https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

NMVOC emission optimization in China through assimilating formaldehyde retrievals from multiple satellite products

Canjie Xu[1], Jianbing Jin*[1], Ke Li[1], Yinfei Qi[2], Ji Xia[1], Hai Xiang Lin[3,4], and Hong Liao[1]

[1]Joint International Research Laboratory of Climate and Environment Change, Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, School of Environmental Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China
[2]College of Geography and Remote Sensing, Hohai University, Nanjing, Jiangsu, China
[3]Institute of Environmental Sciences, Leiden University, Leiden, The Netherlands
[4]Delft Institute of Applied Mathematics, Delft University of Technology, Delft, the Netherlands

Correspondence: Jianbing Jin (jianbing.jin@nuist.edu.cn)

Abstract. Non-methane volatile organic compounds (NMVOCs) serve as key precursors to ozone and secondary organic aerosols. Given that China is a major source of NMVOCs, the emission inventory is crucial for understanding and controlling atmospheric pollution. Mainstream inventories are constructed using bottom-up approaches, which cannot accurately reflect the spatiotemporal characteristics of NMVOCs, resulting in poor model outcomes. This study performed monthly optimization of NMVOC emissions in China by assimilating formaldehyde retrievals from the latest satellite products. A semi-variogram spatial analysis is conducted before assimilation, highlighting the advantages of using Tropospheric Monitoring Instrument (TROPOMI) and Ozone Mapping and Profiler Suite (OMPS) formaldehyde products for estimating high-resolution NMVOCs compared to Ozone Monitoring Instrument (OMI) retrievals. The emission optimization is performed based on a self-developed 4DEnVar-based system. A positive increment of NMVOC emissions was obtained by assimilating OMPS formaldehyde, with annual anthropogenic emissions rising from 22.40 to 41.32 Tg, biogenic emissions increasing from 16.56 to 28.01 Tg, and biomass burning emissions rising from 0.29 to 0.65 Tg. Our model simulations, driven by the posterior inventories, demonstrate superior performance compared to the prior. This is validated through comparisons against the independent satellite measurements and the surface ozone measurements. The RMSE of the posterior formaldehyde columns decreased from 0.49 to 0.45 $\times 10^{16}$ molec/cm^2 nationwide. In the severe-polluted NCP, it was improved effectively, reaching levels comparable to TROPOMI, with the RMSE dropping from 0.52 to 0.37 $\times 10^{16}$ molec/cm^2. Validation using surface ozone observations also yielded favorable results, especially in NCP.

1 Introduction

Non-methane volatile organic compounds (NMVOCs) are significant components of the atmosphere, serving as key precursors to ozone (O_3) and secondary organic aerosols (SOA) (Liu et al., 2017). They engage in numerous photochemical reactions, exerting a considerable influence on atmospheric oxidative capacity and air quality (Zhu et al., 2021). Moreover, NMVOCs such as benzene, trichloroethylene, and chloroform are recognized for their toxicity (Billionnet et al., 2011; Lerner et al.,

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

2012), and prolonged exposure to elevated concentrations can pose significant health risks (He et al., 2015). China has seen a rapid anthropogenic NMVOC emissions increase over the last three decades (Li et al., 2019), became the major source region globally. Investigating NMVOC dynamics and their emission distributions is critical for addressing air pollution challenges in China (Yuan et al., 2013; Hao and Xie, 2018).

NMVOCs are primarily released through anthropogenic activities, biogenic emissions and biomass burning processes. Huge efforts have been devoted to constructing inventories recording these emissions in a bottom-up way, such as the global Community Emission Data System (CEDS) (Hao and Xie, 2018), the regional Multi-resolution Emission Inventory for China (MEIC) (Li et al., 2019) and the Model of Emissions of Gases and Aerosols from Nature v2.1 (MEGAN) (Guenther et al., 2012). These NMVOC emission inventories coupled with chemical transport models like GEOS-Chem (Ito et al., 2007) and WRF-Chem (Azmi et al., 2022), are capable of reproducing the complex processes including transport, deposition, and chemical reactions. This not only helps to better quantify the environmental impact of NMVOCs, but also provides essential tools for predicting future trends and making emission reduction strategies. However, the NMVOC emission factors required in the bottom-up method have large temporal and spatial variations, and this information is usually not widely available (Bo et al., 2008; Sharma et al., 2015). Additionally, due to the implementation of ever-stricter control measures targeting major industries, residential life, and transportation sectors in recent years (Wu et al., 2016; Li et al., 2017; Zheng et al., 2018), NMVOC emission patterns in China have undergone significant changes. As a result, these bottom-up inventories inherently have considerable uncertainties (Li et al., 2014; Qiu et al., 2014). For example, bottom-up estimates of China's annual NMVOC emissions for 2012 range between 18 and 27 Tg as reported by different studies (Kurokawa et al., 2013; Wu et al., 2016; Stavrakou et al., 2017). This significant uncertainty poses challenges for accurately modeling and assessing the environmental impacts of NMVOCs in China (Han et al., 2013; Wang et al., 2014).

There are numerous well-established techniques for measuring the concentrations of various volatile organic compounds in the atmosphere. These include gas chromatography, mass spectrometry, Fourier transform infrared spectroscopy, and non-dispersive infrared analysis. While these methods are highly effective for meeting the requirements of experimental studies and real-time monitoring, their complexity and the associated high labor costs pose significant challenges for long-term measurements or assessments across large spatial scales. Among the various NMVOCs, the optical properties of formaldehyde (HCHO) and glyoxal (CHOCHO) make them particularly suitable for detection via remote sensing technologies. These properties enable formaldehyde and glyoxal to be among the few NMVOCs that can be monitored from satellites. Remote sensing observations of these compounds typically rely on spectral channels in the ultraviolet-visible (UV-Vis) range, with their primary absorption features occurring between 330 and 460 nm (Platt, 1979; Lerot et al., 2010; De Smedt et al., 2012).

Compared to glyoxal, satellite products for formaldehyde is more well-established. Satellite observations of glyoxal began later with initial identifications made using Aura Ozone Monitoring Instrument (OMI), and its retrieval is more challenging than that of formaldehyde (Kurosu et al., 2005; Chan Miller et al., 2014). Recent years have seen further advancements in satellite observational instruments and algorithms for formaldehyde (Abad, 2022; De Smedt et al., 2017), leading to new satellite observation products with improvements in both accuracy and resolution. In contrast, while the glyoxal retrieval algorithm has been updated for OMI product (Alvarado et al., 2014), its satellite products continue to face substantial uncertainties.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

Notably, the latest version of Sentinel-5 Precursor Tropospheric Monitoring Instrument (TROPOMI) glyoxal products is no longer publicly available on the official website.

Formaldehyde measurements from instruments such as the Global Ozone Monitoring Experiment-2 (GOME-2) (De Smedt et al., 2012), OMI (González Abad et al., 2015), Ozone Mapping and Profiler Suite (OMPS) (Li et al., 2015) and TROPOMI (De Smedt et al., 2018) have been used widely for estimating NMVOC emissions through data assimilation. The core of the methodology is to calculate the most likely NMVOC emissions given the formaldehyde observations and the prior information. For instance, Fu et al. (2007) used six years of continuous satellite measurements of formaldehyde columns from GOME (1996–2001) to improve regional emission estimates of reactive NMVOCs, including isoprene, olefins, formaldehyde, and xylene, for East Asia and South Asia. Similarly, formaldehyde data from the GOME-2A satellite were used to constrain NMVOC emissions in India for 2009 (Chaliyakunnel et al., 2019). Souri et al. (2020) used observations from OMPS satellites during the KORUS-AQ campaign to estimate NOx and NMVOC emissions in East Asia from May to June 2016. Kaiser et al. (2018) also utilized high-resolution formaldehyde retrieval data from OMI instrument to quantify isoprene emissions at the ecosystem scale in the southeastern United States during August–September 2013. Those promising results have demonstrated that formaldehyde measurements could be utilized to optimize the existing NMVOC emissions that were established in a bottom-up manner.

Studies focusing on top-down NMVOC emission optimization over China remain relatively limited in recent years. Shim et al. (2005) assimilated formaldehyde observations from the GOME using a global Bayesian inversion to constrain isoprene emissions. Although China was included within their East Asia region, the analysis lacked region-specific focus and did not provide detailed characterization of emission patterns over China. Furthermore, the coarse spatial resolution (4° × 5°) in that study limited the ability to resolve subregional emission features. Stavrakou et al. (2016) conducted a regional inversion in Eastern China using multi-year satellite formaldehyde data from GOME and OMI to constrain VOC emissions during the post-harvest burning period. Their indicated that the crop burning fluxes of VOCs in June exceed by a factor of 2 the combined emissions from other anthropogenic activities in NCP region from 2005 to 2012. Cao et al. (2018) conducted a relatively systematic satellite-based emission inversion study over China. They used a 4DVar method and assimilated OMI and GOME-2A formaldehyde products to estimate monthly NMVOC emissions over China in 2007, with a coarse spatial resolution of 4° × 5°. Considering the increasingly stringent air pollution control policies and significant interannual variability in NMVOC emissions in recent years, e.g., the biomass buring emission is now reduced to a relatively low level, there is an urgent need for high-resolution top-down emission optimization over China.

In terms of the assimilation observation sources, OMI product now remains as one of the most widely used formaldehyde products in related studies. However, the retrievals over China are reported to suffer from the so-called "row anomaly" issue (González Abad et al., 2015), which may reduce the assimilation accuracy especially in the high-resolution configuration. To address this issue, a semi-variogram spatial analysis is performed before assimilation, which highlights the added value of using TROPOMI and OMPS formaldehyde products for estimating high-resolution NMVOCs compared to OMI retrievals. Subsequently, the monthly NMVOC emission optimization in China is conducted. This is achieved by assimilating formaldehyde observations from OMPS and TROPOMI, based on the emission inversion system that couples the four-dimensional ensem-

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

ble variational (4DEnVar) data assimilation algorithm and GEOS-Chem model. The effectiveness of this emission inversion system has been evaluated in our recent study (Jin et al., 2023; Xia et al., 2025).

This paper is organized as follows: Section 2 describes the dataset and methodology, focusing on GEOS-Chem model, input emission sources (anthropogenic, biogenic, and biomass burning), and the satellite and ground-based observations utilized. Section 3 provides an analysis of the assimilation results, including the estimation of posterior NMVOC emissions and the validation of both formaldehyde columns and ground-level O_3 simulations. Section 4 summarizes the key findings and concludes the study.

2 Data and methods

This chapter begins by introducing GEOS-Chem model utilized for simulations in Section 2.1. Section 2.2 presents an overview of the input emission sources for the model, including anthropogenic sources, biogenic sources, and biomass burning inventories. Section 2.3 introduces the three satellite observations employed in the analysis in this study. In Section 2.4, the ground observations used for O_3 validation are presented. Section 2.5 outlines a semi-variogram algorithm for a preliminary assessment of satellite observations quality, while Section 2.6 introduces the 4D-Var algorithm used for data assimilation.

2.1 Model simulation

GEOS-Chem is a chemical transport model driven by meteorological data from Goddard Earth Observing System (GEOS) of NASA's Global Modeling and Assimilation Office (GMAO) (Bey et al., 2001). In this study, we use GEOS-Chem Classic (GCC) v14.1.1 to simulate NMVOC emissions over China, as shown in Figure 1. The global simulation has a horizontal resolution of 2° latitude × 2.5° longitude, with boundary conditions updated every 3 hours. The nested region (72°–136° E, 17.5°–54° N) has a horizontal resolution of 0.5° latitude × 0.625° longitude and 47 vertical layers. Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) meteorological fields (Gelaro et al., 2017) are used to drive GEOS-Chem. Each simulation includes a 6-month spin-up period.

This model version incorporates detailed O_3-HO_x-NO_x photochemistry and fully coupled aerosol-O_3-NO_x-VOCs chemistry representation (Park et al., 2004), coupled with a scheme for primary carbonaceous aerosols, dust, sea salt, and secondary inorganic species (sulfates, nitrates, and ammonium) and their distribution. To better simulate oxidant-aerosol reactions in the stratosphere and troposphere, GEOS-Chem v14.1.0 introduced a new KPP 3.0 chemical solver (Lin et al., 2023), including chemical mechanisms for isoprene, aromatics, ethylene, and acetylene (Bates and Jacob, 2019; Bates et al., 2021; Kwon et al., 2021).

2.2 Prior NMVOC emission inventries

The prior information and its uncertainty is the essential part in the NMVOC emission inversion as will be explained in Section 2.6. The anthropogenic NMVOC emission input into the model mainly comes from the Multi-resolution Emission Inventory for China (MEIC; Zheng et al., 2018). Since the MEIC inventory tailored for the existing chemical species in GEOS-Chem only

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

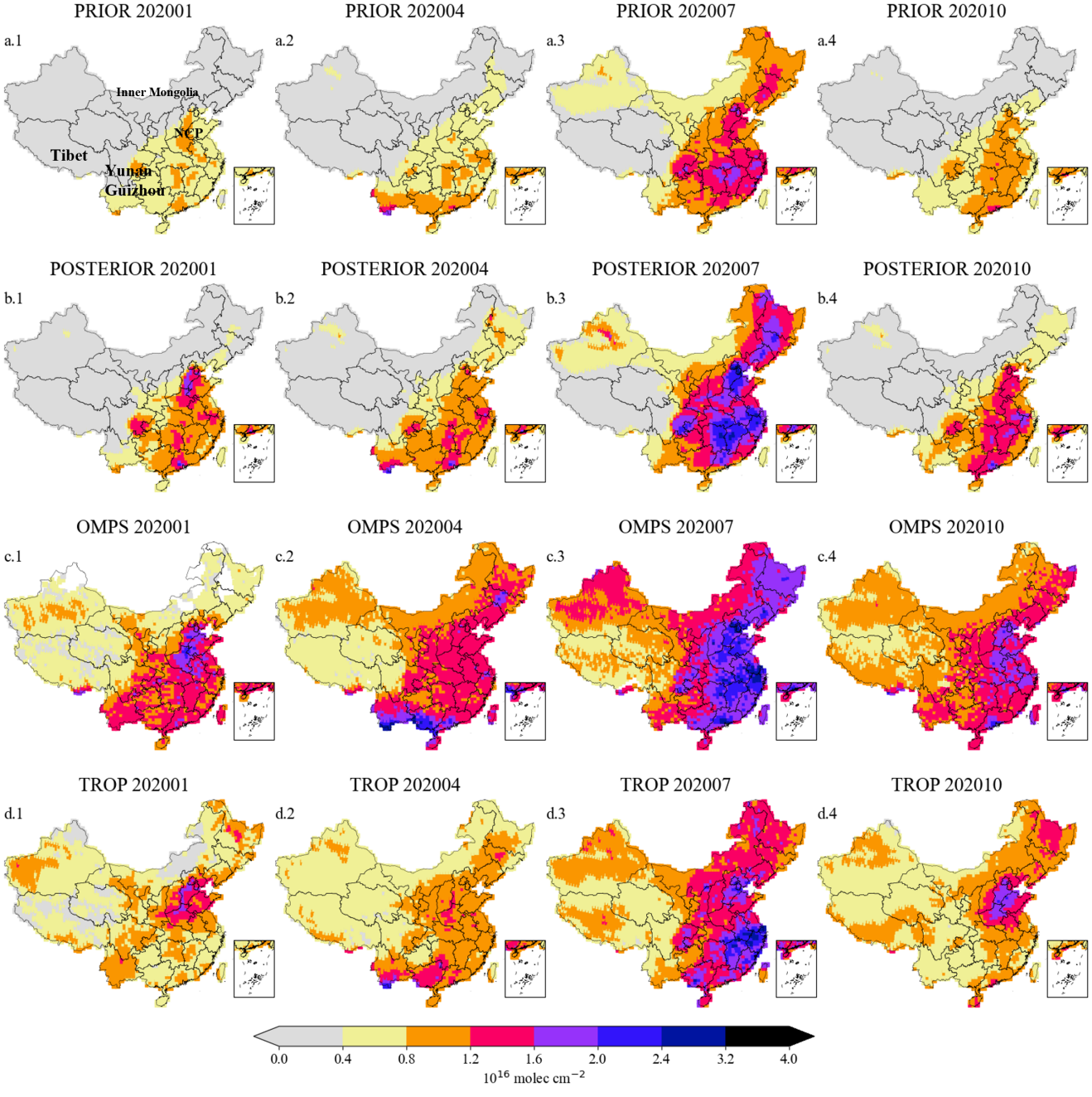

Figure 1. Spatial distributions of the total formaldehyde columns from GEOS-Chem model-simulated prior (a) and posterior (b) results, and from satellite observations by OMPS (c) and TROPOMI (d) in January (a.1)-(d.1), April(a.2)-(d.2), July (a.3)-(d.3), October 2020 (a.4)-(d.4).

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

extends to 2017, the MEIC inventory used in this study is the 2017 emission inventory. This inventory has a spatial resolution of 0.25° latitude × 0.25° longitude and includes industrial, transportation, power generation, and residential emissions. For chemical species used in GEOS-Chem but not included in MEIC and anthropogenic NMVOC emissions outside China, we use the CEDS global inventory as a supplement. The variations in NMVOC emissions mainly originate from biogenic sources. The prior estimates of biogenic NMVOC emissions in this study are obtained from the MEGAN 2.1 model (Guenther et al., 2012). Field straw burning is considered a major seasonal source of NMVOCs in China (Huang et al., 2012; Liu et al., 2015; Stavrakou et al., 2016), and the biomass burning emissions in this study come from the GFED4 global inventory (Van Der Werf et al., 2017). Before these prior emissions are used to drive GEOS-Chem simulations, the spatial resolution is coarsened to an average value on a 0.5° × 0.625° grid resolution consistent with the model configuration as used in Section 2.1.

Figure 2 (a) presents the prior NMVOC emission inventories for 2020, which primarily relies on the anthropogenic emission inventory from MEIC, supplemented by the CEDS inventory for species not included in MEIC. Additionally, the biogenic emission inventory from MEGAN (offline calculation) and the biomass burning inventory from GFED4 are incorporated. The optimization of NMVOC emissions through the assimilation of formaldehyde observations will be conducted using these combined prior inventories.

2.3 Formaldehyde Satellite measurements

2.3.1 NOAA-20 OMPS

Ozone Mapping and Profiler Suite (OMPS) was launched on Suomi National Polar-orbiting Partnership (SNPP) satellite on October 28, 2011, and on the JPSS-1 satellite (now known as NOAA-20) on November 18, 2017. OMPS/SNPP consists of three instruments: the nadir mapper (OMPS-NM), the profile mapper (OMPS-NP), and the limb profiler (OMPS-LP), while OMPS/NOAA-20 includes only the nadir package (OMPS-NM and OMPS-NP). This study uses OMPS-N20 Level 2 NM formaldehyde Total Column swath orbital Version 1 product (Abad, 2022). OMPS-NM is a hyperspectral nadir viewing spectrometer that measures backscattered light with a spectral resolution of approximately 1 nm. The NOAA-20 spectral measurement range is 300–420 nm. The instrument employs a 2-D CCD array detector in a pushbroom geometry, observing the two-dimensional field below the satellite's orbit over a swath width of about 2800 km. With 14 or 15 orbits per day, OMPS-NM provides daily global coverage of trace gas columns in the early afternoon local time, with an equatorial crossing time of approximately 13:30. The spatial resolution of OMPS/NOAA-20 was 17 km × 17 km until February 13, 2019, when it was changed to 12 km × 17 km (Flynn et al., 2014; Pan et al., 2017; Seftor et al., 2014).

In this study, the quality control scheme recommended in OMPS product documentation was applied when using OMPS data. Data points with formaldehyde column densities exceeding 2e17 molecules/cm^2 were excluded to minimize the impact of outliers. Furthermore, the geometric air mass factors were defined as follows:

$$AMF_G = sec(SZA) + sec(VZA) \quad (1)$$

Here, SZA represents the solar zenith angle and VZA denotes the viewing zenith angle. After removing outliers, we further filtered out data points where the product of formaldehyde columns and three times the observation uncertainty was less than

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

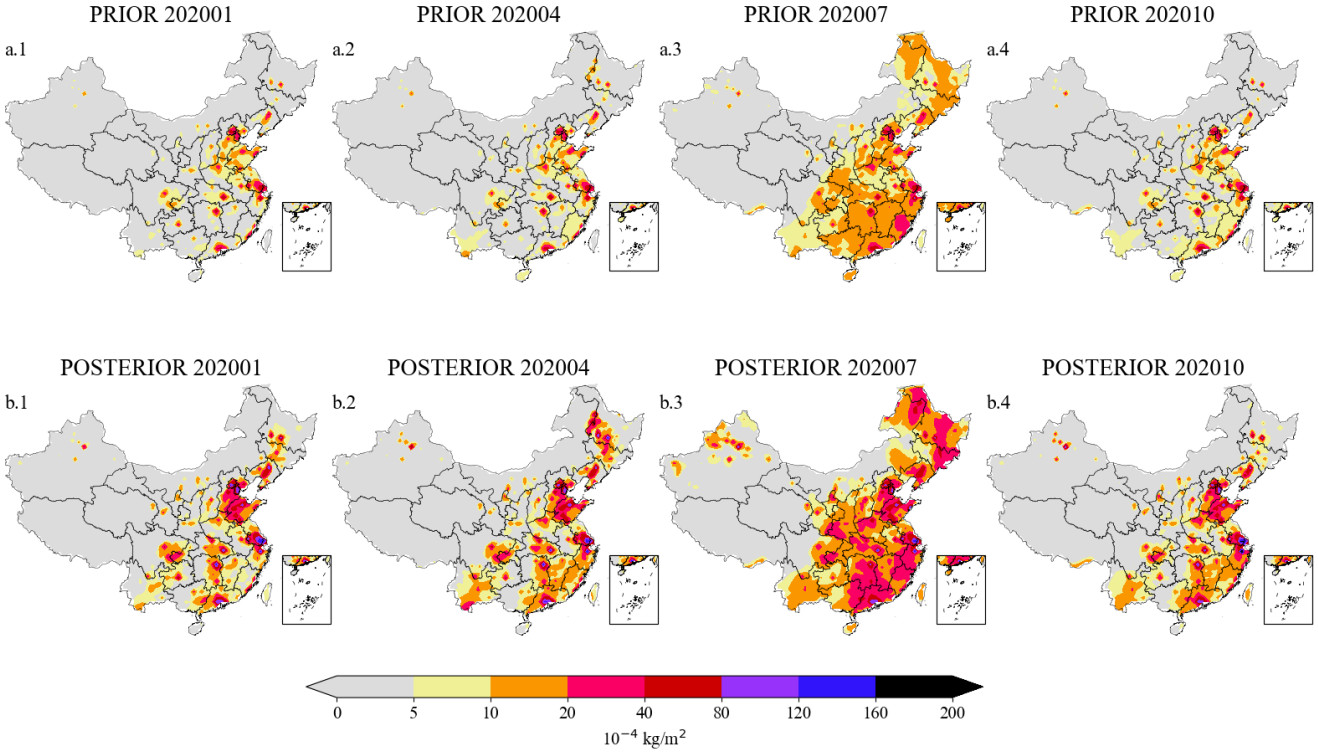

Figure 2. Spatial distributions of the total NMVOC emissions from the prior (a) and posterior (b) results in January (a.1, b.1), April (a.2, b.2), July (a.3, b.3), October (a.4, b.4).

zero. Subsequently, data points were excluded if SZA exceeded 70°, the air mass factor was less than 0.1, the geometric air mass factor exceeded 5, or the cloud fraction surpassed 0.4. Snapshots of filtered OMPS formaldehyde columns is shown in Figure 1(c).

2.3.2 Sentinel-5P TROPOMI

Sentinel-5 Precursor (Sentinel-5P) is a member of the European Space Agency's (ESA) Sentinel satellite series. It is in a low-Earth afternoon polar orbit with a swath of 2600 km, allowing for daily global coverage (Veefkind et al., 2012). Its sole payload is Tropospheric Monitoring Instrument (TROPOMI), a nadir-viewing, 108-degree field-of-view push-broom grating hyperspectral spectrometer. TROPOMI covers the ultraviolet-visible (UV-VIS, 270 nm to 495 nm), near-infrared (NIR, 675 nm to 775 nm), and shortwave infrared (SWIR, 2305 nm to 2385 nm) spectral ranges. Its Level 2 products include vertical columns of O_3, sulfur dioxide, nitrogen dioxide, formaldehyde, carbon monoxide, and methane, as well as O_3 profiles, aerosol

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

layer height, cloud information, and aerosol index. The initial spatial resolution was 3.5 km × 7 km, which was improved to 3.5 km × 5.5 km on August 6, 2019.

The retrieval algorithm for TROPOMI formaldehyde is based on the DOAS method and is directly inherited from OMI QA4ECV product retrieval algorithm (De Smedt et al., 2017). This study uses Sentinel-5P TROPOMI Level 2 Tropospheric formaldehyde Version 2 product (Copernicus Sentinel data processed by ESA, German Aerospace Center (DLR), 2020). Vigouroux et al. (2020) evaluated this TROPOMI formaldehyde product in 2020 using ground-based solar-absorption FTIR (Fourier-transform infrared) measurements, demonstrating the good quality. In 2021, De Smedt et al. further assessed TROPOMI formaldehyde using OMI observations and MAX-DOAS network column measurements, also showing favorable results. When using Level 2 TROPOMI formaldehyde data for the validation in this paper, we excluded only negative values and excessively large outliers to ensure data coverage. Examples of filtered TROPOMI formaldehyde columns are shown in Figure 1(d).

2.3.3 Aura OMI

Ozone Monitoring Instrument (OMI) is an important satellite instrument mounted on Aura satellite, launched on July 15, 2004. Its purpose is to monitor and study the composition of atmospheric gases, aerosols, and clouds to enhance our understanding of atmospheric chemistry and climate change. OMI provides daily global coverage with high spatial resolution (approximately 13 × 24 km), and its wide swath (2600 km) allows it to achieve global coverage in one day. It overpasses at 13:45 LT each day. OMI sensor contains three spectral channels: UV-1, UV-2, and VIS, covering the wavelength ranges of 264–311 nm, 307–383 nm, and 349–504 nm, respectively. These channels enable the observation of various trace gases, including O_3, nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and formaldehyde (Zhang et al., 2019).

OMI/Aura formaldehyde Total Column Daily L2 Global Version 3 product (Chance, 2014) is also used for the observation sources in this paper. The retrieval algorithm for this product is based on a nonlinear least-squares fitting technique, which calculates the slant column density (SCD). The SCD is then converted to vertical column density (VCD) using air mass factors (AMF). Since atmospheric formaldehyde is primarily concentrated in the troposphere, the total VCD can be regarded as the tropospheric VCD of formaldehyde (Duncan et al., 2010). In practical applications, data with a cloud fraction greater than or equal to 0.3 are further excluded.

2.4 O_3 ground station observation

This study aims to constrain the NMVOC emissions in China by assimilating multiple formaldehyde satellite products. As aforementioned, formaldehyde is an important precursor to O_3, the optimization of the NMVOC emission inventories and concentrations are supposed to improve the O_3 simulation simultaneously. To evaluate the magnitude and quality of this impact, the ground level O_3 concentrations from the National Urban Air Quality Real-time Publishing Platform of the China National Environmental Monitoring Center (CNEMC, last access: May 15, 2024) are used in the validation. The O_3 measurements utilized in this study are from 1,602 sites across China. The MDA8 values of surface O_3 observations are calculated based on the hourly data before they are compared against the model simulation. Results of the comparison will be described in Section 3.4.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

2.5 Semi-variogram algorithm

The semi-variogram function describes how the similarity between variables varies with changes in spatial distance and is widely applied in geospatial interpolation and spatial autocorrelation analysis. Samples of both remotely sensed data (satellite or airborne multispectral scanning systems, digitized aerial photography, etc.) and ground data (biomass, soil moisture, etc.) can be used to construct semi-variograms in remote sensing research (Curran, 1988). When a variable is distributed in space, it is called a regionalized variable, and the semi-variance function is the mathematical expectation of the square of the increment between regionalized variable values $Z(x_i)$ and $Z(x_i+h)$, representing the variance of increments of regionalized variables (variograms). $\gamma(h)$ is the semi-variogram, representing the semi-variance value at distance h. Its calculation formula is:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left(Z(x_i) - Z(x_i+h)\right)^2 \tag{2}$$

Where $N(h)$ is the number of sample pairs at distance h, $Z(x_i)$ is the variable value at location x_i, $Z(x_i+h)$ is the variable value at location x_i+h. A larger semi-variance value indicates greater variability between variable values at two points, suggesting lower spatial correlation. A smaller semi-variance value indicates greater similarity between variable values at two points, suggesting higher spatial correlation.

In previous studies, Souri et al. combined the semi-variogram function to investigate the spatial variability of OMI satellite observations. They subsequently used this method to compare OMI data with TROPOMI data, demonstrating the superiority of TROPOMI data (Johnson et al., 2022). In this study, we calculated the semi-variance values of formaldehyde simulations from GEOS-Chem, and formaldehyde observations from TROPOMI, OMPS, and OMI at different spatial resolutions (ranging from 0.5° × 0.5°, 1° × 1°, 2° × 2°, and 4° × 4°). With these, analysis would be carried out in Section 3.1 to evaluate whether these several simulations and observations accurately reflect the spatial variability of the atmospheric formaldehyde.

2.6 Assimilation algorithm

This study employs the four-dimensional ensemble variational (4DEnVar) methodology to assimilate formaldehyde observations to constrain NMVOC emissions. The goal of this assimilation is to find the most likely estimate of the state vector, which is the monthly NMVOC emission inventories $\boldsymbol{f}$ over the entire model domain. The prior estimate $\boldsymbol{f}_b$ is from the inventories described in Section 2.2, and the formaldehyde concentration observations $\boldsymbol{y}$ are described in Section 2.3. Mathematically, assimilation is performed via minimizing the cost function J as follows:

$$\mathcal{J}(\boldsymbol{f}) = \frac{1}{2}(\boldsymbol{f}-\boldsymbol{f}_b)^T \mathbf{B}^{-1}(\boldsymbol{f}-\boldsymbol{f}_b) + \frac{1}{2}\{\boldsymbol{y}-\mathcal{HM}(\boldsymbol{f})\}^T \mathbf{O}^{-1}\{\boldsymbol{y}-\mathcal{HM}(\boldsymbol{f})\} \tag{3}$$

The cost function $\mathcal{J}$ is sum of two part: background and observation penal term. The background term quantifies the difference between the optimal $\boldsymbol{f}$ and the prior emission inventories $\boldsymbol{f}_b$, while the observation term calculates the difference between the simulation driven by $\boldsymbol{f}$ and the satellite observations $\boldsymbol{y}$. In addition to the $\boldsymbol{f}_b$ that represents the prior NMVOC emission

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

vector calculated from the anthropogenic, biogenic, and biomass burning sources as been illustrated in Section 2.2. The uncertainty in the NMVOCs simulation is assumed to be attributed to errors in the emission inventories, and can be compensated using a spatially varying tuning factor α:

$$f(i) = f_b(i) \cdot \alpha(i) \tag{4}$$

in here $f_b(i)$ denotes the NMVOC emission rate in the given grid cell i. The α values are defined to be random variables with a mean of 1.0 and a standard deviation σ_α = 0.2. This empirical value was found to provide sufficient spaces for resolving the observation-minus-simulation errors. A background covariance $\mathbf{B}_\alpha$ is formulated as a product of the constant standard deviation and a spatial correlation matrix $\mathbf{C}$:

$$\mathbf{B}_\alpha(i,j) = \sigma_\alpha \cdot \mathbf{C}(i,j) \tag{5}$$

where $\mathbf{C}(i,j)$ represents a distance-based spatial correlation between two α in the grid cell i and j, and is defined as:

$$\mathbf{C}(i,j) = e^{-(d_{i,j}/l)^2/2} \tag{6}$$

where $d_{i,j}$ represents the distance between two grid cells i and j. l here denotes the correlation length scale which controls the spatially variability freedom of the α. A small l means more errors in fine scale could be resolved using the assimilation, while however requires more ensemble runs to represent the model realization from emission to simulation. An empirical parameterl = 300 km which is used in Jin et al. (2023) to nudge the ammonia emission that has a rapid spatially variability is also taken in this study. With the covariance matrix $\mathbf{B}_\alpha$, the NMVOC emission background covariance $\mathbf{B}$ is obtained via a Schur Product:

$$\mathbf{B} = \mathbf{B}_\alpha \circ \mathbf{C} \tag{7}$$

In the observation term, $\boldsymbol{y}$ is the observation vector, representing satellite observations, $\mathcal{M}$ is GEOS-Chem model driven by emissions $\boldsymbol{f}$, $\mathcal{H}$ is the observation operator that transfer the three dimensional concentration into the observational space, and $\mathbf{O}$ is the observation covariance matrix. The satellite formaldehyde observations are assumed to be independent, therefore $\mathbf{O}$ is a diagonal matrix. The diagonal value here is calculated as:

$$\sigma_{\text{total}} = \sqrt{\sigma^2_{\text{instrument}} + \sigma^2_{\text{represent}}} \tag{8}$$

In the Equation 8, σ_{total} is defined as the total uncertainty, which is the square root of the sum of the squares of the instrument uncertainty $\sigma_{\text{instrument}}$ from the formaldehyde observations and the representative uncertainty $\sigma_{\text{represent}}$ introduced when

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

processing the data into monthly averages. The representative uncertainty $\sigma_{\text{represent}}$ is represented by the standard deviation of the data.

With the assimilation-based emission inversion system above, we conducted three sets of experiments to explore the benefits to emission optimization. These experiments involved assimilating OMPS data and validating the assimilation results using TROPOMI, assimilating TROPOMI data, and finally assimilating the combined OMPS and TROPOMI data by averaging them. In the subsequent results, we primarily analyze the results of the first set of experiments, while the detailed inventories of the latter two experiments are archived in the supplementary material.

3 Results and discussion

Section 3.1 performs the semi-variogram analysis either based on GEOS-Chem simulation, or based on the formaldehyde measurements data from OMPS, TROPOMI and OMI products at different horizontal resolutions. Section 3.2 discusses the posterior NMVOC emissions through assimilating OMPS formaldehyde observations. In Section 3.3, the posterior formaldehyde column driven by the posterior NMVOC emission inventories, are validated. Finally, Section 3.4 presents an independent validation of the posterior ground-level O_3 simulations using ground-based O_3 observations.

3.1 Semi-variogram analysis

The original formaldehyde column data products from multiple sources are available at varying horizontal resolutions: OMI data at 0.25° × 0.25°, TROPOMI data at 5.5 km × 3.5 km, OMPS data at 17 km × 12 km, and GEOS-Chem prior simulation at 0.5° × 0.625°. To ensure a fair comparison during the semi-variogram analysis, linear interpolation is applied to these four formaldehyde datasets to generate averages at resolutions of 0.5° × 0.5°, 1° × 1°, 2° × 2°, and 4° × 4°, as shown in Figure 3.

As shown in the four interpolated results in Figure 3, the spatial distribution of high formaldehyde values is consistently captured across different horizontal resolutions, either by the satellite observations in Figure 3(a, b, c) or by GEOS-Chem simulation in panel (d). These hot spots are particularly prominent in the North China Plain (NCP) and Jiangsu-Zhejiang-Shanghai regions. However, at the higher resolution of 0.5° × 0.5°, OMI formaldehyde data exhibits noticeable noise all over China and lacks the spatial continuity observed in TROPOMI, OMPS, and GEOS-Chem datasets. The significant spatial variability in the NMVOC emission field might account for the discontinuity observed in OMI formaldehyde data. However, this discontinuity contradicts the model simulation and the other two satellite products obtained from the more advanced instruments. Moreover, such discontinuities are not observed in OMI formaldehyde retrievals over the United States, where (Kaiser et al., 2018) demonstrated continuous and high-quality data. Therefore, the discrepancies observed in China may be attributed to uncertain input parameters, such as aerosols and surface albedo. OMI formaldehyde retrievals with larger spatial grid intervals (2° × 2° and 4° × 4°) exhibit increased continuity and smoothness, as shown in Figure 3(a). This improvement is attributed to spatial averaging, which effectively filters out white noise (Lee, 1980).

The differences in spatial variability among OMI, OMPS, TROPOMI, and GEOS-Chem simulations are clearly illustrated in the semi-variogram plot shown in Figure 4. In this plot, smaller semi-variance values indicate stronger spatial autocorrelation

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

and smaller random noise. We define the area below the yellow line in Figure 4 as the "stable zone" of the semi-variance plot, which represents regions where the data effectively capture the true spatial dynamics of atmospheric formaldehyde. The semi-variance plots demonstrate that the semi-variance function curves of GEOS-Chem, TROPOMI, and OMPS formaldehyde column data consistently fall within the stable zone across all spatial grids. In contrast, the semi-variance function curve of OMI formaldehyde data only enters the stable zone at coarser spatial resolutions of $2° \times 2°$ or even $4° \times 4°$. This observation suggests that at these coarser spatial resolutions, OMI formaldehyde data for China achieves sufficient continuity and reliability. This finding is consistent with study in Cao et al. (2018), which constrained NMVOC emissions in China at a coarse resolution of $4° \times 5°$.

Overall, at higher spatial resolutions, OMI formaldehyde data exhibit more white noise, with higher semi-variance values and weaker spatial autocorrelation. In contrast, the semi-variance values of OMPS and TROPOMI align more closely with the trends observed in GEOS-Chem, making them more suitable for high-resolution (<1°) emission optimization.

3.2 NMVOC emissions

NOAA-20 OMPS formaldehyde observations were assimilated, resulting in updated monthly NMVOC emission inventories in 2020. To better illustrate the seasonal pattern, the prior and posterior total NMVOC emission inventories for four months (January, April, July, and October) are presented in Figure 2. These months were selected as typical examples representing four different seasons. To better illustrate the increments that the assimilation made in different source, the prior and the posterior NMVOC emissions from anthropogenic activities, biogenic plus biomass burning are separately plotted in Figure 6.

The prior annual NMVOC emissions and the increments obtained in the assimilation are plotted in Figure 5. In general, our top-down posterior emission estimates show that anthropogenic emissions increased from 22.40 Tg to 41.32 Tg, biogenic emissions rose from 16.56 Tg to 28.01 Tg, and biomass burning emissions increased from 0.29 Tg to 0.65 Tg. The increments in emissions from the prior to the posterior estimates are primarily observed around the hot spots identified in the prior estimates, with significant concentration increases in the North China Plain and the Yangtze River Delta (YRD), as well as notable increases in smaller areas in Chongqing, Hubei, Hunan, Guangdong, and the northeastern provinces.

As shown in Figure 6 (c.2, d.4), the distribution patterns of biogenic NMVOCs in spring and autumn are quite similar, predominantly concentrated in southern China due to the dense vegetation cover in that region. Isoprene and terpenes are the primary biogenic NMVOC emissions, resulting in higher emissions in April and October. Conversely, in January during winter, the dieback of vegetation and snow cover leads to a sharp decrease in leaf area index (LAI), significantly reducing biogenic emissions (Wu et al., 2020). The inter-monthly timing profile of the anthropogenic NMVOC emissions shows little variation as can be seen in Figure 6 (a). In spring, autumn and winter, anthropogenic emissions are generally higher than biogenic emissions while in summer biogenic sources are dominant. In January, April, and October, the posterior estimates indicate that changes in total NMVOC emissions, constrained primarily by anthropogenic sources, are most prominent. Notably, in April and October, the southeastern coastal areas and Yunnan Province exhibit significantly elevated emission levels due to higher vegetation cover, with emissions in these regions approximately 10 to 20 $\times 10^{-4} kg/m^2$ higher in January.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

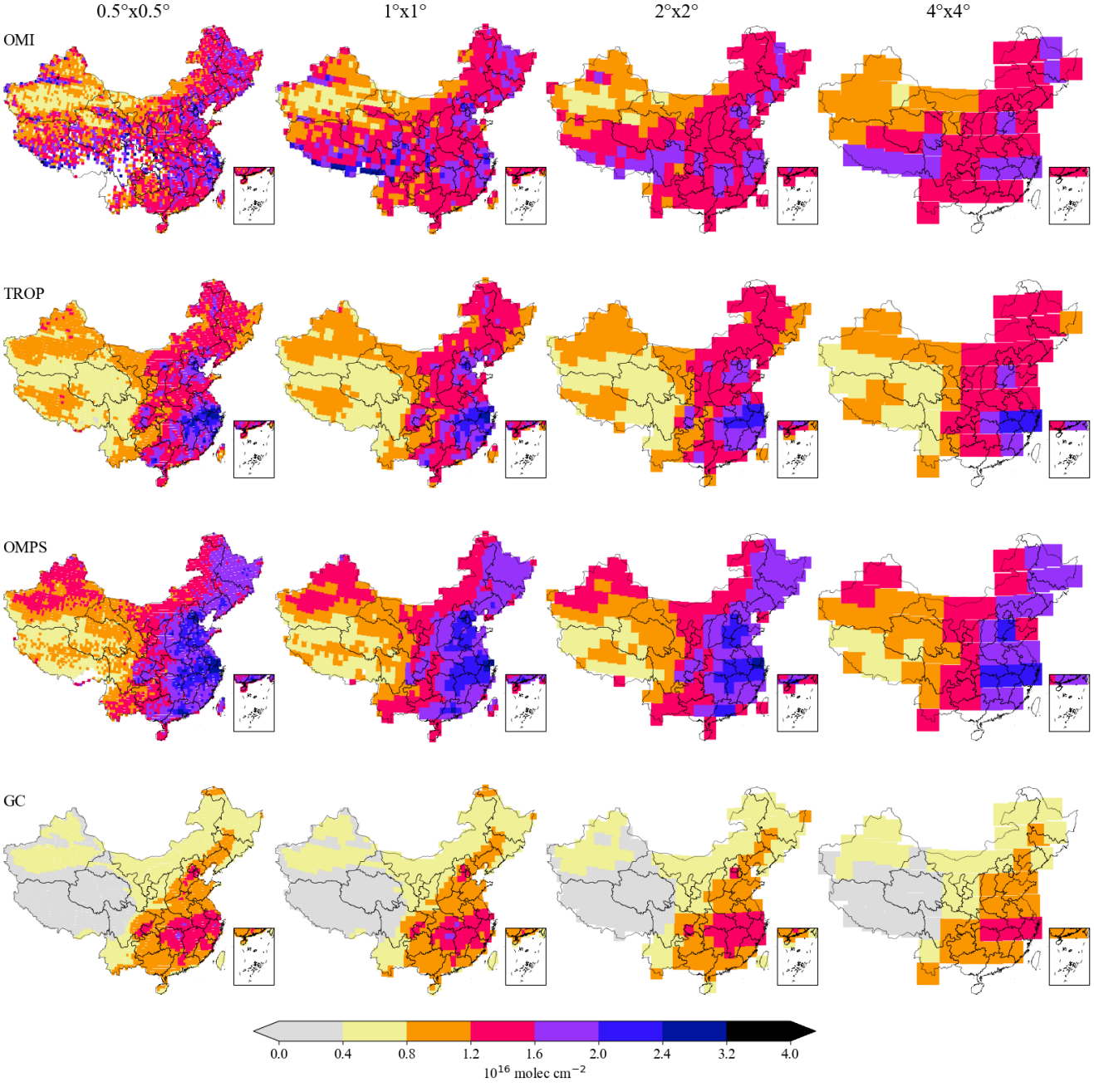

Figure 3. Spatial distributions of total formaldehyde columns from four different sources (OMI, TROPOMI, OMPS, GEOS-Chem) across China, displayed at four different grid resolutions (0.5°x0.5°, 1°x1°, 2°x2°, 4°x4°).

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

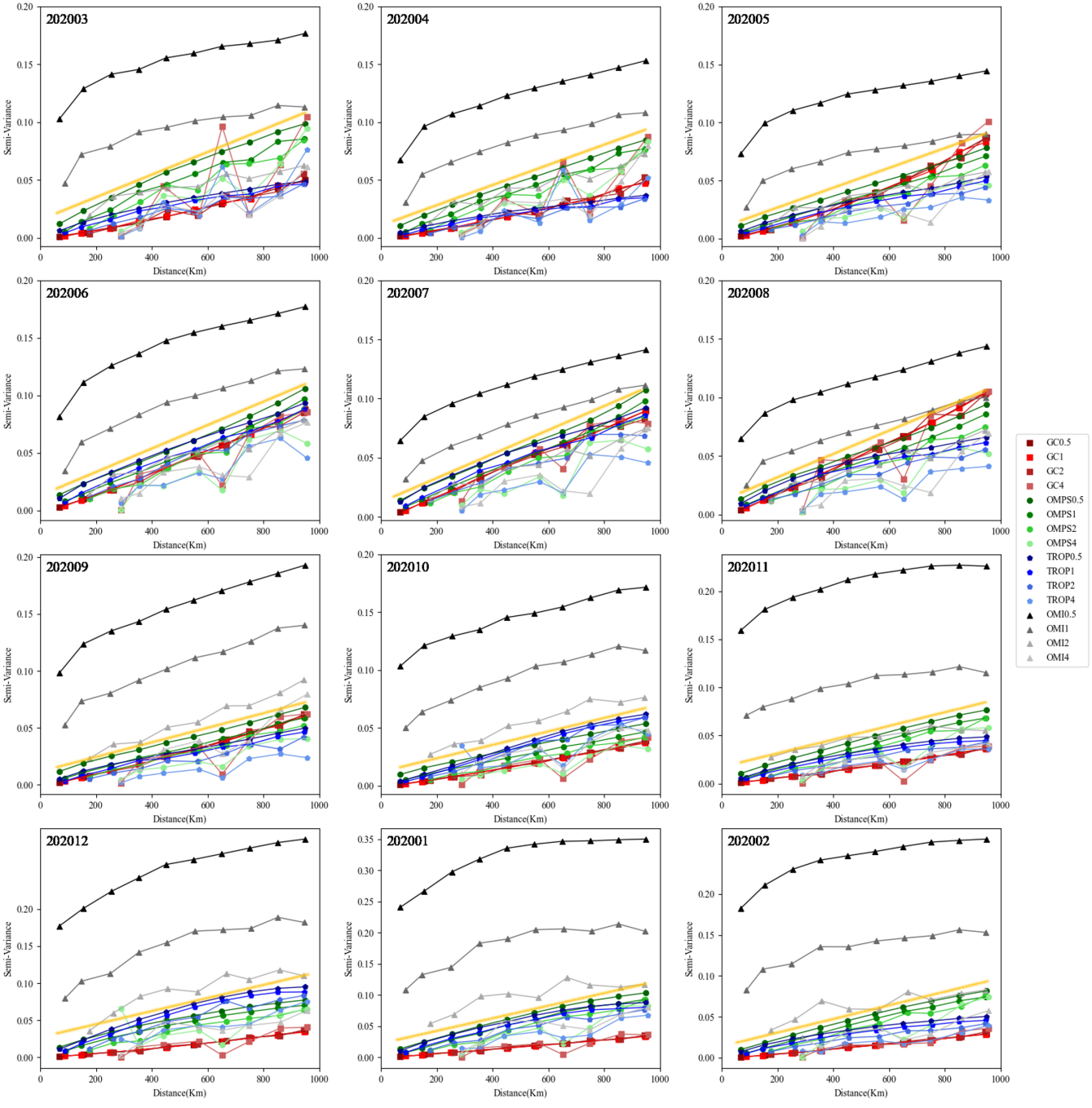

Figure 4. Semi-variogram analysis curves for the four data sources (OMI, TROPOMI, OMPS, GEOS-Chem) in 2020. We refer to the region in the semi-variogram with relatively small changes in semi-variance values as the stable zone (as indicated by the area below the yellow line in the figure).

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

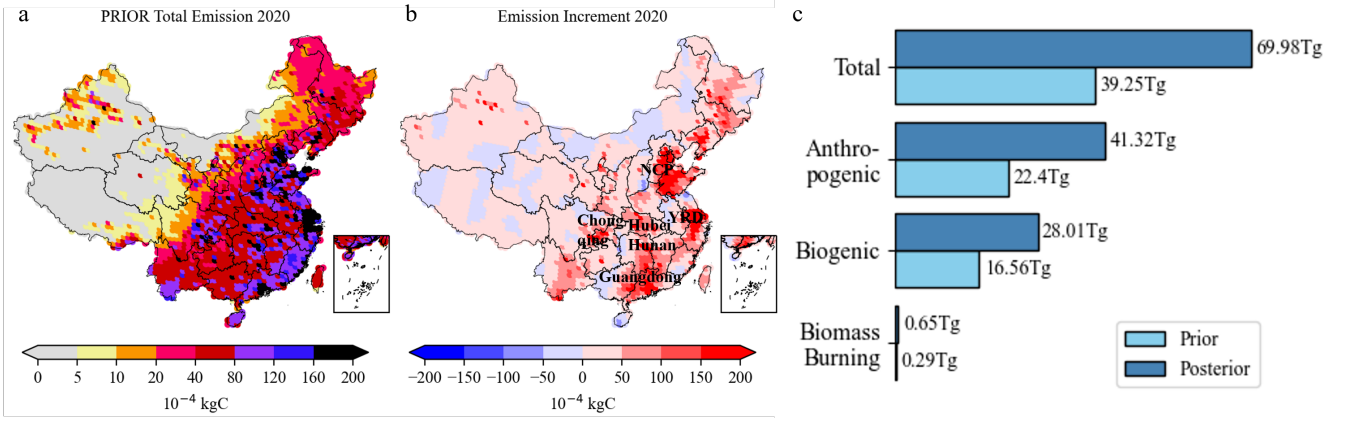

Figure 5. Spatial distributions of the total annual prior NMVOC emissions (a) and the increments obtained in the assimilation (b). Panel (c) compares the prior and posterior annual NMVOC emissions, including total emissions, anthropogenic, biogenic, and biomass burning.

In contrast to the other three seasons, summer features high temperatures, intense radiation, and vigorous vegetation growth, which greatly increases biogenic emissions of isoprene and terpenes. Therefore, as shown in Figure 6 (c.3, d.3), biogenic NMVOC emissions in China peak in July (Wu et al., 2020), with a significant expansion in the area covered by high values compared to other months. The total NMVOC emissions in eastern China in July increase from approximately 5 $\times 10^{-4}$ - 20 $\times 10^{-4} kg/m^2$ in the prior estimates to about 20 $\times 10^{-4}$ - 80 $\times 10^{-4} kg/m^2$ in the posterior estimates.

3.3 Formaldehyde total columns evaluation

As shown in Figure 1 (a) and (b), GEOS-Chem simulated the prior and posterior estimates of formaldehyde for four months of the year 2020 over China. In Figure 1 (a), the prior results exhibit a spatial distribution similar to satellite observations. When compared to OMPS and TROPOMI satellite observations, the prior results accurately reproduce high-value features in most regions, including Yunnan-Guizhou, Guangxi-Guangdong, NCP, the southeastern coast, and the northeast. However, the previous simulation did not accurately represent the actual formaldehyde levels. Specifically, it underestimated formaldehyde concentrations to varying degrees across different regions. By assimilating OMPS formaldehyde columns, improvement was obtained steadily in the posterior simulations. Nationwide, the posterior formaldehyde columns were raised by approximately 50%. Comparing to TROPOMI data used as independent measurements, the formaldehyde levels in the NCP region were raised from less than 1.2 $\times 10^{16}$ molec/cm^2 to around 2.4 $\times 10^{16}$ molec/cm^2 in January, closer to the observed values either from OMPS and TROPOMI. The high-value features in Yunnan-Guizhou became more prominent in April, and significant improvements were also observed in the southeastern coast, NCP, and the northeast in July and October. However, changes in Inner Mongolia, Tibet, and the northwest were minimal. This is because the assigned background uncertainty, which is proportional to the prior emission intensity in these regions, was particularly low, leaving limited flexibility for adjustments in the assimilation. This can be best seen in Figure 2, these areas are depicted in gray on the map, with annual total emissions

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

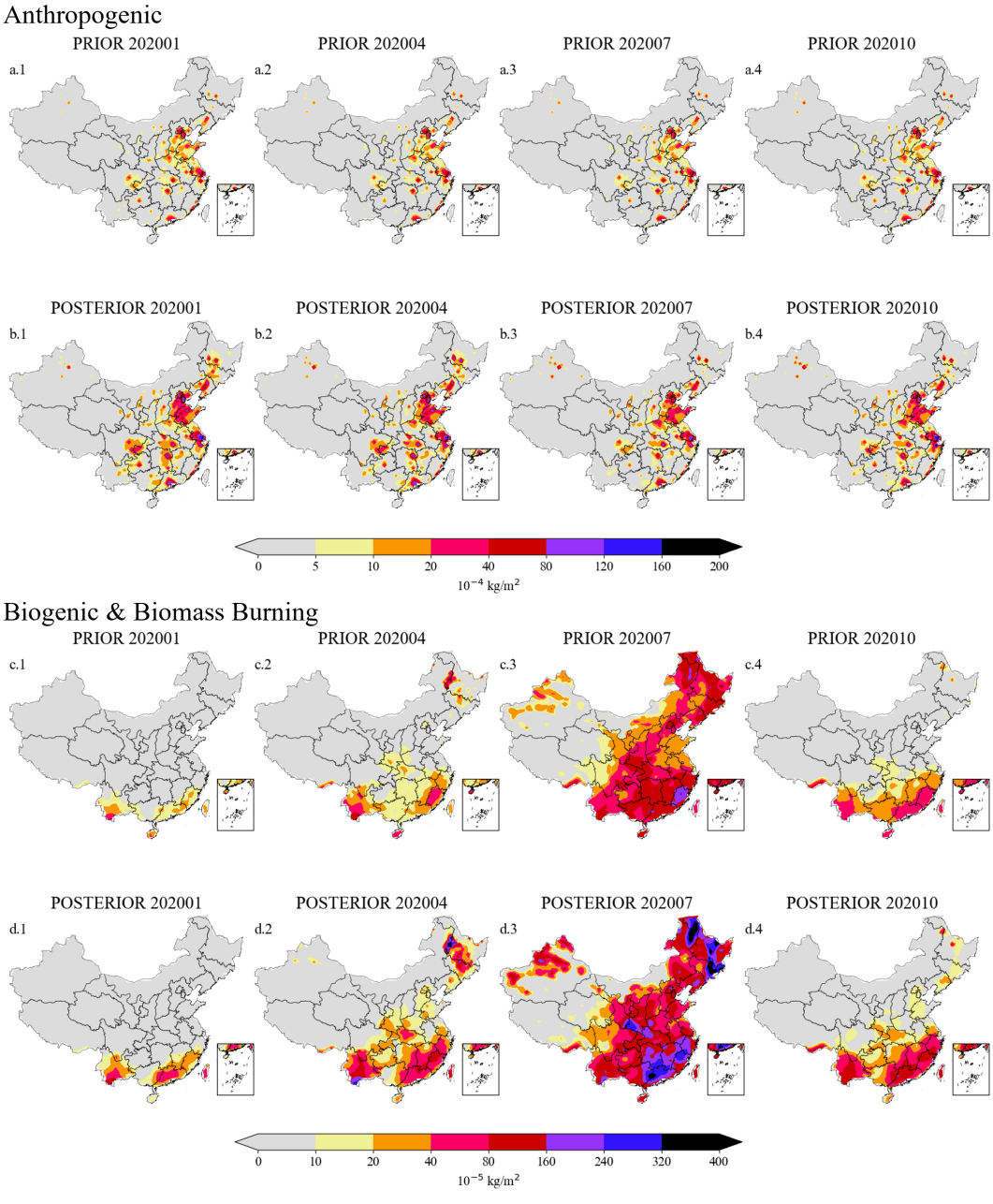

Figure 6. The upper figures show the spatial distributions of the anthropogenic NMVOC emissions form the prior (a) and posterior (b) results in January (a.1, b.1), April (a.2, b.2), July (a.3, b.3), October (a.4, b.4). The lower figures show the spatial distributions of the biogenic and biomass burning NMVOC emissions from the prior (c) and posterior (d) results in January (c.1, d.1), April (c.2, d.2), July (c.3, d.3), October (c.4, d.4).

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

below 5 $\times 10^{-4} kg/m^2$, and the actual values for some grid points in these regions are even less than 1 $\times 10^{-4} kg/m^2$. Though assimilating OMPS observations, the posterior simulation in these regions still remains low, resulting in little change. As illustrated in Figure 7 (a.2, a.3), the minimal changes in these areas also affected the national observation-minus-simulation discrepancies (root mean square error , RMSE), which decreased from 0.49 $\times 10^{16}$ molec/cm^2 to 0.46 $\times 10^{16}$ molec/cm^2. However, as shown in Figure 7 (b.2)-(b.3), when focused on the NCP region, the RMSE decreased from 0.53 $\times 10^{16}$ molec/cm^2 to 0.37 $\times 10^{16}$ molec/cm^{2}effectively.

As indicated in the time series plot in Figure 7 (a.1), due to the minimal changes in formaldehyde columns in Inner Mongolia, Tibet, and the northwest, the nationwide monthly average formaldehyde concentrations in the posterior results are still not high enough compared to TROPOMI data. However, when we plot the time series for the NCP region alone, the results are better. The posterior results show a significant improvement compared to the prior, with concentrations generally matching TROPOMI data. Especially from March to June, the posterior and TROPOMI satellite observations curves for formaldehyde columns align closely.

3.4 Impact of Formaldehyde Assimilation on O_3 Surface Concentration

Formaldehyde is a key precursor for O_3 and secondary organic aerosols, and its variations in the atmosphere are primarily driven by NMVOC emissions. Optimizing NMVOC emissions is expected to improve O_3 concentration simulations. To validate this, we compare the O_3 simulation from different NMVOC emission inventories with the O_3 surface measurements in the following study.

The spatial distribution prior and posterior MDA8 surface O_3 concentration simulation, as well as the corresponding observations, are plotted in Figure 8. When comparing the prior MDA8 O_3 surface concentration results with data from 1602 stations nationwide, the prior results generally capture the main hot spots of MDA8 O_3 surface concentrations, especially in July, where they closely match the ground observations. However, in the other three months shown here, the performance is not as accurate as in July.

Figure 9 illustrates the evaluation of O_3 simulation in terms of different metrics, including probability density (a), scatter plots (b), and correlation coefficients (c).As shown in Figure 9 (a), the frequency distribution histogram clearly demonstrates that the prior ground-level O_3 simulations systematically underestimate the observed values, whereas the posterior results effectively reduce this bias. This indicates that both the overall and regional simulations of surface O_3 concentrations have been significantly improved, especially in the NCP region. As can be seen in Figure 9 (b), the RMSE (root mean square error) between the observed and simulated values decreases from 24.44 μg/m^3 to 22.79 μg/m^3 for China and from 28.0 μg/m^3 to 20.22 μg/m^3 for the NCP region. As shown in Figure 9 (c), the correlation coefficients calculated based on the monthly average time series in the NCP region show a slight improvement, while the increase is more noticeable in the southern and northeastern regions.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

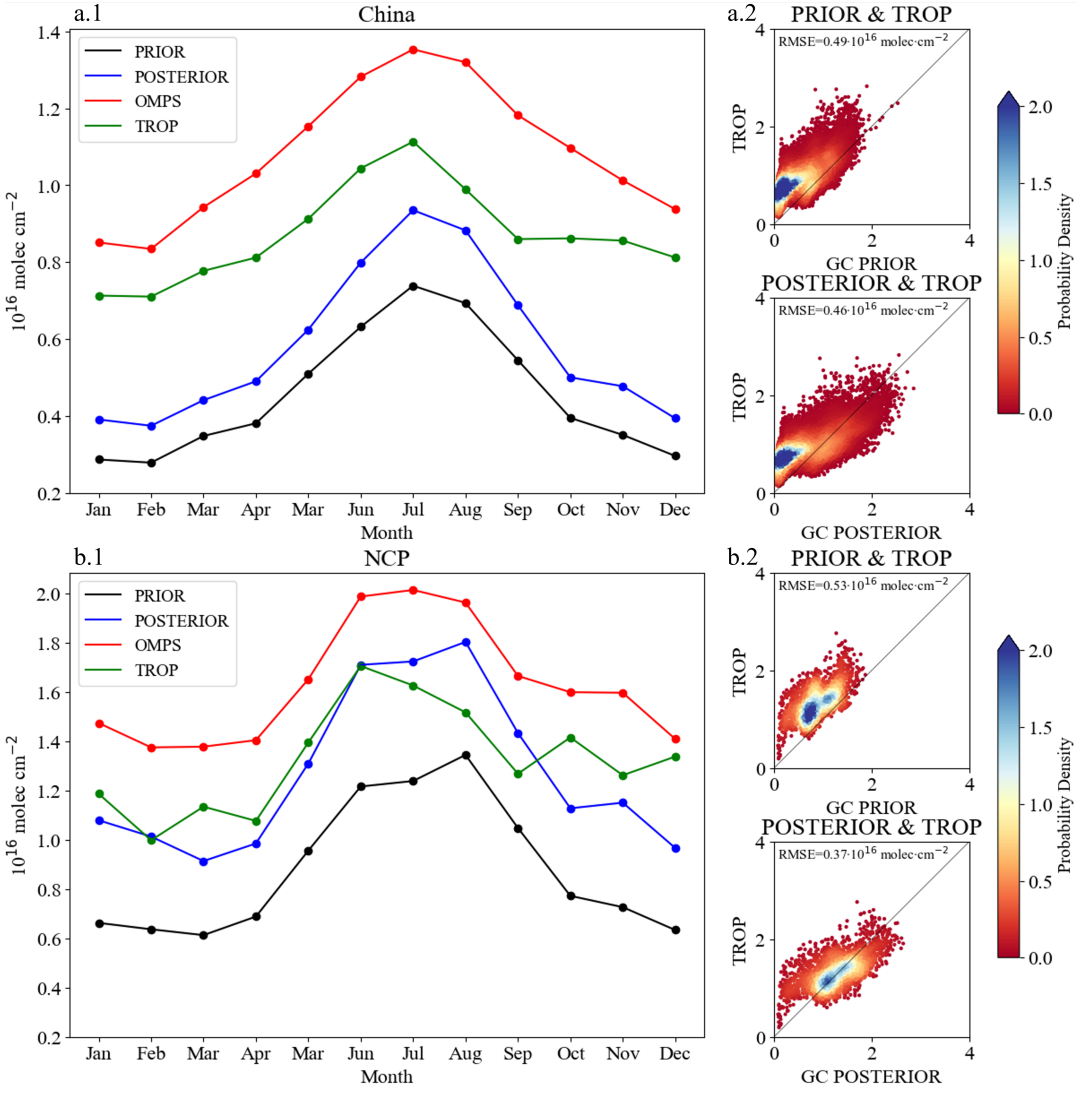

Figure 7. Monthly variations of total formaldehyde columns over China (a.1) and North China Plain (b.1) for prior, posterior, OMPS and TROPOMI in 2020. Scatter plots of TROPOMI satellite observations vs. the simulated total formaldehyde columns over China and North China Plain either using the prior (a.2, b.2) or using the posterior (a.3, b.3) in 2020.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

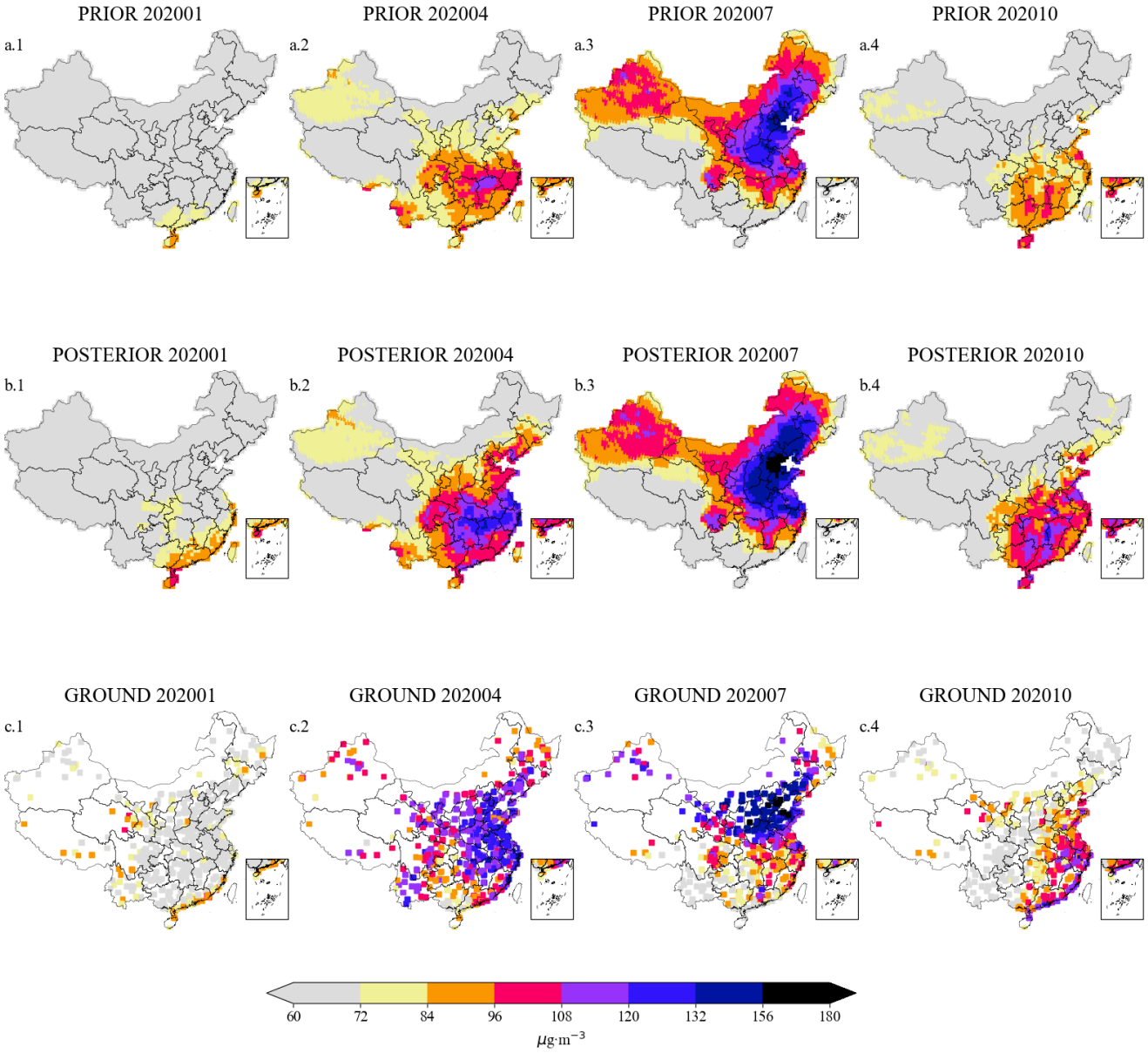

Figure 8. Spatial distributions of the MDA8 surface O_3 concentrations from GEOS-Chem model-simulated prior (a) and posterior (b) results, and from ground O_3 monitoring sites (c) in January (a.1)-(c.1), April (a.2)-(c.2), July (a.3)-(c.3), October 2020 (a.4)-(c.4).

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

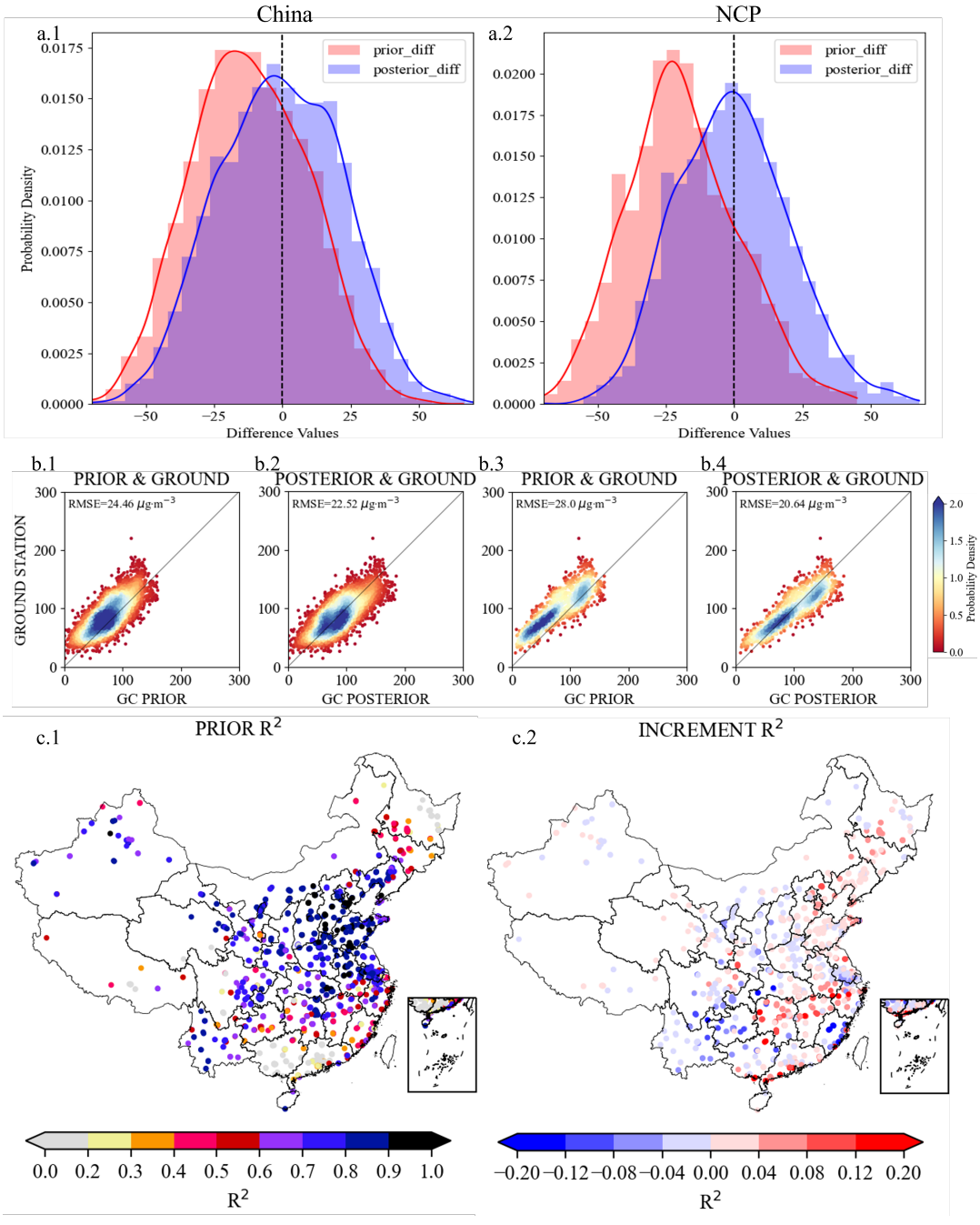

Figure 9. Frequency histogram of the difference between the ground monitoring observed and simulated O_3 surface concentrations over China and North China Plain in 2020 (a.1, a.2) and scatter plot of the observed vs. simulated O_3 surface concentrations using either prior data (b.1, b.3) or posterior data (b.2, b.4). The correlation coefficients calculated based on prior monthly averaged time series in 2020 (c.1), and the difference between posterior and prior correlation coefficients (c.2).

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

4 Summary and conclusion

In recent decades, NMVOC emissions in China have been increasing steadily, leading to a series of environmental and health issues. Currently, the NMVOC emission inventories established are generally based on bottom-up methods and exhibit considerable uncertainty. Utilizing top-down methods to optimize pollutant emissions has become an effective approach. Due to the limited types of NMVOCs species that can be observed by satellite instruments, only formaldehyde and glyoxal can be detected over long timescales within the observable spectral bands. Previous studies have confirmed the feasibility of using satellite formaldehyde retrievals to constrain NMVOC emissions. In this study, we carried out the monthly NMVOC emission optimization through assimilating formaldehyde retrievals from the latest satellite products in 2020 in China.

Before using satellite observations to optimize NMVOC inventories, we used the semi-variogram to perform a quality evaluation of different formaldehyde remote sensing products. The semi-variance values of NOAA-20 OMPS, Sentinel-5P TROPOMI, and Aura OMI satellite observations at grid resolutions of $0.5°\times0.5°$, $1°\times1°$, $2°\times2°$, and $4°\times4°$ are calculated and analyzed respectively. The results indicate that, compared to the other three data, OMI satellite observations exhibit weaker spatial autocorrelation and higher levels of white noise at higher resolutions. In contrast, the semi-variance values for the newest OMPS and TROPOMI align more closely with the trends observed in GEOS-Chem simulation, making them more suitable for high-resolution emission optimization.

The monthly NMVOC emission inventories in China were then estimated by assimilating OMPS formaldehyde column observations. In general, the posterior results show that anthropogenic emissions increased from 22.40 Tg to 41.32 Tg, biogenic emissions rose from 16.56 Tg to 28.01 Tg, and biomass burning emissions increased from 1.72 Tg to 2.11 Tg. The posterior NMVOC emission inventories were then used to drive GEOS-Chem model, and the simulated posterior formaldehyde columns were compared against independent TROPOMI observations. Nationwide, the posterior formaldehyde columns increased by approximately 50%. There remains an underestimation compared to satellite observations, with the RMSE decreasing from 0.49 to 0.45 $\times10^{16}$ molec/cm^2. In the NCP region, the formaldehyde columns were effectively modified, reaching levels comparable to those observed by TROPOMI, with the RMSE dropping from 0.52 to 0.37 $\times10^{16}$ molec/cm^2. Independent validation against the ground-level O_3 observations yielded favorable results, with the RMSE between observed and simulated values decreasing from 24.44 μg/m^3 to 22.79 μg/m^3 across China, and from 28.0 μg/m^3 to 20.22 μg/m^3 in the North China Plain region. This indicates that the estimated NMVOC inventories using a top-down approach can effectively optimize O_3 simulations.

Optimizing NMVOC emission inventories with formaldehyde observations resulted in significant improvements in the NMVOC simulations, while independent validation with ground-level O_3 observations yielded favorable results. This indicates that the NMVOC inventories constructed through this top-down approach are effective. The study still has much space for improvement. For instance, the study only utilized formaldehyde measurements, and incorporating glyoxal data could enhance the inversion results.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

Code and data availability

The 4DEnVar emission inversion system is in the Python environment and is archived on Zenodo.(https://doi.org/10.5281/zenodo.14633919; Canjie Xu, 2025). OMPS-N20 Level 2 NM formaldehyde Total Column swath orbital Version 1 product (https://doi.org/10.5067/CIYXT9A4I2F4, Abad, 2022). Sentinel-5P TROPOMI Level 2 Tropospheric formaldehyde Version 2 product (https://doi.org/10.5270/S5P-vg1i7t0, Copernicus Sentinel data processed by ESA, German Aerospace Center (DLR), 2020). OMI/Aura formaldehyde Total Column Daily L2 Global Version 3 product (https://doi.org/10.5067/Aura/OMI/DATA2016,Chance, 2014). National Urban Air Quality Real-time Publishing Platform of the China National Environmental Monitoring Center (CNEMC, https://air.cnemc.cn:18007/, last access: May 15, 2024).

Acknowledgments

We thank for the technical support of the National Large Scientific and Technological Infrastructure“Earth System Numerical Simulation Facility”(https://cstr.cn/31134.02.EL).

Financial support

This study was supported by the National Key Research and Development Program of China [grant number 2022YFE0136100] and the National Natural Science Foundation of China [grant number 42475150].

Author contribution

JJ conceived the study and designed the emission inversion method. JJ wrote the code of the emission inversion. CX carried out the analysis and evaluation. KL, YQ, JX, ZC, HXL and HL provided useful comments on the paper. CX and JJ prepared the manuscript with contributions from all other co-authors.

Competing interests

The authors declare that they have no conflict of interest.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

References

Abad, G. G.: OMPS-N20 L2 NM Formaldehyde (HCHO) Total Column swath orbital V1, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/CIYXT9A4I2F4, accessed: [Data Access Date], 2022.

Alvarado, L., Richter, A., Vrekoussis, M., Wittrock, F., Hilboll, A., Schreier, S., and Burrows, J.: An improved glyoxal retrieval from OMI measurements, Atmospheric Measurement Techniques, 7, 4133–4150, 2014.

Azmi, S., Sharma, M., and Nagar, P. K.: NMVOC emissions and their formation into secondary organic aerosols over India using WRF-Chem model, Atmospheric Environment, 287, 119 254, 2022.

Bates, K. H. and Jacob, D. J.: A new model mechanism for atmospheric oxidation of isoprene: global effects on oxidants, nitrogen oxides, organic products, and secondary organic aerosol, Atmospheric Chemistry and Physics, 19, 9613–9640, 2019.

Bates, K. H., Jacob, D. J., Wang, S., Hornbrook, R. S., Apel, E. C., Kim, M. J., Millet, D. B., Wells, K. C., Chen, X., Brewer, J. F., et al.: The global budget of atmospheric methanol: new constraints on secondary, oceanic, and terrestrial sources, Journal of Geophysical Research: Atmospheres, 126, e2020JD033 439, 2021.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, Journal of Geophysical Research: Atmospheres, 106, 23 073–23 095, 2001.

Billionnet, C., Gay, E., Kirchner, S., Leynaert, B., and Annesi-Maesano, I.: Quantitative assessments of indoor air pollution and respiratory health in a population-based sample of French dwellings, Environmental research, 111, 425–434, 2011.

Bo, Y., Cai, H., and Xie, S.: Spatial and temporal variation of historical anthropogenic NMVOCs emission inventories in China, Atmospheric Chemistry and Physics, 8, 7297–7316, 2008.

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., et al.: Adjoint inversion of Chinese non-methane volatile organic compound emissions using space-based observations of formaldehyde and glyoxal, Atmospheric Chemistry and Physics, 18, 15 017–15 046, 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, 10 525–10 545, 2019.

Chan Miller, C., Gonzalez Abad, G., Wang, H., Liu, X., Kurosu, T., Jacob, D., and Chance, K.: Glyoxal retrieval from the ozone monitoring instrument, Atmospheric Measurement Techniques, 7, 3891–3907, 2014.

Chance, K.: OMI/Aura Formaldehyde (HCHO) Total Column Daily L2 Global Gridded 0.25 degree x 0.25 degree V3, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/Aura/OMI/DATA2016, accessed: [Data Access Date], 2014.

Copernicus Sentinel data processed by ESA, German Aerospace Center (DLR): Sentinel-5P TROPOMI Tropospheric Formaldehyde HCHO 1-Orbit L2 5.5km x 3.5km, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5270/S5P-vg1i7t0, accessed: [Data Access Date], 2020.

Curran, P. J.: The semivariogram in remote sensing: an introduction, Remote sensing of Environment, 24, 493–507, 1988.

De Smedt, I., Van Roozendael, M., Stavrakou, T., Müller, J.-F., Lerot, C., Theys, N., Valks, P., Hao, N., and Van Der A, R.: Improved retrieval of global tropospheric formaldehyde columns from GOME-2/MetOp-A addressing noise reduction and instrumental degradation issues, Atmospheric Measurement Techniques, 5, 2933–2949, 2012.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

De Smedt, I., Yu, H., Richter, A., Beirle, S., Eskes, H., Boersma, K., Van Roozendael, M., Van Geffen, J., Lorente, A., and Peters, E.: QA4ECV HCHO tropospheric column data from OMI (Version 1.1)[Data set], Royal Belgian Institute for Space Aeronomy, 2017.

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., et al.: Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project, Atmospheric Measurement Techniques, 11, 2395–2426, 2018.

De Smedt, I., Pinardi, G., Vigouroux, C., Compernolle, S., Bais, A., Benavent, N., Boersma, F., Chan, K.-L., Donner, S., Eichmann, K.-U., et al.: Comparative assessment of TROPOMI and OMI formaldehyde observations and validation against MAX-DOAS network column measurements, Atmospheric Chemistry and Physics, 21, 12 561–12 593, 2021.

Duncan, B. N., Yoshida, Y., Olson, J. R., Sillman, S., Martin, R. V., Lamsal, L., Hu, Y., Pickering, K. E., Retscher, C., Allen, D. J., et al.: Application of OMI observations to a space-based indicator of NOx and VOC controls on surface ozone formation, Atmospheric Environment, 44, 2213–2223, 2010.

Flynn, L., Long, C., Wu, X., Evans, R., Beck, C., Petropavlovskikh, I., McConville, G., Yu, W., Zhang, Z., Niu, J., et al.: Performance of the ozone mapping and profiler suite (OMPS) products, Journal of Geophysical Research: Atmospheres, 119, 6181–6195, 2014.

Fu, T.-M., Jacob, D. J., Palmer, P. I., Chance, K., Wang, Y. X., Barletta, B., Blake, D. R., Stanton, J. C., and Pilling, M. J.: Space-based formaldehyde measurements as constraints on volatile organic compound emissions in east and south Asia and implications for ozone, Journal of Geophysical Research: Atmospheres, 112, 2007.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of climate, 30, 5419–5454, 2017.

González Abad, G., Liu, X., Chance, K., Wang, H., Kurosu, T., and Suleiman, R.: Updated Smithsonian astrophysical observatory ozone monitoring instrument (SAO OMI) formaldehyde retrieval, Atmospheric Measurement Techniques, 8, 19–32, 2015.

Guenther, A., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T. a., Emmons, L., 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, 2012.

Han, K., Park, R., Kim, H., Woo, J., Kim, J., and Song, C.: Uncertainty in biogenic isoprene emissions and its impacts on tropospheric chemistry in East Asia, Science of the total environment, 463, 754–771, 2013.

Hao, Y. and Xie, S.: Optimal redistribution of an urban air quality monitoring network using atmospheric dispersion model and genetic algorithm, Atmospheric Environment, 177, 222–233, 2018.

He, Z., Li, G., Chen, J., Huang, Y., An, T., and Zhang, C.: Pollution characteristics and health risk assessment of volatile organic compounds emitted from different plastic solid waste recycling workshops, Environment international, 77, 85–94, 2015.

Huang, X., Li, M., Li, J., and Song, Y.: A high-resolution emission inventory of crop burning in fields in China based on MODIS Thermal Anomalies/Fire products, Atmospheric Environment, 50, 9–15, 2012.

Ito, A., Sillman, S., and Penner, J. E.: Effects of additional nonmethane volatile organic compounds, organic nitrates, and direct emissions of oxygenated organic species on global tropospheric chemistry, Journal of Geophysical Research: Atmospheres, 112, 2007.

Jin, J., Fang, L., Li, B., Liao, H., Wang, Y., Han, W., Li, K., Pang, M., Wu, X., and Lin, H. X.: 4DEnVar-based inversion system for ammonia emission estimation in China through assimilating IASI ammonia retrievals, Environmental Research Letters, 18, 034 005, 2023.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

Johnson, M. S., Philip, S., Kumar, R., Naeger, A., Souri, A. H., Geddes, J., Judd, L., Janz, S., and Sullivan, J.: Satellite remote-sensing capability to assess tropospheric column ratios of formaldehyde and nitrogen dioxide: case study during the LISTOS 2018 field campaign, Atmospheric Measurement Techniques Discussions, 2022, 1–41, 2022.

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., et al.: High-resolution inversion of OMI formaldehyde columns to quantify isoprene emission on ecosystem-relevant scales: application to the southeast US, Atmospheric Chemistry and Physics, 18, 5483–5497, 2018.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Janssens-Maenhout, G., Fukui, T., Kawashima, K., and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia (REAS) version 2, Atmospheric Chemistry and Physics, 13, 11 019–11 058, 2013.

Kurosu, T., Chance, K., and Volkamer, R.: Global measurements of BrO, HCHO, and CHOCHO from the Ozone Monitoring Instrument on EOS Aura, in: AGU Fall Meeting Abstracts, vol. 2005, pp. A54B–01, 2005.

Kwon, H.-A., Park, R. J., Oak, Y. J., Nowlan, C. R., Janz, S. J., Kowalewski, M. G., Fried, A., Walega, J., Bates, K. H., Choi, J., et al.: Top-down estimates of anthropogenic VOC emissions in South Korea using formaldehyde vertical column densities from aircraft during the KORUS-AQ campaign, Elem Sci Anth, 9, 00 109, 2021.

Lee, J.-S.: Digital image enhancement and noise filtering by use of local statistics, IEEE transactions on pattern analysis and machine intelligence, pp. 165–168, 1980.

Lerner, J. C., Sanchez, E. Y., Sambeth, J. E., and Porta, A. A.: Characterization and health risk assessment of VOCs in occupational environments in Buenos Aires, Argentina, Atmospheric environment, 55, 440–447, 2012.

Lerot, C., Stavrakou, T., De Smedt, I., Müller, J.-F., and Van Roozendael, M.: Glyoxal vertical columns from GOME-2 backscattered light measurements and comparisons with a global model, Atmospheric Chemistry and Physics, 10, 12 059–12 072, 2010.

Li, C., Joiner, J., Krotkov, N. A., and Dunlap, L.: A new method for global retrievals of HCHO total columns from the Suomi National Polar-orbiting Partnership Ozone Mapping and Profiler Suite, Geophysical Research Letters, 42, 2515–2522, 2015.

Li, M., Zhang, Q., Streets, D., He, K., Cheng, Y., Emmons, L., Huo, H., Kang, S., Lu, Z., Shao, M., et al.: Mapping Asian anthropogenic emissions of non-methane volatile organic compounds to multiple chemical mechanisms, Atmospheric Chemistry and Physics, 14, 5617–5638, 2014.

Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., et al.: Anthropogenic emission inventories in China: a review, Natl. Sci. Rev., 4, 834–866, 2017.

Li, M., Zhang, Q., Zheng, B., Tong, D., Lei, Y., Liu, F., Hong, C., Kang, S., Yan, L., Zhang, Y., et al.: Persistent growth of anthropogenic non-methane volatile organic compound (NMVOC) emissions in China during 1990–2017: drivers, speciation and ozone formation potential, Atmospheric Chemistry and Physics, 19, 8897–8913, 2019.

Lin, H., Long, M. S., Sander, R., Sandu, A., Yantosca, R. M., Estrada, L. A., Shen, L., and Jacob, D. J.: An Adaptive Auto-Reduction Solver for Speeding Up Integration of Chemical Kinetics in Atmospheric Chemistry Models: Implementation and Evaluation in the Kinetic Pre-Processor (KPP) Version 3.0. 0, Journal of Advances in Modeling Earth Systems, 15, e2022MS003 293, 2023.

Liu, C., Zhang, C., Mu, Y., Liu, J., and Zhang, Y.: Emission of volatile organic compounds from domestic coal stove with the actual alternation of flaming and smoldering combustion processes, Environmental pollution, 221, 385–391, 2017.

Liu, M., Song, Y., Yao, H., Kang, Y., Li, M., Huang, X., and Hu, M.: Estimating emissions from agricultural fires in the North China Plain based on MODIS fire radiative power, Atmospheric Environment, 112, 326–334, 2015.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

Pan, C., Weng, F., and Flynn, L.: Spectral performance and calibration of the Suomi NPP OMPS Nadir Profiler sensor, Earth and Space Science, 4, 737–745, 2017.

Park, R. J., Jacob, D. J., Field, B. D., Yantosca, R. M., and Chin, M.: Natural and transboundary pollution influences on sulfate-nitrate-ammonium aerosols in the United States: Implications for policy, Journal of Geophysical Research: Atmospheres, 109, 2004.

Platt, C.: Remote sounding of high clouds: I. Calculation of visible and infrared optical properties from lidar and radiometer measurements, Journal of Applied Meteorology and Climatology, 18, 1130–1143, 1979.

Qiu, K., Yang, L., Lin, J., Wang, P., Yang, Y., Ye, D., and Wang, L.: Historical industrial emissions of non-methane volatile organic compounds in China for the period of 1980–2010, Atmospheric Environment, 86, 102–112, 2014.

Seftor, C., Jaross, G., Kowitt, M., Haken, M., Li, J., and Flynn, L.: Postlaunch performance of the Suomi National Polar-orbiting partnership Ozone Mapping and Profiler Suite (OMPS) nadir sensors, Journal of Geophysical Research: Atmospheres, 119, 4413–4428, 2014.

Sharma, S., Goel, A., Gupta, D., Kumar, A., Mishra, A., Kundu, S., Chatani, S., and Klimont, Z.: Emission inventory of non-methane volatile organic compounds from anthropogenic sources in India, Atmospheric Environment, 102, 209–219, 2015.

Shim, C., Wang, Y., Choi, Y., Palmer, P. I., Abbot, D. S., and Chance, K.: Constraining global isoprene emissions with Global Ozone Monitoring Experiment (GOME) formaldehyde column measurements, Journal of Geophysical Research: Atmospheres, 110, 2005.

Souri, A. H., Nowlan, C. R., González Abad, G., Zhu, L., Blake, D. R., Fried, A., Weinheimer, A. J., Wisthaler, A., Woo, J.-H., Zhang, Q., et al.: An inversion of NO x and non-methane volatile organic compound (NMVOC) emissions using satellite observations during the KORUS-AQ campaign and implications for surface ozone over East Asia, Atmospheric Chemistry and Physics, 20, 9837–9854, 2020.

Souri, A. H., Chance, K., Sun, K., Liu, X., and Johnson, M. S.: Dealing with Spatial Heterogeneity in Pointwise to Gridded Data Comparisons, Atmospheric Measurement Techniques Discussions, 2021, 1–33, 2021.

Stavrakou, T., Müller, J.-F., Bauwens, M., De Smedt, I., Lerot, C., Van Roozendael, M., Coheur, P.-F., Clerbaux, C., Boersma, K., Van Der A, R., et al.: Substantial underestimation of post-harvest burning emissions in the North China Plain revealed by multi-species space observations, Scientific Reports, 6, 32 307, 2016.

Stavrakou, T., Muller, J., Bauwens, M., and De Smedt, I.: Sources and long-term trends of ozone precursors to Asian Pollution, Air Pollution in Eastern Asia: an integrated perspective, eds. Bouarar, I., Wang, X., Brasseur, G, 2017.

Van Der Werf, G. R., Randerson, J. T., Giglio, L., Van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., Van Marle, M. J., Morton, D. C., Collatz, G. J., et al.: Global fire emissions estimates during 1997–2016, Earth System Science Data, 9, 697–720, 2017.

Veefkind, J. P., Aben, I., McMullan, K., Förster, H., De Vries, J., Otter, G., Claas, J., Eskes, H., De Haan, J., Kleipool, Q., et al.: TROPOMI on the ESA Sentinel-5 Precursor: A GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, Remote sensing of environment, 120, 70–83, 2012.

Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., Cheng, Z., De Mazière, M., De Smedt, I., Grutter, M., Hannigan, J. W., Jones, N., et al.: TROPOMI–Sentinel-5 Precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations, Atmospheric Measurement Techniques, 13, 3751–3767, 2020.

Wang, F., An, J., Li, Y., Tang, Y., Lin, J., Qu, Y., Chen, Y., Zhang, B., and Zhai, J.: Impacts of uncertainty in AVOC emissions on the summer RO x budget and ozone production rate in the three most rapidly-developing economic growth regions of China, Advances in Atmospheric Sciences, 31, 1331–1342, 2014.

Wu, K., Yang, X., Chen, D., Gu, S., Lu, Y., Jiang, Q., Wang, K., Ou, Y., Qian, Y., Shao, P., et al.: Estimation of biogenic VOC emissions and their corresponding impact on ozone and secondary organic aerosol formation in China, Atmospheric Research, 231, 104 656, 2020.

https://doi.org/10.5194/egusphere-2025-140
Preprint. Discussion started: 7 May 2025

Wu, R., Bo, Y., Li, J., Li, L., Li, Y., and Xie, S.: Method to establish the emission inventory of anthropogenic volatile organic compounds in China and its application in the period 2008–2012, Atmospheric environment, 127, 244–254, 2016.

Xia, J., Zhou, Y., Fang, L., Qi, Y., Li, D., Liao, H., and Jin, J.: South Asia ammonia emission inversion through assimilating IASI observations, EGUsphere, 2025, 1–22, 2025.

Yuan, B., Hu, W., Shao, M., Wang, M., Chen, W., Lu, S., Zeng, L., and Hu, M.: VOC emissions, evolutions and contributions to SOA formation at a receptor site in eastern China, Atmospheric Chemistry and Physics, 13, 8815–8832, 2013.

Zhang, C., Liu, C., Hu, Q., Cai, Z., Su, W., Xia, C., Zhu, Y., Wang, S., and Liu, J.: Satellite UV-Vis spectroscopy: implications for air quality trends and their driving forces in China during 2005–2017, Light: Science & Applications, 8, 100, 2019.

Zheng, B., Tong, D., Li, M., Liu, F., Hong, C., Geng, G., Li, H., Li, X., Peng, L., Qi, J., et al.: Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions, Atmospheric Chemistry and Physics, 18, 14 095–14 111, 2018.

Zhu, B., Huang, X.-F., Xia, S.-Y., Lin, L.-L., Cheng, Y., and He, L.-Y.: Biomass-burning emissions could significantly enhance the atmospheric oxidizing capacity in continental air pollution, Environmental Pollution, 285, 117 523, 2021.