

Supplementary material

Supplementary material text

S1. Analysis of mutual influence among source regions. In the inversion process, it assumed that the boundary areas provided enough distance to eliminate the influence among source regions; hence, the pollutant emissions from one region would not have a large effect on the surrounding source areas. To demonstrate the accuracy of this assumption in this simulation, seven simulation regions and one buffer region were selected as the inversion regions in this study. The purpose of the buffer region was to reduce the influence of emission sources outside the study area on the NO₂ or HCHO concentrations in that region, and minimized the influence of boundary conditions on the inversion regions as well. The sensitivity of HCHO concentration to VOCs emissions and NO₂ concentrations to NO_x emissions for each simulated source area were obtained. The sensitivity of NO₂ or HCHO concentration to VOCs or NO_x emissions in each area was calculated as a proportion of the total sensitivity (**Figs. S2 and S3**).

As shown in **Fig. S2**, NO_x emissions from each region had the largest effect on NO₂ concentrations within that region, whereas the effect on other regions was only near the adjacent boundaries of the two regions, and the contribution to other regions did not exceed 10%. Similarly, the VOCs emissions in each region in **Fig. S3** had the largest effect on the HCHO concentration in that region, whereas the effect on other regions was small. The spatial distribution of the sensitivity percentage of VOCs

emissions to HCHO was more regional than that of NO_x emissions to NO₂ concentrations. In general, the influence of the contribution from the boundary area was not significant, providing a reasonable separation to eliminate the boundary effects in the inversion. The locally emitted NO_x or VOCs in each inversion region were the main contributors to the NO₂ or HCHO concentrations in each inversion region, ensuring the reliability of the inversion results.

S2. Effects of changes in the covariance matrix of the pseudodata test. The pseudodata test enabled simultaneously checking the effects of the changes in the covariance matrix of A_i and R_i in the latest method on the inversion results. It showed that the emission adjustment factors of NO_x and VOCs for the simulation area could be infinitely close to 1 ($\delta < 0.01$) after four to five iterations, and the inversion after the first iteration were already very close to the final results. Considering the computational cost and the difficulty of simulation, the response of the first iteration of the test inversion process was sufficient to obtain the influence of the covariance matrices A_i and R_i on inversion results. In addition, two pollutants, NO_x and VOCs, were both involved with different covariance matrices, A_i and R_i . When testing the covariance matrices of A_i and R_i , it was necessary to consider the effects when those matrices of two pollutants changed simultaneously. It was also necessary to consider the effect of separate changes on the covariance matrices of a particular A_i and R_i on the inversion results, while keeping the other A_i and R_i remain constant.

43 To determine the sensitivities of the inversion results to the simulated and measured
44 errors, the covariance matrices of A_i and R_i of both NO_x and VOCs changed
45 simultaneously. The NC region was selected as the test area, and the trends of the
46 adjustment factors for NO_x and VOCs were obtained after applying same changes to
47 the $\sigma(\text{obs})$ and $\sigma(\text{E})$ (**Fig. S4**). The trend showed that as the value of $\sigma(\text{obs})$ increasing
48 and the value of $\sigma(\text{E})$ decreasing, the adjustment factor became larger, especially
49 when $\sigma(\text{obs})$ greater than $1.2 \times 10^{15} \text{ molec}\cdot\text{cm}^{-2}$ and $\sigma(\text{E})$ less than 30%. However, in
50 general, the effects of $\sigma(\text{E})$ and $\sigma(\text{obs})$ fluctuations on the inversion results were
51 relatively small.

52 In the subsequent tests, four scenarios were set up in which the other three errors were
53 remain constant when a certain error was varied, to determine the effects of the
54 variation of the covariance matrix of a certain A_i and R_i on the final inversion
55 results (**Fig. S5**). **Figure S5a** showed when fixing the $\sigma(\text{obs})$ of VOCs at 0.8×10^{15}
56 $\text{molec}\cdot\text{cm}^{-2}$ and the $\sigma(\text{E})$ of NO_x and VOCs at 200%, the variations of NO_x emission
57 adjustment factors with $0.1 \times 10^{15} \text{ molec}\cdot\text{cm}^{-2}$ interval of the $\sigma(\text{obs})$ of NO_2 between
58 $0.5 \times 10^{15} \text{ molec}\cdot\text{cm}^{-2}$ to $1.5 \times 10^{15} \text{ molec}\cdot\text{cm}^{-2}$. **Figure S5b** showed when fixing the
59 $\sigma(\text{obs})$ of NO_x at $0.8 \times 10^{15} \text{ molec}\cdot\text{cm}^{-2}$ and the $\sigma(\text{E})$ of NO_x and VOCs at 200%, the
60 variations of VOCs emission adjustment factors with $0.1 \times 10^{15} \text{ molec}\cdot\text{cm}^{-2}$ interval
61 of the $\sigma(\text{obs})$ of HCHO between $0.5 \times 10^{15} \text{ molec}\cdot\text{cm}^{-2}$ to $1.5 \times 10^{15} \text{ molec}\cdot\text{cm}^{-2}$.
62 Among them, the adjustment factors of VOC emissions in the SC region fluctuated
63 slightly more than that of other regions, but in general, the overall variations were

insignificant. Similarly, **Figs. S5c** and **S5d** showed when fixing the $\sigma(\text{obs})$ of NO_x and VOCs at $0.8 \times 10^{15} \text{ molec} \cdot \text{cm}^{-2}$ and fixing the $\sigma(\text{E})$ of NO_x or VOCs at 200%, the variations of adjustment factors with $\sigma(\text{E})$ of VOCs or NO_x varied from 100% to 200%.

Under the scenario with fixed $\sigma(\text{E})$, the adjustment factors of NO_x emissions were not sensitive to the changes in the $\sigma(\text{obs})$ from $0.5 \times 10^{15} \text{ molec} \cdot \text{cm}^{-2}$ to $1.5 \times 10^{15} \text{ molec} \cdot \text{cm}^{-2}$. The maximum difference of adjustment factors among source areas was approximately -0.022 . The variations of VOCs emission adjustment factors were slightly larger in changing of $\sigma(\text{obs})$, especially for SC and CC regions, with gradually increased trend (approximately 0.050 and 0.323) when $\sigma(\text{obs})$ changing from $0.5 \times 10^{15} \text{ molec} \cdot \text{cm}^{-2}$ to $1.5 \times 10^{15} \text{ molec} \cdot \text{cm}^{-2}$. However, the adjustment factors in other regions had smaller variations. Under the scenario of fixed $\sigma(\text{obs})$, the adjustment factors of NO_x and VOCs emissions were not sensitive to the change of $\sigma(\text{E})$, and almost remained the same during the variations of $\sigma(\text{E})$ from 100% to 200%.

Comparisons of the effects of $\sigma(\text{E})$ and $\sigma(\text{obs})$ on the adjustment factors under various scenarios showed that some source regions were more sensitive to changes in $\sigma(\text{obs})$ than that in other parameters. Since the inversion process required for several iterations, the influences of $\sigma(\text{E})$ and $\sigma(\text{obs})$ would gradually decrease in each iteration, and the adjustment factors could return to the error-permissible range (**Fig. S5**).

S3. Model performance of meteorological parameters. The WRF model was used in this study to provide the meteorological field for the CAMx model. The simulated meteorological variables, such as temperature, relative humidity, barometric pressure, and wind speed were validated with the observed values from 1 June to 30 September 2019 in this section. The simulated daily meteorological results of the model grid where the meteorological stations located were extracted and compared to monitored meteorological elements.

A comparative analysis of simulated versus observed meteorological data, including wind speed, temperature, relative humidity, and atmospheric pressure, across all monitoring stations in China (**Figs. S6 to S8**) was conducted. The results showed high consistency between simulation results and observed data in most regions. Correlation analysis (**Fig. S6**) demonstrated that the model effectively captured the meteorological trends, with correlation coefficients (R) for temperature, humidity, and atmospheric pressure consistently exceeding 0.8, indicating a robust alignment with observational data. Although the correlation for wind speed was marginally lower, it showed satisfactory agreement in the central and eastern parts of China. The distribution of the normalized mean bias (NMB) and normalized mean error (NME) highlighted systematic biases within certain regions (**Figs. S7 and S8**). Specifically, atmospheric pressure and temperature showed minimal bias in the central and eastern regions, suggesting high fidelity in simulation accuracy. Conversely, the temperature was slightly underestimated in the southwestern regions, potentially attributable to the

complex topography and distinct climatic characteristics of the area. Wind speed was overestimated nationwide. Relative humidity was found to be underestimated in the northeastern regions, whereas a slight overestimation was observed in the southern regions.

On a daily average basis, the verification outcomes of meteorological elements across various regions of this study were statistically compiled (**Tables S2 to S5**), showing a high degree of congruence between simulated temperature and atmospheric pressure and their observed counterparts throughout the simulation period. The NMB for atmospheric pressure and temperature in each region was negative, suggesting a minor underestimation relative to observed data. Nevertheless, the R for atmospheric pressure exceeded 0.90, and those for temperature were above 0.70. Furthermore, the simulated relative humidity demonstrated a reasonable alignment with observations, with an averaged NMB within $\pm 14.5\%$, and R were approximately above 0.70. The averaged NMB for wind speed was approximately 37%, with all R exceeding 0.5. Collectively, these results indicated that the model was capable of replicating the observed meteorological conditions for most of the simulation period, with relatively minor errors in model performance assessment.

S4. Impacts of inversion on ground-level NO₂ concentrations. Figure S9 showed scatter plots of comparisons between simulated and monitored NO₂ concentrations across key regions in China. The analysis found that both prior and posterior simulations generally underestimated NO₂ levels, with particularly

pronounced underestimations in the Northwest (NWC) region in China. Simulations with posterior emissions showed even lower NO₂ concentrations comparing to their prior counterparts, possibly attributing to the typically low column NO₂ concentrations observed by the satellite in NWC. In addition, the wind speed was mostly overestimated in the meteorological simulation as showed in Fig S6-S8, coupled with large uncertainties of VOCs emissions in these areas. In studies that generally use inversion models to optimize emission inventories, it is common to observe that the simulated surface NO₂ concentrations are lower than the observed values and the results are even lower than the a priori emission. Wang et al. (2020) found in their inversion of NO_x emissions in China that the NO_x emissions decreased significantly after inversion, and the simulated surface NO₂ concentrations in some regions were lower than the observed values. Cheng et al. (2021) also pointed out in their study on the Beijing-Tianjin-Hebei region that even after optimizing emissions using the 3D-VAR method, the simulation of surface NO₂ was still underestimated.

The core reasons for this phenomenon can be summarized into three aspects: First, there is a bias in the vertical allocation of emissions in the a priori emission inventory, which is manifested as an underestimation of the actual lifting height of elevated sources. This leads to an excessive amount of NO_x being incorrectly allocated to the upper boundary layer. Such structural errors in the vertical direction result in significantly lower simulated NO₂ at surface stations compared to observations, even when the total emission amount is correct. Second, there are differences in sensitivity between satellite

147 observations and surface processes. Satellite data mainly constrain the total column
148 concentration but are insufficient in capturing near-surface vertical mixing and small-
149 scale emission hotspots. Moreover, ground-based stations are over-represented in urban
150 high-emission areas, while satellite retrievals capture contributions from rural
151 distributed sources. The scale difference between the two leads to the necessity of
152 reducing the total emissions during the inversion process to address the high observed
153 values at urban stations (assuming overestimation in the a priori emissions), making
154 final inversion results being even lower than the a priori emission inventory.

155 However, the joint inversion approach had notable improvements in the simulation
156 performance for NO₂ concentrations in the EC and CC regions. These enhancements
157 were evidenced by the adjusted simulation results from postinversions, which aligned
158 more closely with ground observations. **Table S6** corroborated this trend, showing a
159 significant improvement in model performance metrics by the joint inversion, while
160 maintaining the integrity of the underlying trends.

161 A comparative analysis was performed on satellite-derived and simulated daily NO₂
162 and HCHO column concentrations from June to September 2019, with statistical
163 metrics including R, NMB, and NME (**Table S7**). For HCHO, the joint inversion
164 method performed prominently in most regions. The R values in most regions were
165 improved, except in the EC and SC regions, but the declines were relatively
166 insignificant. Meanwhile, the NMB in all regions decreased, especially in the SC
167 region dropping from -48.21% to -19.13%. The NME decreased in all regions after

inversions. For NO_2 , the joint inversion method also showed good performance. The
 R values in most regions were improved, with only slight decreases in the EC and SC
 regions. The NMB decreased, especially in the CC region, changed from 29.84% to -
 13.64%. The NME in most regions decreased, with only a slight increase in the NEC
 region. Due to cloud interference, daily satellite data is usually difficult to achieve full
 coverage of the interested domain. Therefore, long-term averaged values often used
 for comparison with modeled data in the inversion processes. However, this approach
 tends to overlook the diurnal variations in emissions, introducing a certain degree of
 uncertainty into the inversion results. Based on the daily comparison results, the joint
 inversion method in general outperformed the priori and individual inversion methods
 in simulating both HCHO and NO_2 in most source regions.

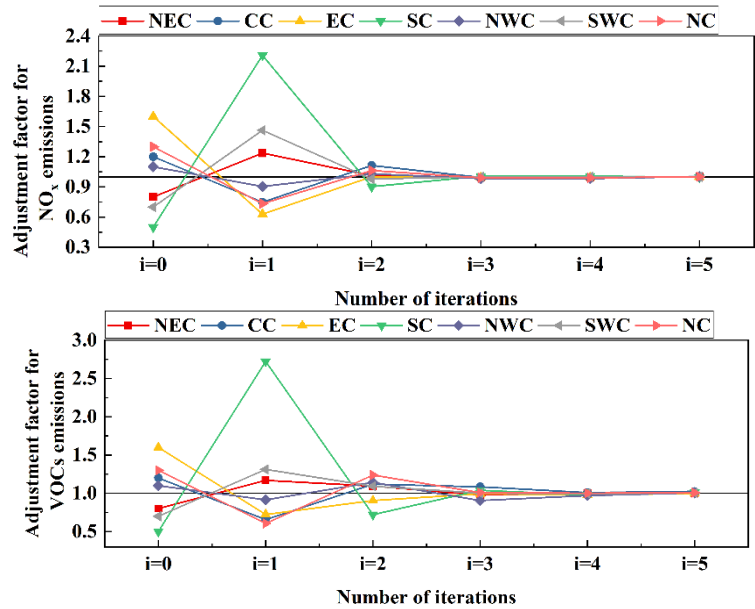


Fig. S1. Pseudodata testing multiple iterations of emission adjustment factors for NO_x and VOCs emissions in various regions

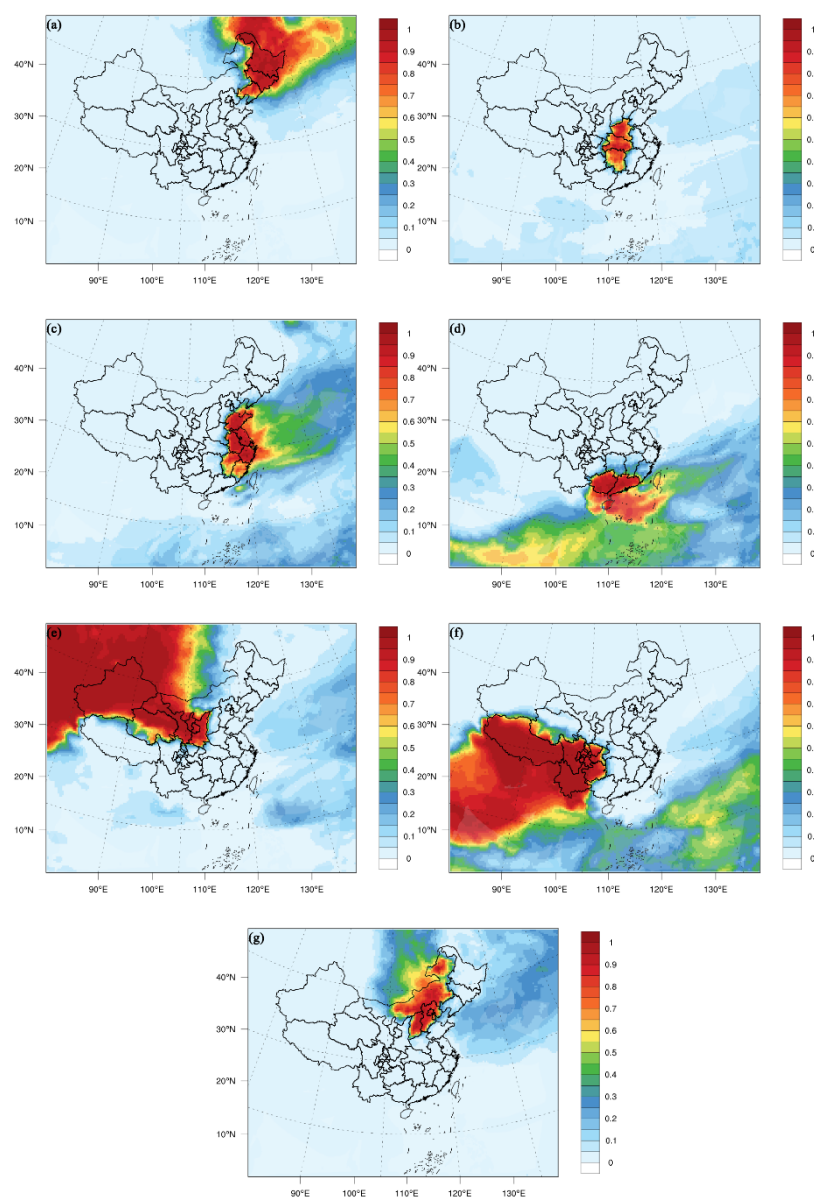


Fig. S2. Spatial distribution of the sensitivity percentages of simulated column NO₂ concentrations to local NO_x emissions for each simulated region

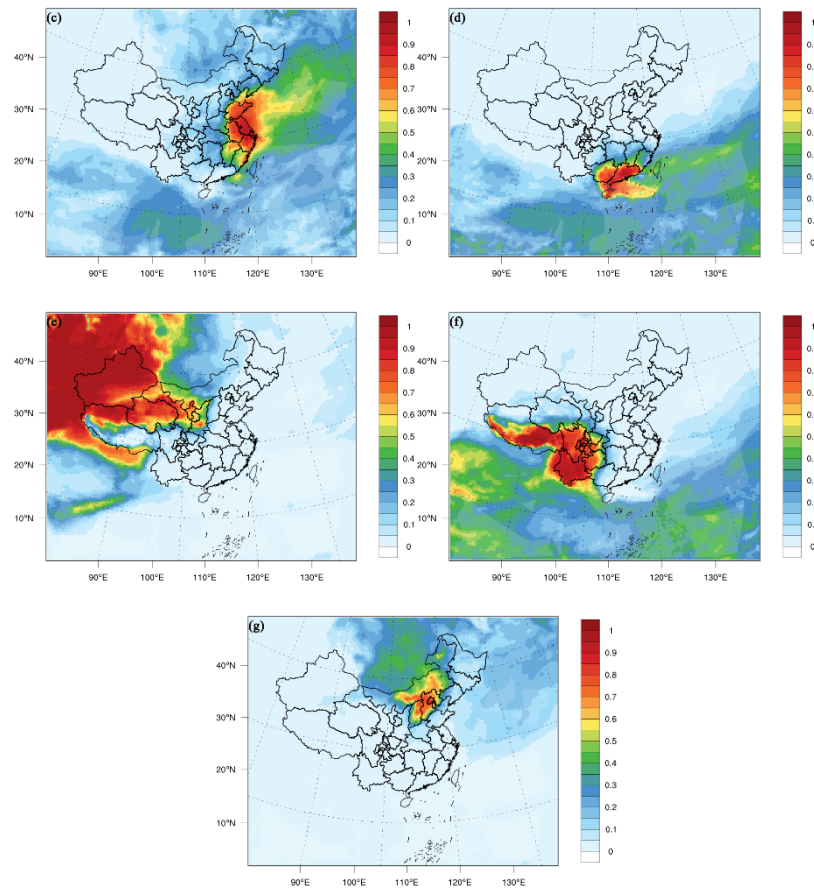
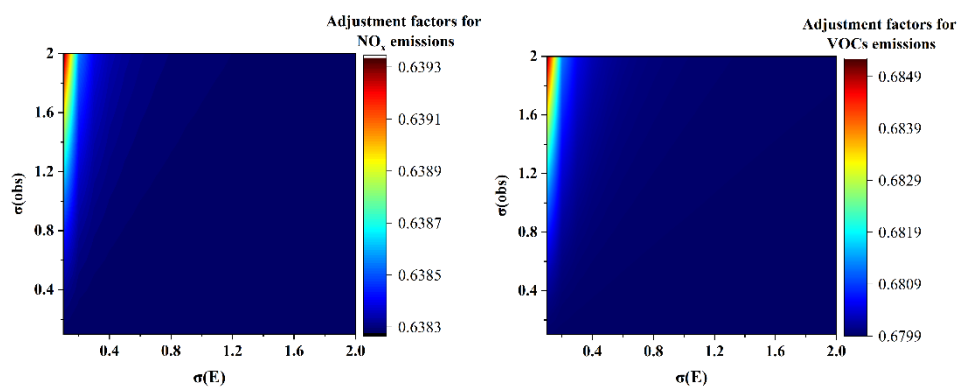


Fig. S3, Spatial distribution of the sensitivity percentages of simulated column HCHO concentrations to local VOCs emissions for each simulated region



ig. S4. Sensitivities of inversion results to A_i and R_i values

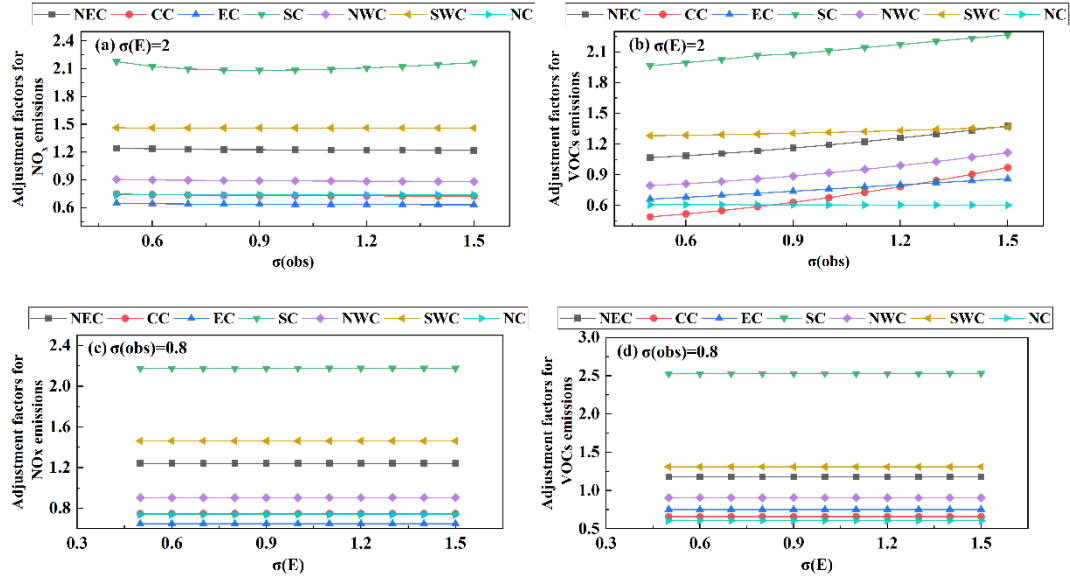


Fig. S5. Sensitivities of inversion results to A_i and R_i values under various emission and observation errors

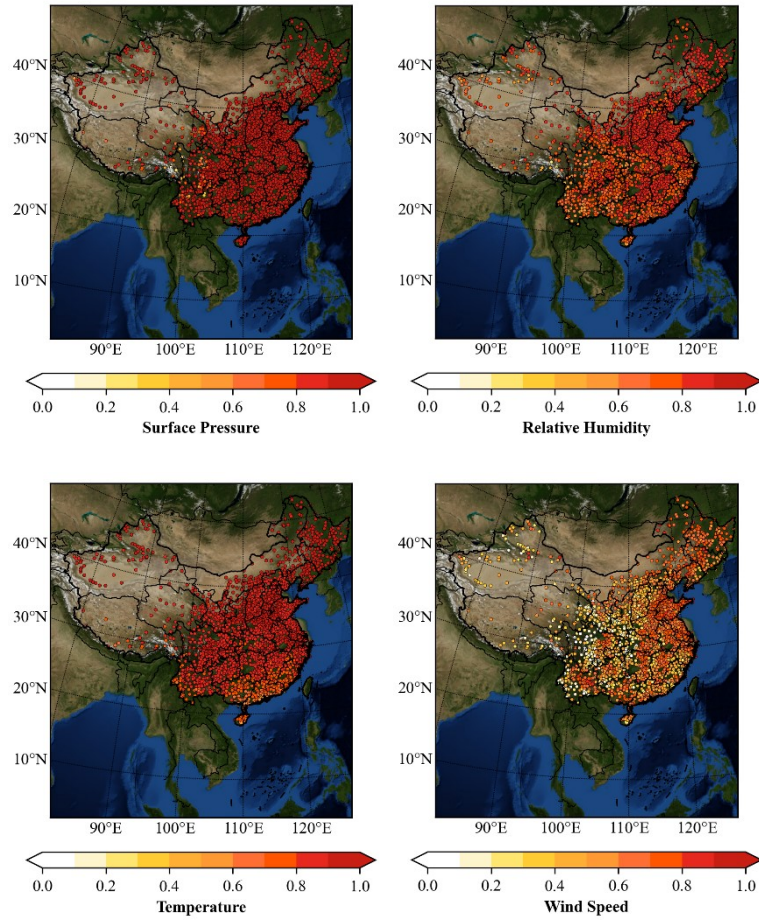
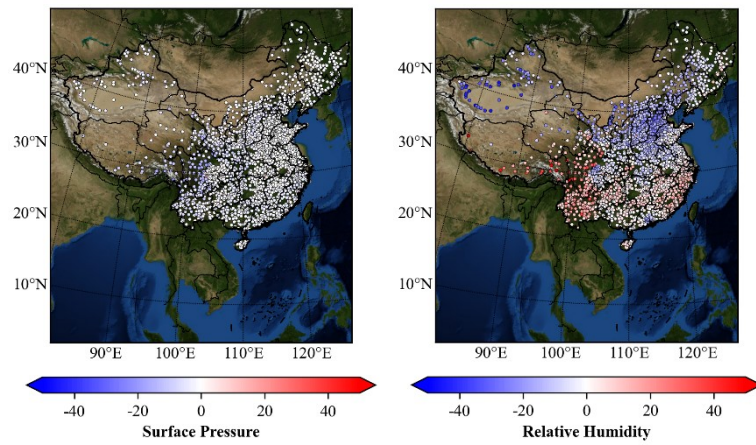
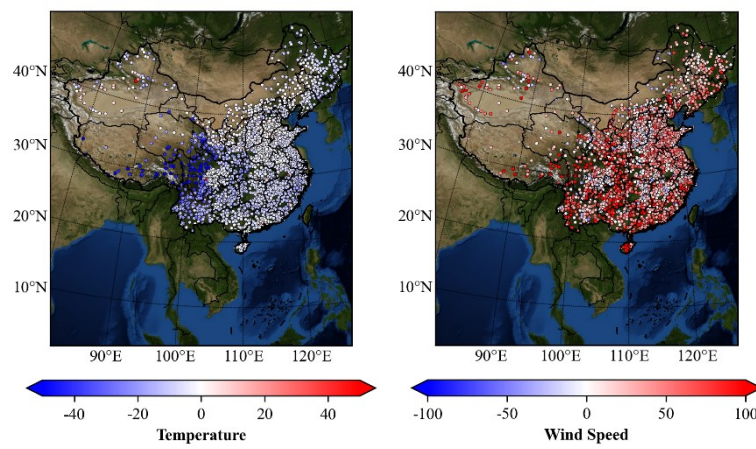


Fig. S6. Spatial distribution of correlation coefficients (R) for simulated vs. observed meteorological parameters in China from June to September 2019. (The background map is from NASA Blue Marble)

210



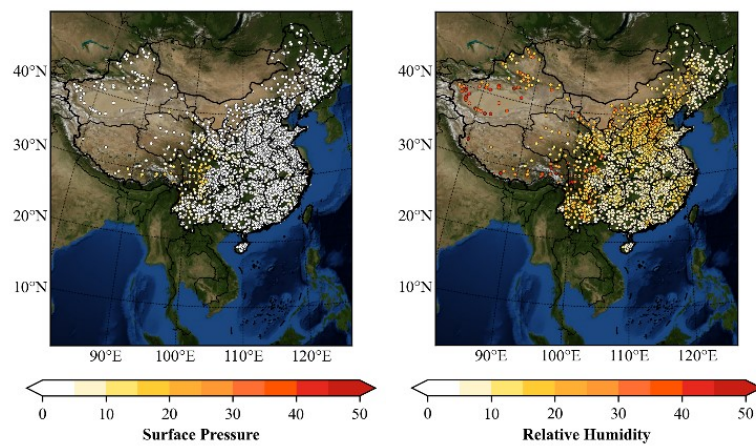
211



212

213 **Fig. S7. Spatial distribution of normalised mean bias (NMB) for simulated vs. observed**
 214 **meteorological parameters in China from June to September 2019. (The background map is from**
 215 **NASA Blue Marble)**

216



217

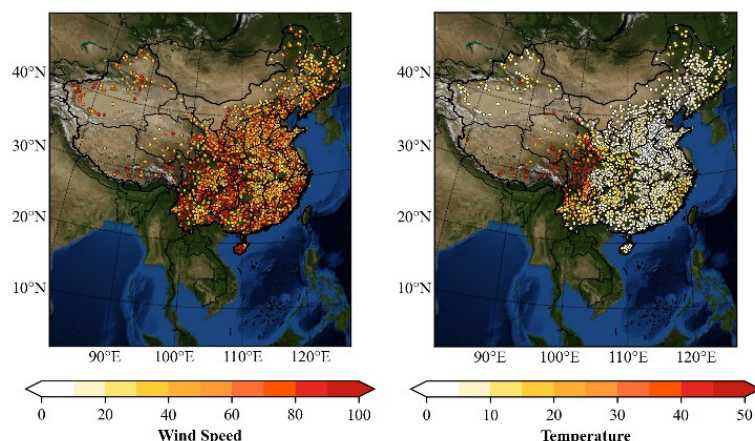


Fig. S8. Spatial distribution of normalised mean error (NME) for simulated vs. observed meteorological parameters in China from June to September 2019. (The background map is from NASA Blue Marble)

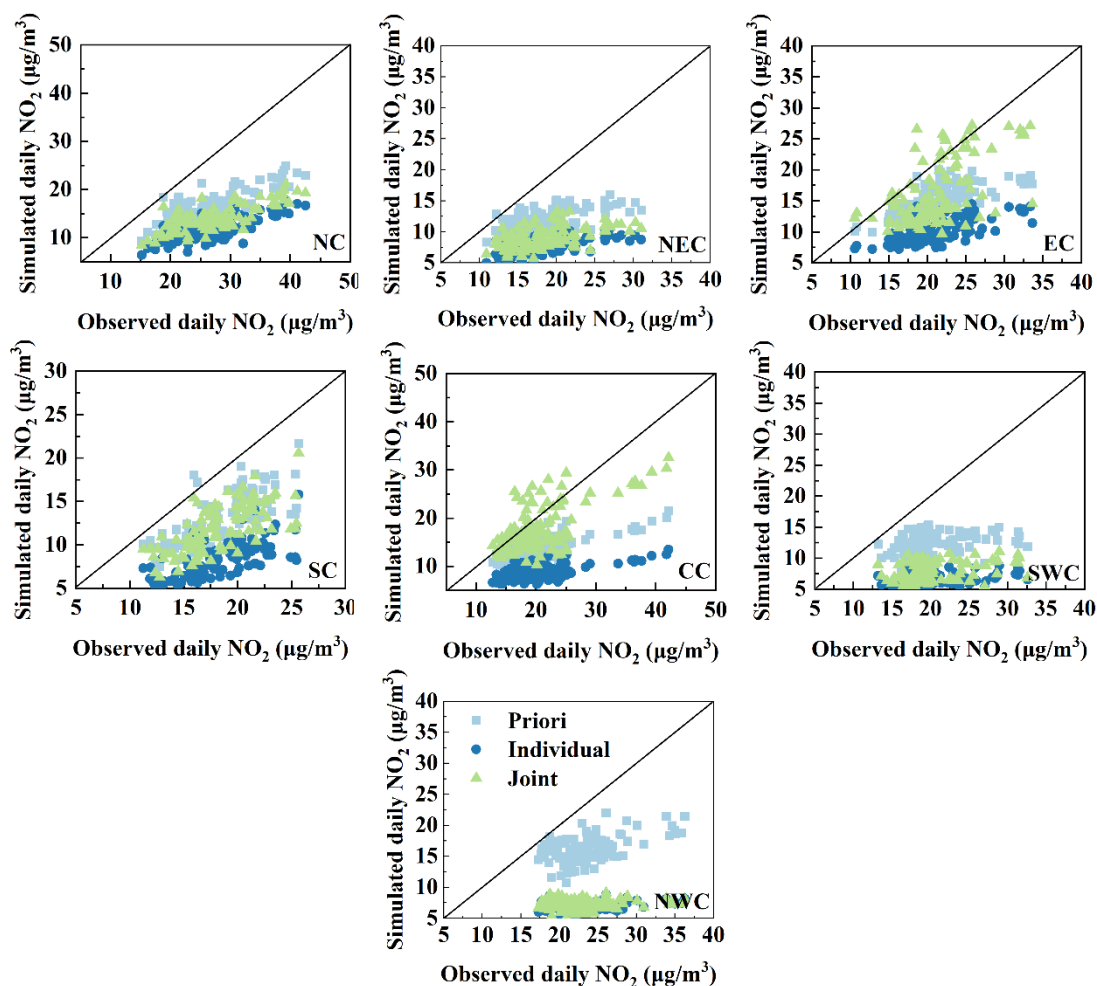


Fig. S9. Comparisons between simulated ground-level NO₂ concentrations by a priori, joint inversed, and individual inversed emissions, and the observed data from monitoring stations

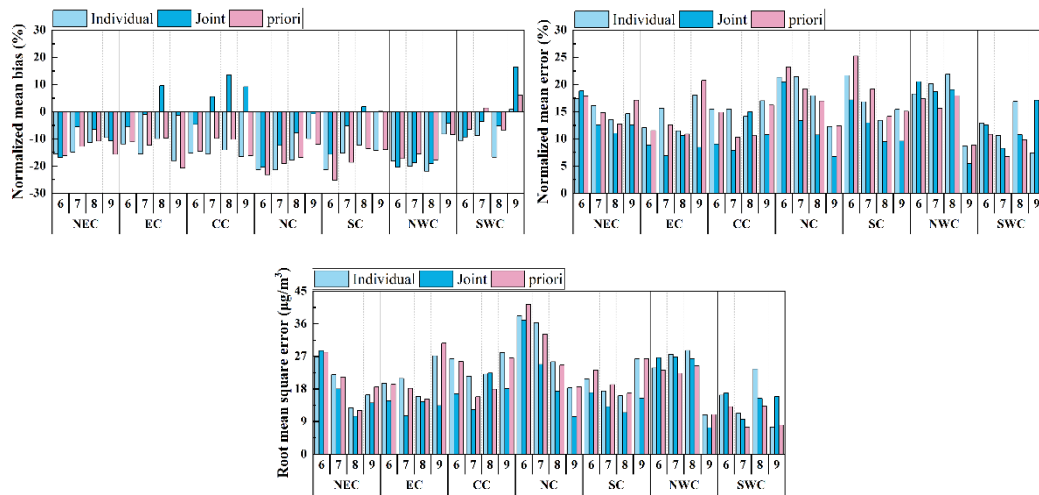


Fig. S10. Monthly variations in normalised mean bias, normalised mean error, and root mean square error for MDA8 O₃ concentrations in key regions of China from June to September 2019

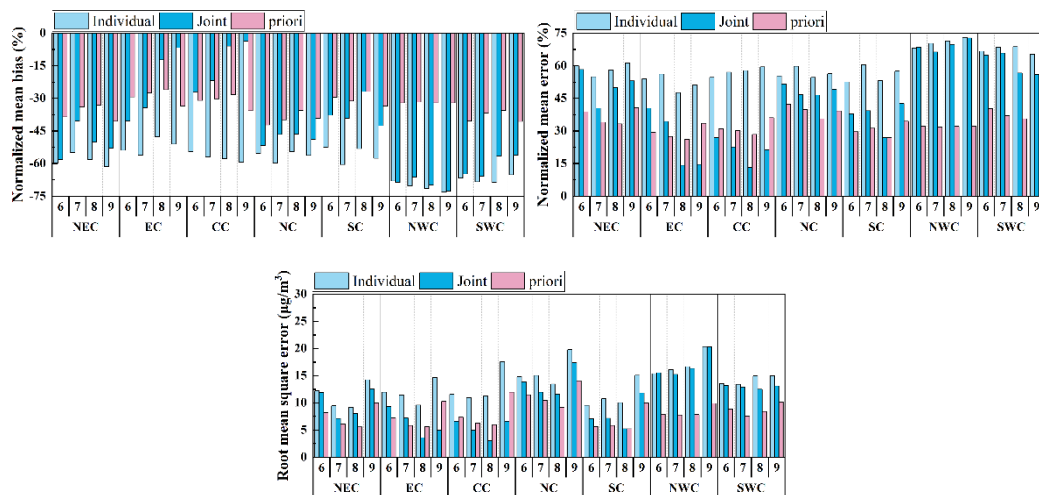


Fig. S11. Monthly variations in normalised mean bias, normalised mean error, and root mean square error for NO₂ concentrations in key regions of China from June to September 2019

Table S1. Regional emission adjustment factors in Pseudodata test

Region	NO _x	VOC
NC	1.3	1.3
NEC	0.8	0.8
CC	1.2	1.2
EC	1.6	1.6
SC	0.5	0.5
NWC	1.1	1.1
SWC	0.7	0.7

238 **Table S2. Validation metrics for simulated and observed atmospheric pressure across different**
239 **regions**

Region	R	NMB	NME	RMSE
CC	0.99	−0.9	1.3	12.4
EC	0.98	−0.7	1.1	10.9
NC	0.99	−1.2	1.3	12.3
NEC	0.99	−0.8	0.9	8.6
NWC	0.96	−2.8	3.3	27.8
SC	0.98	−1.4	1.5	14.8
SWC	0.92	−3.9	4.3	35.2

240 **Table S3. Validation metrics for simulated and observed temperature across different regions**

Region	R	NMB	NME	RMSE
CC	0.88	−4.9	7.3	2.2
EC	0.87	−5.9	7.8	2.3
NC	0.91	−1.9	6.2	1.7
NEC	0.94	−6.3	8.0	1.9
NWC	0.90	−8.1	13.1	2.6
SC	0.72	−4.9	6.6	2.1
SWC	0.85	−16.9	18.1	4.0

241

242 **Table S4. Validation metrics for simulated and observed relative humidity across different**
243 **regions**

Region	R	NMB	NME	RMSE
CC	0.82	−4.7	11.5	10.1
EC	0.82	1.5	9.8	9.1
NC	0.85	−14.5	17.4	13.2
NEC	0.85	−2.7	8.5	8.1
NWC	0.79	−10.8	17.5	12.5
SC	0.76	0.4	7.7	7.7
SWC	0.73	7.8	13.8	11.8

244

Table S5. Validation metrics for simulated and observed wind speed across different regions

Region	R	NMB	NME	RMSE
CC	0.62	29.0	51.6	1.2
EC	0.66	37.5	55.3	1.3
NC	0.61	23.4	51.2	1.3
NEC	0.68	31.9	48.6	1.4
NWC	0.52	37.7	63.8	1.6
SC	0.59	56.4	70.9	1.6
SWC	0.57	44.0	66.9	1.4

Table S6. Model performance statistics for simulated NO₂ concentrations against observations over China (*R* – correlation, NMB – normalized absolute bias, NME – normalized absolute error, MB – mean bias, ME – mean error, RMSE – root mean square error, Obs. – observation mean, Mod. – model mean)

		<i>R</i>	NMB	NME	MB	ME	RMSE	Obs.	Mod.
Individual		0.57	−58.5	58.5	−10.9	10.9	11.3	18	8
Joint	NEC	0.51	−50.4	50.4	−9.4	9.4	9.9	18	9
Priori		0.58	−36.6	36.6	−6.8	6.9	7.4	18	12
Individual		0.61	−56.5	56.5	−15.3	15.3	15.8	27	12
Joint	NC	0.60	−48.5	48.5	−13.2	13.2	13.7	27	14
Priori		0.61	−39.2	39.2	−10.6	10.6	11.3	27	16
Individual		0.66	−52.2	52.2	−11.3	11.3	11.9	22	10
Joint	EC	0.66	−23.5	25.9	−4.9	5.5	6.3	22	17
Priori		0.68	−29.1	29.1	−6.4	6.4	7.2	22	15
Individual		0.74	−55.9	55.9	−10.8	10.8	11.3	19	8
Joint	SC	0.75	−36.6	36.6	−7.1	7.1	7.8	19	12
Priori		0.75	−30.3	30.5	−5.9	5.9	6.6	19	13
Individual		0.44	−57.2	57.2	−12.1	12.1	12.8	21	9
Joint	CC	0.45	−14.6	21.0	−2.9	4.4	5.3	21	18
Priori		0.45	−31.3	31.4	−6.7	6.7	7.9	21	14
Individual		0.42	−70.7	70.7	−16.8	16.8	17.1	24	7
Joint	NWC	0.39	−69.3	69.3	−16.5	16.5	16.8	24	7
Priori		0.46	−32.0	32.0	−7.6	7.6	8.3	24	16
Individual		0.35	−67.3	67.3	−13.7	13.7	14.2	20	7
Joint	SWC	0.35	−60.8	60.8	−12.3	12.3	12.9	20	8
Priori		0.34	−38.4	38.4	−7.8	7.8	8.7	20	13

254 **Table S7. Model performance statistics for simulated daily NO₂ column concentrations and daily**
 255 **HCHO column concentrations against TROPOMI satellite data over China from June to**
 256 **September (*R* – correlation, NMB – normalized absolute bias, NME – normalized absolute error)**

		HCHO			NO ₂		
		<i>R</i>	NMB	NME	<i>R</i>	NMB	NME
Priori		0.62	-41.97	42.03	0.41	-12.40	20.07
Individual	NEC	0.69	-31.34	32.10	0.15	-38.09	38.67
Joint		0.80	-30.08	30.61	0.57	-25.92	26.86
Priori		0.51	-32.92	32.92	0.20	-35.08	36.14
Individual	NC	0.61	-24.81	25.23	0.34	-58.47	58.47
Joint		0.63	-22.20	22.89	0.34	-52.16	52.22
Priori		0.73	-44.67	44.67	0.27	-26.27	28.83
Individual	EC	0.77	-31.26	31.26	0.21	-59.49	59.49
Joint		0.67	-26.27	27.03	0.13	-52.68	52.91
Priori		0.68	-48.21	48.21	0.34	18.62	27.48
Individual	SC	0.67	-34.18	34.18	0.29	-37.27	38.99
Joint		0.38	-19.13	26.47	0.21	-20.51	29.29
Priori		-0.06	-35.14	36.83	0.17	29.84	50.27
Individual	CC	0.22	-17.95	28.87	0.43	-33.77	37.27
Joint		0.88	-21.93	24.27	0.83	-13.64	20.03
Priori		0.20	-36.25	40.69	0.65	10.46	21.33
Individual	NWC	0.41	-20.20	31.17	0.22	-30.64	36.66
Joint		0.74	-10.02	21.27	0.75	-9.79	18.82
Priori		0.22	-35.24	38.14	0.05	-47.30	48.97
Individual	SWC	0.21	-35.61	38.54	0.10	-55.63	56.20
Joint		0.13	-34.20	37.51	0.10	-54.23	54.83