteorology–pollution interactions and its efficiency on multi-year, multi-city datasets, the analysis aims to identify the dominant factors governing regional air-quality evolution. The structure of this paper is as follows. Section **2** introduces the datasets used in this study, including national ground-based PM observations (2.1), GEOS-FP meteorological reanalysis fields (2.2), and the CEDS anthropogenic emission inventory (2.3). Section 3 describes the methodological framework. Section 3.1 presents the extraction of city-level meteorological and emission variables from gridded datasets. Section 3.2 outlines the LightGBM modeling workflow, including the feature-engineering strategy (3.2.1), the model training and leave-one-year-out cross-validation procedure (3.2.2), and the SHAP-based interpretation approach (3.2.3). The interannual trend estimation method is detailed in Section 3.3, and the counterfactual framework used to separate meteorological and emission contributions is introduced in Section 3.4. Section 4 reports the main results, including the interannual evolution of $PM_{2.5}$ and PM_{10} (4.1), machine-learning model performance and key predictor importance (4.2), and the quantified meteorological and emission contributions (4.3). Section 5 provides a broader discussion of the identified driving mechanisms in the context of chemical formation pathways and emission-control policies. Finally, Section 6 summarizes the key findings and presents implications for future air-quality management and research.

2 Data

2.1 Observational data from national monitoring sites

The ground-level air pollutant data for the YRD and BTH regions were acquired from the China National Environmental Monitoring Center (CNEMC) network (https://www.cnemc.cn/, last accessed: December 31, 2022), comprising hourly measurements of $PM_{2.5}$, PM_{10}, SO_2, NO_2, CO, and O_3 concentrations from 2015 to 2022. Observations from multiple monitoring stations within the same city were averaged to derive city-level pollutant concentrations (site-specific details are provided in Table S1 and Fig. S1). The monitoring network includes 80 stations in the BTH region, covering major cities and areas in Beijing, Tianjin, and Hebei Province, and 197 stations in the YRD region, spanning Shanghai, Jiangsu, Zhejiang, and adjacent provinces. To avoid ambiguity between two cities sharing the same English spelling "Taizhou", Taizhou City in Jiangsu Province is denoted as TaizhouJS, while Taizhou City in Zhejiang Province is denoted as TaizhouZJ throughout this study.

All national monitoring stations strictly follow the Technical Specifications for Automatic Ambient Air Quality Monitoring (Ministry of Environmental Protection of China, 2013a), ensuring standardized field operation and quality control procedures. City-level $PM_{2.5}$ and PM_{10} concentrations are released in accordance with the national reference gravimetric method(Ministry of Environmental Protection of China, 2011), which provides the calibration and traceability framework for automated particulate measurements across the monitoring network. Gaseous pollutants (SO_2, NO_2, CO, and O_3) are measured using ultraviolet fluorescence, chemiluminescence, non-dispersive infrared absorption, and ultraviolet photometric analysis, respectively, following the national specifications for continuous gaseous monitoring(Ministry of Environmental Protection of China, 2013b). These unified procedures ensure accuracy, comparability, and long-term stability of pollutant observations.

2.2 GEOS-FP meteorological data

Meteorological fields for 2015–2022 were obtained from the GEOS Forward Processing (GEOS-FP) product (http://geoschemdata.wustl.edu/ExtData/, last accessed: 31 December 2020) at a spatial resolution of 0.25° × 0.3125°. GEOS-FP provides hourly assimilated meteorological variables with relatively high spatial and temporal resolution, enabling refined characterization of mesoscale atmospheric processes over the BTH and YRD regions. Its near–real-time data assimilation framework has been widely applied in regional atmospheric pollution and transport studies, supporting accurate representation of dynamic meteorological conditions(Yin et al., 2021b, 2022a, b). Its near–real-time data assimilation system further improves the accuracy of reanalysis-based meteorological fields and enhances representation of dynamic atmospheric processes (Sun et al., 2021a, b; Yin et al., 2019, 2020, 2021a). The meteorological parameters used in this study include total cloud fraction (CLDTOT), precipitation flux (PRECTOT), 2-m specific humidity (QV2M), 2-m air temperature (T2M), sea-level pressure (SLP), surface downward shortwave radiation (SWGDN), and 10-m zonal (U10M) and meridional (V10M) wind components.

2.3 CEDS emission inventory

Anthropogenic emission data for 2015–2022 were obtained from the Community Emissions

Data System (CEDS), which provides monthly mean fluxes at a spatial resolution of 0.5° × 0.5°. In this study, we use emissions of CH_2O, CO, NH_3, NO_x, SO_2, BC, OC, and paraffinic reactive primary emissions (PRPE). Emissions were further categorized into eight sectors: non-combustion agriculture, energy transformation and extraction, industrial combustion and processes, surface transportation, residential and commercial fuel use, solvents, waste treatment and disposal, and international shipping. Anthropogenic emission data for 2015–2022 were derived from the CEDS, a global inventory providing temporally resolved sector-specific emissions. The CEDS framework supports climate change projections and quantifies human-driven interactions between air pollutants and climate systems, critical for assessing health and ecosystem impacts.

3 methodology

3.1 Emission and meteorological data extraction

City-level emission and meteorological variables were derived from gridded emission inventories and meteorological reanalysis datasets.

For each city, polygon boundaries were obtained from the GADM Level-2 shapefile, and all grid cells whose centers fell within the city polygon were identified.

For emission data, which are expressed as surface fluxes ($kg \cdot m^{-2} \cdot s^{-1}$), the city-scale total emission $E_{\text{city}}(t)$ was obtained by physical integration over all intersecting grid cells:

$$E_{\text{city}}(t) = \sum_{i,j} F_{ij}(t) A_{ij} r_{ij} \qquad (1)$$

where $F_{ij}(t)$ is the emission flux of grid cell (i,j) at time t ; A_{ij} is the nominal grid-cell area; and r_{ij} is the fractional overlap between the grid cell and the city polygon.

This area-weighted integration converts surface fluxes into physically consistent city-total emissions.

For meteorological variables, which represent intensive state quantities such as near-surface temperature (T2M), specific humidity (QV2M), and wind components (U10M/V10M), city-level averages were computed as the arithmetic mean of all grid-cell centers located inside the city polygon:

$$X_{\text{city}}(t) = \frac{1}{N} \sum_{i,j} X_{ij}(t) \qquad (2)$$

where $X_{ij}(t)$ is the variable value in cell (i,j), and N is the number of valid grid cells within the city.

This center-based averaging is computationally efficient and provides a representative estimate of the mean meteorological condition. Because the study region lies mainly in the mid-to-low latitudes of China, where grid-cell area variation with latitude is minor, this approximation introduces negligible bias compared with full area-weighted averaging.

Both extraction procedures ensure spatial consistency between emission and meteorological datasets and yield temporally continuous, city-level time series for subsequent modeling.

3.2 LightGBM modeling

Light Gradient Boosting Machine (LightGBM) is an efficient and scalable implementation of gradient boosting, extensively applied to regression, classification, and ranking tasks (Yin et al., 2021c). By using a histogram-based decision-tree algorithm, the LightGBM model drastically reduces both computation time and memory usage compared to traditional gradient-boosting methods such as XGBoost (Bian et al., 2023; Zhang et al., 2017). It supports the direct handling of categorical variables without one-hot encoding, improving efficiency when processing high-dimensional datasets. During training, LightGBM grows trees in a leaf-wise (best-first) manner, generating deeper splits along the branch that achieves the largest loss reduction.

In contrast, XGBoost and classical Gradient Boosting Decision Trees (GBDT) use a level-wise growth strategy, which provides stability but can become computationally slower for large-scale data. LightGBM also offers extensive hyperparameter controls, such as maximum tree depth, minimum data in leaf, and feature fraction, to balance model complexity and generalization (Ke et al., 2017). Due to its high predictive accuracy, efficient splitting mechanism, and robust computational capability, LightGBM has become one of the most widely used gradient-boosting frameworks in environmental modeling (Liu et al., 2023; Wang et al., 2022; Zhang et al., 2022b).

Prior to model training, all hourly emission and meteorological inputs were aggregated to monthly means, and the LightGBM models were trained on these monthly-resolved datasets. This temporal aggregation ensured that all variables shared the same temporal frequency, preventing inconsistencies between hourly and monthly features. It also matched the model input scale with the monthly trend analysis period (2015–2022), thereby avoiding time-scale mismatches between predictor variables and the evaluation framework.

The model performance was evaluated using three widely recognized regression metrics: the correlation coefficient (R), indicating the linear relationship between predicted and observed concentrations. the coefficient of determination (R^2), measuring the proportion of variance in observations explained by the model; and the root-mean-square error (RMSE), representing the average magnitude of prediction errors. Higher R and R^2 and lower RMSE indicate stronger predictive ability.

3.2.1 Feature Engineering

To ensure interpretability and reduce redundancy, all anthropogenic emissions were aggregated into a single total value for each species (e.g., NOx_total, SO2_total). Rather than treating sector-resolved emissions separately, emissions from all sectors were summed into one unified variable per pollutant. This consolidation reduces feature dimensionality, mitigates multicollinearity among sectoral components, and preserves the dominant emission signal, resulting in a more compact and interpretable model structure.

To incorporate temporal information, three seasonal descriptors were introduced: (1) a pair of harmonic terms (month_sin and month_cos) representing the cyclic annual progression; (2) a categorical season indicator to reflect broad seasonal regimes. These features provide smooth, physically meaningful representations of seasonality without imposing sharp discontinuities between months.

In addition, two derived emission indicators were constructed to characterize temporal variability and multi-scale coupling:

Seasonal difference (sdiff) represents interannual seasonal changes over a 12-month interval. For each feature x, the seasonal difference is defined as:

$$sdiff_t(x) = \frac{x_t - x_{t-1}}{2} \tag{3}$$

where x_t denotes the monthly mean value at time t.This feature highlights annual-scale variations and helps the model capture year-to-year emission variability under identical seasonal conditions.

Rolling detrended residual (detr) represents short-term deviations from a 12-month moving-mean trend. It is expressed as:

$$detr_t(x) = x_t - \mu_t^{(12)}(x) \tag{4}$$

$$\mu_t^{(12)}(x) = \frac{1}{k_t}\sum_{i=t-11}^{t} x_i \cdot 1_{\{x_t\ \text{valid}\}}, \quad k_t \geq 3 \tag{5}$$

Where $\mu_t^{(12)}(x)$ is the rolling mean computed over the past 12 months, requiring at least three valid data points ($k_t \geq 3$). This feature isolates short-term fluctuations by removing low-frequency seasonal trends.

In practice, the final predictor set comprised 35 variables: eight monthly meteorological parameters, eight species-level aggregated emission totals (e.g., NOx_total, SO₂_total, BC_total), three temporal descriptors (month_sin, month_cos, season), and two derived indicators (sdiff and detr) for each emission species. Pollutant concentrations and explicit calendar identifiers (year, month, date) were excluded from the input space to avoid information leakage and to ensure consistent temporal treatment. A complete list of variables used in the model is provided in Table S2.

3.2.2 Model Training and Cross-Validation Procedure

Before model fitting, all months with missing $PM_{2.5}$ or PM_{10} observations were removed to ensure consistency between predictors and targets. The derived emission features (sdiff and detr) inherently contain missing values during their initial 12-month computation window; these were addressed using a simple mean-imputation strategy applied within each training fold. Specifically, missing entries for each feature were replaced with the corresponding feature mean computed solely from the training subset of that fold, thereby preventing temporal leakage and ensuring that imputation relied exclusively on information available before prediction.

The LightGBM model was independently trained for each city, pollutant type ($PM_{2.5}$ and PM_{10}), and cross-validation fold. A leave-one-year-out (LOGO) cross-validation scheme was adopted, whereby data from a single year were held out for testing while the remaining seven years were used for training. This process was repeated sequentially so that each year between 2015 and 2022 served once as the held-out test period. The resulting predictions thus constitute out-of-sample estimates for every year, providing a conservative and temporally robust assessment of model generalizability and avoiding within-year information leakage.

Within each training subset, a randomized search procedure was used to identify the optimal hyperparameter configuration. Ten candidate parameter sets were sampled from predefined distributions and evaluated using three-fold internal cross-validation, with the negative mean squared error (MSE) serving as the scoring metric. The optimized parameters included learning_rate, n_estimators, num_leaves, max_depth, min_child_samples, subsample, colsample_bytree, reg_alpha, and reg_lambda, all drawn from physically plausible ranges. Other settings, such as objective = regression, metric = rmse, early_stopping_rounds = 20, and random_state = 42, were held constant across folds to ensure comparability. A consolidated summary of the optimized hyperparameter ranges is presented in Table S3.

3.2.3 SHAP Interpretation Analysis

To interpret the LightGBM outputs and quantify the contribution of individual predictors, the Shapley Additive explanations (SHAP) framework was applied. For each trained model, SHAP values $s_{i,j}$ represent the marginal contribution of feature j to the prediction of sample i.They are derived from cooperative game theory and satisfy the local additivity principle:

$$f(x_i) = E[f(x)] + \sum_{j=1}^{m} s_{i,j} \tag{6}$$

where $f(x_i)$ is the model prediction for sample i, and $E[f(x)]$ is the expected model output over all samples.

Because this study involves multiple cities across the BTH and YRD regions, a unified regional SHAP ranking was derived by weighting each SHAP importance of city by its corresponding number of valid monthly samples. The weighted importance for feature j is defined as:

$$I_j^{(w)} = \frac{\sum_{c=1}^{N} I_{j,c}\, n_c}{\sum_{c=1}^{N} n_c} \tag{7}$$

Here $I_{j,c}$ is the SHAP-based importance of feature j in city c, and n_c is the number of valid samples for that city, N is the total number of cities.

This weighting strategy ensures that cities with longer and more complete observational records have proportionally greater influence on the regional-level interpretability.

3.3 Interannual trend analysis method

To quantify the interannual trends of $PM_{2.5}$ and PM_{10} concentrations from 2015 to 2022, a linear regression model was employed in this study. For each city, the relationship between annual mean concentration y and year x was modeled as:

$$y = \beta_0 + \beta_1 x + \epsilon \tag{8}$$

where β_0 represents the intercept (baseline concentration), and ϵ denotes the error term. The slope β_1, reflecting the annual rate of concentration change, was estimated via the ordinary least squares (OLS) method. Specifically, the parameters were optimized by minimizing the residual sum of squares (RSS):

$$\arg\min_{\beta_0,\beta_1} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \tag{9}$$

where n is the sample size (e.g., n=6 for the period 2015–2022), x_i denotes the year, and y_i represents the corresponding annual mean concentration.

The slope β_1 was derived as:

$$\beta_1 = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)} \tag{10}$$

The sign of β_1 indicates the direction of concentration trends (negative for decreasing, positive for increasing), while its absolute value quantifies the magnitude of change.

3.4 Methodology for disentangling meteorological and emission contributions

To quantify meteorological and anthropogenic emission contributions, the trained models were applied in a parallel prediction experiment. For each year from 2016 to 2022, anthropogenic emission features were held fixed at their 2015 levels, while meteorological and temporal predictors were retained at their actual contemporaneous values. This produced a counterfactual concentration series driven solely by meteorological variability. This yielded pollutant concentrations driven solely by meteorological variations (denoted as $ML_{2022\mathrm{met}}$ for 2022). The contribution metrics related to 2015 were calculated as follows:

Meteorological contribution ($ML_{2022\mathrm{met}}$):

$$ML_{2022\mathrm{met}} = ML_{15-22} - ML_{2015} \tag{11}$$

$ML_{15-22}$is the non-emission condition unchanged, the emission condition is fixed as the model prediction result in 2015, and $ML_{2015}$is the model prediction result with unchanged meteorological and emission conditions.

Emission contribution ($ML_{2022\mathrm{emis}}$):

$$ML_{2022\mathrm{emis}} = (Obs_{2022} - Obs_{2015}) - ML_{2022\mathrm{met}} \tag{12}$$

$Obs_{2022}$and Obs_{2015}: Observed concentrations in 2022 and 2015, respectively.

4 Results

4.1 Interannual trends of ground-level $PM_{2.5}$ and PM_{10}

Fig. 1 illustrates the interannual evolution of ground-level $PM_{2.5}$ and PM_{10} concentrations across the BTH and YRD regions from 2015 to 2022, with corresponding annual mean spatial distributions shown in Fig. S2–S3. Both pollutants exhibit persistent downward trajectories across nearly all cities. The monthly time series and associated OLS trend lines (Fig. S4) further confirm that these declines remain stable throughout the study period, with the most pronounced reductions occurring before 2020. Statistical diagnostics from linear regression (Table S4) indicate that most cities display significant negative trends ($p < 0.05$), underscoring the robustness and consistency of these decreases.

For $PM_{2.5}$, all cities show negative trends, with annual rates ranging from approximately −1.53 to −9.46 μg m^{-3} yr^{-1}. Within BTH, the most rapid declines occur in Baoding (−9.46 ± 0.74 μg m^{-3} yr^{-1}), Hengshui (−8.36 ± 0.67 μg m^{-3} yr^{-1}), and Xingtai (−8.02 ± 0.56 μg m^{-3} yr^{-1}), while Chengde (−1.98 ± 0.35 μg m^{-3} yr^{-1}) and Zhangjiakou (−2.00 ± 0.42 μg m^{-3} yr^{-1}) exhibit more modest declines. In the YRD region, substantial decreases are observed in Huzhou (−5.69 ± 1.22 μg m^{-3} yr^{-1}), Hefei (−4.83 ± 0.33 μg m^{-3} yr^{-1}), and Chuzhou (−4.28 ± 1.12 μg m^{-3} yr^{-1}), whereas Zhoushan (−2.11 ± 0.14 μg m^{-3} yr^{-1}) and Taizhou-ZJ (−2.72 ± 0.26 μg m^{-3} yr^{-1}) experience comparatively smaller reductions. These regional contrasts closely align with the initial concentration levels shown in Fig. S2, where inland BTH cities began with substantially higher baselines (e.g., Baoding and Hengshui exceeded 105 and 98 μg m^{-3} in 2015), enabling larger absolute declines.

For PM_{10}, all cities also exhibit significant decreases, with annual trends ranging from roughly −1.84 to −13.79 μg m^{-3} yr^{-1}. The largest reductions occur in BTH, particularly Hengshui (−13.79 ± 1.82 μg m^{-3} yr^{-1}), Baoding (−13.07 ± 1.46 μg m^{-3} yr^{-1}), and Xingtai (−12.76 ± 1.29 μg m^{-3} yr^{-1}). In contrast, Chizhou (−1.84 ± 2.03 μg m^{-3} yr^{-1}) and Zhoushan (−2.77 ± 0.30 μg m^{-3} yr^{-1}) show the smallest declines, and the PM_{10} trend in Chizhou is not statistically significant ($p = 0.40$), consistent with its minimal decline rate. As with $PM_{2.5}$, cities starting with the highest PM_{10} levels—such as Baoding (174.6 μg m^{-3} in 2015) and Hengshui (175.9 μg m^{-3})—exhibit the steepest reductions, whereas cities with lower initial levels (e.g., Zhoushan at ~46.8 μg m^{-3}) show correspondingly smaller absolute decreases.

Overall, both $PM_{2.5}$ and PM_{10} decrease more rapidly in BTH than in YRD, reflecting regional differences in initial emissions, industrial structure, and the intensity of mitigation policies. The widespread statistical significance of the trends (Table S4) supports the conclusion that both regions experienced sustained and robust improvements in air quality during 2015–2022.

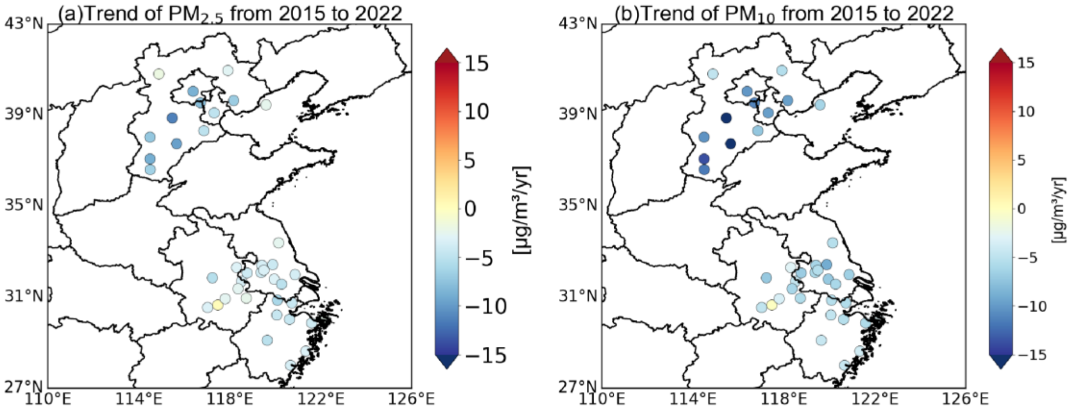

Fig. 1 Interannual variation trends of PM$_{2.5}$ (a) and PM$_{10}$ (b) in each city during 2015-2022.

4.2 Machine learning model performance and variable importance

Fig. 2 shows the predictive performance of the LightGBM models for PM$_{2.5}$ and PM$_{10}$ across all cities. The density scatterplots (Fig. 2a and 2c) indicate strong agreement between predictions and observations, with R/R^2 = 0.82/0.67 for PM$_{2.5}$ and 0.81/0.65 for PM$_{10}$. Consistent performance is also observed at the city level (Fig. 2b and 2d) , and the full statistics provided in Table S5 demonstrate that most cities achieve R > 0.70, confirming the robustness of the models across diverse urban settings.

Fig. 3 summarizes the SHAP-based overall feature importance. For PM$_{2.5}$, although T2M ranks as the single most influential variable in Fig. 3a , the broader pattern shows that emission-related predictors dominate the upper portion of the importance distribution, with SO$_2$_total, BC_total, and NOx_total collectively contributing a larger share than any single meteorological factor. This implies that PM$_{2.5}$ variability during 2015–2022 remains strongly controlled by precursor emissions and their long-term reductions. In contrast, for PM$_{10}$ (Fig. 3b) , meteorological variables, especially QV2M and T2M, occupy the highest ranks, reflecting the greater role of humidity and temperature in modulating coarse particles through hygroscopic growth, boundary-layer dynamics, and seasonal resuspension processes. While NOx_total and SO$_2$_total remain important, their relative influence is smaller than in PM$_{2.5}$, consistent with PM$_{10}$'s stronger coupling to meteorology.

Region-specific SHAP rankings (Fig. S5) further show that PM$_{2.5}$ in BTH is more strongly influenced by SO$_2$_total and BC_total, whereas YRD exhibits a greater sensitivity to meteorological factors such as temperature and humidity. For PM$_{10}$, QV2M is the dominant driver in both regions, but precursor importance (e.g., NOx_total) remains higher in BTH. These spatial contrasts highlight differing emission structures and seasonal behaviors between regions, underscoring the need for differentiated control strategies for $PM_{2.5}$ and PM_{10}.

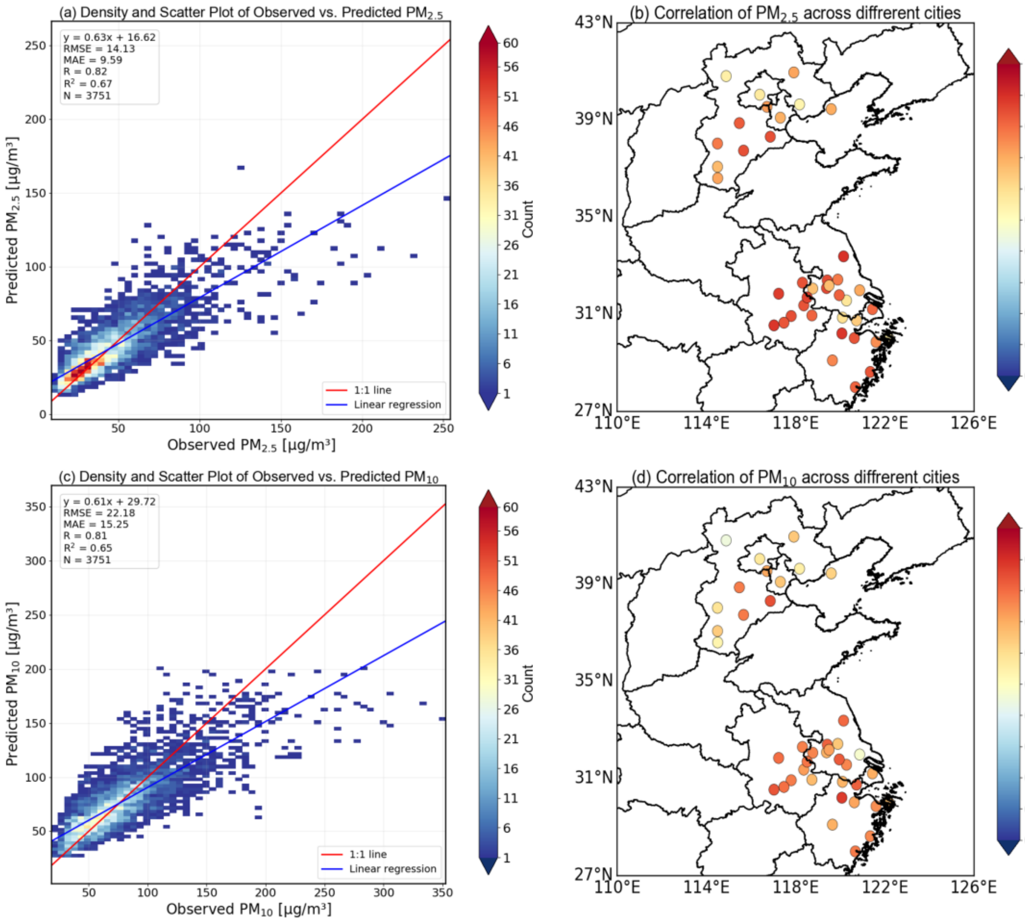

Fig. 2 The density scatter plots of $PM_{2.5}$ (a) and PM_{10} (c) concentrations observed and predicted, respectively. The correlation of $PM_{2.5}$ (b) and PM_{10} (c) in each city over BTH and YRD regions, respectively.

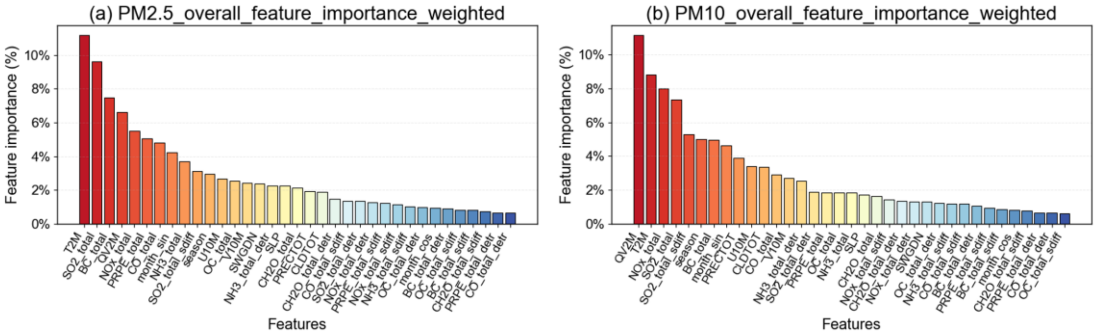

Fig. 3 The feature importance of rankings of $PM_{2.5}$ (a) and PM_{10} (b) in ML model prediction, respectively.

4.3 Contributions of emissions and meteorology

The spatial patterns of average emission and meteorological contributions (Fig. 4) reveal clear regional contrasts between northern cities in BTH and southern cities in YRD. For PM$_{2.5}$, emission contributions are strongly negative across most BTH cities such as Baoding, Hengshui, and Xingtai, often exceeding −30 μg m^{-3}, reflecting substantial reductions from historically elevated emission levels. In contrast, YRD cities such as Suzhou, Hangzhou, and Ningbo display more moderate negative contributions, typically around −10 to −20 μg m^{-3}, consistent with their lower baseline emissions. Meteorological contributions also exhibit north–south gradients: several coastal or humid subtropical YRD cities (e.g., Ningbo, Wenzhou) show slightly positive meteorological influences, whereas continental BTH cities (e.g., Beijing, Tianjin) show predominantly negative effects associated with winter stagnation, shallow boundary layers, and lower humidity. Similar spatial patterns are observed for PM$_{10}$ (Fig. 4b, d), where emission-driven reductions remain larger in BTH than in YRD, mirroring regional differences in emission intensity.

The interannual evolution shown in Fig. 5 further clarifies these relationships. Emission contributions remain consistently negative from 2016 to 2022 for both pollutants and intensify over time. For PM$_{2.5}$, the reductions increase from −9.13 μg m^{-3} in 2016 to −31.39 μg m^{-3} in 2022, while PM$_{10}$ reductions strengthen from −9.77 to −42.91 μg m^{-3} over the same period. These progressively stronger negative contributions align with downward trends in major emission species, including NOx_total, SO_total$_2$, and BC_total. In contrast, meteorological contributions remain positive throughout all years, approximately 2–4 μg m^{-3} for PM$_{2.5}$ and 0.5–3 μg m^{-3} for PM$_{10}$, indicating that annual meteorological conditions generally acted to increase particulate levels rather than suppress them. Although PM$_{10}$ is often considered more responsive to meteorological variability, its smaller meteorological contributions here largely reflect the substantially faster decline in PM$_{10}$ emissions during 2016–2022, which leaves less residual variability to be attributed to meteorology. In addition, PM$_{2.5}$ formation is strongly modulated by temperature- and humidity-dependent secondary processes, such as nitrate formation and hygroscopic growth, whereas PM$_{10}$ is dominated by mechanically generated particles that respond more weakly to these meteorological drivers.

Temporal relationships between individual emission indicators and their contributions are shown in Fig. 6 For PM$_{2.5}$, monthly variations in NOx_total, SO$_2$_total, and BC_total exhibit strong correlations with their respective emission contributions (R = 0.89–0.95). Similar correspondence is found for PM_{10} (R = 0.89–0.93). These high correlations indicate that the estimated emission contributions track the month-to-month emission variability in a physically consistent way: months with higher emissions correspond to greater positive contributions to particle levels, whereas months with lower emissions correspond to diminished contributions. This coherence holds despite nonlinear model responses, which adjust contribution magnitudes according to the concurrent meteorological and chemical environment. Meteorological drivers display opposite temporal behavior (Fig. 7). Temperature and humidity show strong negative correlations with their meteorological contributions for both pollutants (R ≈ −0.96 to −0.98), indicating that lower temperatures or reduced moisture tend to increase particulate levels, while warmer or wetter periods are associated with weaker contributions. This behavior is consistent with well-understood seasonal differences in atmospheric mixing and humidity-driven processes.

Taken together, the dominance of increasingly negative emission contributions (Fig. 5), combined with the generally positive meteorological contributions, indicates that improvements in $PM_{2.5}$ and PM_{10} were achieved despite meteorological conditions that tended to enhance particle levels. Emissions therefore play the central role in explaining the observed multi-year declines across both pollutants.

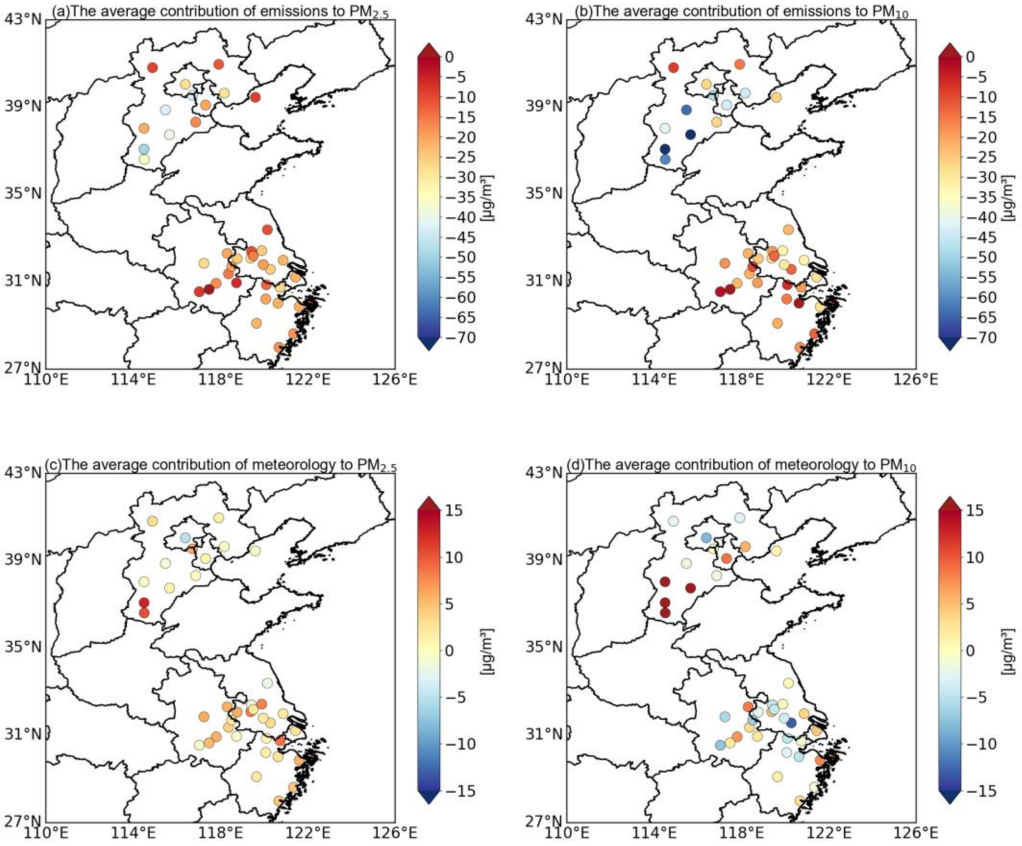

Fig. 4 The average contributions of emissions and meteorological variables to $PM_{2.5}$ (for (a) and (c)) and PM_{10} (for (b) and (d)), respectively.

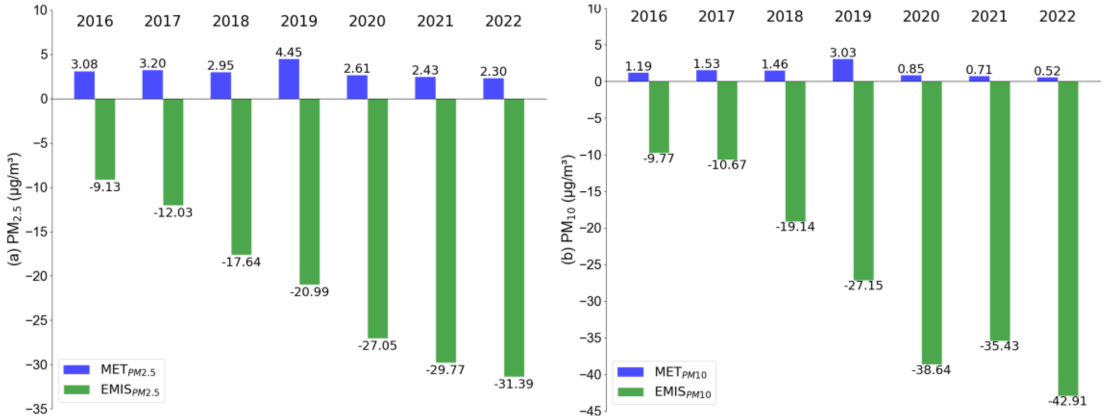

Fig. 5 The averaging the emission or meteorological contributions to $PM_{2.5}$ (a) and PM_{10} (b) of each year relative to 2015.

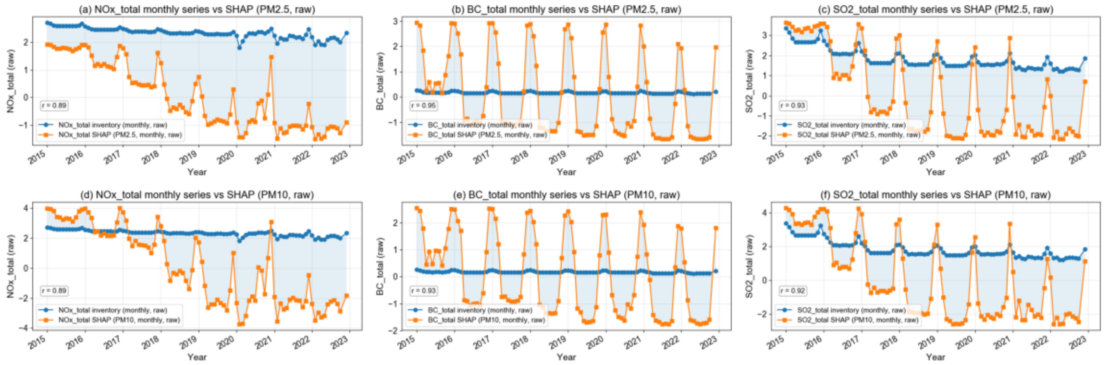

Fig. 6 Monthly emission inventories of NOx_total, BC_total, and SO2_total and their corresponding SHAP values for PM2.5 and PM10 from 2015 to 2022. Panels (a)–(c) show NOx_total, BC_total, and SO2_total versus SHAP contributions for PM2.5, while panels (d)–(f) present the corresponding relationships for PM10.

Fig. 7 Monthly meteorological variables (T2M and QV2M) and their corresponding SHAP values for $PM_{2.5}$ and PM_{10} from 2015 to 2022. Panels (a)–(b) show the relationships for $PM_{2.5}$, and panels (c)–(d) show the corresponding results for PM_{10}.

5 Discussions

To support the physical interpretation of the SHAP-derived attributions, we first examine the spatial distributions and statistical relationships of key precursors. Fig. 8 shows that the ambient concentrations of CO, NO_2, and SO_2 exhibit pronounced spatial contrasts across the BTH and YRD regions. Heavy industrial cities such as Tangshan, Xingtai, Handan, and Baoding maintain the highest levels of all three pollutants, whereas coastal YRD cities such as Zhoushan and TaizhouZJ display substantially lower concentrations due to cleaner energy structures and reduced coal-intensive activities. These spatial patterns provide important context for the correlation results summarized in Table 1 CO exhibits the strongest association with $PM_{2.5}$ (mean R $\approx$ 0.72) and

PM_{10} (mean R ≈ 0.62), followed by NO_2 (R ≈ 0.59/0.57) and SO_2 (R ≈ 0.49/0.48). The dominance of $PM_{2.5}$–CO covariation reflects their shared combustion origins, including vehicle exhaust, industrial boilers, and winter heating, and aligns with recent multi-platform and top-down studies showing that CO, NO_x, carbonaceous aerosols, and combustion-related $PM_{2.5}$ arise from tightly coupled fuel-burning processes and jointly contribute to CO_2 and black carbon budgets(Tiwari et al., 2025; Wang et al., 2021, 2025). In contrast, consistently weaker SO_2 correlations, despite elevated SO_2 levels in several northern cities, reflect the substantial effectiveness of nationwide desulfurization policies, which have suppressed sulfate formation while shifting secondary aerosol chemistry toward nitrate-dominated regimes. NO_2 correlations fall between those of CO and SO_2, indicating persistent contributions from traffic and industrial sources but increasing modulation by evolving chemical pathways and long-term emission controls.These correlation structures also contextualize the temperature, NOx_total interaction patterns. Across the observational range, lower temperatures correspond to markedly stronger NOx-related contributions to $PM_{2.5}$, consistent with enhanced ammonium nitrate stability and reduced boundary-layer mixing during cold-season stagnation. At higher temperatures, the influence of NOx_total progressively weakens due to thermodynamic limitations on nitrate formation and increased vertical dilution. This interaction pattern, illustrated in Fig. 9, extends naturally from the emission–$PM_{2.5}$ relationships revealed in Fig. 8 and Table 1, showing that NOx_total remains an influential precursor but its chemical effectiveness is strongly conditioned by temperature. A comparable interaction is not examined for PM_{10}, as coarse particles are more strongly shaped by mechanical dust, resuspension, and surface-driven processes, leading to less consistent temperature-dependent chemical behavior. Likewise, we do not present a QV2M–precursor interaction panel: humidity influences multiple competing processes such as hygroscopic growth, aqueous-phase reactions, boundary-layer suppression, and wet removal. These intertwined effects obscure a singular mechanistic relationship with SO_2 at the seasonal scale, making the interaction less suitable for clear physical interpretation.

Taken together, these results outline a coherent emission–chemical transformation–meteorological coupling framework: combustion-related pollutants such as CO, NO_2, and SO_2 co-vary with $PM_{2.5}$ because they originate from shared fuel-burning sources; evolving chemical pathways, most notably the post-desulfurization transition toward nitrate-dominated secondary formation, modulate the strength of these associations; and meteorological conditions, particularly temperature, further amplify or suppress the efficiency of secondary inorganic aerosol production. This integrated perspective underscores the need for $PM_{2.5}$ mitigation strategies that jointly consider precursor reductions, seasonal chemical regimes, and meteorological variability.

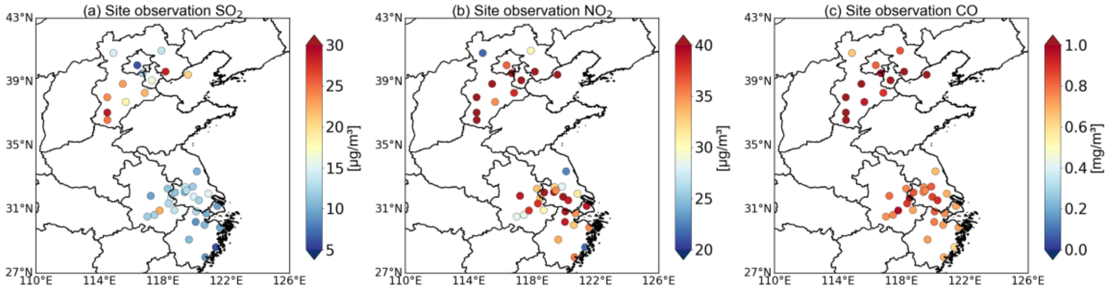

Fig. 8 The average concentrations of SO_2 (a), NO_2 (b), CO (c), respectively, during 2015 to 2020 over BTH and YRD regions.

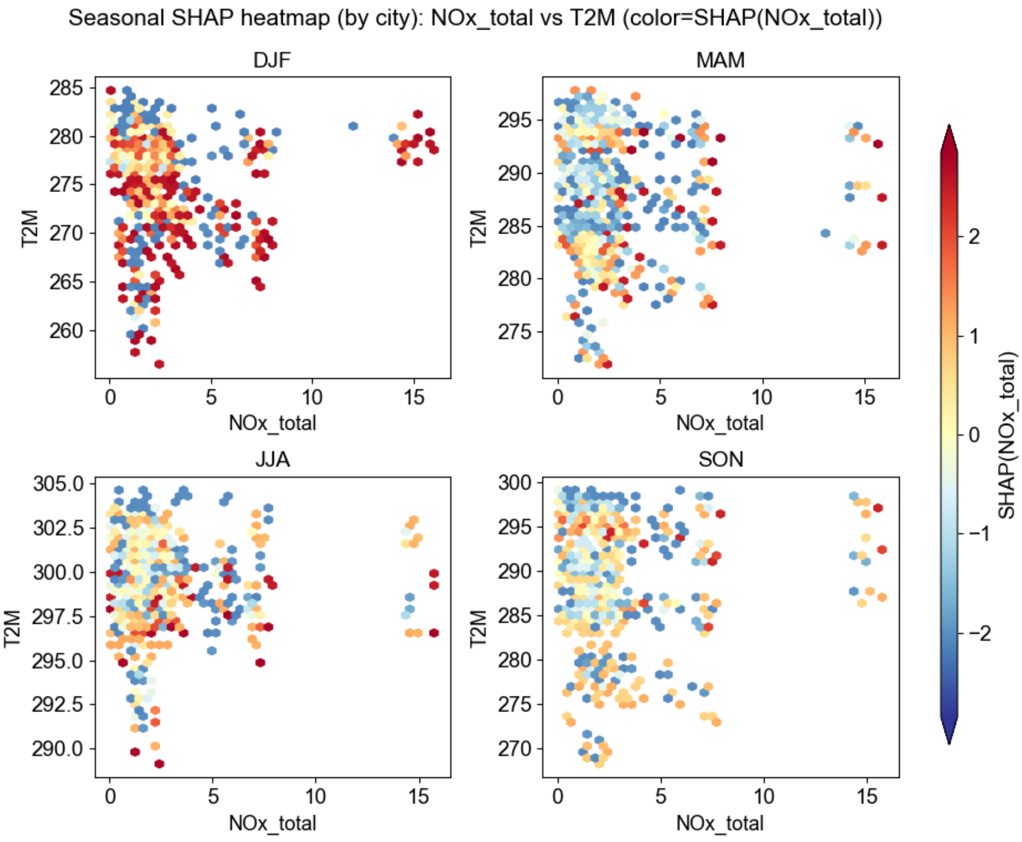

Fig. 9 Seasonal scatter plots of NOx_total versus T2M across all cities, with point colors representing the SHAP contribution of NOx_total. The four panels correspond to the DJF, MAM, JJA, and SON seasons.

Table 1 The correlation values among SO_2, NO_2, CO and $PM_{2.5}$ / PM_{10}, respectively.

	SO_2	NO_2	CO

$PM_{2.5}$	0.49	0.59	0.72
PM_{10}	0.48	0.57	0.62

6 Conclusions

This study integrates multi-source emission inventories, GEOS-FP meteorological fields, and ground-based observations to investigate the drivers of $PM_{2.5}$ and PM_{10} variability across the BTH and YRD regions over 2015–2022 using a unified LightGBM modeling and attribution framework. Both pollutants exhibit clear and statistically significant declines across most cities, supported by strong model performance under a rigorous leave-one-year-out cross-validation design (R/R^2 = 0.82/0.67 for $PM_{2.5}$; 0.81/0.65 for PM_{10}). The consistent agreement between predictions and observations across regions and years demonstrates the robustness of the machine learning representation of emission–meteorology–PM interactions.

A key contribution of this work lies in the incorporation of SHAP-based interpretability diagnostics, which allow the model to remain physically traceable while leveraging nonlinear learning capabilities. The attribution experiments clearly indicate that anthropogenic emission reductions are the dominant driver of PM improvements. For $PM_{2.5}$, emission-driven decreases intensify from $-9.13\ \mu g\ m^{-3}$ in 2016 to $-31.39\ \mu g\ m^{-3}$ in 2022; for PM_{10}, corresponding reductions strengthen from -9.77 to $-42.91\ \mu g\ m^{-3}$. These progressively stronger negative contributions are consistent with sustained declines in NO_x, SO_2, and BC emissions. In contrast, meteorological contributions remain positive in all years, indicating that prevailing weather conditions generally acted to increase particle levels rather than suppress them. The monthly coherence between precursor emissions and their SHAP-derived contributions, together with physically interpretable temperature and humidity meteorological patterns, provides additional internal validation of the model's mechanistic consistency.

These results highlight several strengths of the proposed framework. The use of harmonized monthly emission totals, multi-scale temporal descriptors (sdiff and detr), and SHAP interaction assessment enables a physically grounded interpretation of complex emission–meteorology coupling. The explicit construction of counterfactual predictions offers a transparent means of separating meteorological and anthropogenic influences. Collectively, these methodological elements provide a reproducible pathway for extending machine-learning-based attribution studies to other regions or pollutants.

Despite these advantages, several limitations remain. The monthly temporal resolution cannot fully resolve short-term meteorological or chemical processes; uncertainties in bottom-up emission inventories and meteorological reanalyses may influence the absolute magnitudes of estimated contributions; and, as with all empirical models, causality cannot be inferred directly from statistical relationships. Future work may integrate higher-frequency observations, additional chemical precursors (e.g., VOCs), or hybrid machine-learning–chemical transport approaches to further improve process interpretability and mechanistic fidelity.

Overall, this study demonstrates that multi-sector anthropogenic emission reductions rather than meteorological variability primarily explain the observed decreases in $PM_{2.5}$ and PM_{10} from 2015 to 2022, and that physically informed machine-learning tools, coupled with SHAP interpretability, provide a powerful framework for diagnosing multi-pollutant air-quality evolution and supporting evidence-based emission control strategies.

Code and data availability

The code and data for this study can be found on https://doi.org/10.5281/zenodo.17779780.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgements

This work is jointly supported by Excellent Young Scientists Fund of the National Natural Science Foundation of China (62322514), Anhui Science Fund for Distinguished Young Scholars (2308085J25), and National Key Research and Development Program of China (2023YFC3709502, 2022YFC3700100).

Financial support

This work has been supported by Excellent Young Scientists Fund of the National Natural Science Foundation of China (62322514), Anhui Science Fund for Distinguished Young Scholars

(2308085J25), and National Key Research and Development Program of China (2023YFC3709502, 2022YFC3700100).

Author contributions

HY and YWS designed this study. ZFP wrote the paper with help from HY and YWS. ZFP contributed to analysis of the data for this study. All co-authors commented on this study.

References

China publishes action plan to improve air quality: https://english.www.gov.cn/policies/latestreleases/202312/08/content_WS65724f25c6d0868f4e8e1fcb.html, last access: 16 May 2025.

Determination of atmospheric articles PM10 and PM2.5 in ambient air by gravimetric method: https://english.mee.gov.cn/Resources/standards/Air_Environment/air_method/201111/t20111101_219390.shtml, last access: 1 December 2025.

Multi-scale analysis of the impacts of meteorology and emissions on PM2.5 and O3 trends at various regions in China from 2013 to 2020 2. Key weather elements and emissions - ScienceDirect: https://www.sciencedirect.com/science/article/pii/S0048969722009391#s0030, last access: 16 July 2025.

Specifications and Test Procedures for Ambient Air Quality Continuous Automated Monitoring System for SO2, NO2, O3 and CO: https://english.mee.gov.cn/Resources/standards/Air_Environment/air_method/201308/t20130816_257557.shtml, last access: 1 December 2025.

Specifications and Test Procedures for PM10 and PM2.5 Sampler: https://english.mee.gov.cn/Resources/standards/Air_Environment/air_method/201308/t20130816_257558.shtml, last access: 1 December 2025.

The State Council rolls out a three-year action plan for clean air: https://english.mee.gov.cn/News_service/news_release/201807/t20180713_446624.shtml, last access: 16 May 2025.

WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide: https://www.who.int/publications/i/item/9789240034228, last access: 6 May 2025.

Bian, L., Qin, X., Zhang, C., Guo, P., and Wu, H.: Application, interpretability and prediction of machine learning method combined with LSTM and LightGBM-a case study for runoff simulation in an arid area, J. Hydrol., 625, 130091, https://doi.org/10.1016/j.jhydrol.2023.130091, 2023.

Chen, Y., Su, W., Xing, C., Yin, H., Lin, H., Zhang, C., Liu, H., Hu, Q., and Liu, C.: Kilometer-level glyoxal retrieval via satellite for anthropogenic volatile organic compound emission source and secondary organic aerosol formation identification, Remote Sens. Environ., 270, 112852, https://doi.org/10.1016/j.rse.2021.112852, 2022.

Dai, H., Zhu, J., Liao, H., Li, J., Liang, M., Yang, Y., and Yue, X.: Co-occurrence of ozone and PM2.5 pollution in the Yangtze River Delta over 2013–2019: Spatiotemporal distribution and meteorological conditions, Atmospheric Res., 249, 105363, https://doi.org/10.1016/j.atmosres.2020.105363, 2021.

Dai, H., Liao, H., Li, K., Yue, X., Yang, Y., Zhu, J., Jin, J., Li, B., and Jiang, X.: Composited analyses of the chemical and physical characteristics of co-polluted days by ozone and $PM_{2.5}$ over 2013–2020 in the Beijing–Tianjin–Hebei region, Atmospheric Chem. Phys., 23, 23–39, https://doi.org/10.5194/acp-23-23-2023, 2023.

Ding, A., Huang, X., Nie, W., Chi, X., Xu, Z., Zheng, L., Xu, Z., Xie, Y., Qi, X., Shen, Y., Sun, P., Wang, J., Wang, L., Sun, J., Yang, X.-Q., Qin, W., Zhang, X., Cheng, W., Liu, W., Pan, L., and Fu, C.: Significant reduction of $PM_{2.5}$ in eastern China due to regional-scale emission control: evidence from SORPES in 2011–2018, Atmospheric Chem. Phys., 19, 11791–11801, https://doi.org/10.5194/acp-19-11791-2019, 2019.

Dominici, F., Greenstone, M., and Sunstein, C. R.: Particulate Matter Matters, Science, 344, 257–259, https://doi.org/10.1126/science.1247348, 2014.

Feng, X., Tian, Y., Xue, Q., Song, D., Huang, F., and Feng, Y.: Measurement report: Spatiotemporal and policy-related variations of $PM_{2.5}$ composition and sources during 2015–2019 at multiple sites in a Chinese megacity, Atmospheric Chem. Phys., 21, 16219–16235, https://doi.org/10.5194/acp-21-16219-2021, 2021.

Franklin, M., Koutrakis, P., and Schwartz, P.: The role of particle composition on the association between PM2.5 and mortality, Epidemiol. Camb. Mass, 19, 680–689, https://doi.org/10.1097/ede.0b013e3181812bb7, 2008.

Fu, L., Guo, Y., Zhu, Q., Chen, Z., Yu, S., Xu, J., Tang, W., Wu, C., He, G., Hu, J., Zeng, F., Dong, X., Yang, P., Lin, Z., Wu, F., Liu, T., and Ma, W.: Effects of long-term exposure to ambient fine particulate matter and its specific components on blood pressure and hypertension incidence, Environ. Int., 184, 108464, https://doi.org/10.1016/j.envint.2024.108464, 2024.

Jerrett, M.: Atmospheric science: The death toll from air-pollution sources, Nature, 525, 330–331, https://doi.org/10.1038/525330a, 2015.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Advances in Neural Information Processing Systems, 2017.

Kioumourtzoglou, M.-A., Schwartz, J. D., Weisskopf, M. G., Melly, S. J., Wang, Y., Dominici, F., and Zanobetti, A.: Long-term PM2.5 Exposure and Neurological Hospital Admissions in the

Northeastern United States, Environ. Health Perspect., 124, 23–29, https://doi.org/10.1289/ehp.1408973, 2016.

Leung, D. M., Tai, A. P. K., Mickley, L. J., Moch, J. M., van Donkelaar, A., Shen, L., and Martin, R. V.: Synoptic meteorological modes of variability for fine particulate matter ($PM_{2.5}$) air quality in major metropolitan regions of China, Atmospheric Chem. Phys., 18, 6733–6748, https://doi.org/10.5194/acp-18-6733-2018, 2018.

Liu, C., Yang, Y., Wang, H., Ren, L., Wei, J., Wang, P., and Liao, H.: Influence of Spatial Dipole Pattern in Asian Aerosol Changes on East Asian Summer Monsoon, J. Clim., 36, 1575–1585, https://doi.org/10.1175/JCLI-D-22-0335.1, 2023.

Peng, Z., Zhang, B., Wang, D., Niu, X., Sun, J., Xu, H., Cao, J., and Shen, Z.: Application of machine learning in atmospheric pollution research: A state-of-art review, Sci. Total Environ., 910, 168588, https://doi.org/10.1016/j.scitotenv.2023.168588, 2024.

Seaton, A., MacNee, W., Donaldson, K., and Godden, D.: Particulate air pollution and acute health effects, Lancet Lond. Engl., 345, 176–178, https://doi.org/10.1016/s0140-6736(95)90173-6, 1995.

Shen, L., Mickley, L. J., and Murray, L. T.: Influence of 2000–2050 climate change on particulate matter in the United States: results from a new statistical model, Atmospheric Chem. Phys., 17, 4355–4367, https://doi.org/10.5194/acp-17-4355-2017, 2017.

Song, X.-H., Yan, L., Liu, W., He, J.-Y., Wang, Y.-C., Huang, T.-L., Li, Y.-Y., Chen, M., Meng, J.-J., and Hou, Z.-F.: [Spatiotemporal Distribution Characteristics of Co-pollution of PM2.5 and Ozone over BTH with Surrounding Area from 2015 to 2021], Huan Jing Ke Xue Huanjing Kexue, 44, 1841–1851, https://doi.org/10.13227/j.hjkx.202205089, 2023.

Sun, Y., Yin, H., Liu, C., Zhang, L., Cheng, Y., Palm, M., Notholt, J., Lu, X., Vigouroux, C., Zheng, B., Wang, W., Jones, N., Shan, C., Qin, M., Tian, Y., Hu, Q., Meng, F., and Liu, J.: Mapping the drivers of formaldehyde (HCHO) variability from 2015 to 2019 over eastern China: insights from Fourier transform infrared observation and GEOS-Chem model simulation, Atmospheric Chem. Phys., 21, 6365–6387, https://doi.org/10.5194/acp-21-6365-2021, 2021a.

Sun, Y., Yin, H., Liu, C., Mahieu, E., Notholt, J., Té, Y., Lu, X., Palm, M., Wang, W., Shan, C., Hu, Q., Qin, M., Tian, Y., and Zheng, B.: The reduction in C_2H_6 from 2015 to 2020 over Hefei, eastern China, points to air quality improvement in China, Atmospheric Chem. Phys., 21, 11759–11779, https://doi.org/10.5194/acp-21-11759-2021, 2021b.

Tiwari, P., Cohen, J. B., Lu, L., Wang, S., Li, X., Guan, L., Liu, Z., Li, Z., and Qin, K.: Multi-platform observations and constraints reveal overlooked urban sources of black carbon in Xuzhou and Dhaka, Commun. Earth Environ., 6, 38, https://doi.org/10.1038/s43247-025-02012-x, 2025.

Wang, S., Cohen, J. B., Deng, W., Qin, K., and Guo, J.: Using a New Top-Down Constrained Emissions Inventory to Attribute the Previously Unknown Source of Extreme Aerosol Loadings Observed Annually in the Monsoon Asia Free Troposphere, Earths Future, 9, e2021EF002167, https://doi.org/10.1029/2021EF002167, 2021.

Wang, S., Cohen, J. B., Guan, L., Lu, L., Tiwari, P., and Qin, K.: Observationally constrained global NOx and CO emissions variability reveals sources which contribute significantly to CO2 emissions, Npj Clim. Atmospheric Sci., 8, 87, https://doi.org/10.1038/s41612-025-00977-2, 2025.

Wang, X., Xue, Y., Jin, C., Sun, Y., and Li, N.: Spatial downscaling of surface ozone concentration calculation from remotely sensed data based on mutual information, Front. Environ. Sci., 10, https://doi.org/10.3389/fenvs.2022.925979, 2022.

Wu, P.-C. and Huang, K.-F.: Tracing local sources and long-range transport of PM10 in central Taiwan by using chemical characteristics and Pb isotope ratios, Sci. Rep., 11, 7593, https://doi.org/10.1038/s41598-021-87051-y, 2021.

Yin, H., Sun, Y., Liu, C., Zhang, L., Lu, X., Wang, W., Shan, C., Hu, Q., Tian, Y., Zhang, C., Su, W., Zhang, H., Palm, M., Notholt, J., and Liu, J.: FTIR time series of stratospheric NO_2 over Hefei, China, and comparisons with OMI and GEOS-Chem model data, Opt. Express, 27, A1225–A1240, https://doi.org/10.1364/OE.27.0A1225, 2019.

Yin, H., Sun, Y., Liu, C., Lu, X., Smale, D., Blumenstock, T., Nagahama, T., Wang, W., Tian, Y., Hu, Q., Shan, C., Zhang, H., and Liu, J.: Ground-based FTIR observation of hydrogen chloride (HCl) over Hefei, China, and comparisons with GEOS-Chem model data and other ground-based FTIR stations data, Opt. Express, 28, 8041–8055, https://doi.org/10.1364/OE.384377, 2020.

Yin, H., Sun, Y., Wang, W., Shan, C., Tian, Y., and Liu, C.: Ground-based high-resolution remote sensing of sulphur hexafluoride (SF_6) over Hefei, China: characterization, optical misalignment, influence, and variability, Opt. Express, 29, 34051–34065, https://doi.org/10.1364/OE.440193, 2021a.

Yin, H., Liu, C., Hu, Q., Liu, T., Wang, S., Gao, M., Xu, S., Zhang, C., and Su, W.: Opposite impact of emission reduction during the COVID-19 lockdown period on the surface concentrations of PM2.5 and O3 in Wuhan, China, Environ. Pollut., 289, 117899, https://doi.org/10.1016/j.envpol.2021.117899, 2021b.

Yin, H., Lu, X., Sun, Y., Li, K., Gao, M., Zheng, B., and Liu, C.: Unprecedented decline in summertime surface ozone over eastern China in 2020 comparably attributable to anthropogenic emission reductions and meteorology, Environ. Res. Lett., 16, 124069, https://doi.org/10.1088/1748-9326/ac3e22, 2021c.

Yin, H., Sun, Y., Notholt, J., Palm, M., Ye, C., and Liu, C.: Quantifying the drivers of surface ozone anomalies in the urban areas over the Qinghai-Tibet Plateau, Atmospheric Chem. Phys., 22, 14401–14419, https://doi.org/10.5194/acp-22-14401-2022, 2022a.

Yin, H., Sun, Y., You, Y., Notholt, J., Palm, M., Wang, W., Shan, C., and Liu, C.: Using machine learning approach to reproduce the measured feature and understand the model-to-measurement discrepancy of atmospheric formaldehyde, Sci. Total Environ., 851, 158271, https://doi.org/10.1016/j.scitotenv.2022.158271, 2022b.

Zhang, H., Si, S., and Hsieh, C.-J.: GPU-acceleration for Large-scale Tree Boosting, https://doi.org/10.48550/arXiv.1706.08359, 26 June 2017.

Zhang, Q., Meng, X., Shi, S., Kan, L., Chen, R., and Kan, H.: Overview of particulate air pollution and human health in China: Evidence, challenges, and opportunities, The Innovation, 3, 100312, https://doi.org/10.1016/j.xinn.2022.100312, 2022a.

Zhang, R., Wang, G., Guo, S., Zamora, M. L., Ying, Q., Lin, Y., Wang, W., Hu, M., and Wang, Y.: Formation of urban fine particulate matter, Chem. Rev., 115, 3803–3855, https://doi.org/10.1021/acs.chemrev.5b00067, 2015.

Zhang, Y., Yu, S., Chen, X., Li, Z., Li, M., Song, Z., Liu, W., Li, P., Zhang, X., Lichtfouse, E., and Rosenfeld, D.: Local production, downward and regional transport aggravated surface ozone pollution during the historical orange-alert large-scale ozone episode in eastern China, Environ. Chem. Lett., 20, 1577–1588, https://doi.org/10.1007/s10311-022-01421-0, 2022b.

Zhao, R., Gu, X., Xue, B., Zhang, J., and Ren, W.: Short period PM2.5 prediction based on multivariate linear regression model, PLOS ONE, 13, e0201011, https://doi.org/10.1371/journal.pone.0201011, 2018.

Zhao, X. J., Zhao, P. S., Xu, J., Meng, Pu, W. W., Dong, F., He, D., and Shi, Q. F.: Analysis of a winter regional haze event and its formation mechanism in the North China Plain, Atmospheric Chem. Phys., 13, 5685–5696, https://doi.org/10.5194/acp-13-5685-2013, 2013.