



*Supplement of*

## **Constraining non-methane VOC emissions with TROPOMI HCHO observations: impact on summertime ozone simulation in August 2022 in China**

**Shuzhuang Feng et al.**

*Correspondence to:* Fei Jiang (jiangf@nju.edu.cn)

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

## **Sect. S1**

Meteorological processes play a dominant role in the transport, mixing, and chemical reactions of pollutants. Therefore, their simulation accuracy significantly affects emission inversion. To evaluate the performance of WRF simulations quantitatively, we utilized the surface meteorological measurements from 400 stations with 3-hour intervals, including temperature at 2 m (T2), relative humidity at 2 m (RH2), and wind speed at 10 m (WS10), and planetary boundary layer height (PBLH) measured by sounding from 84 stations with 12-hour intervals in this study. These surface and sounding data were obtained from the National Climate Data Center (NCDC) integrated surface database (<http://www.ncdc.noaa.gov/oa/ncdc.html>, last access: 25 May 2022) and the website of the University of Wyoming (<http://weather.uwyo.edu/upperair/sounding.html>, last access: 25 May 2022), respectively. The observed PBLH were calculated through the bulk Richardson number method with sounding data. Here, three basic statistical measures, mean bias, root-mean-square error (RMSE), and correlation coefficient (CORR), are applied to evaluate the results (Table S3). The results showed that the WRF model satisfactorily reproduced T2, RH2, WS10, and PBLH (Figure S11), with small biases of  $-0.5$  °C,  $-5.3\%$ ,  $0.3$  m/s, and  $-42.4$  m, respectively.

## **Sect. S2**

The prior NMVOC emissions were found to be overestimated relative to the top-down constraints from TROPOMI HCHO retrievals. The results of the top-down inversion may be susceptible to uncertainties related to the inversion configuration and observational data. Particularly, background error settings affect the relative weighting of prior and observation to posterior emissions, which may potentially introduce considerable uncertainty into the spatial patterns and magnitudes of the NMVOC emission inversion. Another critical concern pertains to HCHO retrieval errors. Correcting the low TROPOMI HCHO column biases would result in elevated posterior

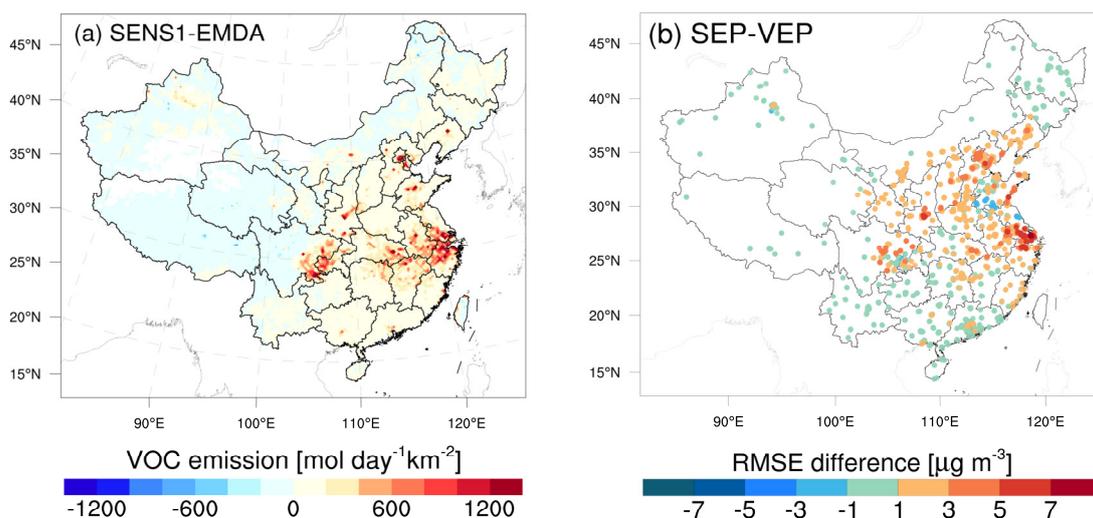
emissions, while the opposite holds true. Due to the spatiotemporal variability in retrieval errors, directly incorporating observations into an inversion system remains a challenging task. Based on the biases outlined in Vigouroux et al. (2020), one sensitivity experiment (SENS1) addressed the existing biases in TROPOMI HCHO by reducing measurements by 25% ( $<2.5 \times 10^{15}$  molec cm<sup>-2</sup>) in clean regions and increasing them by 30% ( $\geq 8 \times 10^{15}$  molec cm<sup>-2</sup>) in polluted regions. Figure S1 shows that bias-corrected HCHO columns resulted in a slight decrease in NMVOC emissions in the low-pollution regions of western China, whereas emissions increased in the high-pollution regions of eastern and central China, particularly in the SCB and the vicinity of the YRD. In comparison to the EMDA experiment, the posterior emissions from SENS1 increased by 12.8% (decreased by 43.9% compared to prior emissions), indicating that the existing retrieval error in HCHO measurements likely exerts an influence on the estimation of NMVOC emissions, especially in heavily polluted regions. The results highlight the significance of a thorough data validation for the HCHO column product. However, the emissions increase in the SENS1 has slightly deteriorated the performance of O<sub>3</sub> simulations.

To investigate the impact of background error on emission inversion, the other sensitivity experiment (SENS2) was conducted, doubling the background error to 80%. Compared with the base inversion, the SENS2 produced a noticeable increase in posterior NMVOC emissions in southwestern China, especially in Tibet. In contrast, emissions in eastern China exhibited a slight decrease (Figure S2). This can be expected, as the inversion is more inclined to deviate from the a priori due to decreased confidence. However, at a national scale, the difference between the two posterior emissions was nearly negligible. The substantial disparities over the Tibetan Plateau between the two inversions can be attributed to the horizontal HCHO inhomogeneity caused by mountain terrain and the relatively low signal-to-noise ratio in the TROPOMI satellite data in the background atmosphere (Cheng et al., 2023), resulting in the inclusion of more outliers in the inversion (Su et al., 2020). Nevertheless, the discrepancies in NMVOC emission estimates amounted to a mere 0.2%, suggesting that the posterior

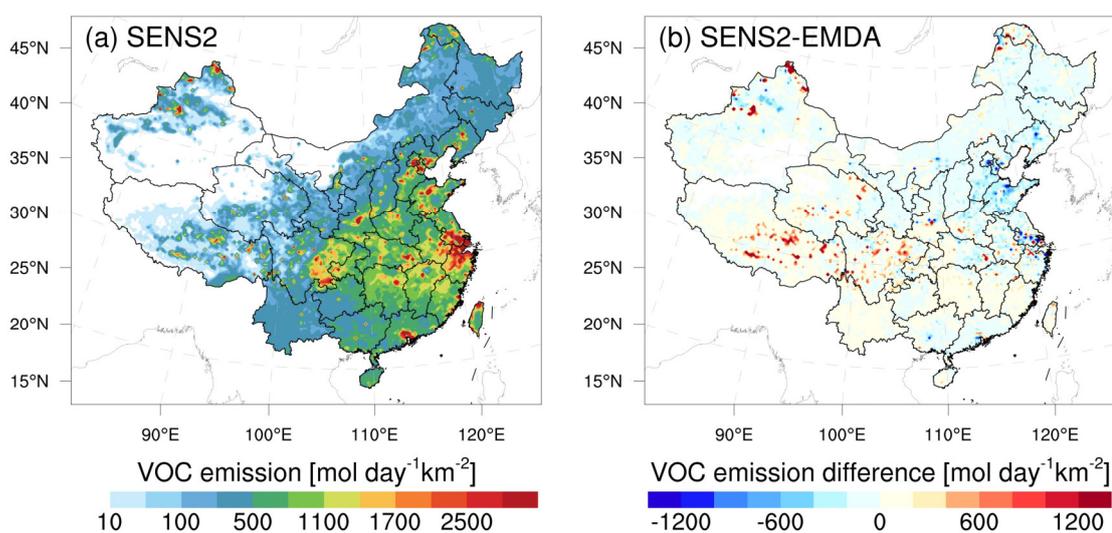
emission estimates were not largely affected by the background error setting. This can be primarily attributed to the superiority of the ‘two-step’ inversion strategy employed within the RAPAS system.

### **Sect. S3**

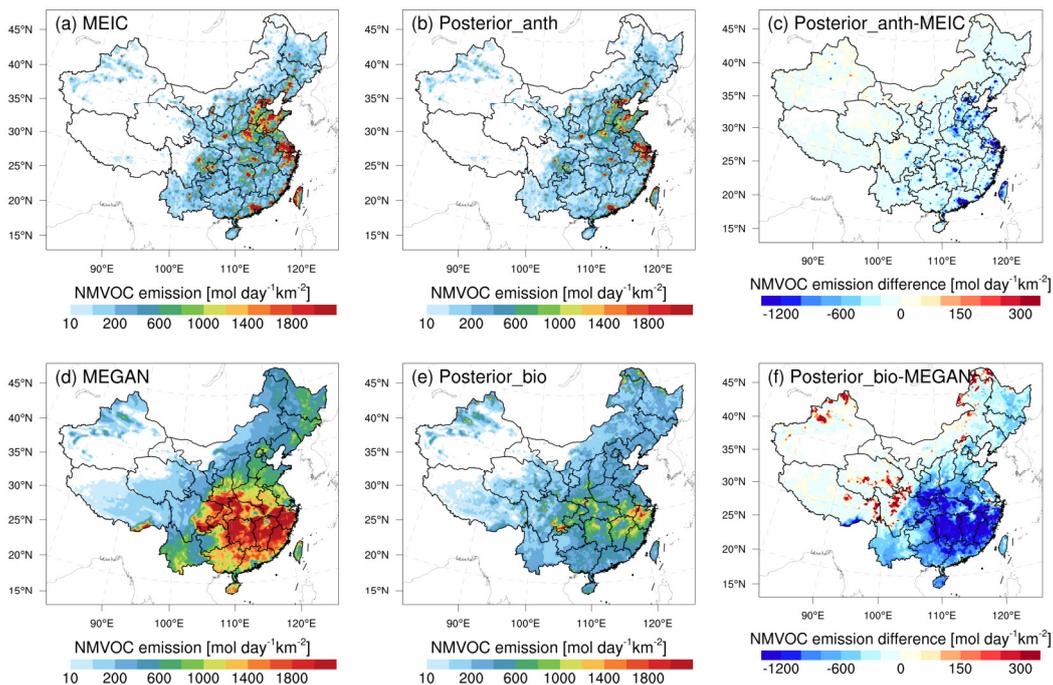
Figure S4 shows the spatial distribution of temporally averaged prior and posterior NO<sub>x</sub> emissions, along with their differences in emissions for August 2022. Higher emissions were predominantly concentrated in central and eastern China, especially in the NCP, Yangtze River Delta (YRD), and PRD, all of which are significant industrial or high-density urban areas. Lower emissions were primarily occurred across Northwest and Southern China. Compared with the prior emissions, posterior NO<sub>x</sub> emissions exhibited a significant decrease in the NCP, YRD, and parts of Central China. Furthermore, it was observed that most key urban areas tended to have their emissions overestimated in the prior inventory, while their surrounding areas had their emissions underestimated. Owing to the absence of precise emission location data, spatial proxies like population density, GDP, and road networks, are utilized for allocating total emissions across residential, industrial, and transportation emission sectors, which introduces uncertainties in spatial disaggregation. Overall, the total national NO<sub>x</sub> emissions decreased by 10.2% to 47.3 Gg in the posterior emissions. This disparity arises from uncertainties in the prior inventory and changes in China's NO<sub>x</sub> emissions reduction from 2020 to 2022.



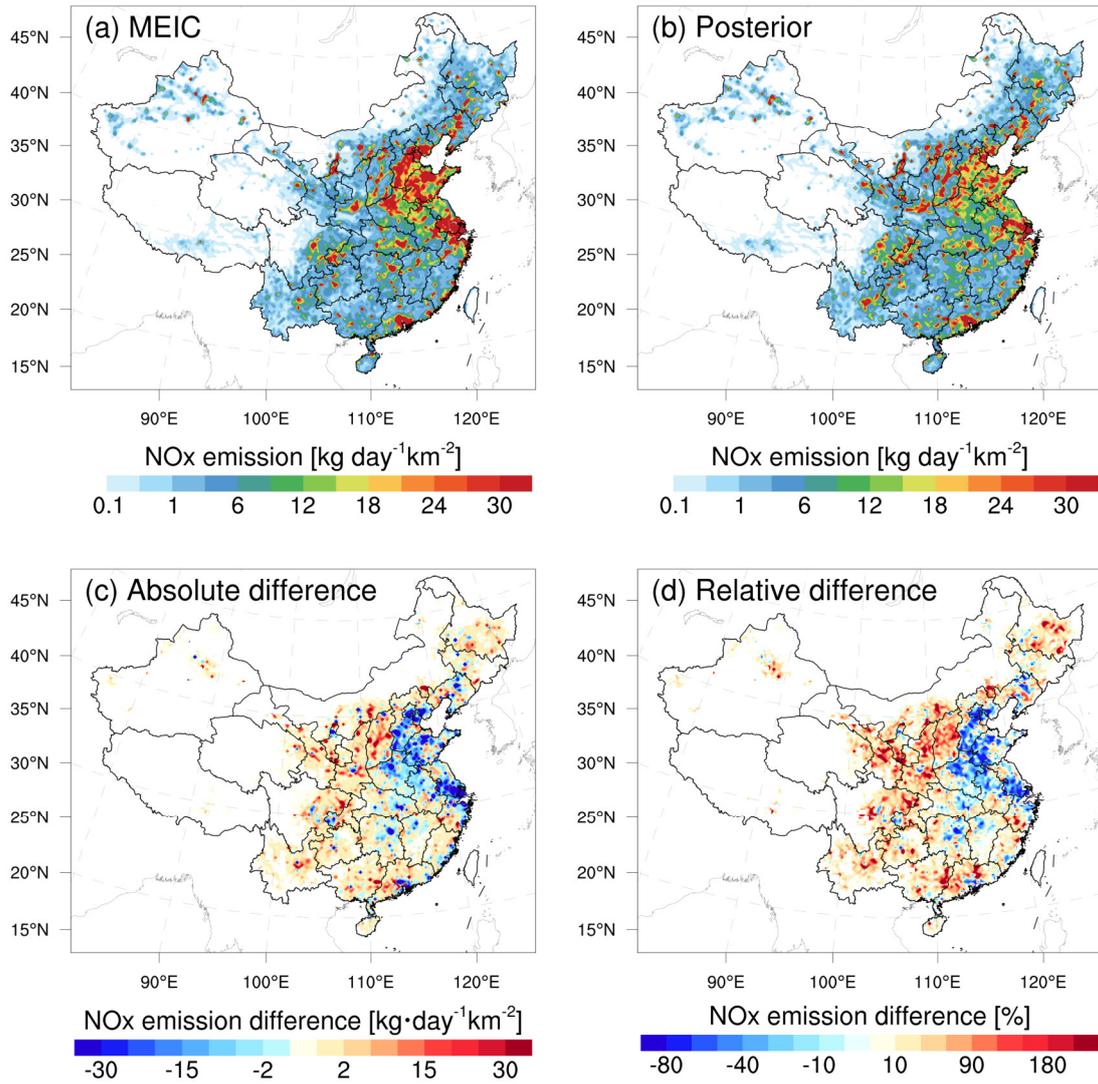
**Figure S1.** Spatial distribution of (a) differences in posterior emissions between the SENS1 and EMDA, and (b) differences in RMSE between the posterior simulations (SEP) of SENS1 and VEP. Compared with EMDA, SENS1 reduced the TROPOMI HCHO measurements by 25% ( $< 2.5 \times 10^{15}$  molec  $\text{cm}^{-2}$ ) in clean regions and increased them by 30% ( $> 8 \times 10^{15}$  molec  $\text{cm}^{-2}$ ) in polluted regions.



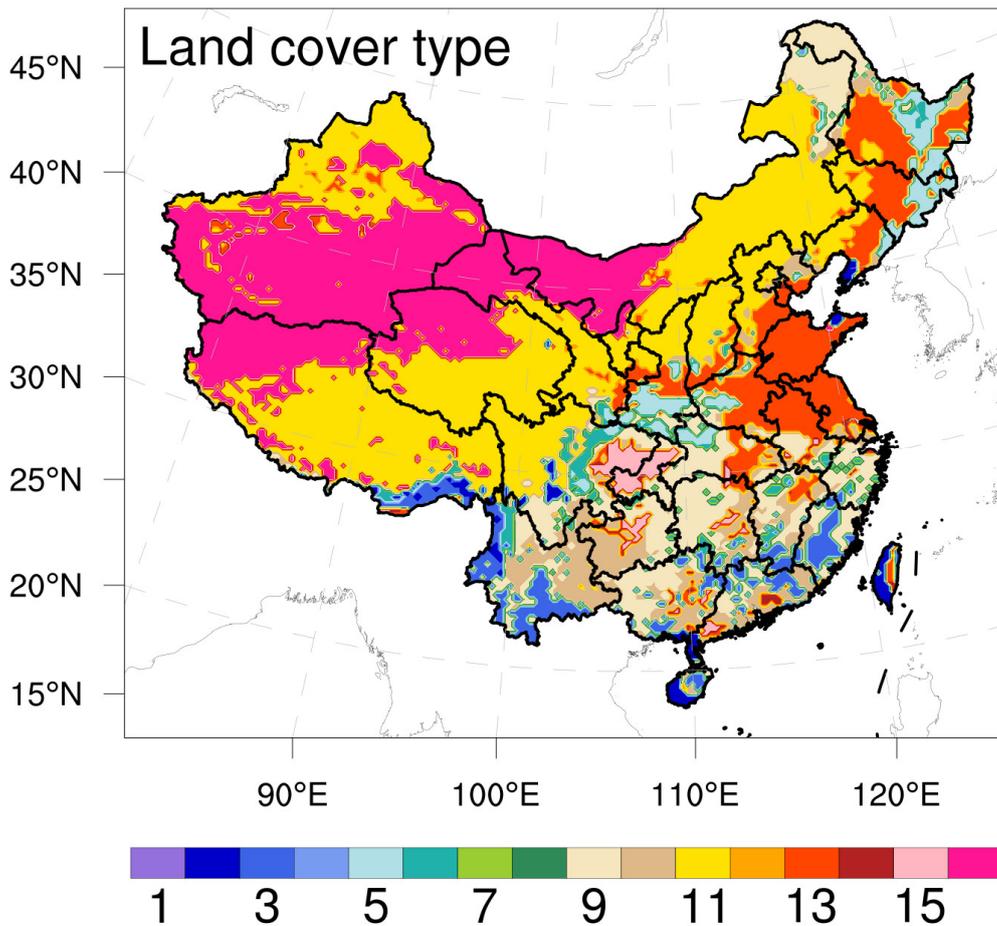
**Figure S2.** Spatial distribution of the time-averaged (a) posterior emissions in SENS2 and (b) differences in posterior emissions between SENS2 and EMDA (SENS2-EMDA). SENS2 increased background error from 40% to 80% compared to EMDA.



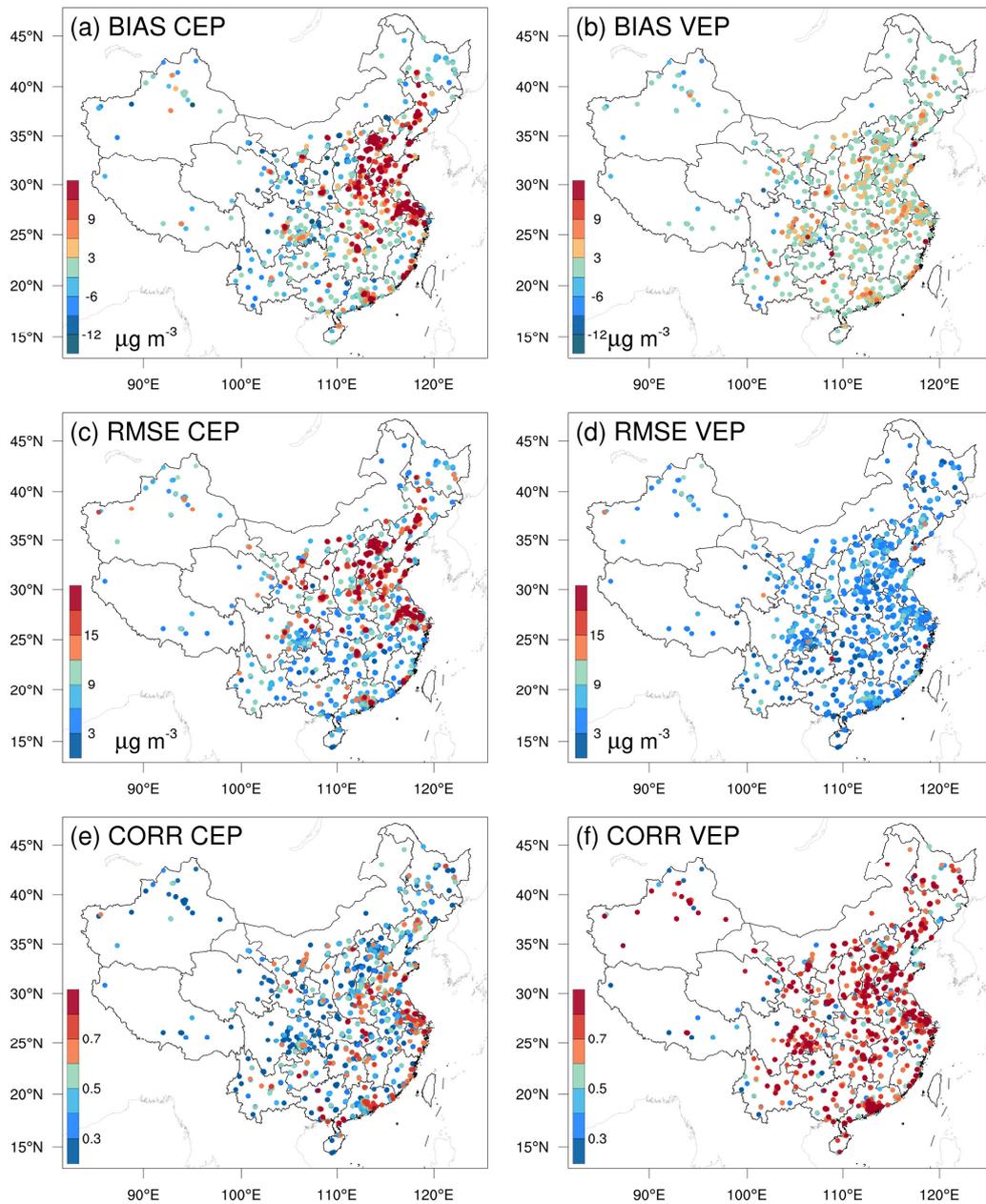
**Figure S3.** Spatial distribution of the time-averaged (a) prior anthropogenic emissions (MEIC 2020), (b) posterior anthropogenic emissions, (c) difference between prior and posterior anthropogenic emissions (posterior minus prior), (d) prior biogenic emissions (MEGAN), (e) posterior biogenic emissions, (f) difference between prior and posterior biogenic emissions (posterior minus prior) of NMVOCs over China.



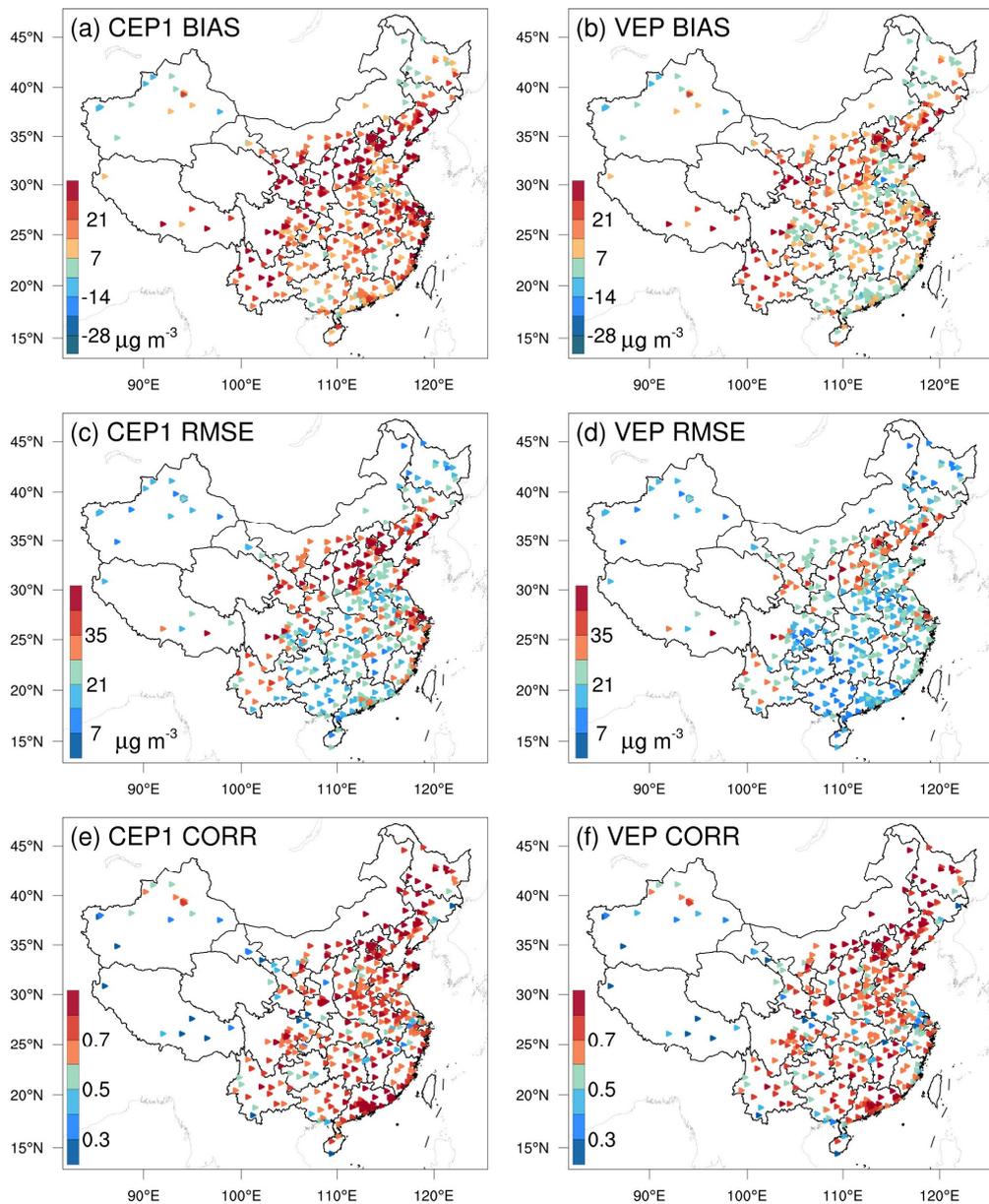
**Figure S4.** Spatial distribution of the time-averaged (a) prior emissions (MEIC 2020), (b) posterior emissions, (c) absolute difference (posterior minus MEIC), and (d) relative difference of NO<sub>x</sub> over China.



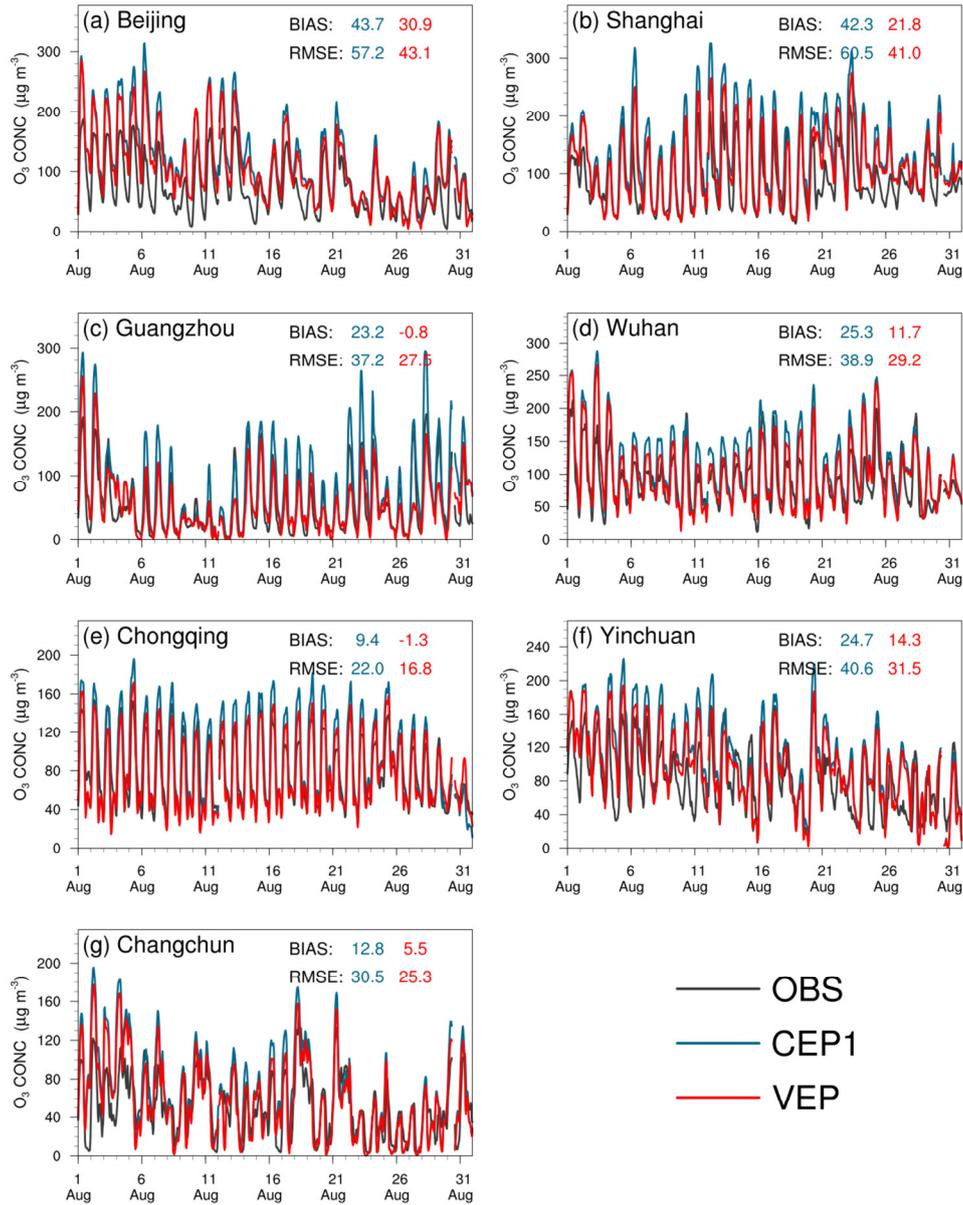
**Figure S5.** Combined MODIS International Geosphere-Biosphere Programme (IGBP) data from the MCD12C1 product, 2020 (<https://lpdaac.usgs.gov/products/mcd12c1v061/>). 1, Evergreen needleleaf forests; 2, Evergreen broadleaf forests; 3, Deciduous needleleaf forests; 4, Deciduous broadleaf forests; 5, Mixed forests; 6, Closed shrublands; 7, Open shrublands; 8, Woody savannas; 9, Savannas; 10, Grasslands; 11, Permanent wetlands; 12, Croplands; 13, Urban and built-up; 14, Cropland-natural vegetation mosaics; 15, Snow and ice; 16, Sparsely vegetated.



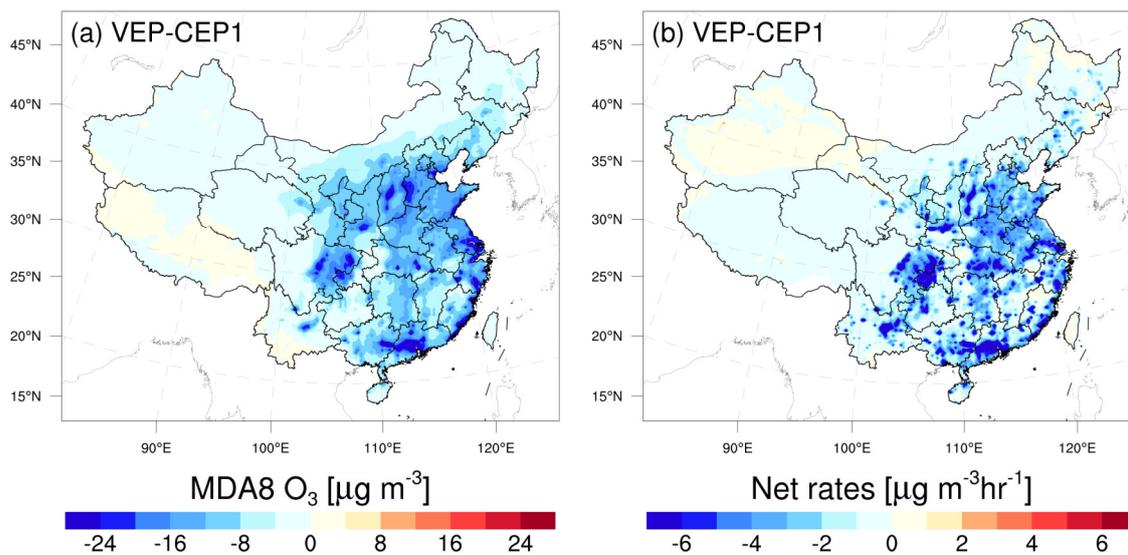
**Figure S6.** Spatial distribution of mean bias (BIAS, a and b), root mean square error (RMSE, c and d), and correlation coefficient (CORR, e and f) for simulated NO<sub>2</sub> using prior (left, CEP) and posterior (right, VEP) emissions, respectively.



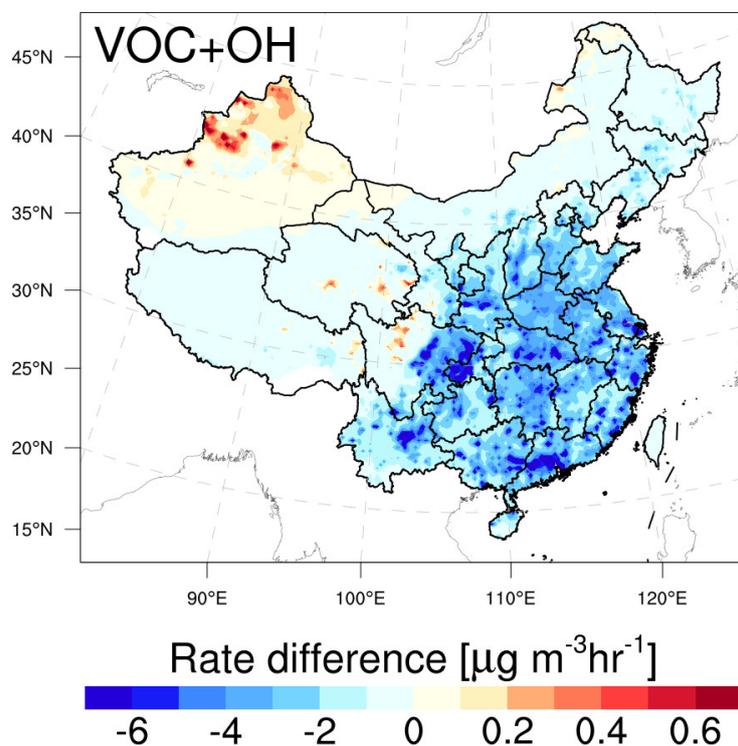
**Figure S7.** Spatial distribution of mean bias (BIAS, a and b), root mean square error (RMSE, c and d), and correlation coefficient (CORR, e and f) for simulated O<sub>3</sub> using prior (left, CEP1) and posterior (right, VEP) emissions against independent observations.



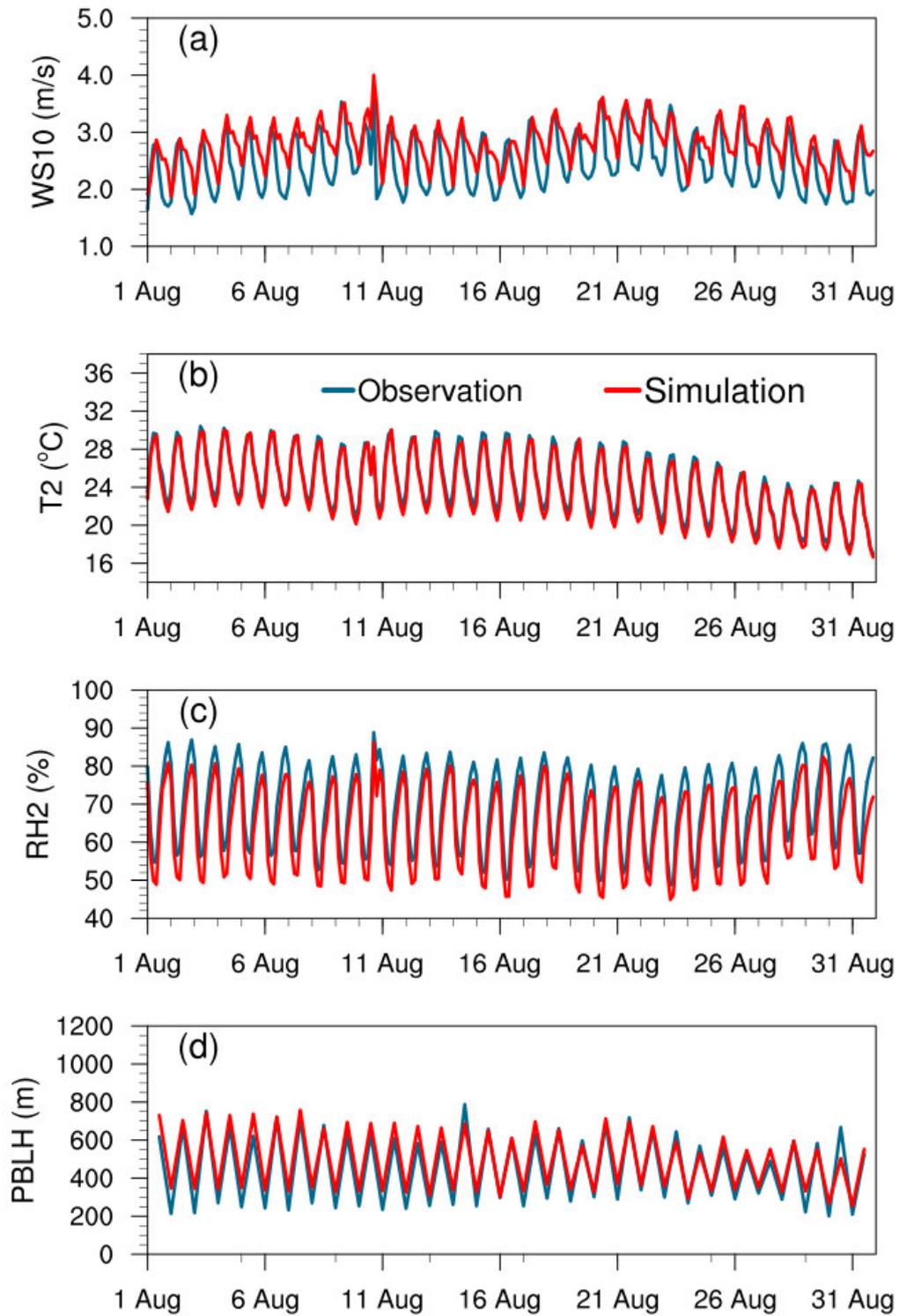
**Figure S8.** Time series comparison of simulated and observed hourly surface O<sub>3</sub> concentrations (µg m<sup>-3</sup>) from CEP1 and VEP experiments over (a) Beijing, (b) Shanghai, (c) Guangzhou, (d) Wuhan, (e) Chongqing, (f) Yinchuan, and (g) Changchun, representing key cities in North China, East China, South China, Central China, Southwest China, Northwest China, and Northeast China, respectively.



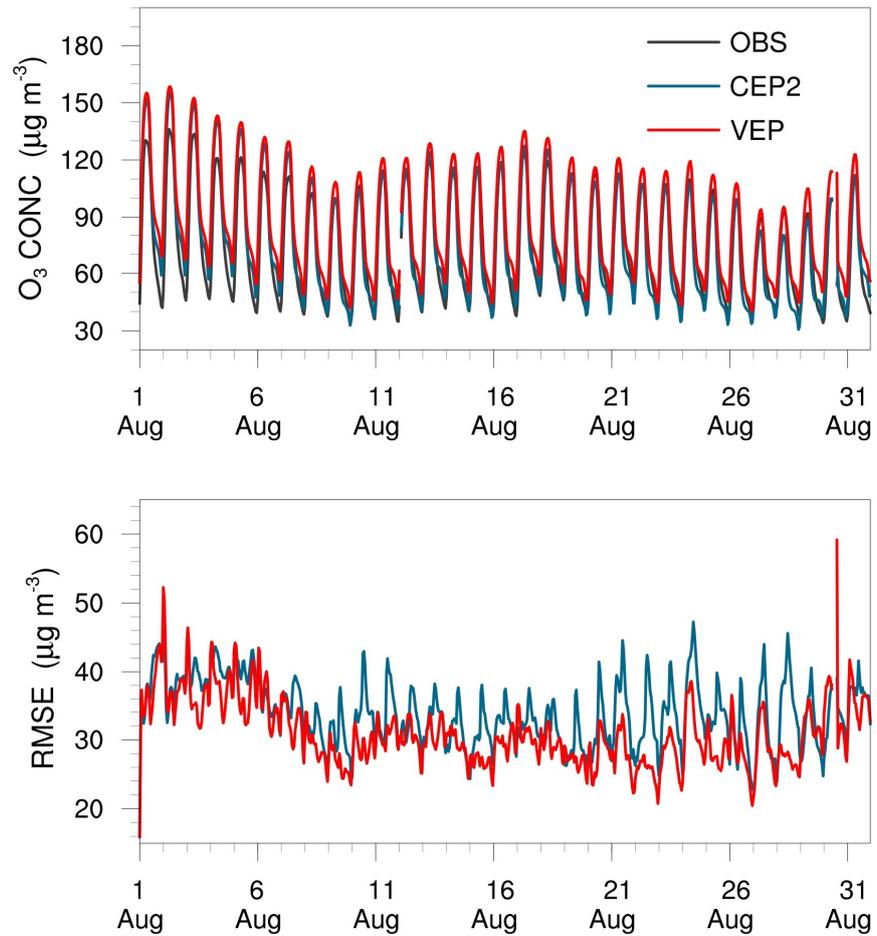
**Figure S9.** Differences in simulated (a) maximum daily 8-hour average (MDA8) O<sub>3</sub> concentrations and (b) net reaction rates between CEP1 and VEP experiments.



**Figure S10.** Differences in reaction rates of VOC + OH between CEP1 and VEP experiments at the surface. Time period: August 2022, 12:00–18:00 CST.



**Figure S11.** Time series of observed and simulated wind speeds at 10 m (WS10, m/s), temperature at 2 m (T2, °C), relative humidity at 2 m (RH2, %), and planetary boundary layer height (PBLH, m) over mainland China.



**Figure S12.** Time series of (a) hourly surface O<sub>3</sub> concentrations ( $\mu\text{g m}^{-3}$ ) and (b) RMSE ( $\mu\text{g m}^{-3}$ ) obtained from VEP and CEP2 experiments, CEP2 was simulated using the posterior emissions disregarding the uncertainty in NO<sub>x</sub> emissions.

**Table S1.** Configuration options of WRF/CMAQ

WRF		CMAQ	
Parameter	Scheme	Parameter	Scheme
Microphysics	WSM6	Horizontal/Vertical advection	yamo/wrf
Longwave	RRTM	Horizontal/Vertical diffusion	multiscale/acm2
Shortwave	Goddard	Deposition	m3dry
Boundary layer	ACM	Chemistry solver	EBI
Cumulus	Kain-Fritsch	Photolysis	phot_inline
Land-surface	Noah	Aerosol module	AERO6
Surface layer	Revised	Cloud module	cloud_acm_ae6
Urban canopy	No	Gas-phase chemistry	CB05tucl

**Table S2.** Statistics comparison of simulated and observed 10-m wind speed (WS10), 2-m temperature (T2), 2-m relative humidity (RH2), and planetary boundary layer height (PBLH).

Variable Met.	No. of sites	Mean Obs.	Mean Sim.	BIAS	RMSE	CORR
WS10 (m/s)	400	2.5	2.8	0.3	0.5	0.73
T2 (°C)	400	24.5	24.0	-0.5	0.6	1.00
RH2 (%)	400	69.0	63.8	-5.3	5.5	0.99
PBLH (m)	84	451.2	493.6	42.4	68.6	0.96

\* BIAS, mean bias; RMSE, root mean square error; CORR, correlation coefficient