Diagnosing drivers of $PM_{2.5}$ simulation biases from meteorology, chemical composition, and emission sources using an efficient machine learning method

Shuai Wang[1], Mengyuan Zhang[1], Yueqi Gao[1], Peng Wang[2,3], Qingyan Fu[4], Hongliang Zhang[1,3,5]

[1]Department of Environmental Science and Engineering, Fudan University, Shanghai 200438, China

[2]Department of Atmospheric and Oceanic Sciences, and Institute of Atmospheric Sciences, Fudan University, Shanghai, 200438, China

[3]IRDR ICoE on Risk Interconnectivity and Governance on Weather/Climate Extremes Impact and Public Health, Fudan University, Shanghai, China;

[4]Shanghai Environmental Monitoring Center, Shanghai 200235, China

[5]Institute of Eco-Chongming (IEC), Shanghai 200062, China

Correspondence to: Hongliang Zhang (zhanghl@fudan.edu.cn)

Figure

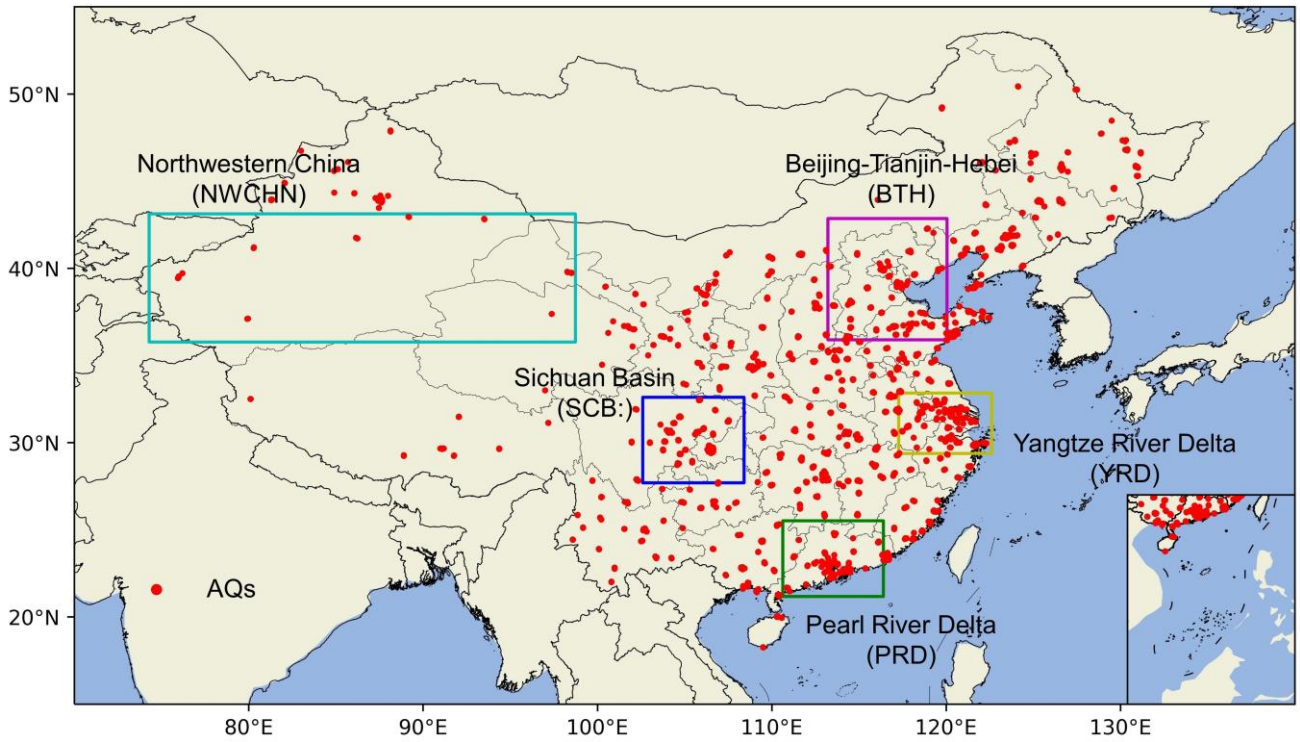

Figure S1. Location of observation sites and five key regions in China, AQs: air quality sites.

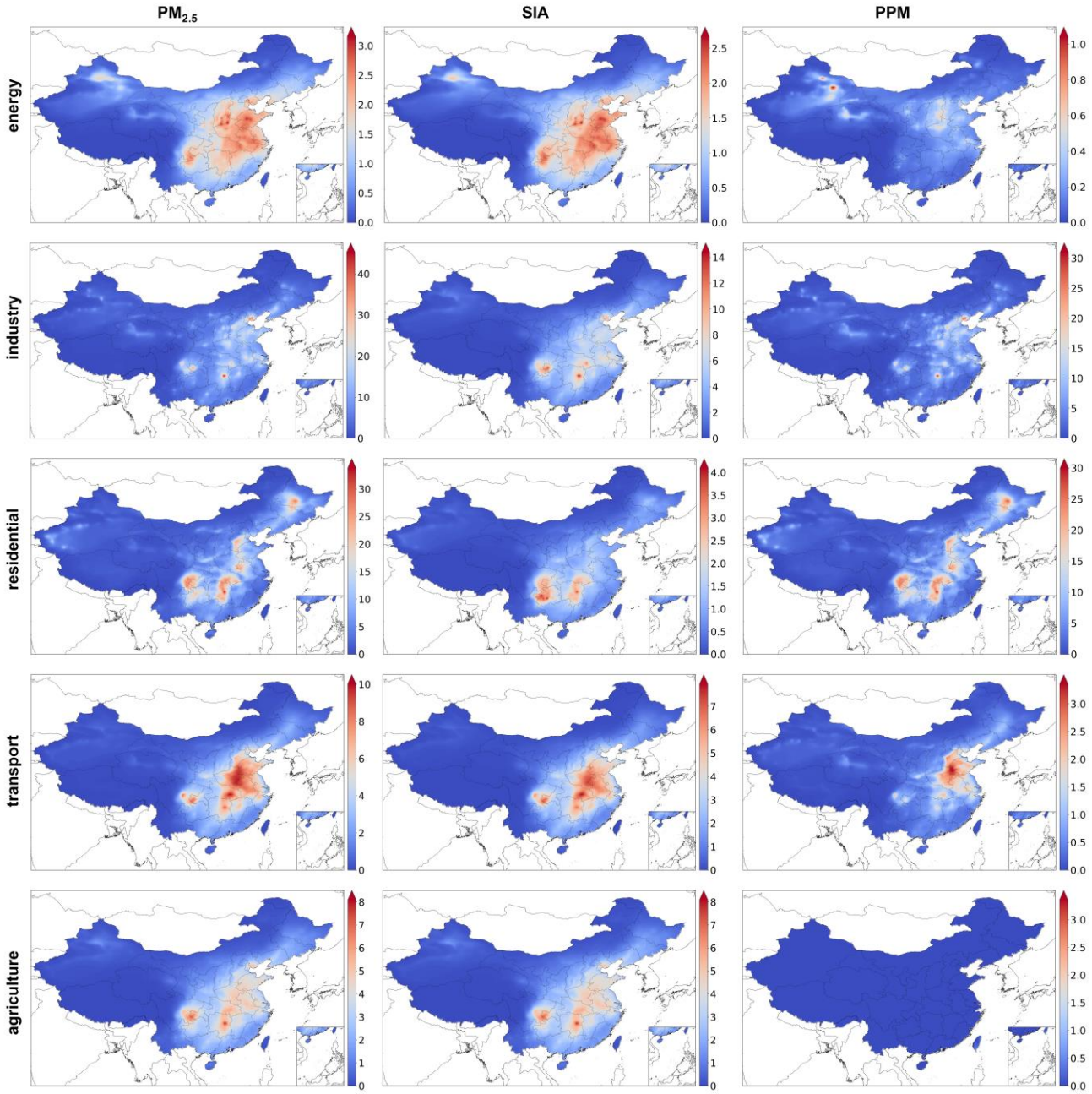

Figure S2. Annual average contributions (μg/m^3) from different sources to PM2.5, PPM, and SIA in China.

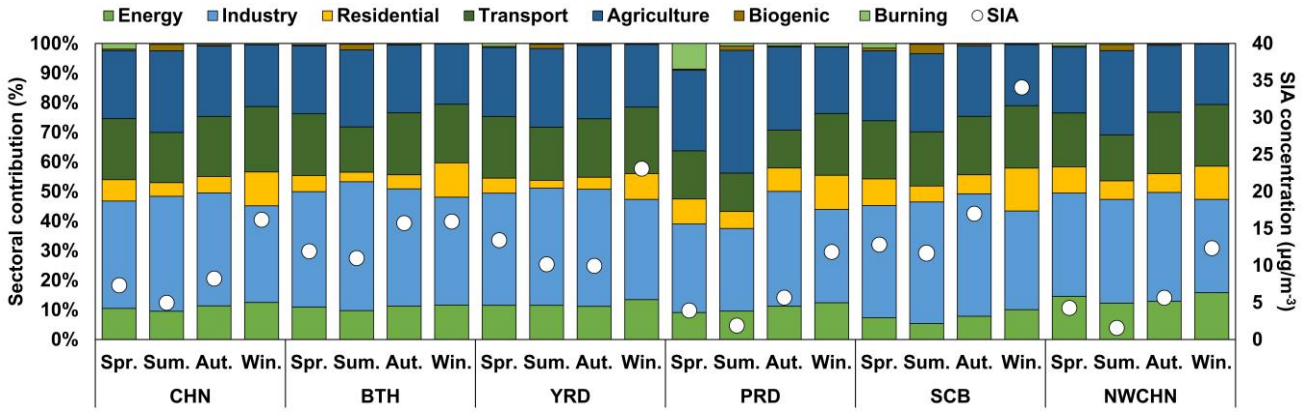

Figure S3. Seasonal average fractional contributions from different sources to secondary PM2.5 concentration (white circle on the right-hand axis) in China and five regions.

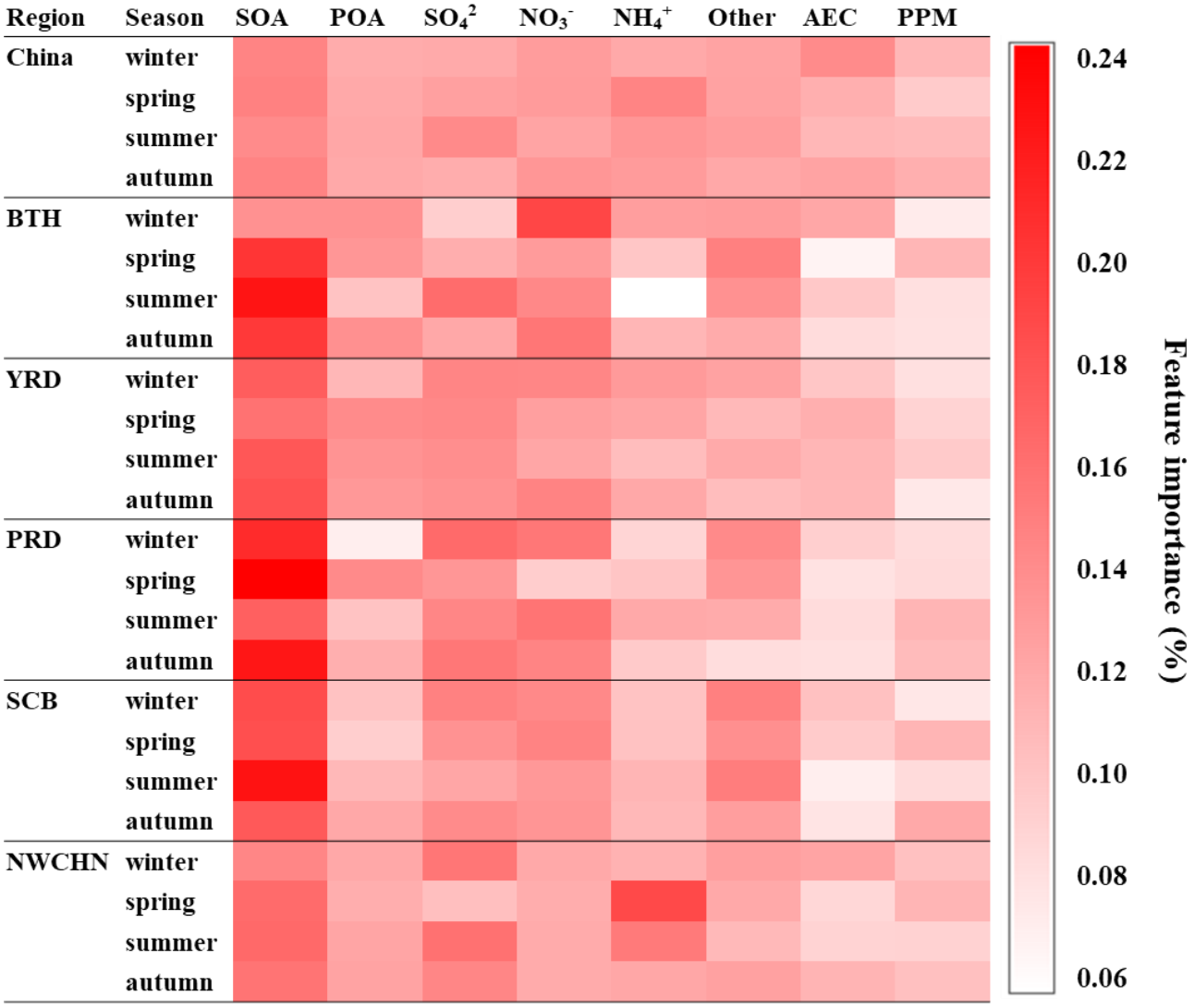

Figure S4. Contribution (%) of each $PM_{2.5}$ components to CMAQ simulation biases by region and season.

Table

Table S1. Summary of the WRF model variables used in this study.

Variable	Description	Spatial Resolution	Temporal Resolution
PRSFC	surface pressure	36 km × 36 km	daily
PBL	Planetary boundary layer height		
TEMP2	temperature at 2 m		
RH	relative humidity at 2 m		
WSPD10	wind speed at 10 m		
WDIR10	wind direction at 10 m		
CFRAC	total cloud fraction		
SOIM	volumetric soil moisture in top		
GLW	longwave radiation at ground		
RSTOMI	bulk stomatal resistance		

Table S2. The CMAQ model performance in China and key regions in 2019 with a spatial resolution of 36 km (OBS is mean observation; PRE is mean prediction; MNB is mean normalized bias; MNE is mean normalized error; MFB is mean fractional bias; MFE is mean fractional error.

Region	Index	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Criteria
China	OBS	66.68	64.93	57.97	48.88	41.93	42.08	39.46	28.74	37.20	45.28	47.76	54.37	
	PRE	44.79	38.93	38.06	24.57	23.91	29.56	25.76	26.01	35.02	36.38	42.47	50.97	
	MNB	-0.12	-0.18	-0.22	-0.41	-0.34	-0.16	-0.27	0.00	0.01	-0.09	0.00	0.19	≤±0.6
	MNE	0.41	0.44	0.40	0.47	0.47	0.56	0.43	0.48	0.36	0.35	0.35	0.48	≤0.75
	MFB	-0.27	-0.34	-0.36	-0.61	-0.53	-0.41	-0.43	-0.18	-0.10	-0.20	-0.09	0.03	
	MFE	0.47	0.53	0.49	0.66	0.61	0.64	0.54	0.49	0.36	0.40	0.35	0.42	
BTH	OBS	87.55	83.38	56.34	48.20	38.08	38.92	36.17	27.40	35.70	45.14	52.12	65.04	
	PRE	57.85	48.95	39.29	26.51	25.16	30.13	27.07	30.18	37.74	38.12	47.53	60.93	
	MNB	-0.19	-0.25	-0.22	-0.37	-0.23	-0.09	-0.18	0.25	0.14	-0.04	0.01	0.14	≤±0.6
	MNE	0.42	0.43	0.37	0.44	0.44	0.53	0.41	0.62	0.42	0.38	0.38	0.49	≤0.75
	MFB	-0.34	-0.41	-0.34	-0.55	-0.39	-0.31	-0.32	-0.01	0.01	-0.15	-0.10	-0.03	
	MFE	0.51	0.54	0.45	0.60	0.53	0.57	0.48	0.51	0.37	0.40	0.37	0.45	
YRD	OBS	68.03	54.49	50.08	38.69	32.48	28.07	24.63	23.69	27.80	39.41	40.68	55.27	
	PRE	49.19	42.16	36.92	31.49	23.86	28.31	26.30	29.65	22.63	28.77	28.86	48.51	
	MNB	-0.17	-0.05	-0.18	-0.12	-0.22	0.14	0.14	0.25	-0.16	-0.22	-0.24	-0.05	≤±0.6
	MNE	0.45	0.45	0.38	0.43	0.39	0.56	0.50	0.56	0.37	0.37	0.37	0.38	≤0.75
	MFB	-0.35	-0.22	-0.31	-0.27	-0.35	-0.09	-0.04	0.05	-0.28	-0.35	-0.36	-0.17	
	MFE	0.53	0.47	0.45	0.47	0.48	0.48	0.46	0.46	0.44	0.47	0.46	0.41	
PRD	OBS	45.08	24.55	27.90	22.52	19.67	13.48	17.73	21.59	28.73	35.47	40.74	41.85	
	PRE	34.40	21.08	24.31	24.65	17.06	12.78	12.18	20.20	22.93	25.64	25.27	31.88	
	MNB	-0.05	0.06	-0.06	0.24	-0.09	0.02	-0.31	0.01	-0.11	-0.21	-0.33	-0.08	≤±0.6
	MNE	0.52	0.48	0.35	0.45	0.34	0.34	0.35	0.42	0.41	0.39	0.38	0.45	≤0.75
	MFB	-0.25	-0.11	-0.16	0.10	-0.18	-0.07	-0.41	-0.12	-0.23	-0.33	-0.46	-0.24	
	MFE	0.53	0.46	0.38	0.35	0.37	0.34	0.44	0.42	0.45	0.45	0.50	0.48	
SCB	OBS	78.16	53.53	44.85	33.25	33.35	23.35	19.99	24.43	28.15	25.02	38.87	63.99	
	PRE	98.44	60.28	43.83	31.03	29.49	40.63	32.93	38.19	36.02	34.38	56.15	80.33	
	MNB	0.38	0.30	0.10	0.01	-0.01	1.07	**0.82**	**0.75**	0.50	0.59	0.59	0.41	≤±0.6
	MNE	0.56	0.49	0.47	0.35	0.44	1.24	**0.95**	**0.93**	0.72	**0.77**	0.73	0.61	≤0.75
	MFB	0.19	0.14	-0.06	-0.09	-0.16	0.40	0.38	0.32	0.20	0.28	0.32	0.19	
	MFE	0.42	0.38	0.44	0.36	0.45	0.63	0.56	0.56	0.49	0.50	0.49	0.44	
NWCHN	OBS	89.32	74.70	50.50	36.88	41.38	22.87	24.54	23.51	23.88	33.08	53.19	75.09	
	PRE	27.28	31.38	19.98	13.81	11.65	11.99	12.01	13.05	16.17	15.18	23.12	29.29	
	MNB	-0.61	-0.48	-0.46	-0.51	-0.55	-0.37	-0.31	-0.26	-0.20	-0.39	-0.42	-0.46	≤±0.6
	MNE	0.66	0.55	0.57	0.57	0.59	0.51	0.45	0.51	0.47	0.51	0.56	0.60	≤0.75
	MFB	-1.00	-0.78	-0.75	-0.79	-0.87	-0.59	-0.49	-0.47	-0.39	-0.63	-0.70	-0.79	
	MFE	1.03	0.83	0.82	0.83	0.90	0.69	0.60	0.63	0.57	0.72	0.78	0.88	

$$\mathrm{MNB} = \frac{\sum_{i=1}^{N}(P_i - O_i)}{\sum_{i=1}^{N} O_i}$$

$$\mathrm{MNE} = \frac{\sum_{i=1}^{N} |P_i - O_i|}{\sum_{i=1}^{N} O_i}$$

$$\mathrm{MFB} = \frac{1}{N} \sum_{i=1}^{N} \frac{(P_i - O_i)}{(P_i + O_i)/2}$$

$$\mathrm{MFE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|P_i - O_i|}{(P_i + O_i)/2}$$

* i represents the pairing of N observations O and predictions P by site and time.

Table S3. Meteorological Model performance in China 2019 (OBS is mean observation; PRE is mean prediction; MB is mean bias; GE is gross error; and RMSE is root mean square error). The value that do not meet the benchmark are shown as bold.

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Benchmark*
T2 (K)	OBS	272.00	274.99	282.81	290.19	292.85	296.99	299.56	299.18	294.79	288.01	280.94	274.67	
	PRE	275.07	277.25	283.24	289.62	292.72	297.15	299.22	299.08	295.14	288.77	282.69	277.48	
	MB	**3.06**	**2.25**	0.41	**-0.59**	-0.15	0.14	-0.36	-0.12	0.34	**0.74**	**1.73**	**2.80**	≤±0.5
	GE	**5.36**	**5.67**	**5.04**	**4.78**	**4.89**	**4.86**	**4.70**	**4.53**	**4.47**	**4.36**	**4.54**	**5.05**	≤2.0
	RMSE	6.89	7.31	7.05	6.65	6.82	6.74	6.79	6.41	6.15	5.93	5.94	6.49	
WS (ms^{-1})	OBS	3.09	3.30	3.34	3.54	3.54	3.42	3.24	3.26	3.18	3.29	3.26	2.92	
	PRE	4.47	4.51	4.55	4.55	4.57	4.17	4.21	4.25	4.22	4.55	4.63	4.50	
	MB	**1.39**	**1.20**	**1.21**	**1.01**	**1.03**	**0.75**	**0.96**	**0.99**	**1.03**	**1.26**	**1.37**	**1.59**	≤±0.5
	GE	**2.20**	**2.13**	**2.14**	**2.08**	**2.15**	**2.01**	1.97	2.00	**2.04**	**2.19**	**2.27**	**2.33**	≤2.0
	RMSE	**2.91**	**2.81**	**2.76**	**2.65**	**2.76**	**2.57**	**2.58**	**2.63**	**2.67**	**2.85**	**2.95**	**3.04**	≤2.0
WD (°)	OBS	167.32	147.60	164.78	160.21	170.01	167.11	168.12	154.88	151.97	164.15	171.68	176.45	
	PRE	144.64	131.84	162.14	157.96	163.20	162.35	168.64	137.28	127.28	142.09	147.78	157.83	
	MB	**-16.11**	**-10.59**	-3.42	-2.56	-5.13	-3.81	1.81	**-10.41**	**-14.39**	**-17.22**	**-18.44**	**-12.78**	≤±10
	GE	**74.45**	**67.43**	**73.59**	**72.09**	**73.60**	**77.26**	**75.04**	**74.95**	**71.70**	**73.80**	**72.70**	**74.10**	≤±30
	RMSE	92.25	86.05	90.23	88.83	90.19	93.33	91.69	91.10	88.07	90.90	90.22	91.84	
RH (%)	OBS	82.22	82.04	73.75	67.41	62.01	61.95	60.03	60.34	63.45	70.50	72.80	76.99	
	PRE	77.23	81.01	76.17	69.82	61.49	61.40	61.89	61.40	61.27	67.06	69.54	76.26	
	MB	-5.00	-1.03	2.41	2.41	-0.53	-0.55	1.86	1.06	-2.18	-3.44	-3.26	-0.73	
	GE	11.85	12.17	14.94	17.04	19.40	16.85	14.99	16.20	18.34	17.37	17.28	13.32	
	RMSE	15.05	15.46	19.32	22.23	24.91	22.23	19.83	21.11	23.15	22.24	21.89	17.09	

Note: * are benchmarks limits suggested by (Emery et al., 2001)

The formulas used in the table are as follows:

$$\mathrm{MB} = \frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)$$

$$\mathrm{GE} = \frac{1}{N} \sum_{i=1}^{N} |P_i - O_i|$$

$$\mathrm{RMSE} = \left[\frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)^2 \right]^{1/2}$$

* i represents the pairing of N observations O and predictions P by site and time.

Table S4. CMAQ simulation bias under dry and wet days in China and key regions, unit: μg/m^3. L_1: dry days (RH < 60 %); L_1: wet days (RH > 80 %)

region	winter	spring	summer	autumn

	L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2
China	-27.87	2.34	-14.56	1.24	-7.09	3.93	-7.11	-0.79
BTH	-22.86	12.90	-13.64	-6.57	-4.74	1.10	-1.66	5.08
YRD	-6.46	-6.23	-6.94	-3.40	-2.83	5.39	-7.09	-3.23
PRD	-2.77	-9.54	-21.55	-0.84	-6.34	-3.56	-15.21	-5.18
SCB	4.17	38.71	0.17	0.02	-2.07	29.41	-12.76	20.22
NWCHN	-44.64	-52.98	-27.29	-20.14	-14.06	-5.83	-22.72	-8.30

Table S5. Contribution (%) of each sectoral source to CMAQ simulation biases by region and season. $PM_{2.5\_res}$: residential, $PM_{2.5\_ene}$: energy, $PM_{2.5\_tra}$: transportation, $PM_{2.5\_arg}$: agriculture, $PM_{2.5\_ind}$: industry, Other: other $PM_{2.5}$ components.

region	$PM_{2.5\_res}$	$PM_{2.5\_ene}$	$PM_{2.5\_tra}$	$PM_{2.5\_arg}$	$PM_{2.5\_ind}$	EC	Other
BTH	0.20	0.16	0.14	0.13	0.10	0.14	0.12
china	0.16	0.16	0.14	0.14	0.13	0.13	0.14
FWP	0.18	0.18	0.13	0.12	0.12	0.14	0.14
NWCHN	0.18	0.15	0.12	0.15	0.13	0.13	0.14
PRD	0.17	0.14	0.14	0.13	0.13	0.14	0.15
SCB	0.16	0.15	0.16	0.14	0.12	0.13	0.13
YRD	0.18	0.16	0.15	0.13	0.12	0.14	0.12

Reference

Emery, C., Tai, E., and Yarwood, G.: Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes, Prepared for the Texas natural resource conservation commission, by ENVIRON International Corporation, 2001.