



Supplement of

Technical note: Reconstructing missing surface aerosol elemental carbon data in long-term series with ensemble learning

Qingxiao Meng et al.

Correspondence to: Yunjiang Zhang (yjzhang@nuist.edu.cn)

The copyright of individual parts of the supplement might differ from the article licence.

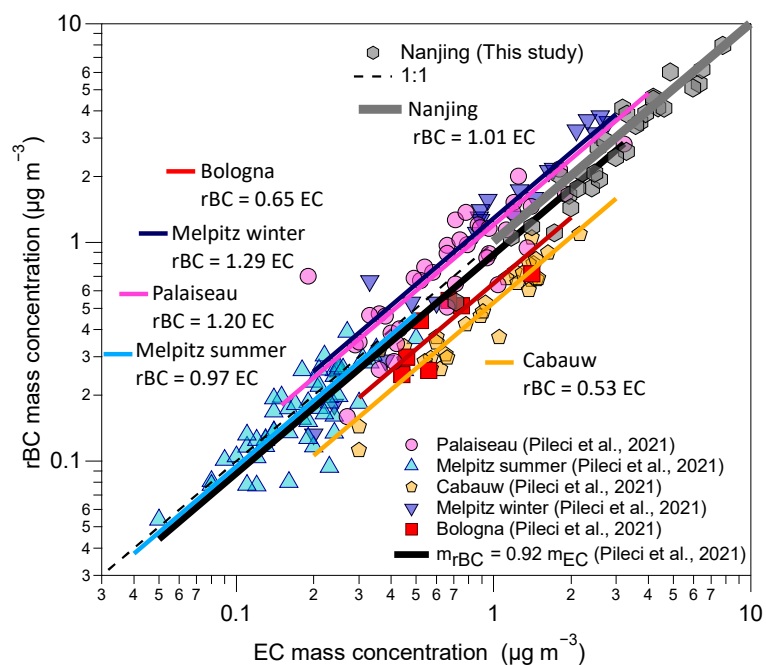


Figure S1. Comparison of rBC and EC across different regions. The rBC data for the comparison in this study was obtained from our previous work for Nanjing (Yang et al., 2019), while the comparison results in other cities were sourced from Pileci et al. (2021).

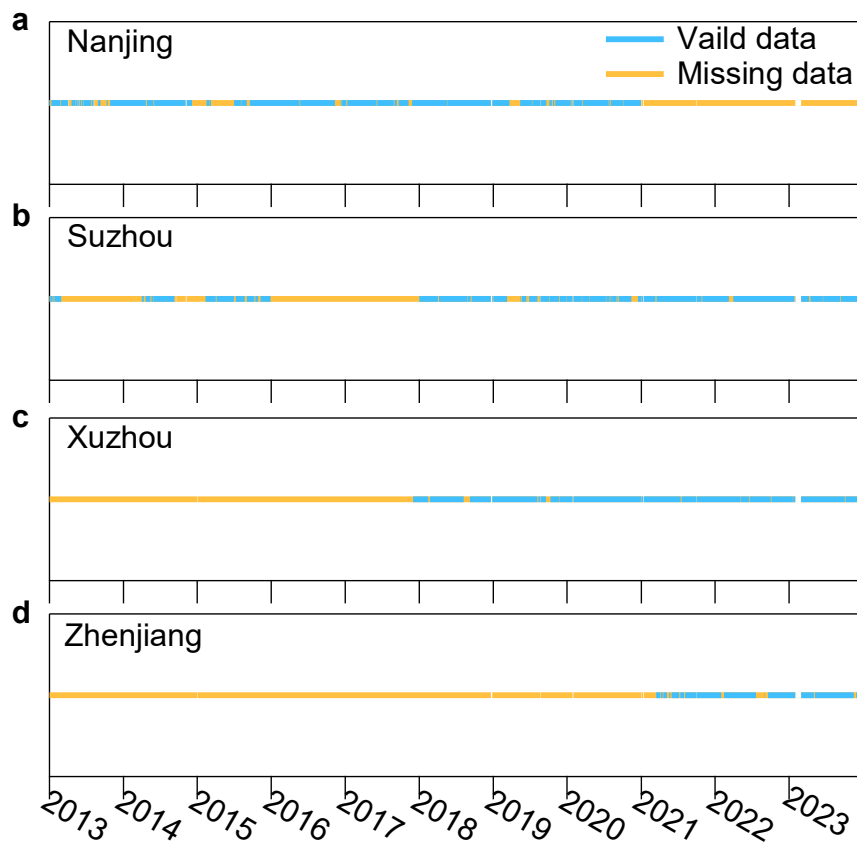


Figure S2. The amount of the vaild data and missing data from 2013 to 2023 in the four cities: a. Nanjing, b. Suzhou, c Xuzhou, and d. Zhenjiang. The blue line represents the amount of valid ground observation data, while the yellow line represents the amount of missing data reconstructed using the ensemble learning model.

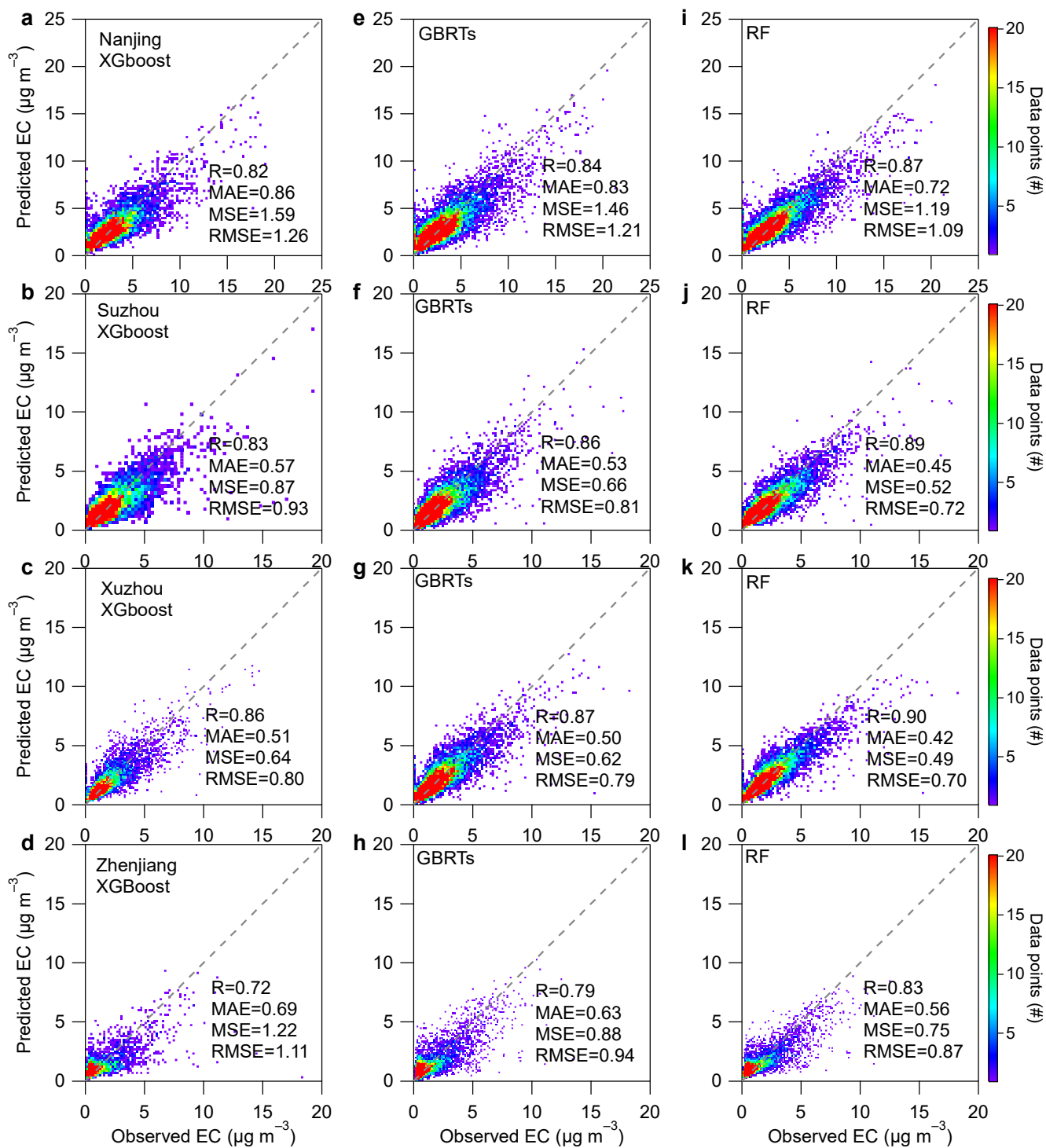


Figure S3. Performance evaluation of three machine learning models using the test set for reconstructing EC observation data (a-d, XGBoost; e-h, GBRTs; i-l, Random Forests).

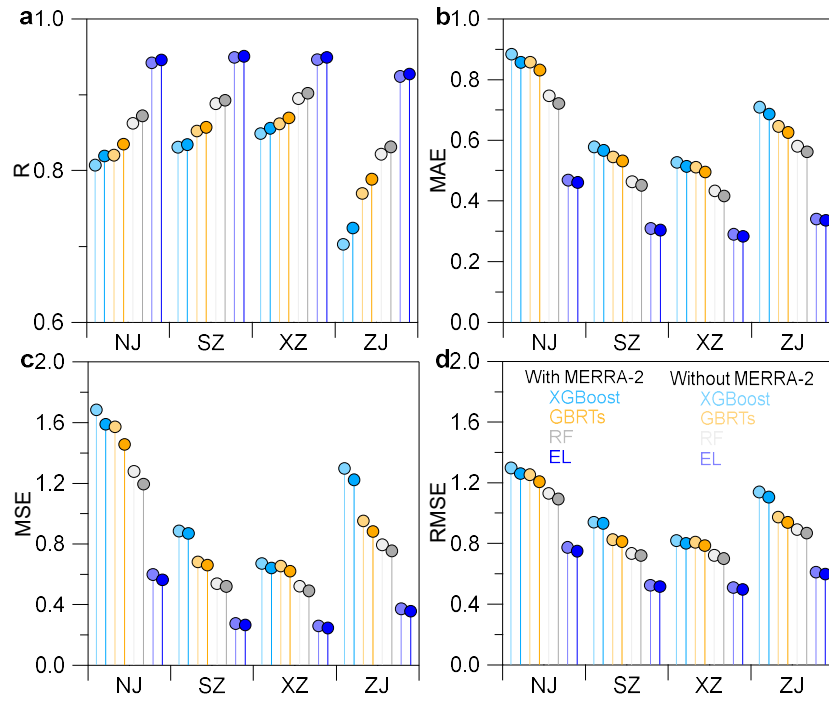


Figure S4. Comparison of the model performance parameters for four cities, training the machine learning model with or without using MERRA-2 BCC as a predictor variable (The darker color represents with MERRA-2, while the lighter color represents without MERRA-2.).

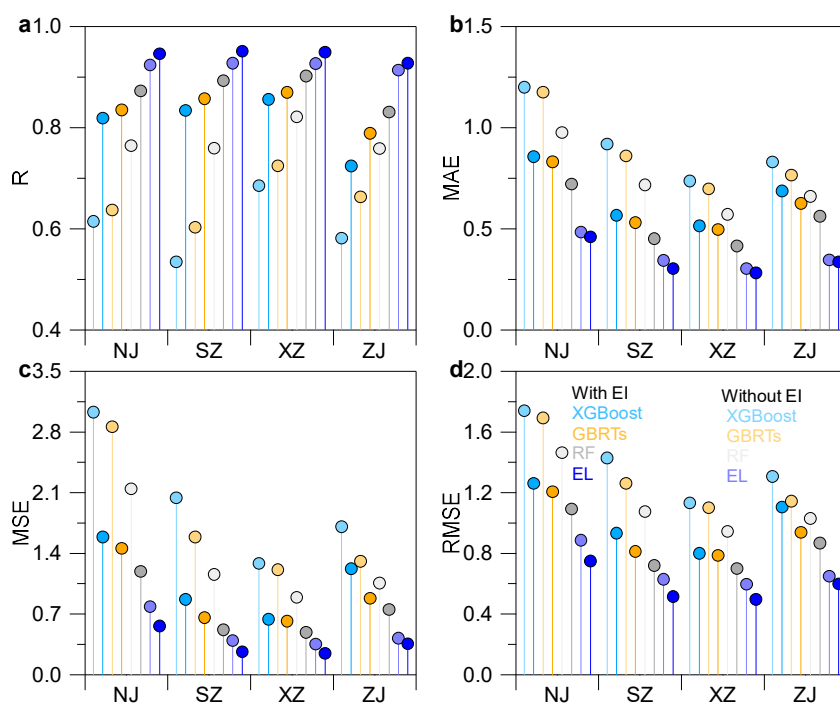


Figure S5. Comparison of model performance metrics for the four cities when training the machine learning model with and without all emission indicators (EI) as predictor variables, such as SO₂, NO₂, CO, and BCC. Darker bars represent model performance with the inclusion of EI, while lighter bars indicate performance without EI.

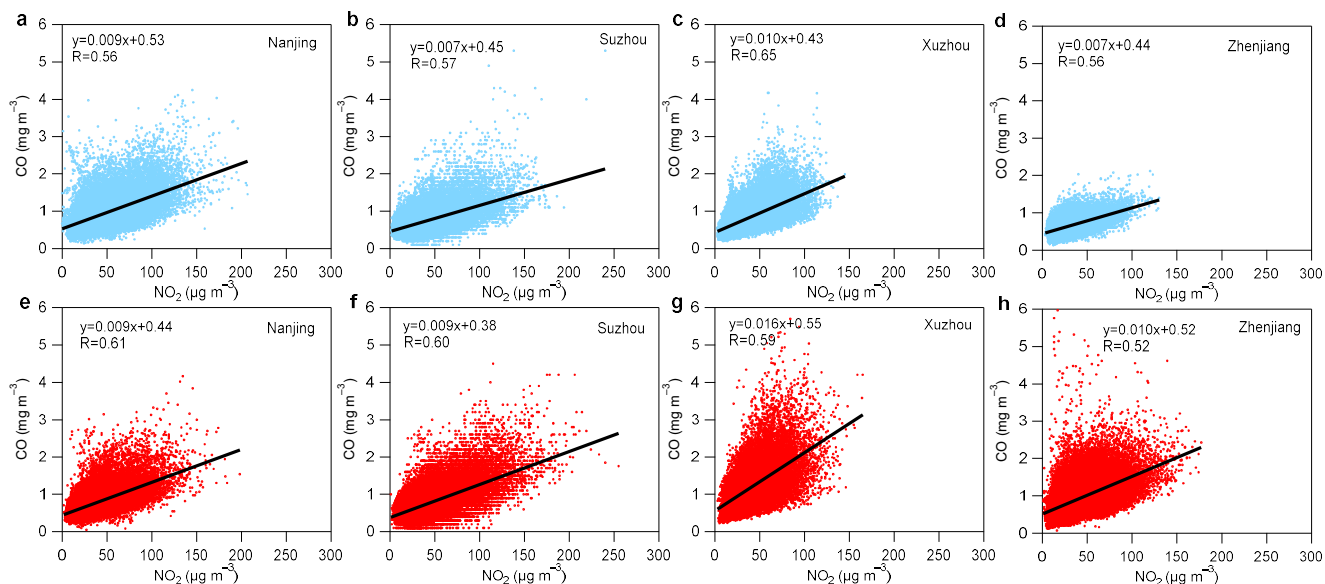


Figure S6. Correlation between CO and NO₂ concentrations. Panels (a–d, in blue) represent periods with available ground-based EC observations, while panels (e–h, in red) correspond to periods with missing EC observations.

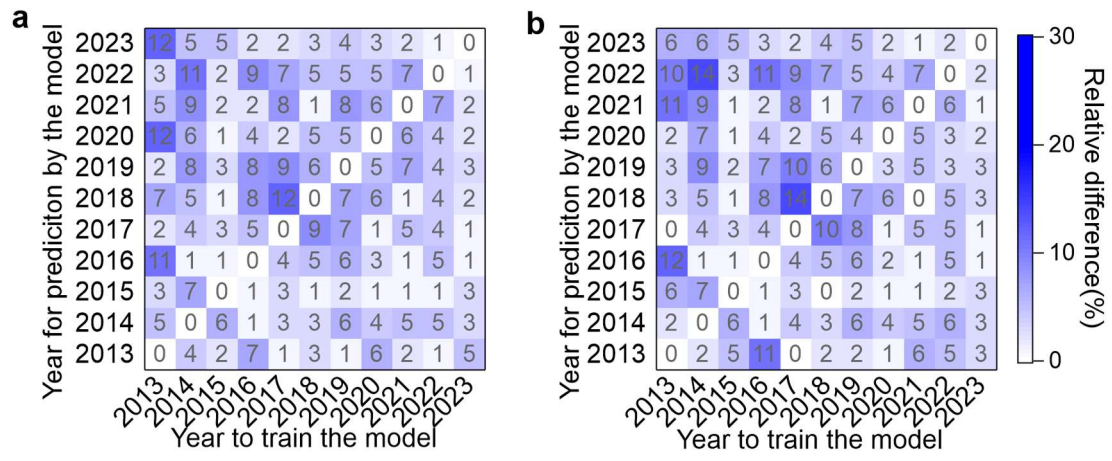


Figure S7. Cross-matrix relative difference analysis of the FEA method for Nanjing. (a) Relative difference in FEA analysis results based on the dataset reconstructed using observational data from 2013 to 2020. (b) Relative difference in FEA analysis results based on the dataset reconstructed using observational data from 2014 to 2019.

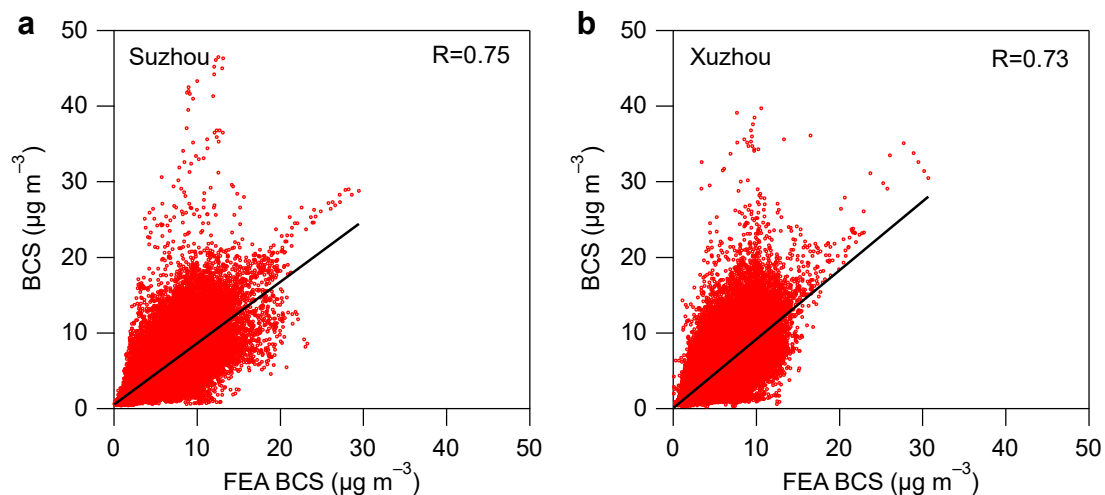


Figure S8. Correlation analysis between the BCS data in MERRA2 and its BCS driven only by meteorological conditions. (a. Suzhou and b. Xuzhou). FEA BCS refers to the Black Carbon Surface concentration data driven by meteorological variations. The calculation method for FEA BCS is as follows: a model is trained using data from 2013, and then this model is applied along with meteorological data from 2013 to 2023 to predict the concentrations for the same period.

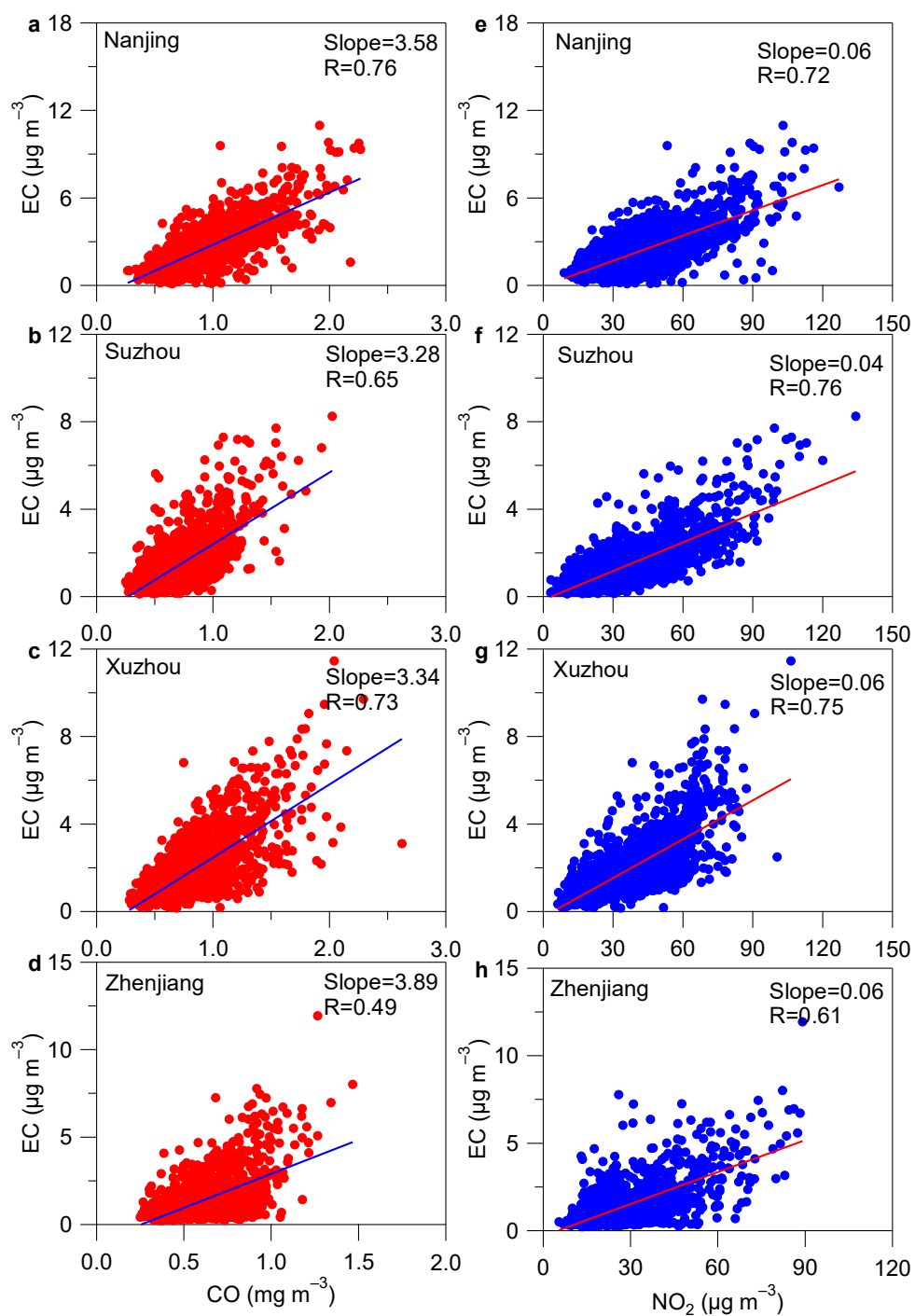


Figure S9. Correlation analysis between observed EC and CO and NO₂ during periods when EC observations were available.

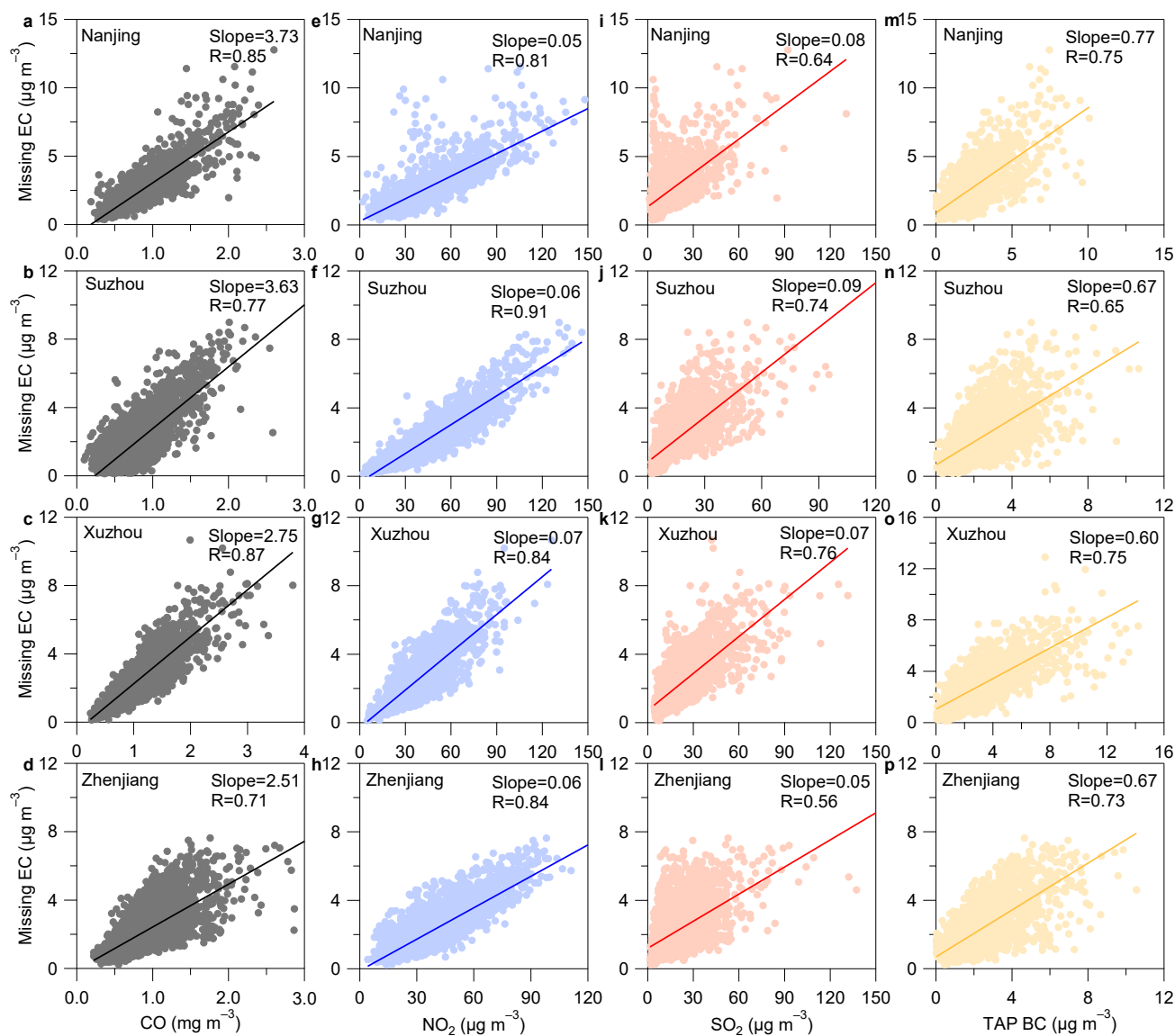


Figure S10. Correlation between reconstructed EC (referring to the originally missing EC observations reconstructed using the ensemble learning model) and observed concentrations of CO, SO_2 , NO_2 , and TAP BC.

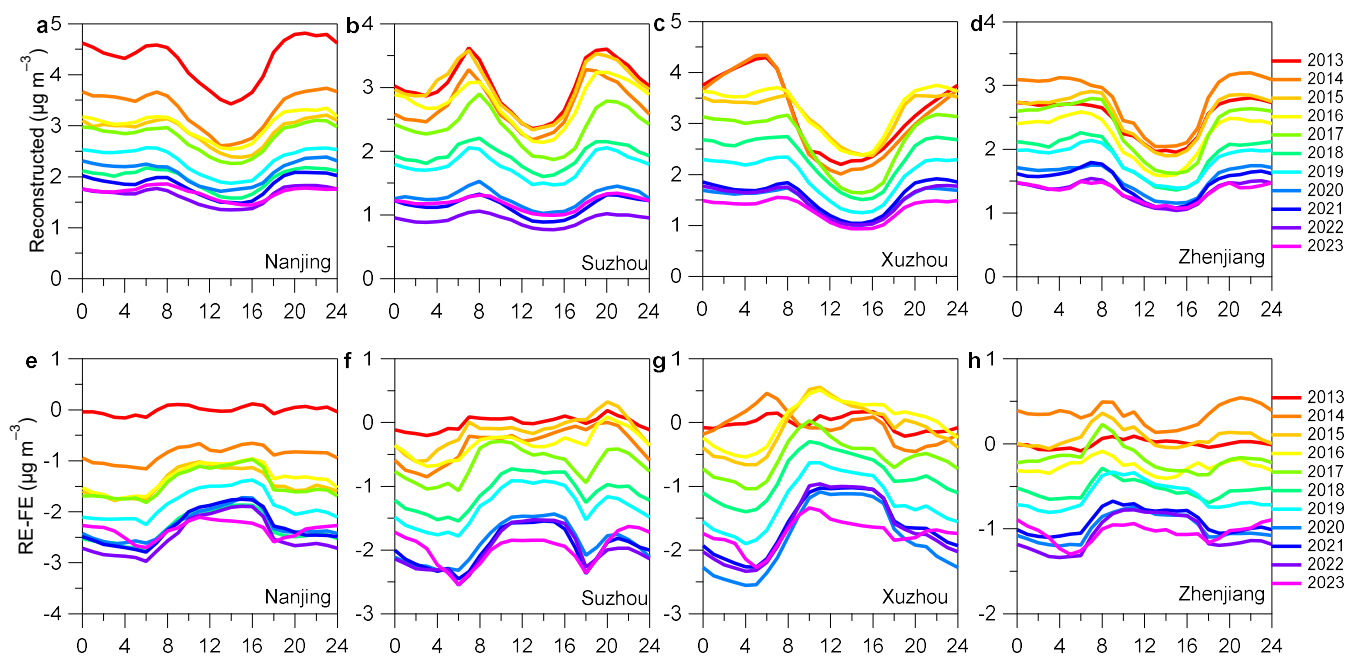


Figure S11. Diurnal variation of EC concentration. a–d. Diurnal variations of EC concentrations for each year from 2013 to 2023. e–h. Diurnal variations of EC concentrations for each year from 2013 to 2023 driven only by emission control. RE-FE represents the contribution of emission control to EC concentrations. It is calculated by training a model using data from 2013 and then applying this model to meteorological data from 2013 to 2023 to predict concentrations. The difference between the reconstructed concentrations and these predictions yields the RE-FE values.

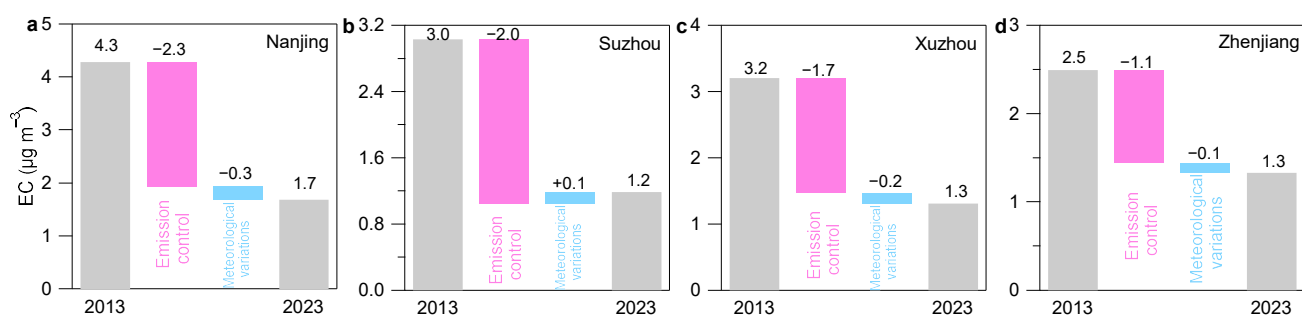


Figure S12. Contributions to EC concentration changes from anthropogenic emission control and meteorological conditions in four cities between 2013 and 2023.

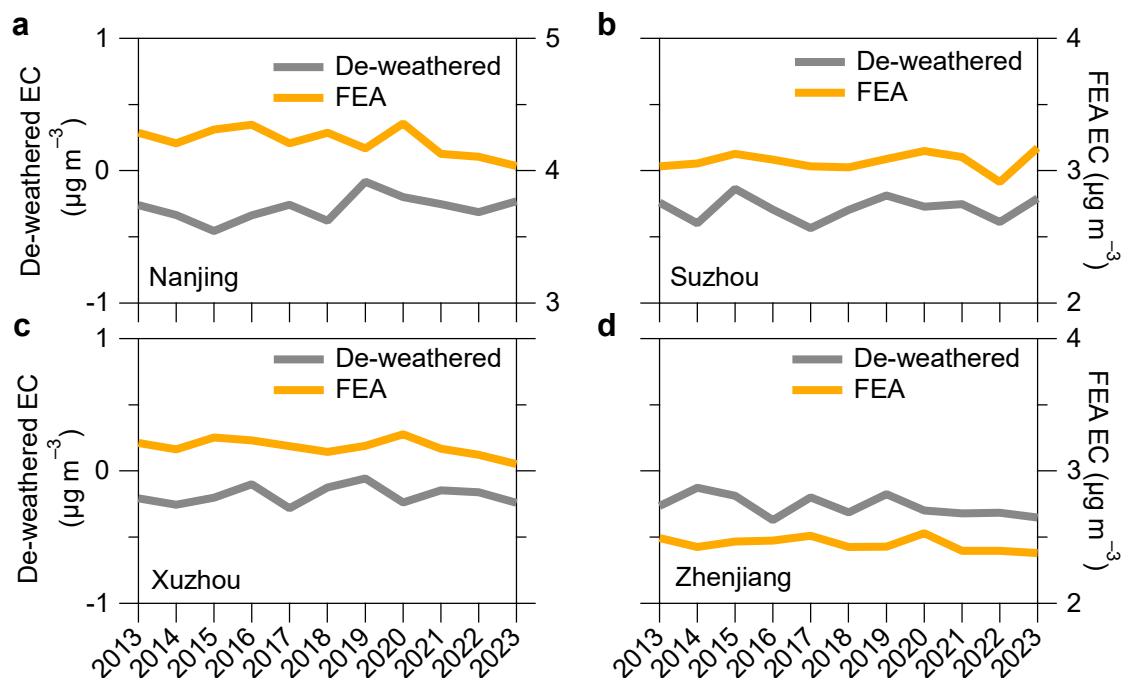


Figure S13. Trend in meteorological-driven EC from 2013 to 2023. Comparison of the results obtained from two different methods, including the FEA method developed in this study and the widely used de-weathered method. The EC concentrations driven only by meteorological conditions were calculated using the FEA method. In contrast, the de-weathered method first estimates EC concentrations influenced only by emission controls. By subtracting these values from the observed data, an approximation of the concentration changes driven by meteorological variability can be derived.

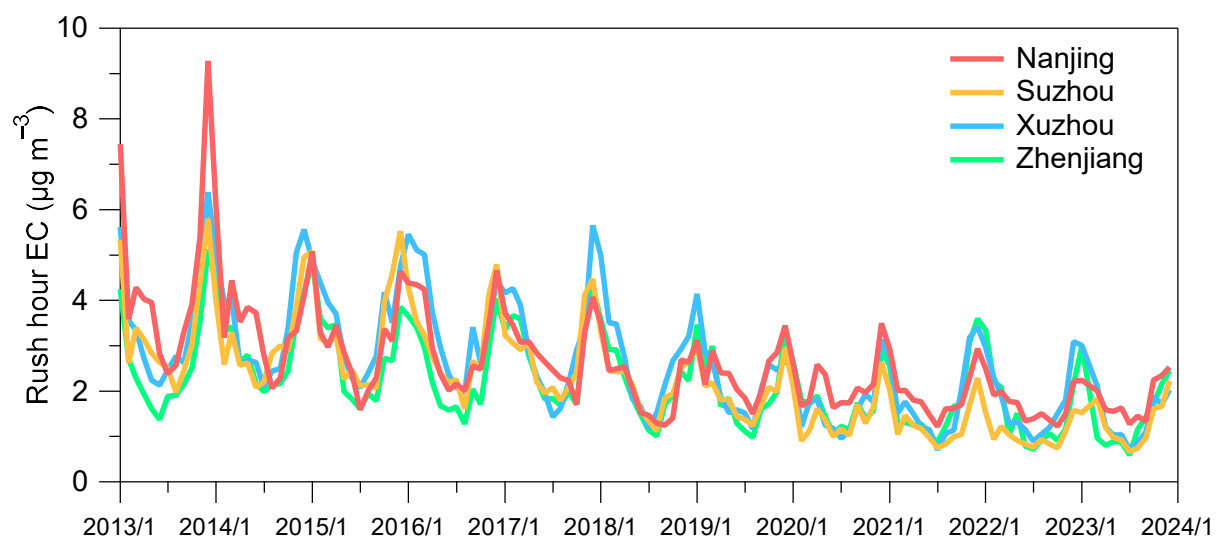


Figure S14. Trends in monthly average EC concentrations during morning rush hours (07:00–09:00) from 2013 to 2023 across the four cities: Nanjing, Suzhou, Xuzhou, and Zhenjiang.

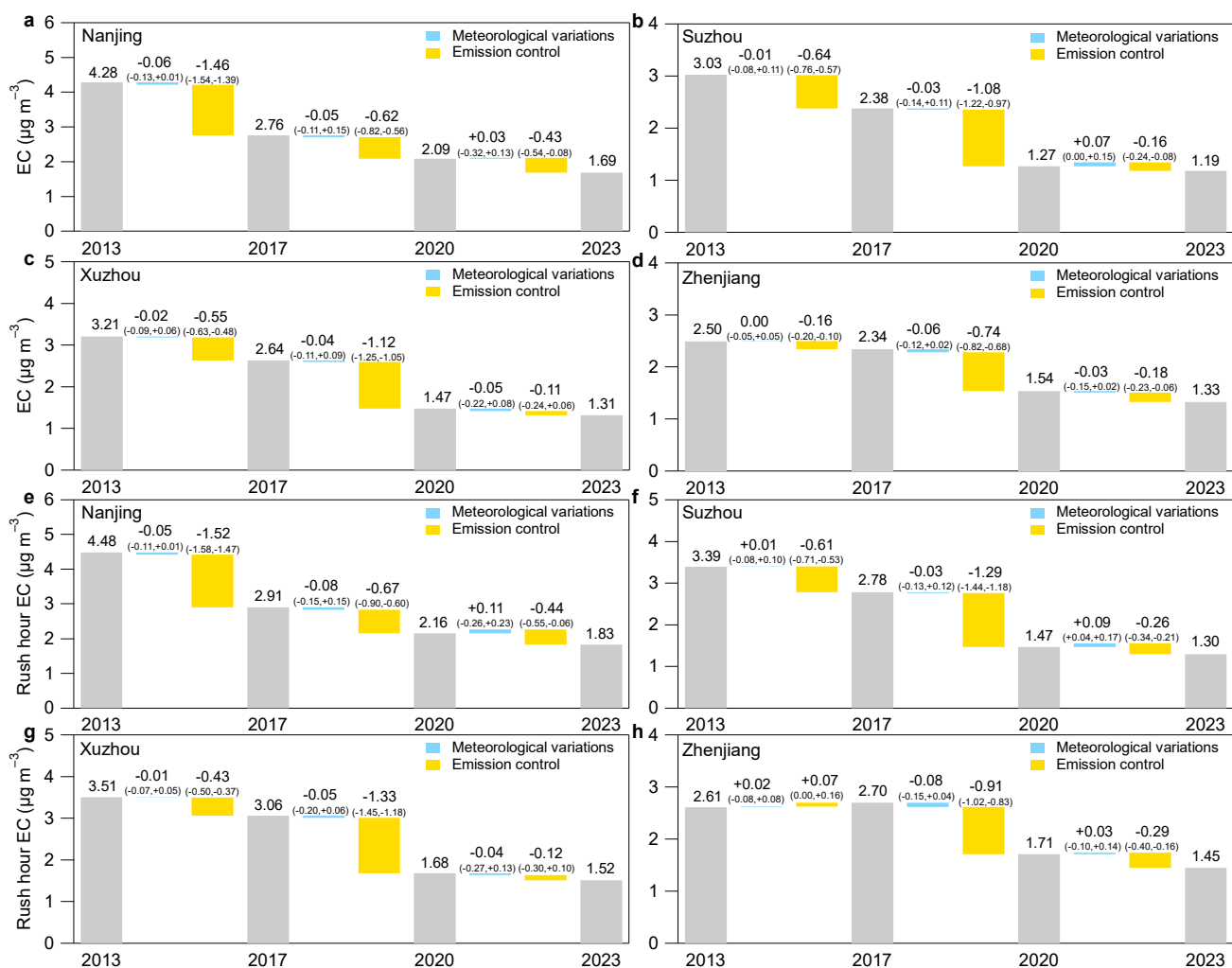


Figure S15. Drivers of the EC trend from 2013 to 2023. **a-d** the contributions of anthropogenic emission control and meteorological variations on the trends in EC concentration in the four cities, **e-h** the contributions of anthropogenic emission control and meteorological variations on the trends in EC concentration during rush hours in the four cities.

Table S1. Summarization of field campaigns at four sites.

City	Longitude	Latitude
Nanjing	118.7538°	32.0551°
Suzhou	120.628°	31.2864°
Xuzhou	117.256°	34.2153°
Zhenjiang	119.6707°	32.1875°

Table S2. Comparison Table of Meteorological and Emission indicator Variables

Variable abbreviations	Meteorological and Emission indicator variables	Unit
U10	10m u-component of wind	m s^{-1}
V10	10m v-component of wind	m s^{-1}
U850	850hPa u-component of wind	m s^{-1}
V850	850hPa v-component of wind	m s^{-1}
W850	850hPa w-component of wind	m s^{-1}
U650	650hPa u-component of wind	m s^{-1}
V650	650hPa v-component of wind	m s^{-1}
W650	650hPa w-component of wind	m s^{-1}
U500	500hPa u-component of wind	m s^{-1}
V500	500hPa v-component of wind	m s^{-1}
W500	500hPa w-component of wind	m s^{-1}
Tmx	Maximum 2m temperature	K
BLH	Boundary layer height	m
RH	Relative Humidity	Dimensionless
SR	Mean surface direct short-wave radiation flux	W m^{-2}
SP	Mean sea level pressure	Pa
TCC	Total cloud cover	Dimensionless
TP	Total precipitation	m
CO	Carbon Monoxide	mg m^{-3}
SO ₂	Sulfur Dioxide	$\mu\text{g m}^{-3}$
NO ₂	Nitrogen Dioxide	$\mu\text{g m}^{-3}$
BCC	Black Carbon Column Mass Density	$\mu\text{g m}^{-2}$

Table S3. Performance evaluation of test set for driving factor analysis model.

City	Method	R	MAE	MSE	RMSE
Nanjing	XGBoost	0.88	0.97	2.02	1.42
Nanjing	GBRTs	0.90	0.86	1.54	1.24
Nanjing	RF	0.92	0.76	1.35	1.16
Nanjing	EL	0.96	0.52	0.62	0.79
Suzhou	XGBoost	0.85	0.72	1.10	1.05
Suzhou	GBRTs	0.87	0.63	0.86	0.93
Suzhou	RF	0.89	0.58	0.81	0.90
Suzhou	EL	0.96	0.38	0.32	0.56
Xuzhou	XGBoost	0.82	0.71	1.17	1.08
Xuzhou	GBRTs	0.86	0.62	0.93	0.97
Xuzhou	RF	0.88	0.56	0.92	0.96
Xuzhou	EL	0.95	0.36	0.33	0.58
Zhenjiang	XGBoost	0.89	0.52	0.50	0.71
Zhenjiang	GBRTs	0.91	0.47	0.40	0.63
Zhenjiang	RF	0.93	0.42	0.34	0.59
Zhenjiang	EL	0.96	0.28	0.17	0.41

Table S4. Correlation analysis of data from the Reconstructed, MERRA-2, and TAP Modeled methods with ground-based observation data.

City	Data	R	Slope
Nanjing	Reconstructed	0.97	0.86
Nanjing	MERRA-2	0.61	0.64
Nanjing	TAP	0.68	0.54
Suzhou	Reconstructed	0.97	0.88
Suzhou	MERRA-2	0.59	0.87
Suzhou	TAP	0.68	0.55
Xuzhou	Reconstructed	0.98	0.89
Xuzhou	MERRA-2	0.72	1.27
Xuzhou	TAP	0.76	0.92
Zhenjiang	Reconstructed	0.97	0.87
Zhenjiang	MERRA-2	0.69	0.95
Zhenjiang	TAP	0.65	0.38

Table S5. Annual mean EC concentration data for the four representative cities in the Yangtze River Delta (Nanjing, Suzhou, Xuzhou, and Yangzhou) from 2013 to 2023.

Year	Nanjing ($\mu\text{g m}^{-3}$)	Suzhou ($\mu\text{g m}^{-3}$)	Xuzhou ($\mu\text{g m}^{-3}$)	Zhenjiang ($\mu\text{g m}^{-3}$)
2013	4.28	3.03	3.21	2.50
2014	3.33	2.72	3.11	2.78
2015	2.89	2.99	3.17	2.54
2016	3.01	2.76	3.26	2.21
2017	2.76	2.38	2.64	2.34
2018	1.97	1.93	2.30	1.89
2019	2.32	1.77	1.94	1.82
2020	2.09	1.27	1.47	1.54
2021	1.84	1.14	1.57	1.48
2022	1.63	0.92	1.49	1.34
2023	1.69	1.19	1.32	1.33

Table S6. The results of the Mann-Kendall Test (MK) in this study are compared to those obtained using the De-weathered method.

City	Method	Slope	P _{value}
Nanjing	De-weathered	-0.208	<0.05
Nanjing	FEA	-0.214	<0.05
Suzhou	De-weathered	-0.221	<0.05
Suzhou	FEA	-0.229	<0.05
Xuzhou	De-weathered	-0.221	<0.05
Xuzhou	FEA	-0.204	<0.05
Zhenjiang	De-weathered	-0.142	<0.05
Zhenjiang	FEA	-0.138	<0.05

Table S7. Analysis of driving factors for CO, SO₂ and NO₂.

City	Year	Pollutant	Emission control	Meteorological variables
Nanjing	2013-2023	CO	-0.42 (mg m ⁻³)	0.01 (mg m ⁻³)
Nanjing	2013-2023	NO ₂	-20.99(μg m ⁻³)	-2.23(μg m ⁻³)
Nanjing	2013-2023	SO ₂	-16.60(μg m ⁻³)	-2.65(μg m ⁻³)
Suzhou	2013-2023	CO	-0.12 (mg m ⁻³)	0.03 (mg m ⁻³)
Suzhou	2013-2023	NO ₂	-34.64(μg m ⁻³)	2.17(μg m ⁻³)
Suzhou	2013-2023	SO ₂	-21.60(μg m ⁻³)	1.00(μg m ⁻³)
Xuzhou	2013-2023	CO	-0.83 (mg m ⁻³)	-0.02 (mg m ⁻³)
Xuzhou	2013-2023	NO ₂	-18.52(μg m ⁻³)	-0.65(μg m ⁻³)
Xuzhou	2013-2023	SO ₂	-42.74(μg m ⁻³)	0.39(μg m ⁻³)
Zhenjiang	2013-2023	CO	-0.57 (mg m ⁻³)	0.03 (mg m ⁻³)
Zhenjiang	2013-2023	NO ₂	-3.18(μg m ⁻³)	-0.37(μg m ⁻³)
Zhenjiang	2013-2023	SO ₂	-28.40(μg m ⁻³)	1.25(μg m ⁻³)