# Constraining Non-Methane VOC Emissions with TROPOMI HCHO observations: Impact on Summertime Ozone Simulation in August 2022 in China

Shuzhuang Feng[1], Fei Jiang[1,2,5*], Tianlu Qian[3], Nan Wang[4], Mengwei Jia[1], Songci Zheng[1], Jiansong Chen[6], Fang Ying[6], Weimin Ju[1,2]

[1] *Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, International Institute for Earth System Science, Nanjing University, Nanjing, 210023, China*

[2] *Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, 210023, China*

[3] *School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing, 210042, China*

[4] *College of Carbon Neutrality Future Technology, Sichuan University, Chengdu, 610207, China*

[5] *Frontiers Science Center for Critical Earth Material Cycling, Nanjing University, Nanjing, 210023, China*

[6] *Hangzhou Municipal Ecology and Environment Bureau, Hangzhou, 310020, China*

* Corresponding author: Tel.: +86-25-83597077; Fax: +86-25-83592288; E-mail address: jiangf@nju.edu.cn

## Abstract

Non-methane volatile organic compounds (NMVOC), serving as crucial precursors of $O_3$, have a significant impact on atmospheric oxidative capacity and $O_3$ formation. However, both anthropogenic and biogenic NMVOC emissions remain subject to considerable uncertainty. Here, we extended the Regional multi-Air Pollutant Assimilation System (RAPAS) with the EnKF algorithm to optimize NMVOC emissions in China in August 2022 by assimilating TROPOMI HCHO retrievals. We also simultaneously optimize $NO_x$ emissions by assimilating in-situ $NO_2$ observations to address the chemical feedback among VOC-$NO_x$-$O_3$. Furthermore, a process-based analysis was employed to quantify the impact of NMVOC emission changes on various chemical reactions related to $O_3$ formation and depletion. NMVOC emissions exhibited a substantial reduction of 50.2%, especially in the middle and lower reaches of the Yangtze River, revealing a prior overestimation of biogenic NMVOC emissions due to extreme heatwave. Compared with the forecast with prior NMVOC emissions, the forecast with posterior emissions significantly improved HCHO simulations, reducing biases by 75.7%, indicating a notable decrease in posterior emission uncertainties. The forecast with posterior emissions also effectively corrected the overestimation of $O_3$ in forecast with prior emissions, reducing biases by 49.3%. This can be primarily attributed to a significant decrease in the $RO_2$ + NO reaction rate and an increase in the $NO_2$ + OH reaction rate in the afternoon, thus limiting $O_3$ generation. Sensitivity analyses emphasized the necessity of considering both NMVOC and $NO_x$ emissions for a comprehensive assessment of $O_3$ chemistry. This study enhances our understanding of the effects of NMVOC emissions on $O_3$ production and can contribute to the development of effective emission reduction policies.

**Keywords**

NMVOC emissions, $O_3$ pollution, Emission inversion, HCHO column retrievals, Data assimilation

## 1 Introduction

Since the Chinese government implemented the Air Pollution Prevention and Control Action Plan in 2013, there has been a notable reduction in $NO_x$ emissions (Zheng et al., 2018). However, despite these advancements, the issue of $O_3$ pollution persists and, in certain cases, has shown signs of worsening (Ren et al., 2022). The increase in $O_3$ concentration can be attributed not only to adverse meteorological conditions but also predominantly to unbalanced joint control of non-methane volatile organic compounds (NMVOCs) and nitrogen oxides ($NO_x$) (Li et al., 2020). NMVOCs are vital precursors of $O_3$ and have a substantial impact on the atmospheric oxidation capacity, thereby altering the lifetimes of other pollutants. Accurately quantifying NMVOC emissions holds significant importance in investigating their impact on $O_3$ chemistry and in formulating emission reduction policies.

Anthropogenic NMVOC emissions have traditionally been estimated using a "bottom-up" method. However, the accuracy and timeliness of these estimations face challenges owing to the scarcity of local measurements for emission factors, the incompleteness and unreliability of activity data, and the diverse range of species and technologies involved (Cao et al., 2018; Hong et al., 2017). Furthermore, uncertainties arise in model-ready NMVOC emissions due to spatial and temporal allocations using various "proxy" data for different source sectors (Li et al., 2017a). Li et al. (2021) reported substantial discrepancies among emission estimates in various studies, ranging 23% to 56%. Biogenic NMVOC emissions are typically estimated using models like the Model of Emissions of Gases and Aerosols from Nature (MEGAN) (Guenther et al., 2012) and the Biogenic Emission Inventory System (BEIS) (Pierce et al., 1998). NMVOC emissions result from the multiplication of plant-specific standard emission rates by dimensionless activity factors. Nonetheless, apart from inaccuracies in the distribution of plant functional types, empirical parameterization, especially concerning responses to temperature and drought stress, can introduce substantial uncertainties (Angot et al., 2020; Seco et al., 2022; Jiang et al., 2018). Warneke et al. (2010) determined isoprene emission rates through field measurements and conducted a comparison with MEGAN and BEIS estimates, revealing a notable tendency for MEGAN to overestimate emissions, while BEIS consistently underestimated them. Similarly, Marais et al. (2014) found that MEGAN's isoprene emission estimates were 5-10 times higher than the canopy-scale flux measurements obtained from African field campaigns.

A top-down approach, utilizing observed data, has been developed for estimating VOCs emissions. For instance, based on aircraft and ground-based field measurements, the source-receptor relationships algorithm with Lagrangian particle dispersion model (Fang et al., 2016), mixed layer gradient techniques (Mo et al., 2020), eddy covariance flux measurements (Yuan et al., 2015), and box model (Wang et al., 2020) have been employed to complement or verify bottom-up results. However, these approaches do not comprehensively consider the complex nonlinear chemical reactions and transport processes that VOCs undergo in the atmosphere. Formaldehyde (HCHO) and glyoxal (CHOCHO) in the atmosphere serve as crucial oxidization intermediates for various VOCs (Hong et al., 2021; Liu et al., 2012). Satellite-based observations can readily detect their presence in the form of vertical column density (VCD) from space, making them widely utilized for estimating NMVOC emissions. A commonly used approach assumes that the observed HCHO/CHOCHO columns are locally linearly correlated with VOC emission rates (Palmer et al., 2006; Liu et al., 2012). However, this approach does not consider the spatial offset resulting from chemistry reactions and transport processes. Chaliyakunnel et al. (2019) conducted a Bayesian analysis to derive an optimal estimate of VOC emissions using HCHO measurements over the Indian subcontinent. Their results indicated that biogenic VOC emissions modeled by MEGANv2.1 were overestimated by approximately 30–60%, whereas anthropogenic VOC emissions derived from the RETRO inventory were underestimated by 13–16%. Cao et al. (2018) employed the GEOS-Chem model and its adjoint, incorporating tropospheric HCHO and CHOCHO column data from the GOME-2A and OMI satellites as constraints, to quantify Chinese NMVOC emissions. They demonstrated a low bias in the MEGAN model, in contrast to the significant overestimation shown in Bauwens et al. (2016), especially in southern China.

Several investigations have been conducted to explore the implications of inverted VOC emissions on surface $O_3$. For instance, using the Eulerian box model, Zhou et al. (2023) employed concurrent VOC measurements to constrain anthropogenic VOC emissions. This led to improved simulations of VOCs and $O_3$, with a reduction in high emissions by 15%–36% in the Pearl River Delta (PRD) region. Local model biases in simulating the oxidation of NMVOCs and $O_3$ are closed related to uncertainties in $NO_x$ emissions (Wolfe et al., 2016; Chan Miller et al., 2017). To tackle these critical questions, Kaiser et al. (2018) applied an adjoint algorithm to estimate isoprene

emission over the southeast US by downwardly adjusting anthropogenic $NO_x$ emissions by 50% to rectify $NO_2$ simulations. Their findings indicated that isoprene emissions from MEGAN v2.1 were overestimated by an average of 40%, slightly lower than the 50% reduction in Bauwens et al. (2016). Souri et al. (2020) simultaneously optimized NMVOC and $NO_x$ emissions utilizing OMPS-NM HCHO and OMI $NO_2$ retrievals in East Asia. They found that predominantly anthropogenic NMVOC emissions from MIX-Asia 2010 increased over the North China Plain (NCP), whereas predominantly biogenic NMVOC emissions from MEGAN v2.1 decreased over southern China after the adjustment. Unfortunately, the posterior simulations exacerbated the overestimation of $O_3$ levels in northern China.

Most studies regarding the inversion of NMVOC emissions or its impact on $O_3$ neglected the uncertainties associated with $NO_x$-dependent production or loss of NMVOC oxidation and $O_3$. An iteratively nonlinear joint inversion of $NO_x$ and NMVOCs using multi-species observations is expected to minimize the uncertainties in their emissions and is well-suited to address the intricate relationship among VOC-$NO_x$-$O_3$. In this study, we extended the Regional multi-Air Pollutant Assimilation System (RAPAS) upon the ensemble Kalman filter (EnKF) assimilation algorithm to enhance the optimization of NMVOC emissions over China, utilizing the TROPOspheric Monitoring Instrument (TROPOMI) HCHO retrievals with high spatial coverage and resolution. To more accurately quantify the impact of NMVOC emissions on $O_3$, $NO_x$ emissions were simultaneously adjusted using nationwide in-situ $NO_2$ observations. Process analysis was subsequently employed to quantify various chemical pathways associated with $O_3$ formation and loss. Through a top-down constraint on both emissions, this study aims to offer a more scientific insight into the consequences of optimizing NMVOC emissions on $O_3$ and contribute to the development of appropriate emission reduction policies.

## 2 Data and Methods

### 2.1 Data Assimilation System

The RAPAS system (Feng et al., 2023) has been developed based on a regional chemical transport model (CTM) and ensemble square root filter (EnSRF) assimilation modules (Whitaker and Hamill, 2002), which are employed for simulating atmospheric compositions and inferring anthropogenic emissions by assimilating surface

observations, respectively (Feng et al., 2022; Feng et al., 2020). The inversion process follows a two-step procedure within each inversion window, in which the emissions are inferred first and then input into the CMAQ model to simulate initial conditions of the next window. Meanwhile, the optimized emissions are transferred to the next window as prior emissions. The two-step inversion strategy facilitates error propagation and iterative emission optimization, which have proven the superiority and robustness of our system in estimating emissions (Feng et al., 2023). In this study, we extended the data frame to include the assimilation of TROPOMI HCHO retrievals for optimizing NMVOC emissions. Concise descriptions of the forecast model, data assimilation approach, and experimental settings follow.

**2.1.1 Atmospheric Transport Model**

The Weather Research and Forecast (WRF v4.0) model (Skamarock and Klemp, 2008) and the Community Multiscale Air Quality Modeling System (CMAQ v5.0.2) (Byun and Schere, 2006) were applied to simulate meteorological conditions and atmospheric chemistry, respectively. WRF simulations were conducted with a 27-km horizontal resolution, covering the entire mainland China on a grid of $225 \times 165$ cells (Figure 1). The CMAQ model was run over the same domain, but with a removal of three grid cells on each side of the WRF domain. The vertical settings in WRF and CMAQ was the same as Feng et al. (2020). To account for the rapid expansion of urbanization, we updated underlying surface information for urban and built-up land using the MODIS Land Cover Type Product (MCD12C1) Version 6.1 of 2022. Chemical lateral boundary conditions for NO, $NO_2$, HCHO, and $O_3$ were extracted from the output of the global CTM (i.e., the Whole Atmosphere Community Climate Model, WACCM) with a resolution of $0.9° \times 1.25°$ at 6-hour intervals (Marsh et al., 2013). Meanwhile, boundary conditions for the other NMVOCs were obtained directly from background profiles. In the first data assimilation (DA) window, chemical initial conditions (excluding NMVOCs) were also derived from the WACCM outputs, whereas in subsequent windows, they were derived through forward simulation using optimized emissions from the previous window. Table S1 lists the detailed physical and chemical configurations. To assess the impact of updated NMVOC emissions on $O_3$ production efficiency, we further decoupled the contribution of the primary chemical processes to the $O_3$ levels using the CMAQ Integrated Reaction Rate (IRR) analysis.

**2.1.2 EnKF Assimilation Algorithm**

The emissions are constrained using the Ensemble Square Root Filter (EnSRF) algorithm introduced by Whitaker and Hamill (2002). This approach fully accounts for temporal and geographical variations in both the transportation and chemical reactions within the emission estimates. During the forecast step, the background ensembles are derived by applying perturbation to the prior emissions. The perturbed samples are typically drawn from Gaussian distributions with a mean of zero and a standard deviation equal to the prior emission uncertainty in each grid cell. Ensemble runs of the CMAQ model were subsequently performed to propagate the background errors with each ensemble sample of state vectors.

In the analysis step, the ensemble mean $\overline{X^a}$ of the analyzed state is regarded as the best estimate of emissions, which is obtained by updating the background ensemble mean through the following equations:

$$\overline{X^a} = \overline{X^b} + \mathbf{K}(\mathbf{y} - H\overline{X^b}) \tag{1}$$

$$\mathbf{K} = \boldsymbol{P^b}H^T(H\boldsymbol{P^b}H^T + \boldsymbol{R})^{-1} \tag{2}$$

where $\boldsymbol{y}$ is the observational vector; $\boldsymbol{H}$ represents the observation operator mapping model space to observation space; The expression $\boldsymbol{y} - \boldsymbol{H}\overline{X^b}$ quantifies the disparities between simulated and observed concentrations; $\boldsymbol{P^b}\boldsymbol{H^T}$ illustrates how uncertainties in emissions relate to uncertainties in simulated concentrations; The Kalman gain matrix $\boldsymbol{K}$, dependent on background error covariance $\boldsymbol{P^b}$ and observation error covariance $\boldsymbol{R}$, determines the relative contributions to the updated analysis.

State variables for emissions include $NO_x$ and NMVOCs. To reduce the degree of freedom in the analysis and avoid the difficulty associated with estimating spatio-temporal variations in background errors for individual species, we focus on optimizing the lumped total NMVOC emissions. During the forecast step, we differentiate individual NMVOC species emissions from the total NMVOC emissions using bottom-up statistical information. For a consistent comparison between simulations and observations, model-simulated $NO_2$ were diagnosed at the time and location of surface $NO_2$ measurements, whereas model-simulated HCHO was horizontally sampled to align with TROPOMI HCHO VCD retrievals, and subsequently integrated vertically.

In this study, the DA window was set to one day and daily TROPOMI HCHO columns were utilized as observational constraints in our inversion framework. The ensemble size was set to 50 to strike a balance between computational cost and inversion accuracy. To reduce the impact of unrealistic long-distance error correlations, the Gaspari and Cohn function (Gaspari and Cohn, 1999) was utilized as covariance localization to ensure the meaningful influence of observations on state variables within a specified cutoff radius, while mitigating their negative impacts on distant state variables. The optimal localization scale is interconnected with factors such as the assimilation window, the dynamic system, and the lifetime of chemical species. Given the average wind speed of 2.8 m/s (Table S2) and a DA window of 1 day, the localization scales for $NO_2$ and HCHO, both characterized as highly reactive species with lifespans of just a few hours, were set to 150 km and 100 km, respectively.

**2.2 Observation Data and Errors**

Considering the availability of HCHO data, we utilized daily offline retrievals of tropospheric HCHO columns from Sentinel-5P (S5P) L3 TROPOMI data obtained through Google Earth Engine (De Smedt et al., 2018). The S5P satellite follows a near-polar sun-synchronous orbit at an altitude of 824 km with a 17-day repeating cycle. It crosses the Equator at 13:30 local solar time (LST) on the ascending node. The spatial resolution at nadir was refined to $3.5 \times 5.5$ km$^2$ on 6 August 2019. Following the recommendations in the S5P HCHO product user manual, we filtered the source data to exclude pixels with qa_value less than 0.5 for HCHO column number density and 0.8 for aerosol index (AER_AI). The remaining high-quality pixels with minimal snow/ice or cloud interference are averaged to 27-km grids. Figure 1b illustrates the coverage and data amount of TROPOMI HCHO retrievals in August 2022 after processing. Although the distribution of filtered data exhibits spatial non-uniformity, most grid cells have observational coverage for over half of the time, particularly in the southern region of China where NMVOC emissions are higher. Based on validation against a global network of 25 ground-based Fourier transform infrared (FTIR) column measurements (Vigouroux et al., 2020), TROPOMI HCHO overestimates by 25% ($<2.5\times10^{15}$ molec cm$^{-2}$) in clean regions and underestimates by 30% ($>=8\times10^{15}$ molec cm$^{-2}$) in polluted regions. Therefore, we set the measurement error to 30%. To evaluate the effect of observational data retrieval errors on emission estimates, we conducted a sensitivity experiment in which HCHO columns were empirically bias-corrected

according to the error characteristics described above (Figure S1). The posterior emissions increased by 12.8% compared to those in the base experiment (EMDA), indicating that the existing retrieval error in HCHO measurements likely exerts an influence on the estimation of NMVOC emissions. The representation error can be disregarded because the model's resolution significantly surpasses that of the TROPOMI pixels.

To address the chemical feedback among VOC-$NO_x$-$O_3$, we also simultaneously optimized $NO_x$ emissions by assimilating in-situ $NO_2$ observations. The extensively covered and high-precision monitoring network can provide sufficient constraints for emission inversion (Figure 1a). Hourly averaged surface $NO_2$ observations from national control air quality stations obtained from the Ministry of Ecology and Environment of the People's Republic of China (http://106.37.208.228:8082/, last access: 5 May 2023). In case where multiple stations are located within the same grid, a random site is chosen for validation, while the remaining sites are averaged to mitigate the impact of error correlation (Houtekamer and Zhang, 2016) for assimilation. In total, 1276 stations were chosen for assimilation and an additional 425 independent stations were selected for verification (Figure 1a). The observation error covariance matrix $\boldsymbol{R}$ incorporates contributions from both measurement and representation errors. The measurement error is defined as $\varepsilon_0 = 1.0 + 0.005 \times \Pi_0$, where $\Pi_0$ represents the observed $NO_2$ concentration. Following the approach of Elbern et al. (2007) and Feng et al. (2018), the representative error is defined as $\varepsilon_r = \gamma \varepsilon_0 \sqrt{\Delta l / L}$, where $\gamma$ is a tunable parameter (here, $\gamma$=0.5), $\Delta l$ is the grid spacing (27 km), and $L$ is the radius (here, $L$=0.5) of the observation's influence area. The total observation error ($r$) was defined as $r = \sqrt{\varepsilon_0^2 + \varepsilon_r^2}$. The observation errors are assumed to be uncorrelated so that $\boldsymbol{R}$ is a diagonal matrix.

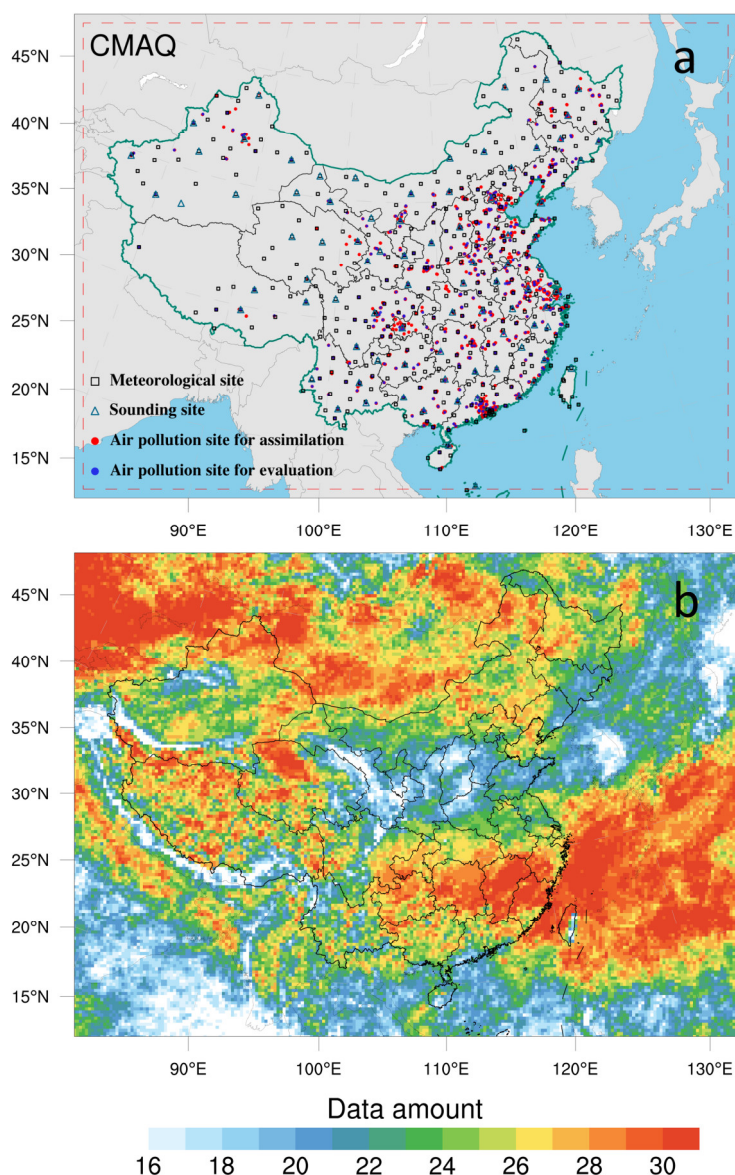



**Figure 1**. Model domain and observation network (a) and data amount of TROPOMI HCHO retrievals during August 2022 in each grid (b). The red dashed frame delineates the CMAQ computational domain; black squares denote surface meteorological measurement sites; navy triangles indicate sounding sites (Text S1), and red and blue dots represent air pollution measurement sites, where red dots are used for assimilation and blue dots for independent evaluation.

**2.3 Prior Emissions and Uncertainties**

The prior anthropogenic $NO_x$ and NMVOC emissions for China were obtained from the most recent Multi-resolution Emission Inventory for China of 2020 (MEIC, http://www.meicmodel.org/, last access: 8 May 2023) (Zhang et al., 2009). For anthropogenic emissions outside China, we utilized the mosaic Asian anthropogenic

emission inventory (MIX) for the base year of 2010 (Li et al., 2017b). The daily emission inventory, which was arithmetically averaged from the combined monthly emission inventory, was employed as the first guess. Ship emissions were derived from the shipping emission inventory model (SEIM) for 2017, which was calculated based on the observed vessel automatic identification system (Liu et al., 2017). Biomass burning emissions were retrieved from the Global Fire Emissions Database version 4.1 (GFEDv4, https://www.globalfiredata.org/, last access: 8 May 2023) (van der Werf et al., 2017; Mu et al., 2011). Biogenic $NO_x$ and NMVOC emissions were calculated using the Model of Emissions of Gases and Aerosols from Nature (MEGAN) developed by Guenther et al. (2012).

As previously mentioned, the optimized emissions are transferred to the next DA window as prior emissions for iterative inversion. For biogenic emissions, it is decomposed into hourly scales based on the daily varying temporal profiles in MEGAN as model inputs. Daily emission variations will largely dominate the uncertainty in emissions. Taking into account compensating for model errors and avoiding filter divergence, we consistently applied an uncertainty of 25% to each model grid of $NO_x$ emissions at each DA window, as in Feng et al. (2020). NMVOC emissions typically exhibit greater uncertainties compared to $NO_x$ emissions (Li et al., 2017b). Based on model evaluation, the uncertainty of NMVOC emissions was set to 40% (Kaiser et al., 2018; Souri et al., 2020; Cao et al., 2018). A sensitivity experiment involving a doubling of the prior uncertainty (80%) revealed that the differences in posterior NMVOC emissions amounted to a mere 0.2% (Figure S2). The implementation of a 'two-step' inversion strategy allows for the timely correction of residual errors from the previous assimilation window in the current window, thus ensuring that the RAPAS system has a relatively low dependence on prior uncertainty settings. This study also addresses uncertainties in emissions for CO, $SO_2$, primary $PM_{2.5}$, and coarse $PM_{10}$ to consider the chemical feedback between different species following Feng et al. (2023).

## 3 Experimental Design

During the summer of 2022, southern China experienced severe heatwave conditions. The combination of high temperatures and drought had a pronounced effect on vegetation growth and NMVOC emissions, thereby influencing $O_3$ production (Wang et al., 2023). Consequently, we opted to focus on August 2022, as it presented an ideal period for testing the capabilities of our DA system. Before implementing the emission

inversion, a relatively perfect initial field is generated at 0000 UTC on August 1 2022 through conducting a 5-day simulation with 6-hour interval 3D-Var data assimilation. Subsequently, daily emissions are continuously updated over the entire month of August (EMDA). Additionally, we designed a sensitivity experiment (EMS) to illustrate the significance of optimizing $NO_x$ emissions in quantifying VOC-$O_3$ chemical reactions. In this experiment, $NO_x$ emissions were not optimized. To validate the posterior emissions of $NO_x$ and NMVOCs in EMDA, we compared two parallel forward simulation experiments, denoted as CEP and VEP, corresponding to prior and posterior emission scenarios, respectively, against $NO_2$ and HCHO measurements. To investigate the impact of optimizing NMVOC emissions on the secondary production and loss of surface $O_3$, a forward simulation experiment (CEP1) was conducted with the prior NMVOC emissions and the posterior $NO_x$ emissions. Another forward modelling experiment (CEP2) used the posterior emissions of EMS to evaluate its performance. All experiments employ identical meteorological fields, as well as the same gas-phase and aerosol modules. Table 1 summarizes the different emission inversion and validation experiments conducted in this study.

**Table 1**. The assimilation, sensitivity, and validation experiments conducted in this study.

| Exp.Type | Exp. Name | NMVOC emissions | $NO_x$ emissions |
|---|---|---|---|
| Assimilation | EMDA | MEIC 2020 and MEGAN for August (the first DA window), optimized emissions of the previous window (other DA windows) | MEIC 2020 and MEGAN for August (the first DA window), optimized emissions of the previous window (other DA windows) |
| Sensitivity | EMS | Same as EMDA | MEIC 2020 and MEGAN for August |
| Validation | CEP | MEIC 2020 and MEGAN for August | MEIC 2020 and MEGAN for August |
| | VEP | Posterior emissions of EMDA | Posterior emissions of EMDA |
| | CEP1 | Same as CEP | Posterior emissions of EMDA |
| | CEP2 | Posterior emissions of EMS | Same as CEP |

## 4 Results

### 4.1 Inverted Emissions

Figure 2 shows the spatial distribution of temporally averaged prior and posterior NMVOC emissions, along with their differences, in NMVOC emissions. Hotspots of prior NMVOC emissions were prevalent across much of central and southern China. However, posterior NMVOC emissions were predominantly concentrated in the NCP, Yangtze River Delta (YRD), PRD, and Sichuan Basin (SCB), characterized by high levels of anthropogenic activity. High emissions are also located in parts of central and southern China with warm climate favorable for emitting biogenic NMVOCs. Employing TROPOMI HCHO observations as constraints led to widespread decreases of approximately 60–70% over these areas, indicating a large substantial of biogenic NMVOC emissions. In northwestern China, there was a moderate increase in NMVOC emissions.

A potential significant TROPOMI retrieval errors in polluted regions could exacerbate the emission decreases (Text S2). Additionally, uncertainties in MEGAN parameterization have significant implications for NMVOC emission estimations, particularly concerning the responses of vegetation in MEGAN to temperature and drought stress (Angot et al., 2020; Jiang et al., 2018). Zhang et al. (2021) highlighted that the temperature-dependent activity factor noticeably increases with rising temperatures in MEGAN. Wang et al. (2021b) pointed out that the missing of a drought scheme is one of the factors causing the overestimation of isoprene emissions in MEGAN. Opacka et al. (2022) optimized the empirical parameter in the MEGANv2.1 soil moisture stress algorithm, resulting in significant reductions in isoprene emissions and providing better agreement between modelled and observed HCHO temporal variability in the central U.S. During the study period, China experienced severe heatwave conditions, which may further hinder the MEGAN's ability to effectively capture the impacts of high temperatures and drought on vegetation, thus resulting in significant overestimation in NMVOC emissions (Wang et al., 2022). Ultimately, the biogenic NMVOC emissions decreased by 53.7%, which was higher than the 43.4% decrease in anthropogenic NMVOC emissions (Figure S3). Overall, the large magnitude of emission decrease of 50.2% in our inversion is comparable to studies in southern China (Bauwens et al., 2016; Zhou et al., 2023), southeastern US (Kaiser et al., 2018), Africa (Marais et al., 2014), India (Chaliyakunnel et al., 2019), Amazonia

(Bauwens et al., 2016), and parts of Europe (Curci et al., 2010), but opposite to the large-scale emission increase over China in Cao et al. (2018). For NO$_x$ (Figure S4), the nationwide total emissions decreased by 10.2%, with the main reductions concentrated in the NCP, YRD, parts of Central China, and most key urban areas.

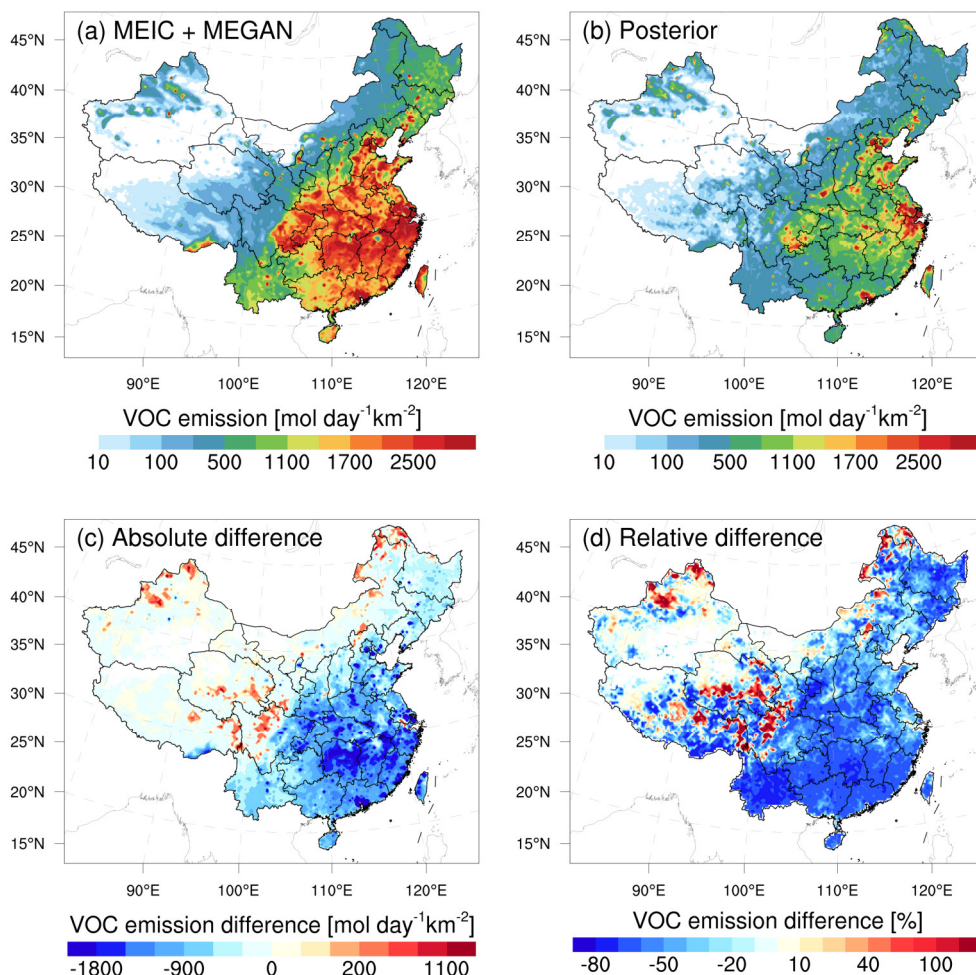



**Figure 2**. Spatial distribution of the time-averaged (a) prior emissions (MEIC 2020 + MEGAN), (b) posterior emissions, (c) absolute difference (posterior minus prior), and (d) relative difference of NMVOCs over China.

Table 2 shows the changes in emissions of biogenic NMVOCs across different land cover types (Figure S5) after inversion. The most significant reduction in biogenic emissions occurred within woody savannas, accounting for 26.9% of the overall reduction, followed by savannas and croplands, accounting for 21.2% and 17.2% respectively. Among all vegetation types, the broadleaf evergreen forests, recognized as the primary source of isoprene emission (Wang et al., 2021a), presented the greatest uncertainty, with NMVOC emissions experiencing a significant reduction of 66.2%. Standard emission rates in MEGAN are derived from leaf- or canopy-scale flux

393 measurements and extrapolated globally across regions sharing similar landcover

394 characteristics, based on very limited observations (Guenther et al., 1995). This

395 methodology introduces biases due to the large variability in emission rates among

396 plant species.

397 **Table 2**. Prior and posterior biogenic NMVOC emissions, as well as their differences

398 for different land cover types.

| Land cover type | Prior Mmol/month | Posterior Mmol/month | Difference Mmol/month (%) |
|---|---|---|---|
| Evergreen needleleaf forests | 955.7 | 549.3 | -406.4 (-42.5) |
| Evergreen broadleaf forests | 13985.1 | 4728.2 | -9256.8 (-66.2) |
| Deciduous needleleaf forests | 46.6 | 48.8 | 2.2 (4.7) |
| Deciduous broadleaf forests | 8335.5 | 3487.4 | -4848.1 (-58.2) |
| Mixed forests | 8731.0 | 3961.7 | -4769.4 (-54.6) |
| Closed shrublands | 9.7 | 3.7 | -6.0 (-61.5) |
| Open shrublands | 21.3 | 8.6 | -12.8 (-59.8) |
| Woody savannas | 39327.2 | 16925.2 | -22402.0 (-57.0) |
| Savannas | 28319.7 | 10629.4 | -17690.3 (-62.5) |
| Grasslands | 16912.7 | 14269.6 | -2643.1 (-15.6) |
| Permanent wetlands | 286.1 | 115.4 | -170.8 (-59.7) |
| Croplands | 25537.8 | 11215.5 | -14322.2 (-56.1) |
| Cropland-natural vegetation mosaics | 10894.7 | 4289.8 | -6605.0 (-60.6) |
| Sparsely vegetated | 1814.7 | 1644.0 | -170.6 (-9.4) |

### 399 4.2 Evaluations for Posterior Emissions

400 The $NO_x$ emissions were first evaluated by indirectly comparing the forward simulated

401 $NO_2$ concentrations with measurements. As shown in Figure S6, the CEP with prior

402 emissions exhibited positive biases in eastern China and negative biases in western

403 China. However, when posterior emissions were used in the VEP, a substantial

404 improvement in simulation performance was observed. Biases were limited to within

405 $\pm 3$ $\mu g$ $m^{-3}$, and correlation coefficients exceeded 0.7 across the entire region. Figure 3

406 presents the simulated HCHO VCDs using prior and posterior NMVOCs emissions,

along with their associated biases. Both experiments showed high VCDs over central and eastern China, especially in the YRD and SCB. However, the CEP displayed substantial overestimation across most of mainland China, with the largest bias reaching $12 \times 10^{15}$ molec cm$^{-2}$ in Central China. Conversely, the VEP demonstrated notable improvements in both the magnitude and spatial distribution of simulated HCHO columns after the inversion compared to TROPOMI retrievals. More than 84% of the areas exhibited biases of less than $1 \times 10^{15}$ molec cm$^{-2}$, and no significant spatial variation was observed. Overall, the biases in simulated HCHO VCDs decreased by 75.7% after the inversion. These results emphasize the efficiency of our system in reducing uncertainty in both NO$_x$ and NMVOC emissions.

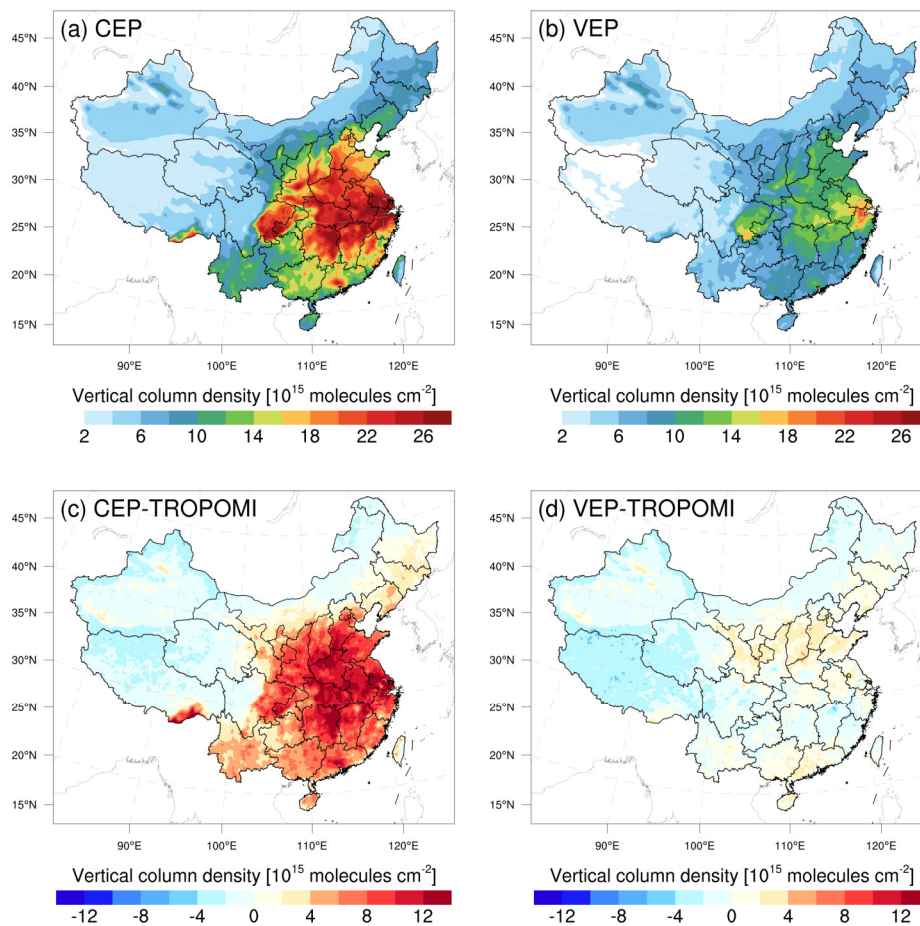



**Figure 3**. Simulated HCHO vertical column densities using prior (a) and posterior (b) NMVOC emissions, along with their biases (c and d) against TROPOMI measurement. All model results were sampled at TROPOMI overpass time.

**4.3 Implications for Surface O$_3$**

Figure 4 shows the spatial distribution of the mean bias (BIAS), root mean square error (RMSE), and correlation coefficient (CORR) for simulated O$_3$ concentrations in the CEP1 and VEP experiments compared to assimilated observations. Beyond the northwestern region of China, the CEP1 exhibited significant overestimation throughout the entire area, with a BIAS of 20.5 μg m$^{-3}$. In the VEP, the modeled O$_3$ chemical production were alleviated, especially in the southern regions of China where NMVOC emissions had significantly decreased. Overall, observation-constrained NMVOC emissions resulted in a 49.3% decrease in the BIAS, bringing it down to 10.4 μg m$^{-3}$. Additionally, the RMSE showed noticeable improvement due to the assimilation of HCHO observation, reducing the value from 30.9 to 23.3 μg m$^{-3}$. Despite a significant reduction in NMVOC emissions after inversion, notable overestimations persisted in northern provinces such as Liaoning, Hebei, Shanxi, and Shaanxi. This may be attributed to limited NMVOC constraints resulting from insufficient observations during the study period (Figures 1b and 3d). The remaining discrepancies between simulations and observations can be attributed to the combined results of intricate urban-rural sensitivity regimes and O$_3$ photochemistry reactions, which may not be comprehensively represented by CMAQ model, masking any potential improvement expected from the constrained emissions (See Sect. 4.4). The CORR was comparable between the CEP1 and VEP experiments, reflecting that the CMAQ model effectively simulated the temporal variation of O$_3$ concentrations. The biases at the independent sites were similar to those at the assimilated sites (Figure S7). In comparison to CEP1, the decreasing ratios in BIAS and RMSE in VEP were 46.7% and 23.4%, respectively.

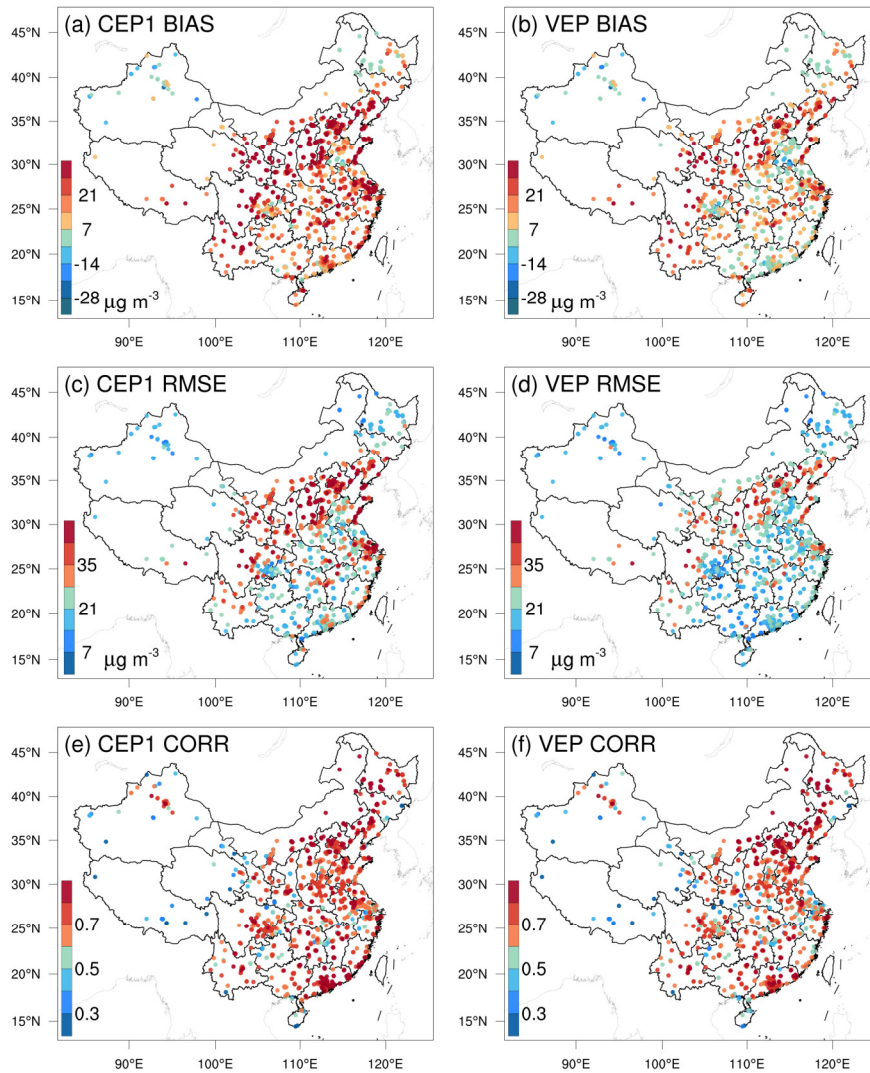



**Figure 4**. Spatial distribution of mean bias (BIAS, a and b), root mean square error (RMSE, c and d), and correlation coefficient (CORR, e and f) for simulated $O_3$ using prior (left, CEP1) and posterior (right, VEP) emissions, respectively, against assimilated observations.

Figure 5 shows the time series of simulated and observed hourly $O_3$ concentrations and their RMSEs, verified against surface monitoring sites. The VEP achieved better representations of diurnal $O_3$ variations compared with those in the CEP1, especially excelling in reproducing elevated $O_3$ concentrations at noon. Constraining the NMVOC emissions also led to better model simulations in terms of RMSE throughout the entire study period. Time-averaged BIAS and RMSE decreased from 20.6 and 37.3 μg m$^{-3}$ to 10.6 and 31.0 μg m$^{-3}$, respectively. We also evaluated the simulation results for seven key cities (i.e., Beijing, Shanghai, Guangzhou, Wuhan, Chongqing, Yinchuan, and Changchun, which represent key cities in North, East, South, Central, Southwest,

Northwest, and Northeast China, respectively), and the biases in the VEP with posterior

emissions all showed a significant reduction (Figure S8). Overall, the assimilation of

HCHO column observations effectively reduced NMVOC emission uncertainties and

consequently improved simulations of HCHO and $O_3$. These improvements hold

promise for further research into the implications of emission optimizations on regional

$O_3$ photochemistry.

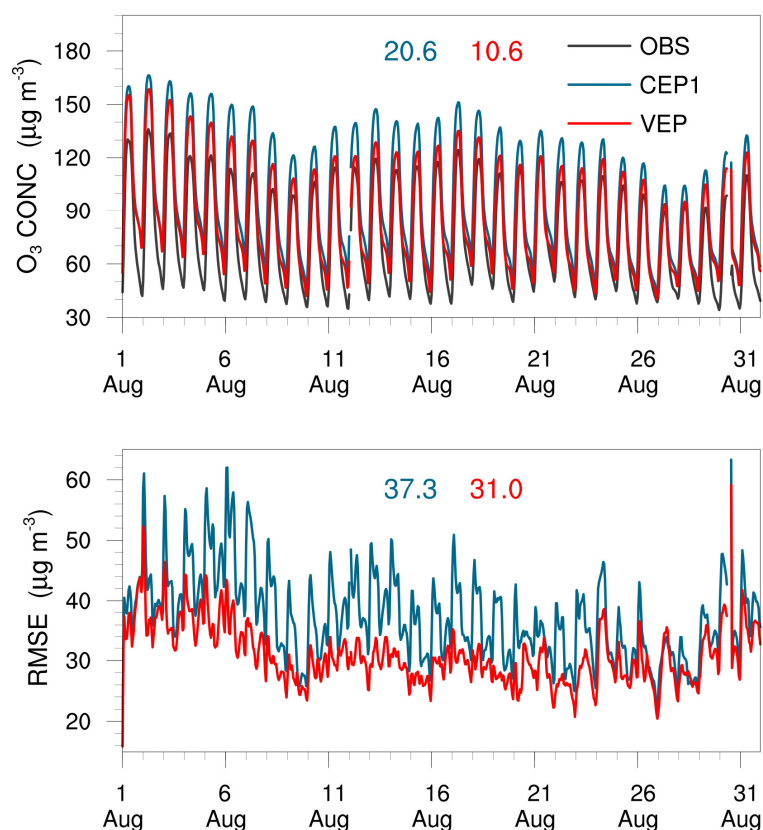



**Figure 5**. Time series comparison of hourly surface $O_3$ concentrations ($\mu g\ m^{-3}$) and

RMSE ($\mu g\ m^{-3}$) from CEP1 and VEP experiments against all observations at 1701

monitoring sites. The blue and red values on the graph represent the time-averaged

statistics in the CEP1 and VEP experiments, respectively.

As crucial $O_3$ precursors, the abundance of NMVOCs plays a significant role in

modulating $O_3$ production. Here we employed the IRRs to elucidate changes related to

$O_3$ production and loss at the surface, stemming from constrained $NO_x$ and NMVOC

emissions. Figure 6 illustrates comparisons of the simulated maximum daily 8-hour

average (MDA8) surface $O_3$ levels and net reaction rates before and after the inversion.

The CEP1 exhibited an overestimation of $O_3$ levels, with a BIAS of 22.6% compared

to observed $O_3$ concentrations. This overestimation corresponded to the high net

479    chemical rates of $O_3$ in these areas (Figure S9). After inversion, $O_3$ net rates mitigated

480    in most regions. Consequently, the VEP experiment yielded results that closely aligned

481    with observations, with a BIAS of 9.2%. Referring to Figure 6e and 6f, differences in

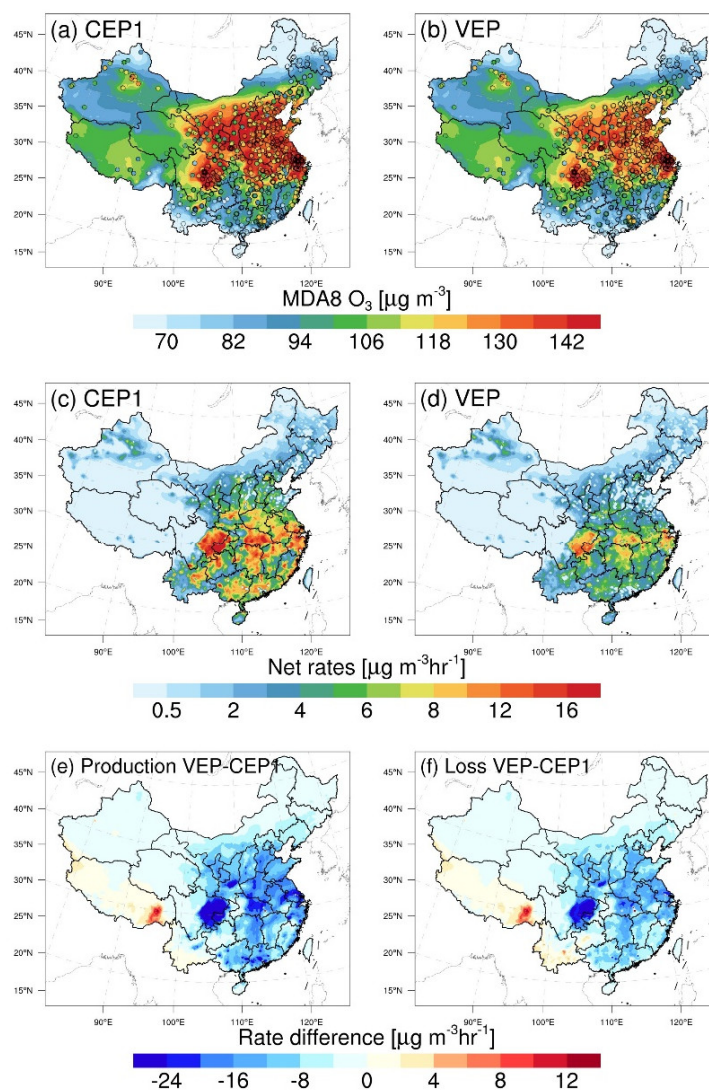
482    production rates of $O_3$ closely track the changes in the NMVOC emissions (Figure 2).

483    The discrepancies in specific regions may be attributed to the complex nonlinear

484    relationships associated with $O_3$ and its precursors, which depend on prevailing

485    chemical regimes and regional transport. Additionally, changes in $O_3$ production

486    predominantly drive the overall decrease in $O_3$ concentrations, outweighing changes in

487    $O_3$ loss.

488

489    **Figure 6**. Comparisons of (a, b) simulated maximum daily 8-hour average (MDA8) $O_3$

490    concentrations, (c, d) net reaction rates, (e, f) and differences in production and loss

491    rates between CEP1 and VEP experiments at the surface. Surface MDA8 $O_3$ values

492    (circles) from the national control air quality stations were overlaid

Figure 7 shows the differences in the six principal pathways responsible for $O_3$ loss and formation, when comparing simulations employing prior and posterior emissions. The reactions of $HO_2 + NO$ and $RO_2 + NO$ are treated as the pathways leading to $O_3$ formation, whereas $O_3$ loss involves reactions including $NO2 + OH$, $O_3 + HO_2$, $O_3 + NMVOCs$, and $O1D + H_2O$ (Wang et al., 2019). Our analysis was focused on the time frame from 12:00 to 18:00 according to China standard time (CST). The differences were computed by subtracting the simulation with posterior emissions from those with prior emissions. Following the emission of NMVOCs, they undergo rapid oxidation by atmospheric hydroxyl (OH) radicals. Due to the substantial decrease in NMVOC emissions, there was a reduction in the production of hydroperoxy radicals ($HO_2$) and organic peroxy radicals ($RO_2$) (Figure S10). Consequently, this reduction in $HO_2/RO_2$ levels, coupled with their reaction with NO, resulted in diminished $O_3$ production (Figures 7a and 7b). A strong correlation was observed between changes in $O_3$ production via the $RO_2 + NO$ reaction and NMVOC emissions (Figure 2), consistent with the findings of Souri et al. (2020). Typically, in NMVOC-rich environments, a decrease in NMVOC emissions boosts OH concentrations. Consequently, we noted an enhancement in the $NO_2 + OH$ reaction in the eastern and central regions of China. In response to heightened $HO_x$ concentrations over these areas, an increased $O_3$ loss through the $O_3 + HO_x$ pathway was observed. Furthermore, we detected a substantial decrease in $O_3$ loss through reactions with NMVOCs, especially in the southern China, where substantial isoprene emissions are prevalent. This reduction was primarily attributable to the decrease in NMVOC and $O_3$ levels. While the $NMVOC + O_3$ reaction proceeds at a substantially slower rate $NMVOC + OH$, this specific chemical pathway remains significant in oxidizing NMVOC and forming $HO_x$ in forests areas (Paulson and Orlando, 1996). The difference in $O1D + H_2O$ is primarily driven by the decrease of $O_3$ photolysis. Although the rate of $O_3$ loss decreases in some chemical pathways, overall, the rate of $O_3$ production dominates the changes in $O_3$ concentration.

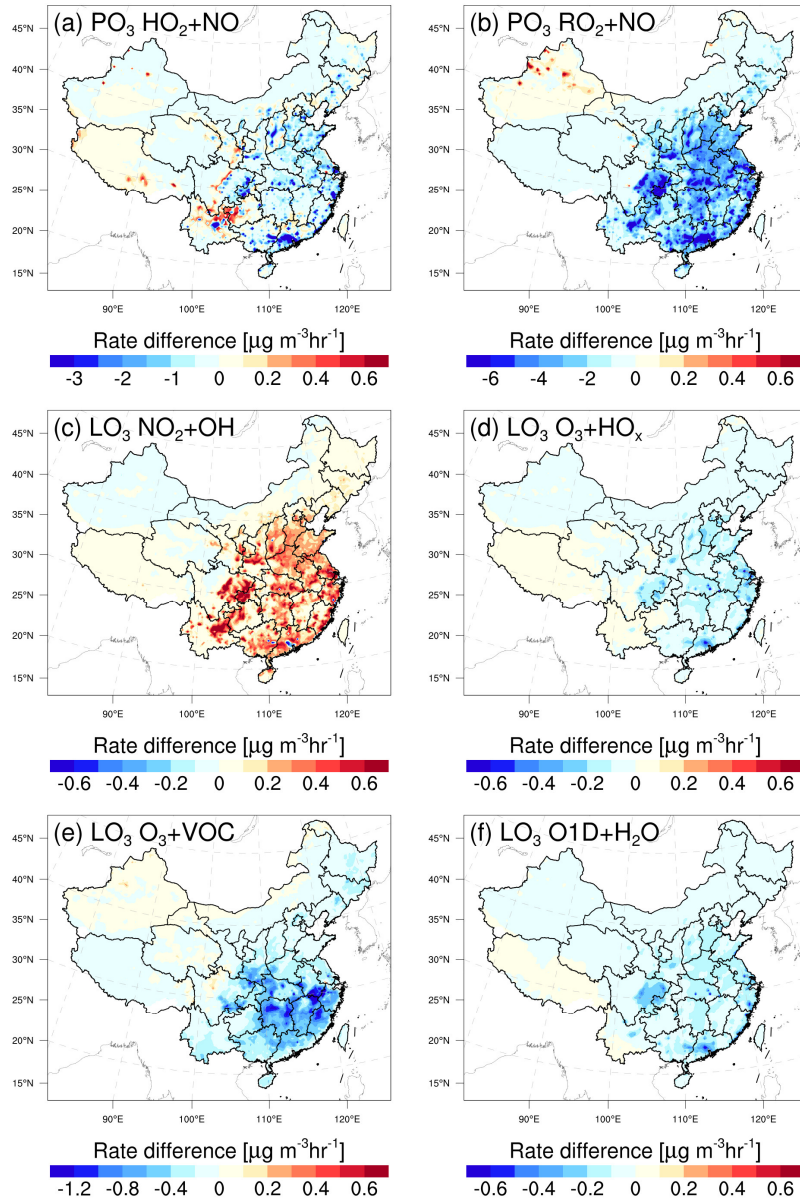



**Figure 7**. Differences in six major pathways of $O_3$ production and loss between CEP1 and VEP experiments at the surface. Time period: August 2022, 12:00–18:00 CST. $PO_3$ and $LO_3$ represent the pathways of $O_3$ formation and loss, respectively.

**4.4 Discussions**

$O_3$ simulations over China have a tendency to be overestimated in studies involving chemical transport modeling. For example, by intercomparing 14 state-of-the-art CTMs with $O_3$ observations within the framework of the MICS-Asia III, Li et al. (2019) identified a substantial overestimation of annual surface $O_3$ in East Asia, ranging from 20 to 60 μg m$^{-3}$. Notably, the NCP exhibited substantial overestimations, with most models overestimating $O_3$ by 100–200% during May–October. Despite our

531    optimization of $O_3$ precursor emissions, the posterior simulations still exhibit some

532    degree of overestimation (Figure 4), suggesting that there may indeed be an effect of

533    systematic bias, such as meteorological fields, spatial resolution, model treatments of

534    nonlinear photochemistry and other physical processes. The WRF can generally

535    reproduce meteorological conditions sufficiently in terms of their temporal variation

536    and magnitude over China (Figure S11), with small biases of -0.5 °C, -5.3%, 0.3 m/s,

537    and -42.4 m for temperature at 2 m, relative humidity at 2 m, and wind speed at 10 m,

538    and planetary boundary layer height, respectively. However, due to the relatively coarse

539    spatial resolution, NO titration effects in urban areas may not be well represented in the

540    model, which can lead to an overestimation of $O_3$ in these areas. Additionally, model

541    inherent errors arising from the model structure, parameterization, and the

542    simplification or lack of chemical mechanisms inevitably affect the $O_3$ simulations. For

543    example, Li et al. (2018) reported that heterogeneous reactions of nitrogen compounds

544    could weaken the atmospheric oxidation capacity and thus reduce surface $O_3$

545    concentration by 20–40 $\mu g\ m^{-3}$ for the polluted regions over China. These reactions

546    have not been fully incorporated in CMAQ chemical mechanisms. However, there is

547    still a lack of reasonable and effective algorithms for addressing model errors through

548    assimilation (Houtekamer and Zhang, 2016).$O_3$ concentration and $NO_x$ (VOC)

549    emissions are positively correlated in the $NO_x$ (VOC)-limited region and negatively

550    correlated in the VOC ($NO_x$)-limited region (Tang et al., 2011). Therefore, the

551    uncertainty in $NO_x$ emissions can affect the model's diagnosis of $O_3$-$NO_x$-VOC

552    sensitivity, thereby introducing substantial model errors in the HCHO yield from VOC

553    oxidation. In the base inversion experiment (EMDA), we simultaneously assimilated

554    $NO_2$ and HCHO observations to optimize $NO_x$ and NMVOC emissions. To evaluate the

555    impact of optimized $NO_x$ emissions on $O_3$-VOC chemistry, EMS disregarded the

556    uncertainty of $NO_x$ and focused on optimizing NMVOC emissions. Compared to the

557    EMDA, in areas where $NO_x$ is significantly overestimated, NMVOC emissions in the

558    EMS have correspondingly decreased (Figure 8b). This might be due to under high-

559    $NO_x$ conditions, HCHO production occurs promptly, thereby compensating for the

560    substantial amount of HCHO already present in the atmosphere by reducing emissions

561    (Chan Miller et al., 2017). Figure S12 shows comparisons of concentrations and RMSE

562    between the simulations using posterior emissions from EMS and EMDA experiments.

563    Compared to VEP, CEP2 showed a larger RMSE, highlighting the necessity for

564    simultaneous optimization of $NO_x$ emissions when evaluating the impact of NMVOC

emission optimization on $O_3$. Additionally, CEP2 using prior $NO_x$ emissions exhibited lower $O_3$ levels over parts of NCP and YRD, as well as some urban areas (Figure 8c), but with larger biases and RMSEs (Figure 8d). The reduction in NMVOC emissions contributed to a partial decrease in $O_3$ concentration. More significantly, these areas typically align with VOC-limited mechanisms (Wang et al., 2019; Wang et al., 2021c). Therefore, the overestimation of $NO_x$ emissions (Figure S4) excessively inhibits $O_3$ accumulation due to the titration effect, thereby disrupting the evaluation of NMVOC contributions to $O_3$. This substantial disparity also seriously affects $O_3$ source apportionment, precursor-sensitive area delineation, and emissions reduction policy formulation.

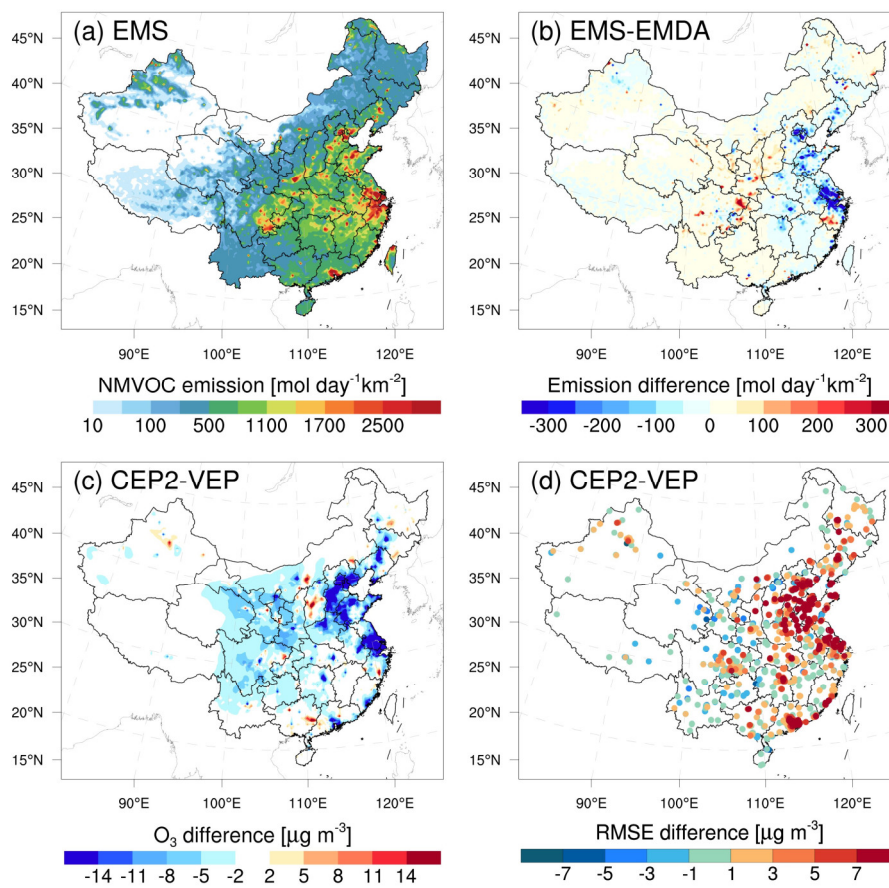



**Figure 8**. Spatial distribution of (a) posterior emissions in the EMS experiment, (b) differences in posterior emissions between EMS and EMDA, and differences in (c) simulated $O_3$ concentrations and (d) RMSE between CEP2 and VEP experiments. EMS did not optimize $NO_x$ emissions compared to EMDA.

## 5 Summary and Conclusions

In this study, we extended the RAPAS assimilation system with the EnKF assimilation algorithm to optimize NMVOC emissions using the TROPOMI HCHO retrievals. Taking the MEIC 2020 for anthropogenic emissions and MEGANv2.1 output for biogenic sources as a priori, NMVOC emissions over China in August 2022 were inferred. Importantly, we implicitly took the chemical feedback among VOC-NO$_x$-O$_3$ into account by simultaneously adjusting NO$_x$ emissions using nationwide in-situ NO$_2$ observations. Furthermore, we quantified the impact of NMVOC emission inversion on surface O$_3$ pollution using the CMAQ-IRR model.

The application of TROPOMI HCHO observations as constraints led to a substantial reduction of 50.2% compared to the prior emissions for NMVOCs in August 2022. A domain-wide significant decrease was found over central and southern China with abundant forests, especially for the broadleaf evergreen forests, implying a considerable overestimation of biogenic NMVOC emissions. Observation-constrained emissions significantly improved the performance of surface NO$_2$ and HCHO column simulations, reducing biases by 97.4% and 75.7%, respectively. This highlights the effectiveness of the RAPAS in reducing uncertainty in NO$_x$ and NMVOC emissions. Isolating the impact of NO$_x$ emission changes, the posterior NMVOC emissions significantly mitigated the overestimation in prior O$_3$ simulations, resulting in a 49.3% decrease in surface O$_3$ biases. This is mainly attributed to a substantial decrease in the RO2 + NO reaction rate (a major pathway for O$_3$ production) and an increase NO$_2$ + OH reaction rate (a major pathway for O$_3$ loss) during the afternoon, resulting in a decrease in the simulated MDA8 surface O$_3$ concentrations by approximately 15 μg m$^{-3}$.

Sensitivity inversions demonstrate the robustness of top-down emissions to variations in prior uncertainty settings, yet they are sensitive to HCHO column biases, highlighting the importance of comprehensive validation studies utilizing available remote-sensing data and, if possible, airborne validation campaigns. Moreover, we found that, in comparison to optimizing NMVOC emissions alone, the joint optimization of NMVOC and NO$_x$ emissions can significantly improve the overall performance of O$_3$ simulations. Ignoring errors in NO$_x$ emissions introduces uncertainty in quantifying the impact of NMVOC emissions on surface O$_3$, especially in areas where overestimated NO$_x$ emissions can unrealistically amplify titration effects,

613    highlighting the necessity of simultaneous optimization of $NO_x$ emissions.

614

## Data availability

616    The observations used for assimilation and the optimized emissions in this study can be

617    accessed at https://doi.org/10.5281/zenodo.10079006 (Feng and Jiang, 2023).

618

## Author contribution

620    SF and FJ conceived and designed the research. SF developed the data assimilation

621    code, analyzed data, and prepared the paper with contributions from all co-authors. FJ

622    supervised and assisted in conceptualization and writing. TQ, NW, MJ, SZ, JC, FY, and

623    WJ reviewed and commented on the paper.

624

## Competing interests

626    The authors declare that they have no conflict of interest.

627

## Acknowledgements

629    This work is supported by the National Key R&D Program of China (Grant No.

630    2022YFB3904801), the National Natural Science Foundation of China (Grant No:

631    42305116 and 42377102), the Natural Science Foundation of Jiangsu Province of China

632    (Grant No: BK20230801), and the Hangzhou Agricultural and Social Development

633    Scientific Research Project (Grant No: 202203B29). The authors also gratefully

634    acknowledge the High-Performance Computing Center (HPCC) of Nanjing University

635    for doing the numerical calculations in this paper on its blade cluster system.

636

## References

638 Angot, H., McErlean, K., Hu, L., Millet, D. B., Hueber, J., Cui, K., Moss, J., Wielgasz, C., Milligan,
639    T., Ketcherside, D., Bret-Harte, M. S., and Helmig, D.: Biogenic volatile organic compound
640    ambient mixing ratios and emission rates in the Alaskan Arctic tundra, Biogeosciences, 17,
641    6219-6236, 10.5194/bg-17-6219-2020, 2020.

642 Bauwens, M., Stavrakou, T., Müller, J. F., De Smedt, I., Van Roozendael, M., van der Werf, G. R.,
643    Wiedinmyer, C., Kaiser, J. W., Sindelarova, K., and Guenther, A.: Nine years of global
644    hydrocarbon emissions based on source inversion of OMI formaldehyde observations, Atmos.
645    Chem. Phys., 16, 10133-10158, 10.5194/acp-16-10133-2016, 2016.

Byun, D., and Schere, K. L.: Review of the governing equations, computational algorithms, and other components of the models-3 Community Multiscale Air Quality (CMAQ) modeling system, Applied Mechanics Reviews, 59, 51-77, 10.1115/1.2128636, 2006.

Cao, H., Fu, T. M., Zhang, L., Henze, D. K., Miller, C. C., Lerot, C., Abad, G. G., De Smedt, I., Zhang, Q., van Roozendael, M., Hendrick, F., Chance, K., Li, J., Zheng, J., and Zhao, Y.: Adjoint inversion of Chinese non-methane volatile organic compound emissions using space-based observations of formaldehyde and glyoxal, Atmos. Chem. Phys., 18, 15017-15046, 10.5194/acp-18-15017-2018, 2018.

Chaliyakunnel, S., Millet, D. B., and Chen, X.: Constraining Emissions of Volatile Organic Compounds Over the Indian Subcontinent Using Space-Based Formaldehyde Measurements, Journal of Geophysical Research: Atmospheres, 124, 10525-10545, 10.1029/2019JD031262, 2019.

Chan Miller, C., Jacob, D. J., Marais, E. A., Yu, K., Travis, K. R., Kim, P. S., Fisher, J. A., Zhu, L., Wolfe, G. M., Hanisco, T. F., Keutsch, F. N., Kaiser, J., Min, K. E., Brown, S. S., Washenfelder, R. A., González Abad, G., and Chance, K.: Glyoxal yield from isoprene oxidation and relation to formaldehyde: chemical mechanism, constraints from SENEX aircraft observations, and interpretation of OMI satellite data, Atmos. Chem. Phys., 17, 8725-8738, 10.5194/acp-17-8725-2017, 2017.

Cheng, S., Cheng, X., Ma, J., Xu, X., Zhang, W., Lv, J., Bai, G., Chen, B., Ma, S., Ziegler, S., Donner, S., and Wagner, T.: Mobile MAX-DOAS observations of tropospheric NO2 and HCHO during summer over the Three Rivers' Source region in China, Atmos. Chem. Phys., 23, 3655-3677, 10.5194/acp-23-3655-2023, 2023.

Curci, G., Palmer, P. I., Kurosu, T. P., Chance, K., and Visconti, G.: Estimating European volatile organic compound emissions using satellite observations of formaldehyde from the Ozone Monitoring Instrument, Atmos. Chem. Phys., 10, 11501-11517, 10.5194/acp-10-11501-2010, 2010.

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., Pedergnana, M., Loyola, D., Beirle, S., Wagner, T., Eskes, H., van Geffen, J., Boersma, K. F., and Veefkind, P.: Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, Atmos. Meas. Tech., 11, 2395-2426, 10.5194/amt-11-2395-2018, 2018.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmospheric Chemistry and Physics, 7, 3749-3769, 10.5194/acp-7-3749-2007, 2007.

Fang, X., Shao, M., Stohl, A., Zhang, Q., Zheng, J., Guo, H., Wang, C., Wang, M., Ou, J., Thompson, R. L., and Prinn, R. G.: Top-down estimates of benzene and toluene emissions in the Pearl River Delta and Hong Kong, China, Atmos. Chem. Phys., 16, 3369-3382, 10.5194/acp-16-3369-2016, 2016.

Feng, S., Jiang, F., Jiang, Z., Wang, H., Cai, Z., and Zhang, L.: Impact of 3DVAR assimilation of surface PM2.5 observations on PM2.5 forecasts over China during wintertime, Atmospheric Environment, 187, 34-49, 10.1016/j.atmosenv.2018.05.049, 2018.

Feng, S., Jiang, F., Wang, H., Wang, H., Ju, W., Shen, Y., Zheng, Y., Wu, Z., and Ding, A.: NOx Emission Changes Over China During the COVID-19 Epidemic Inferred From Surface NO2 Observations, Geophysical Research Letters, 47, 10.1029/2020gl090080, 2020.

Feng, S., Jiang, F., Wang, H., Shen, Y., Zheng, Y., Zhang, L., Lou, C., and Ju, W.: Anthropogenic emissions estimated using surface observations and their impacts on PM2.5 source apportionment over the Yangtze River Delta, China, Science of The Total Environment, 828, 154522, 10.1016/j.scitotenv.2022.154522, 2022.

Feng, S., Jiang, F., Wu, Z., Wang, H., He, W., Shen, Y., Zhang, L., Zheng, Y., Lou, C., Jiang, Z., and Ju, W.: A Regional multi-Air Pollutant Assimilation System (RAPAS v1.0) for emission estimates: system development and application, Geosci. Model Dev., 16, 5949-5977, 10.5194/gmd-16-5949-2023, 2023.

Gaspari, G., and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723-757, 10.1256/smsqj.55416, 1999.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K., and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geoscientific Model Development, 5, 1471-1492, 10.5194/gmd-5-1471-2012, 2012.

Hong, C., Zhang, Q., He, K., Guan, D., Li, M., Liu, F., and Zheng, B.: Variations of China's emission estimates: response to uncertainties in energy statistics, Atmos. Chem. Phys., 17, 1227-1239, 10.5194/acp-17-1227-2017, 2017.

Hong, Q., Liu, C., Hu, Q., Zhang, Y., Xing, C., Su, W., Ji, X., and Xiao, S.: Evaluating the feasibility of formaldehyde derived from hyperspectral remote sensing as a proxy for volatile organic compounds, Atmospheric Research, 264, 105777, 10.1016/j.atmosres.2021.105777, 2021.

Houtekamer, P. L., and Zhang, F.: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 144, 4489-4532, 10.1175/mwr-d-15-0440.1, 2016.

Jiang, X., Guenther, A., Potosnak, M., Geron, C., Seco, R., Karl, T., Kim, S., Gu, L., and Pallardy, S.: Isoprene emission response to drought and the impact on global atmospheric chemistry, Atmospheric Environment, 183, 69-83, 10.1016/j.atmosenv.2018.01.026, 2018.

Kaiser, J., Jacob, D. J., Zhu, L., Travis, K. R., Fisher, J. A., González Abad, G., Zhang, L., Zhang, X., Fried, A., Crounse, J. D., St. Clair, J. M., and Wisthaler, A.: High-resolution inversion of OMI formaldehyde columns to quantify isoprene emission on ecosystem-relevant scales: application to the southeast US, Atmos. Chem. Phys., 18, 5483-5497, 10.5194/acp-18-5483-2018, 2018.

Li, B., Ho, S. S. H., Li, X., Guo, L., Chen, A., Hu, L., Yang, Y., Chen, D., Lin, A., and Fang, X.: A comprehensive review on anthropogenic volatile organic compounds (VOCs) emission estimates in China: Comparison and outlook, Environment International, 156, 106710, 10.1016/j.envint.2021.106710, 2021.

Li, J., Chen, X., Wang, Z., Du, H., Yang, W., Sun, Y., Hu, B., Li, J., Wang, W., Wang, T., Fu, P., and Huang, H.: Radiative and heterogeneous chemical effects of aerosols on ozone and inorganic aerosols over East Asia, Science of The Total Environment, 622-623, 1327-1342, 10.1016/j.scitotenv.2017.12.041, 2018.

Li, J., Nagashima, T., Kong, L., Ge, B., Yamaji, K., Fu, J. S., Wang, X., Fan, Q., Itahashi, S., Lee, H. J., Kim, C. H., Lin, C. Y., Zhang, M., Tao, Z., Kajino, M., Liao, H., Li, M., Woo, J. H., Kurokawa, J., Wang, Z., Wu, Q., Akimoto, H., Carmichael, G. R., and Wang, Z.: Model evaluation and intercomparison of surface-level ozone and relevant species in East Asia in the

context of MICS-Asia Phase III – Part 1: Overview, Atmos. Chem. Phys., 19, 12993-13015, 10.5194/acp-19-12993-2019, 2019.

Li, K., Jacob, D. J., Shen, L., Lu, X., De Smedt, I., and Liao, H.: Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences, Atmos. Chem. Phys., 20, 11423-11433, 10.5194/acp-20-11423-2020, 2020.

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., and He, K.: Anthropogenic emission inventories in China: a review, National Science Review, 4, 834-866, 10.1093/nsr/nwx150, 2017a.

Li, M., Zhang, Q., Kurokawa, J.-i., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H., and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmospheric Chemistry And Physics, 17, 935-963, 10.5194/acp-17-935-2017, 2017b.

Liu, H., Liu, Z., and Lu, F.: A Systematic Comparison of Particle Filter and EnKF in Assimilating Time-Averaged Observations, Journal of Geophysical Research-Atmospheres, 122, 13155-13173, 10.1002/2017jd026798, 2017.

Liu, Z., Wang, Y., Vrekoussis, M., Richter, A., Wittrock, F., Burrows, J. P., Shao, M., Chang, C.-C., Liu, S.-C., Wang, H., and Chen, C.: Exploring the missing source of glyoxal (CHOCHO) over China, Geophysical Research Letters, 39, 10.1029/2012GL051645, 2012.

Marais, E. A., Jacob, D. J., Guenther, A., Chance, K., Kurosu, T. P., Murphy, J. G., Reeves, C. E., and Pye, H. O. T.: Improved model of isoprene emissions in Africa using Ozone Monitoring Instrument (OMI) satellite observations of formaldehyde: implications for oxidants and particulate matter, Atmos. Chem. Phys., 14, 7693-7703, 10.5194/acp-14-7693-2014, 2014.

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., and Polvani, L. M.: Climate Change from 1850 to 2005 Simulated in CESM1(WACCM), Journal of Climate, 26, 7372-7391, 10.1175/JCLI-D-12-00558.1, 2013.

Mo, Z., Huang, S., Yuan, B., Pei, C., Song, Q., Qi, J., Wang, M., Wang, B., Wang, C., Li, M., Zhang, Q., and Shao, M.: Deriving emission fluxes of volatile organic compounds from tower observation in the Pearl River Delta, China, Science of The Total Environment, 741, 139763, 10.1016/j.scitotenv.2020.139763, 2020.

Mu, M., Randerson, J. T., van der Werf, G. R., Giglio, L., Kasibhatla, P., Morton, D., Collatz, G. J., DeFries, R. S., Hyer, E. J., Prins, E. M., Griffith, D. W. T., Wunch, D., Toon, G. C., Sherlock, V., and Wennberg, P. O.: Daily and 3-hourly variability in global fire emissions and consequences for atmospheric model predictions of carbon monoxide, Journal of Geophysical Research-Atmospheres, 116, 10.1029/2011jd016245, 2011.

Opacka, B., Müller, J.-F., Stavrakou, T., Miralles, D. G., Koppa, A., Pagán, B. R., Potosnak, M. J., Seco, R., De Smedt, I., and Guenther, A. B.: Impact of Drought on Isoprene Fluxes Assessed Using Field Data, Satellite-Based GLEAM Soil Moisture and HCHO Observations from OMI, Remote Sensing, 14, 2021, 2022.

Palmer, P. I., Abbot, D. S., Fu, T.-M., Jacob, D. J., Chance, K., Kurosu, T. P., Guenther, A., Wiedinmyer, C., Stanton, J. C., Pilling, M. J., Pressley, S. N., Lamb, B., and Sumner, A. L.: Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column, Journal of Geophysical Research:

776       Atmospheres, 111, 10.1029/2005JD006689, 2006.

777     Paulson, S. E., and Orlando, J. J.: The reactions of ozone with alkenes: An important source of HOx
778        in the boundary layer, Geophysical Research Letters, 23, 3727-3730, 10.1029/96GL03477,
779        1996.

780     Pierce, T., Geron, C., Bender, L., Dennis, R., Tonnesen, G., and Guenther, A.: Influence of increased
781        isoprene emissions on regional ozone modeling, Journal of Geophysical Research:
782        Atmospheres, 103, 25611-25629, 10.1029/98JD01804, 1998.

783     Ren, J., Guo, F., and Xie, S.: Diagnosing ozone–NOx–VOC sensitivity and revealing causes of
784        ozone increases in China based on 2013–2021 satellite retrievals, Atmos. Chem. Phys., 22,
785        15035-15047, 10.5194/acp-22-15035-2022, 2022.

786     Seco, R., Holst, T., Davie-Martin, C. L., Simin, T., Guenther, A., Pirk, N., Rinne, J., and Rinnan, R.:
787        Strong isoprene emission response to temperature in tundra vegetation, Proceedings of the
788        National Academy of Sciences, 119, e2118014119, doi:10.1073/pnas.2118014119, 2022.

789     Skamarock, W. C., and Klemp, J. B.: A time-split nonhydrostatic atmospheric model for weather
790        research and forecasting applications, Journal Of Computational Physics, 227, 3465-3485,
791        10.1016/j.jcp.2007.01.037, 2008.

792     Souri, A. H., Nowlan, C. R., González Abad, G., Zhu, L., Blake, D. R., Fried, A., Weinheimer, A. J.,
793        Wisthaler, A., Woo, J. H., Zhang, Q., Chan Miller, C. E., Liu, X., and Chance, K.: An inversion
794        of NOx and non-methane volatile organic compound (NMVOC) emissions using satellite
795        observations during the KORUS-AQ campaign and implications for surface ozone over East
796        Asia, Atmos. Chem. Phys., 20, 9837-9854, 10.5194/acp-20-9837-2020, 2020.

797     Souri, A. H., Chance, K., Bak, J., Nowlan, C. R., González Abad, G., Jung, Y., Wong, D. C., Mao,
798        J., and Liu, X.: Unraveling pathways of elevated ozone induced by the 2020 lockdown in
799        Europe by an observationally constrained regional model using TROPOMI, Atmos. Chem.
800        Phys., 21, 18227-18245, 10.5194/acp-21-18227-2021, 2021.

801     Su, W., Liu, C., Chan, K. L., Hu, Q., Liu, H., Ji, X., Zhu, Y., Liu, T., Zhang, C., Chen, Y., and Liu,
802        J.: An improved TROPOMI tropospheric HCHO retrieval over China, Atmos. Meas. Tech., 13,
803        6271-6292, 10.5194/amt-13-6271-2020, 2020.

804     Tang, X., Zhu, J., Wang, Z. F., and Gbaguidi, A.: Improvement of ozone forecast over Beijing based
805        on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions,
806        Atmospheric Chemistry And Physics, 11, 12901-12916, 10.5194/acp-11-12901-2011, 2011.

807     van der Werf, G. R., Randerson, J. T., Giglio, L., van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu,
808        M., van Marle, M. J. E., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.:
809        Global fire emissions estimates during 1997-2016, Earth System Science Data, 9, 697-720,
810        10.5194/essd-9-697-2017, 2017.

811     Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., Cheng, Z., De Mazière, M.,
812        De Smedt, I., Grutter, M., Hannigan, J. W., Jones, N., Kivi, R., Loyola, D., Lutsch, E., Mahieu,
813        E., Makarova, M., Metzger, J. M., Morino, I., Murata, I., Nagahama, T., Notholt, J., Ortega, I.,
814        Palm, M., Pinardi, G., Röhling, A., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y.,
815        van Roozendael, M., Wang, P., and Winkler, H.: TROPOMI–Sentinel-5 Precursor
816        formaldehyde validation using an extensive network of ground-based Fourier-transform
817        infrared stations, Atmos. Meas. Tech., 13, 3751-3767, 10.5194/amt-13-3751-2020, 2020.

818     Wang, H., Lu, X., Seco, R., Stavrakou, T., Karl, T., Jiang, X., Gu, L., and Guenther, A. B.: Modeling

Isoprene Emission Response to Drought and Heatwaves Within MEGAN Using Evapotranspiration Data and by Coupling With the Community Land Model, Journal of Advances in Modeling Earth Systems, 14, e2022MS003174, 10.1029/2022MS003174, 2022.

Wang, H., Yan, R., Xu, T., Wang, Y., Wang, Q., Zhang, T., An, J., Huang, C., Gao, Y., Gao, Y., Li, X., Yu, C., Jing, S., Qiao, L., Lou, S., Tao, S., and Li, Y.: Observation Constrained Aromatic Emissions in Shanghai, China, Journal of Geophysical Research: Atmospheres, 125, e2019JD031815, 10.1029/2019JD031815, 2020.

Wang, H., Wu, Q., Guenther, A. B., Yang, X., Wang, L., Xiao, T., Li, J., Feng, J., Xu, Q., and Cheng, H.: A long-term estimation of biogenic volatile organic compound (BVOC) emission in China from 2001–2016: the roles of land cover change and climate variability, Atmos. Chem. Phys., 21, 4825-4848, 10.5194/acp-21-4825-2021, 2021a.

Wang, J., Yan, R., Wu, G., Liu, Y., Wang, M., Zeng, N., Jiang, F., Wang, H., He, W., Wu, M., Ju, W., and Chen, J. M.: Unprecedented decline in photosynthesis caused by summer 2022 record-breaking compound drought-heatwave over Yangtze River Basin, Science Bulletin, 68, 2160-2163, 10.1016/j.scib.2023.08.011, 2023.

Wang, N., Lyu, X., Deng, X., Huang, X., Jiang, F., and Ding, A.: Aggravating O3 pollution due to NOx emission control in eastern China, Science of The Total Environment, 677, 732-744, 10.1016/j.scitotenv.2019.04.388, 2019.

Wang, P., Liu, Y., Dai, J., Fu, X., Wang, X., Guenther, A., and Wang, T.: Isoprene Emissions Response to Drought and the Impacts on Ozone and SOA in China, Journal of Geophysical Research: Atmospheres, 126, e2020JD033263, 10.1029/2020JD033263, 2021b.

Wang, W., van der A, R., Ding, J., van Weele, M., and Cheng, T.: Spatial and temporal changes of the ozone sensitivity in China based on satellite and ground-based observations, Atmos. Chem. Phys., 21, 7253-7269, 10.5194/acp-21-7253-2021, 2021c.

Warneke, C., de Gouw, J. A., Del Negro, L., Brioude, J., McKeen, S., Stark, H., Kuster, W. C., Goldan, P. D., Trainer, M., Fehsenfeld, F. C., Wiedinmyer, C., Guenther, A. B., Hansel, A., Wisthaler, A., Atlas, E., Holloway, J. S., Ryerson, T. B., Peischl, J., Huey, L. G., and Hanks, A. T. C.: Biogenic emission measurement and inventories determination of biogenic emissions in the eastern United States and Texas and comparison with biogenic emission inventories, Journal of Geophysical Research: Atmospheres, 115, 10.1029/2009JD012445, 2010.

Whitaker, J. S., and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Monthly Weather Review, 130, 1913-1924, 10.1175/1520-0493(2002)130<1913:Edawpo>2.0.Co;2, 2002.

Wolfe, G. M., Kaiser, J., Hanisco, T. F., Keutsch, F. N., de Gouw, J. A., Gilman, J. B., Graus, M., Hatch, C. D., Holloway, J., Horowitz, L. W., Lee, B. H., Lerner, B. M., Lopez-Hilifiker, F., Mao, J., Marvin, M. R., Peischl, J., Pollack, I. B., Roberts, J. M., Ryerson, T. B., Thornton, J. A., Veres, P. R., and Warneke, C.: Formaldehyde production from isoprene oxidation across NOx regimes, Atmos. Chem. Phys., 16, 2597-2610, 10.5194/acp-16-2597-2016, 2016.

Yuan, B., Kaser, L., Karl, T., Graus, M., Peischl, J., Campos, T. L., Shertz, S., Apel, E. C., Hornbrook, R. S., Hills, A., Gilman, J. B., Lerner, B. M., Warneke, C., Flocke, F. M., Ryerson, T. B., Guenther, A. B., and de Gouw, J. A.: Airborne flux measurements of methane and volatile organic compounds over the Haynesville and Marcellus shale gas production regions, Journal of Geophysical Research: Atmospheres, 120, 6271-6289, 10.1002/2015JD023242, 2015.

Zhang, M., Zhao, C., Yang, Y., Du, Q., Shen, Y., Lin, S., Gu, D., Su, W., and Liu, C.: Modeling sensitivities of BVOCs to different versions of MEGAN emission schemes in WRF-Chem (v3.6) and its impacts over eastern China, Geosci. Model Dev., 14, 6155-6175, 10.5194/gmd-14-6155-2021, 2021.

Zhang, Q., Streets, D. G., Carmichael, G. R., He, K. B., Huo, H., Kannari, A., Klimont, Z., Park, I. S., Reddy, S., Fu, J. S., Chen, D., Duan, L., Lei, Y., Wang, L. T., and Yao, Z. L.: Asian emissions in 2006 for the NASA INTEX-B mission, Atmos. Chem. Phys., 9, 5131-5153, 10.5194/acp-9-5131-2009, 2009.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., Yan, L., Zhang, Y., Zhao, H., Zheng, Y., He, K., and Zhang, Q.: Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions, Atmospheric Chemistry And Physics, 18, 14095-14111, 10.5194/acp-18-14095-2018, 2018.

Zhou, B., Guo, H., Zeren, Y., Wang, Y., Lyu, X., Wang, B., and Wang, H.: An Observational Constraint of VOC Emissions for Air Quality Modeling Study in the Pearl River Delta Region, Journal of Geophysical Research: Atmospheres, 128, e2022JD038122, 10.1029/2022JD038122, 2023.