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~~ are key precursors of ozone and secondary organic aerosols. ~~Given that China is a major source of NMVOCs, the emission inventory is crucial~~ As one of the world's largest NMVOC emitters, accurate emission inventories are essential for understanding and ~~controlling atmospheric pollution . Mainstream inventories are constructed using~~ mitigating air pollution in China. Commonly-used inventories (e.g., MEIC) are largely based on bottom-up ~~approaches, which cannot accurately reflect the spatiotemporal characteristics of NMVOCs~~methods, which often fail to capture the spatiotemporal variability of NMVOC emissions, resulting in ~~poor model outcomes~~significant model-observation mismatches. 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~~ evaluates the vertical profile structure, qualified-data volume and monthly mean biases of formaldehyde products from OMI, OMPS and TROPOMI satellites, emphasizing the enhanced capabilities of the latest OMPS and TROPOMI retrievals. Monthly NMVOC emissions over China in 2020 are then optimized by independently assimilating formaldehyde retrievals either from OMPS or from TROPOMI, using 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~~ 4DEnVar assimilation emission inversion system. The OMPS- and TROPOMI-driven assimilation

yields consistent seasonal and regional increments in NMVOC emissions in general, but distinctions are also notable. A consistency analysis is introduced to assess the reliability of these two posterior emissions. Highly consistent increments are obtained in the North China Plain (May-June), the Yangtze River Delta and Pearl River Delta (January-March, ~~with the RMSE dropping from 0.52 to 0.37 $\times 10^{16}$ molec~~October-December), and the Sichuan Basin (January, June-December). These adjustments significantly improve surface ozone simulations, with 81.25% of consistent cases demonstrating reduced biases and an average RMSE reduction of 24.7%. These findings highlight the effectiveness of OMPS and TROPOMI formaldehyde assimilation, coupled with consistency analysis, in refining NMVOC emission estimates and enhancing ozone simulation accuracy. Similar promising results are achieved in the OMPS/~~cm^2. Validation using surface ozone observations also yielded favorable results, especially in NCP.~~ TROPOMI-based NMVOC emission inversion in 2019.

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., 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~~, gradually becoming one of the important contributors to global NMVOC emissions (Li et al., 2019). 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~~ For biomass burning, widely used inventories include the Global Fire Emissions Database (GFED) and the Fire INventory from NCAR (FINN) (Wiedinmyer et al., 2011). 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~~ these inventories are widely used to simulate transport, deposition, and chemical ~~reactions. This not only helps to better quantify the environmental impact~~ transformations of NMVOCs, ~~but also provides essential tools for predicting future trends and making emission reduction~~ supporting air quality assessments and emission control 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~~estimates remain highly uncertain because both emission factors and activity data vary greatly in space and time and are often poorly constrained (Bo et al., 2008; Sharma et al., 2015). For

anthropogenic sources, nationwide uncertainties of ±68-78% have been reported due to variable activity data and emission factors under rapid structural transitions in industry, solvent use, and transportation ~~sectors~~ (Li et al., 2017, 2019). Biogenic emissions are even more uncertain, highly sensitive to land-cover, meteorology, and parameterizations, with Chinese BVOC estimates varying from 10 to 58.9 Tg C yr^{-1} (Li et al., 2020; Wang et al., 2021; Pei et al., 2025). Biomass burning emissions also show large discrepancies across inventories (e.g., GFED, FINN, GFAS), largely driven by uncertainties in burned area, fuel loading, and emission factors (He et al., 2011; Hua et al., 2024). In addition, strict air pollution controls implemented 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~~ targeting industry, residential use, and transportation have significantly altered emission patterns (Wu et al., 2016; Li et al., 2017; Zheng et al., 2018). Consequently, bottom-up inventories ~~inherently have considerable~~ carry substantial uncertainties (Li et al., 2014; Qiu et al., 2014). For example, ~~bottom-up~~ estimates of China~~'s annual~~'s total 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., 20~~ ~~. This significant uncertainty poses~~ depending on the inventory used (Kurokawa et al., 2013; Wu et al., 2016; Stavrakou et al., 2017) , posing major challenges for accurately ~~modeling and assessing the environmental impacts~~ assessing the role of NMVOCs in ~~China~~ air quality and climate (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 (Sakdapipanich and Insom, 2006; Cheng et al., 2017; Xing et al., 2022). Among the various NMVOCs, the optical properties of formaldehyde ~~(HCHO) and glyoxal (CHOCHO)~~ and glyoxal 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~~are better 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. 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~~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~~ NO_x 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~~ 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,~~ , and the coarse spatial resolution (4° ~~×~~ × 5°) in that study further 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~~ , and they indicated that the crop burning fluxes of VOCs in June ~~exceed~~ exceeded by a factor of ~~2~~ two the combined emissions from other anthropogenic activities in the NCP region from 2005 to 2012. Cao et al. (2018) conducted a relatively systematic satellite-based emission inversion study over China~~. They used~~ , using a 4DVar method and ~~assimilated~~ assimilating OMI and GOME-2A formaldehyde products to estimate monthly NMVOC emissions ~~over China~~ in 2007, ~~with a coarse spatial resolution of~~ though the spatial resolution (4° ~~×~~ × 5°~~. Considering the increasingly stringent air pollution control policies and significant interannual variability in NMVOC emissionsin recent years, e.g.~~) was still too coarse. Choi et al. (2022) applied a 4DVar system to assimilate TROPOMI formaldehyde over East Asia, demonstrating the capability of high-resolution satellite data to capture regional and seasonal variability in VOC emissions, ~~the biomass buring emission is now reduced to a relatively low level~~but the analysis was conducted only for May-June. Beyond China, a number of important studies have advanced top-down VOC inversion methodologies: Palmer et al. (2003) pioneered the use of GOME formaldehyde observations in a Bayesian framework to constrain global isoprene emissions, laying the foundation for subsequent satellite-based VOC studies; Wells et al. (2020, 2022) further advanced this field by retrieving isoprene emissions from CrIS measurements and providing high-resolution constraints on VOC oxidation chemistry at the global scale; and Oomen et al. (2024) derived weekly top-down VOC fluxes over Europe from TROPOMI formaldehyde data using the MAGRITTEv1.1 model, providing improved constraints on isoprene, biomass burning, and anthropogenic VOC emissions. Considering the increasingly stringent air pollution control in China (Wu et al., 2024), there is an urgent need for high-resolution top-down ~~emission optimization over China~~NMVOC emission optimization.

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 ensure the quality of OMI data, rigorous filtering is usually required; however, the number of valid grid cells remaining after such screening is severely limited, making it insufficient for high-resolution assimilation. In contrast, the TROPOMI and OMPS formaldehyde products for estimating do not suffer from this issue and are therefore more suitable for high-resolution NMVOCs compared to OMI retrievalsemission optimization at the national scale. Subsequently, the monthly NMVOC emission optimization in China is conducted. This is achieved by independently assimilating formaldehyde observations from OMPS and either from OMPS or from TROPOMI, based on the emission inversion system that couples the four-dimensional ensemble 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). studies (Jin et al., 2023; Xia et al., 2025) . In this study, we focus on the year 2020 for the main analysis, while results for 2019 are also presented in the Supplementary Information to provide additional context and support.

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 ozone simulations. Section 4 summarizes the key findings and concludes the study.

2 Data and methods

This chapter

This section begins by introducing the GEOS-Chem model utilized for simulations in Section 2.1. Section 2.2 presents an overview of the input emission sources for the modelemissions used as the prior NMVOC inventories, including anthropogenicsources, biogenicsources, biogenic, 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 ozone 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 introduces the 4DEnVar algorithm used for data assimilation.

2.1 Model simulation

GEOS-Chem is a chemical transport model driven by meteorological data from the 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 formaldehyde columns to constrain 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°–136-136° E, 17.5°–54-54° 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~~ .1 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). Since the satellite overpasses China mainly between 12:00 and 14:00 local time, the model outputs within this time window are sampled to calculate the formaldehyde columns for fair comparison with the satellite observations. Our GEOS-Chem model outputs both total and tropospheric formaldehyde column concentrations, enabling comparison with OMPS total column data and TROPOMI formaldehyde column measurements as will be introduced later. Samples of the formaldehyde tropospheric column simulation are presented in Figure 1 (a).

2.2 ~~Prior NMVOC emission inventries~~

2.2 Prior NMVOC emission inventories

~~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 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 2019 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~~ . In this study, the biomass burning emissions ~~in this study come from the~~ are taken from the GFED version 4 (GFED4~~global inventory~~) global inventory for 2020 (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~~ biogenic emissions are provided by MEGAN (offline calculation) ~~and the biomass burning inventory~~ for the year 2020 with an hourly temporal resolution, directly through the HEMCO emission component of GEOS-Chem; in this study, we did not run the MEGAN model separately. Biomass burning emissions are taken from GFED4~~are incorporated.~~

~~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).~~

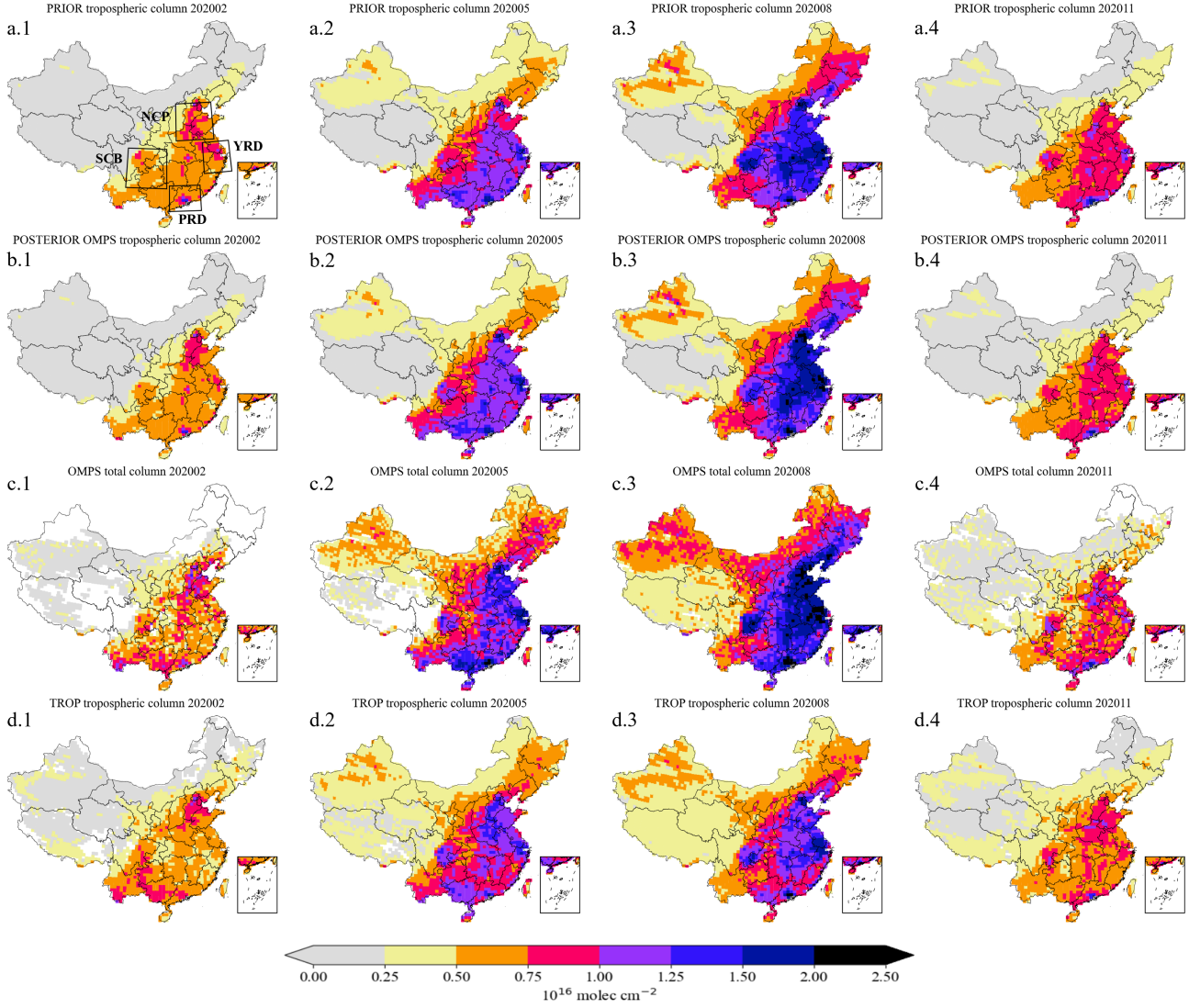

Figure 1. Spatial distributions of formaldehyde columns from GEOS-Chem model-simulated prior tropospheric columns (a) and posterior tropospheric columns constrained by OMPS assimilation (b), satellite observations of OMPS total columns (c), and satellite observations of TROPOMI tropospheric columns (d) in February (a.1-d.1), May (a.2-d.2), August (a.3-d.3), and November (a.4-d.4) of 2020.

~~The optimization of NMVOC emissions through the assimilation of formaldehyde observations will be conducted using these combined prior inventories.~~ . The combination of these three sources is treated as the prior emission inventory used in the following NMVOC emission optimization.

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~~ 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. After removing outliers, we further excluded data points where the sum of formaldehyde column and twice the observation uncertainty was less than zero. Furthermore, the geometric air mass factors (AMF_G) were defined as follows:

$$AMF_G = sec(SZA) + sec(VZA) \tag{1}$$

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 zero. Subsequently, data points were excluded if~~ Additional data screening was applied by excluding observations with SZA ~~exceeded~~ greater than 70°, ~~the~~ an air mass factor ~~was~~ less than 0.1, ~~the~~ a geometric air mass factor ~~exceeded 5, or the cloud fraction surpassed~~ greater than 4, a cloud fraction exceeding 0.4~~. Snapshots of filtered~~ , or with positive snow and ice fractions. All screened data were then averaged to a spatial resolution of 0.5° latitude × 0.625° longitude on a monthly basis, consistent with the GEOS-Chem model configuration. To make a fair comparison between the observed and simulation formaldehyde column concentration in the assimilation, we further imposed constraints on the number of observations within each grid cell. Specifically, two filtering schemes were tested, in which grid cells with fewer than 10 or fewer than 50 original observations were excluded. The OMPS formaldehyde columns ~~is~~ after applying the threshold of 50 are shown in Figure ~~1(c~~1 (c), while the

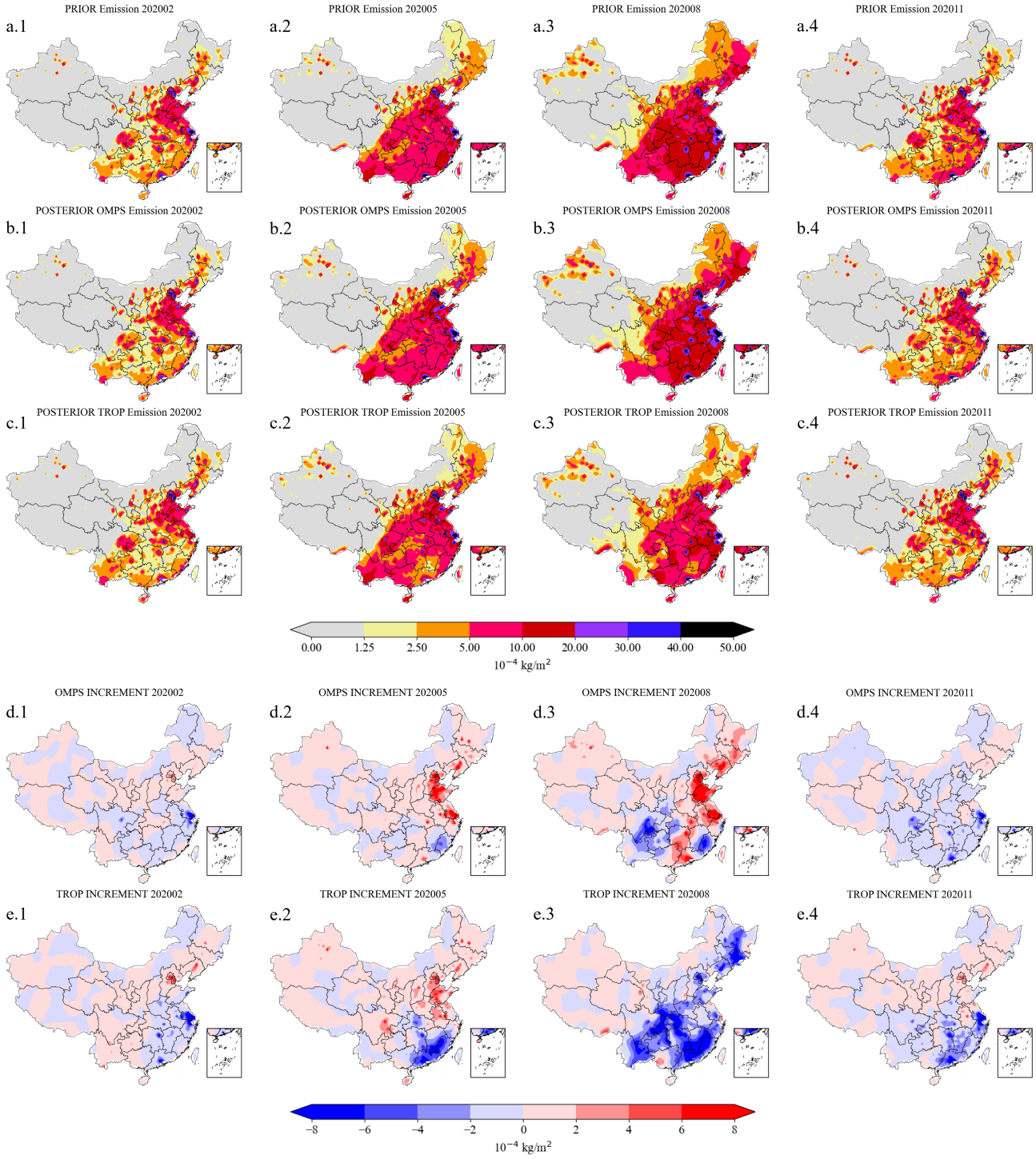

Figure 2. Spatial distributions of the total NMVOC emissions from the prior (a) and posterior (b) results in ~~January~~ February (a.1, b.1), ~~April~~ May (a.2, b.2), ~~July~~ August (a.3, b.3), ~~October~~ November (a.4, b.4) 2020. Panels (d.1-d.4) and (e.1-e.4) show the corresponding emission increments (posterior minus prior) derived from OMPS and TROPOMI assimilation.

results with the threshold of 10 are provided in the Supplement. The differences between the two filtering schemes are minor, particularly across the four study regions considered in this work.

Formaldehyde vertical column densities (VCDs) retrieved from satellite observations are derived using air mass factors (AMF), which strongly depend on the a priori vertical profiles of formaldehyde. Direct comparisons between satellite products and model simulations may be biased if the a priori profiles used in the retrieval differ from the simulated ones. To ensure consistency between the satellite observations and GEOS-Chem simulation, we applied an AMF correction by recalculating the AMF with model-simulated profiles following the method used in Palmer et al. (2001):

$$AMF = \int_{p_s}^{0} w(p) S(p)\, dp \tag{2}$$

The right-hand side of the equation represents the vertically integrated product of the scattering weight $w(p)$ and the shape factor $S(p)$ as a function of pressure p, where $w(p)$ characterizes the sensitivity of the satellite measurement to a given atmospheric layer and $S(p)$ describes the normalized vertical distribution of formaldehyde. The scattering weights $w(p)$ are primarily determined by satellite observational geometry (e.g., solar and viewing zenith angles), surface albedo, and cloud fraction, while the shape factor $S(p)$ depends on the vertical profiles of formaldehyde. The integration is performed over the pressure coordinate from the surface (p_s) to the top of the atmosphere. Figure 3 illustrates the vertical distribution of the shape profile, highlighting the relative contributions of different layers. The vertical column density (VCD) is obtained from the ratio of the slant column density (SCD) to the AMF:

$$VCD = \frac{SCD}{AMF} \tag{3}$$

In the OMPS formaldehyde product, the SCD is derived as the sum of three components: the fitted differential slant column amount (ΔSCD), the reference sector correction (SCD_{Ref}), and the bias correction (SCD_B):

$$SCD = \Delta SCD + SCD_{Ref} + SCD_B \tag{4}$$

Here, ΔSCD represents the differential slant column amount retrieved from the DOAS spectral fitting, SCD_{Ref} is the reference sector correction that accounts for background contributions and instrumental offsets by using clean reference regions, and SCD_B denotes an additional bias correction to mitigate systematic errors.

The processed OMPS satellite observations were ultimately assimilated as total columns, which are presented in Figure 1 (c.1-c.4).

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

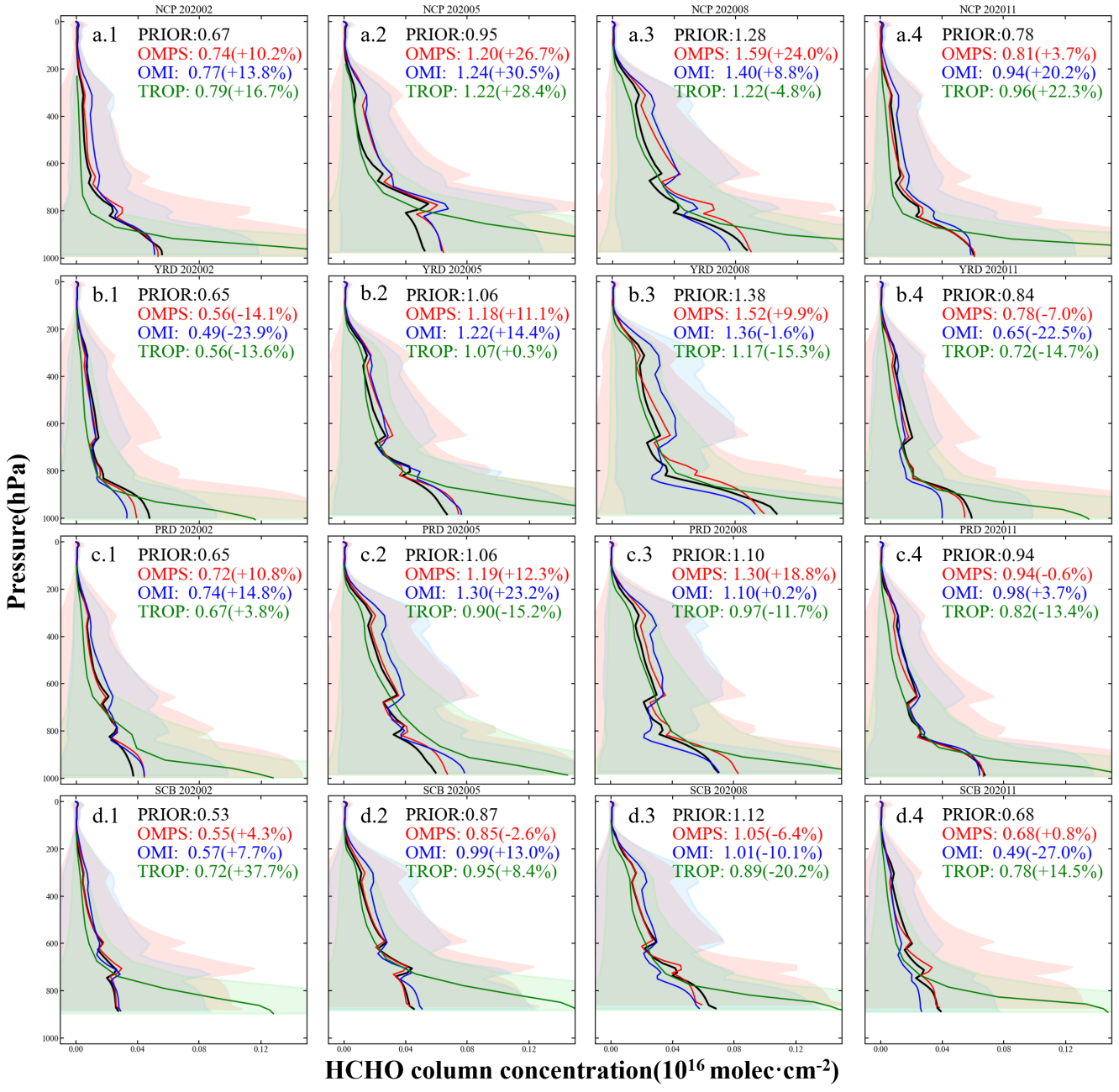

Figure 3. Vertical distributions of the regional mean formaldehyde columns from GEOS-Chem model-simulated prior (black) and satellite observations by OMPS (blue), TROPOMI (red), and OMI (green). Panels (a)-(d) correspond to the North China Plain, Yangtze River Delta, Pearl River Delta, and Northeast China, respectively. Sub-panels (a.1-d.1), (a.2-d.2), (a.3-d.3), and (a.4-d.4) represent February, May, August, and November 2020, respectively. Values in parentheses indicate the biases of satellite observations relative to the prior simulation. Shaded areas denote the observational uncertainties.

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~~ozone, sulfur dioxide, nitrogen dioxide, formaldehyde, carbon monoxide, and methane, as well as ~~O_3~~ ozone profiles, aerosol 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 the OMI QA4ECV product retrieval algorithm (De Smedt et al., 2017). This study uses the 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~~ its good quality. ~~In 2021, De Smedt et al.~~ De Smedt et al. (2021) 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)~~study, we applied the recommended quality assurance filtering by retaining only pixels with a qa value greater than 0.5. This criterion ensures the exclusion of error flags and requires that the cloud radiance fraction at 340 nm is below 0.5, the solar zenith angle (SZA) does not exceed 70°, the surface albedo is below 0.2, no snow or ice warning is present, and the air mass factor (AMF) is larger than 0.1. The TROPOMI product provides vertical information on 34 layers, but the retrieval is primarily sensitive to the troposphere and thus reports the formaldehyde tropospheric column. After filtering, the TROPOMI observations were aggregated to monthly means on a 0.5° × 0.625° grid, ensuring consistency with the resolution used in the GEOS-Chem simulations. In addition, we further constrained the number of observations per grid cell: Figure 1 (d) shows the results after excluding grid cells with fewer than 50 observations, while the results with a threshold of 10 are also provided in the Supplement. The differences between the two filtering schemes are minor, particularly over the study regions.

Beyond the recommended quality screening, a key consideration when comparing TROPOMI formaldehyde retrievals with model outputs is the dependence on the assumed a priori vertical profile. Traditionally, studies have relied on AMF-based corrections, in which AMF is recalculated using model-derived profiles to reduce such discrepancies (Palmer et al., 2001; Boersma et al., 20 . More recently, the availability of averaging kernel (AVK) information in the TROPOMI product has allowed a more consistent comparison by accounting for the impact of the assumed vertical profile shape in the retrieval, following the approach introduced in the IASI NH_3 version 4 product (Clarisse et al., 2023; Xia et al., 2025). In this study, we apply AVK-based correction for TROPOMI formaldehyde by projecting the model profiles onto the satellite pressure grid, thereby achieving a more harmonized comparison with GEOS-Chem simulations. The corrected column is calculated as:

$$\hat{X}^m = \frac{\hat{X}^a - B}{\sum_p A_p^a m_p} + B \tag{5}$$

where $\hat{X}^m$ denotes the formaldehyde column adjusted with the model profile, $\hat{X}^a$ is the retrieved column based on the a priori profile, and B is the background concentration. The term A_p^a represents the AVK at pressure level p, and m_p is the normalized model shape factor at the same level, defined as:

$$m_p = \frac{M_p^m - B_p}{M^m - B} \tag{6}$$

The processed TROPOMI retrievals were assimilated as tropospheric columns, which are presented in Figure 1 (d.1-d.4), with their vertical shape profiles shown in Figure 3 (green line) to illustrate the normalized contribution of each pressure layer to the tropospheric columns. We adopted tropospheric rather than total columns because the retrieval product itself provides tropospheric columns, and recalculating total columns would introduce substantial uncertainties.

2.3.3 Aura OMI

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

In this study, we use the 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~~ . In order to minimize the influence of poor-quality data, we applied strict quality filtering. Only pixels with cloud fraction $\leqslant$ 0.3~~are further excluded~~, solar zenith angle $\leqslant$ 70°, and a main data quality flag = 0 were retained. To avoid poor-quality measurements at large pixel sizes, the five marginal pixels on each side of the swath were discarded, and only pixels within rows 6-55 were used (Zhu et al., 2017; Xue et al., 2020). Because OMI has experienced a row anomaly since 2007, pixels with Xtrack quality flags = 0 were further selected to eliminate its impact. Additionally, given the large uncertainties in formaldehyde retrievals, pixels with a fitting root mean square (RMS) $\leqslant$ 0.003 were retained to remove most outliers (Souri et al., 2017).

The OMI observations are then aggregated to monthly means on a 0.5° × 0.625° grid, consistent with the GEOS-Chem model resolution. To ensure sufficient sampling per grid cell, we also applied two filtering schemes based on the number of

observations, excluding grid cells with fewer than 10 or fewer than 50 valid pixels. Unlike OMPS and TROPOMI, however, OMI shows a strong reduction in data coverage under these constraints, and the product becomes sparse after applying the threshold of 50 observations. This indicates that OMI suffers from insufficient sampling density in China for high-resolution assimilation. The vertical profile correction of OMI formaldehyde was conducted using the same approach as applied to OMPS, by recalculating AMF with model-simulated vertical profiles.

2.4 ~~O_3~~ Ozone 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~~ozone, the optimization of the NMVOC emission inventories and concentrations are supposed to improve the ~~O_3~~ ozone simulation simultaneously. To evaluate the magnitude and quality of this impact, the ground level ~~O_3~~ ozone 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~~ ozone measurements utilized in this study are from 1,602 sites across China. The MDA8 values of surface ~~O_3~~ ozone 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.

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:~~

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

~~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.~~ 3.4.

2.5 Assimilation algorithm

This study employs the four-dimensional ensemble variational (4DEnVar) methodology to ~~assimilate formaldehyde observations to constrain NMVOC emissions~~ optimize NMVOC emissions with satellite formaldehyde observations. The goal of ~~this~~ the 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. Note that $\boldsymbol{f}$ represents the vector of total NMVOC emissions, rather than separately gridded anthropogenic, biogenic, or biomass burning VOC emissions. To optimize emissions from these three sectors, additional observations or a well-defined spatial correlation structure are required, which are not available in this study. 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:

$$\mathcal{J}(\boldsymbol{f}) = \frac{1}{2}(\boldsymbol{f}-\boldsymbol{f}_b)^T\mathbf{B}^{-1}(\boldsymbol{f}-\boldsymbol{f}_b) + \frac{1}{2}\{\boldsymbol{y}-\mathcal{HM}(\boldsymbol{f})\}^T\mathbf{O}^{-1}\{\boldsymbol{y}-\mathcal{HM}(\boldsymbol{f})\} \tag{7}$$

The cost function $\mathcal{J}$ is the sum of two ~~part~~parts: 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 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:

$$f(i) = f_b(i)\cdot\alpha(i) \tag{8}$$

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, a minimum of 0.1 and a standard deviation ~~$\sigma_\alpha = 0.2$~~of 0.4, corresponding to a uniform 120% uncertainty applied to the total NMVOC emissions rather than sector-specific settings as adopted in previous studies (Choi et al., 2022; Jung et al., 2022; Souri e . The rationale for this choice is provided in the Supplement. 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 $\mathbf{C}$:

$$\mathbf{B}_\alpha(i,j) = \sigma_\alpha\cdot\mathbf{C}(i,j) \tag{9}$$

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

$$\mathbf{C}(i,j) = e^{-(d_{i,j}/l)^2/2} \tag{10}$$

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 αs. A small value of l ~~means more errors in fine scale could be resolved using the assimilation, while however requires more~~ indicates that the tuning factors αs are less spatially correlated, thereby enabling emission optimization at a finer spatial scale. However, this also necessitates a larger number of ensemble runs to adequately represent the model realization from emission to simulation. An empirical ~~parameterl~~ parameter l = 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:

$$\mathbf{B} = \mathbf{B}_\alpha \circ \mathbf{C} \tag{11}$$

In the observation term, $\boldsymbol{y}$ is the observation vector, representing satellite observations, $\mathcal{M}$ is the GEOS-Chem model driven by emissions $\boldsymbol{f}$, $\mathcal{H}$ is the observation operator that ~~transfer the three dimensional~~ transfers the three-dimensional concentration into the observational space, and $\mathbf{O}$ is the observation covariance matrix. In this study, the assimilated observations include the OMPS total columns and TROPOMI tropospheric columns. A distinct observation operator $\mathcal{H}$ is configured to enable a fair comparison of the observation-minus-simulation mismatch. The satellite formaldehyde observations are assumed to be independent, therefore $\mathbf{O}$ is a diagonal matrix. The diagonal value here is calculated as:

$$\sigma_{\text{total}} = \sqrt{\sigma^2_{\text{instrument}} + \sigma^2_{\text{represent}}} \tag{12}$$

In ~~the~~ Equation 12, 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 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 spatial distribution of the total uncertainty is provided in Figure S2 in the Supplement.

The assimilation methodology used in this paper is the four-dimensional ensemble variational (4DEnVar). Different from the classic 4DVar that requires adjoint in the cost function minimization, 4DEnVar emulates the GEOS-Chem formaldehyde simulating model using an ensemble-based linear approximation and hence is adjoint-free. The method is first proposed by Liu et al. (2008) and successfully implemented in our recent dust aerosol (Jin et al., 2021) and ammonia emission inversion (Jin et al., 2023; Xia et al., 2025). The detailed procedures for minimizing the cost function Equation 7 are illustrated in section 'Minimization of the Cost Function in 4DEnVar' in 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~~This section first presents the three satellite observations evaluation in terms of the vertical profile structure, qualified-data volume and monthly mean biases. Independent assimilations are then performed by either assimilating the OMPS or assimilating the TROPOMI retrievals independently. Posterior of the NMVOC emission, ~~2 discusses the posterior NMVOC emissions through assimilating OMPS formaldehyde observations. In Section 3.3, the posterior formaldehyde column driven by the posterior NMVOC emissioninventories, are validated. Finally, Section 3.4 presents an independent validation of the posterior ground-level O_3 simulations using ground-based O_3 observations~~formaldehyde concentration results and the impact on ozone simulation are discussed. A consistency analysis is introduced to assess the reliability of the two posterior emission.

3.1 ~~Semi-variogram analysis~~Satellite data evaluation

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~~ results shown in Figure 3 are based on formaldehyde columns before the application of model vertical profile corrections. The total columns of OMI, OMPS and the 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 datasetsto generate averages at resolutions of 0.5° × 0.5°, 1° × 1°, 2° × 2°, and 4° × 4°, as shown in Figure 3.~~ simulations are presented. Their vertical structures generally decrease with altitude but exhibit characteristic minima and maxima at several pressure levels: a trough at approximately 750-850 hPa, followed by a first peak about 30 hPa above; a second pronounced peak around 600-700 hPa; and a third peak at 350 hPa, which is more prominent in April and August but weaker in November and January. Above 350 hPa, concentrations gradually diminish to near zero. These features are broadly consistent with previously reported formaldehyde profile distributions (Zhu et al., 2020). Overall, OMPS exhibits a closer agreement with the model profiles than OMI, whereas OMI shows displaced or poorly defined peaks in several regions, with particularly anomalous enhancements in the upper levels over NCP.

In contrast, the TROPOMI product provides tropospheric columns and displays markedly different vertical structures compared to the other three datasets. Its profiles decrease monotonically with altitude, with only localized perturbations over SCB (around 300 hPa) and PRD (around 800 hPa). The overall trend resembles a logarithmic decay, with especially high near-surface concentrations (Vigouroux et al., 2020). These discrepancies highlight the importance of applying vertical profile corrections in satellite-model comparisons Eskes and Boersma (2003). Specifically, the similarity in vertical structures between OMI, OMPS, and GEOS-Chem supports the use of the AMF approach for correction, while the distinct TROPOMI profiles necessitate the application of AVK.

Uncertainty is a key component in the assimilation process and serves as a crucial indicator of satellite data quality. Figure 3 illustrates the vertical distribution of retrieval uncertainties. In the mid- to upper troposphere (200-600 hPa), OMPS and OMI show comparable levels of uncertainty. However, below 600 hPa, OMPS uncertainties become substantially larger, likely due

to cloud contamination and retrieval algorithm approximations (González Abad et al., 2016; Nowlan et al., 2023). 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~~Supplementary Figure S2, the overall uncertainty of OMPS is significantly higher than that of the other two satellite datasets. At first glance, OMI data may appear superior, but this advantage largely results from strict filtering, which excludes a substantial fraction of problematic data. As illustrated in Supplementary Figure S3 (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 observedin 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°)~~, applying a threshold of 50 observations per grid cell drastically reduces spatial coverage, rendering OMI unsuitable for national-scale assimilation. Previous studies that assimilated OMI over China have typically interpolated the data to coarser resolutions to ensure applicability (Cao et al., 2018; Miyazaki et al., 2020) . Therefore, only OMPS 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 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° × 2° or even 4° × 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° × 5°.~~ TROPOMI formaldehyde columns are assimilated in this study, while OMI is excluded for our high-resolution emission inversion due to the poor data coverage.

Figure 3 also presents satellite retrieval deviations from the prior model estimates. When all three satellite datasets exhibit the same sign of deviation (positive or negative) relative to the model, they are considered consistent. Such consistency is observed, for example, in February, May, and November over NCP and in February over PRD and SCB, where all three datasets show positive deviations; and in February and November over YRD and in August over SCB, where all show negative deviations. These cases indicate stronger reliability. In other situations, when OMPS and TROPOMI exhibit the same bias direction, they are also considered consistent, as in November over PRD and SCB. 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. 10 out of 16 cases (62.5%) exhibit consistency, with higher coherence primarily occurring in the cold season and during spring and autumn months over NCP and SCB. Subsequent analyses will explicitly consider this consistency to enhance the robustness of the conclusions.

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°).

3.2 NMVOC emissions

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. The spatial characteristics of the NMVOC emissions in 2020 are clearly shown in Figure 2 which presents the spatial distribution of four monthly average emissions from the prior simulation (a.1-a.4) and the posterior estimates constrained by OMPS (b.1-b.4) and TROPOMI (c.1-c.4) formaldehyde observations. Significant emission increments relative to the prior estimates are mainly concentrated in eastern and southern China. In most regions, the posterior results constrained by OMPS (d.1-d.4) and TROPOMI (e.1-e.4) exhibit broadly consistent adjustment patterns. However, notable differences between the two posterior estimates can still be observed, particularly over eastern China in August and southern China in May. The results reveal pronounced seasonal variability and regional heterogeneity in emission intensity, with the NCP, YRD, PRD, and SCB identified as major emission hotspots throughout the year.

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).

3.3 NMVOC emissions

NOAA-20 OMPS formaldehyde observations were assimilated, resulting in updated monthly NMVOC emission inventories Although the major high-emission regions can be clearly identified, the complexity of emission source types and the wide range of emission magnitudes render the maps visually dense, making it difficult to directly interpret regional characteristics and seasonal changes. Therefore, subsequent analyses are focused on these four representative regions to enable a more detailed investigation. Figure 4 further displays the monthly and annual totals of NMVOC emissions across China in 2020. To better illustrate the seasonal pattern, the prior and posterior total NMVOC emission inventories for four months(January, April, JulyIn general, the two posterior estimates exhibit good agreement for most months. Specifically, both show consistent decreases during January, February, 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, to December, while simultaneous increases are observed from March to May. But notable discrepancies are observed during the June-September

~~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).~~

Figure 4. Monthly NMVOC emissions in 2020 from the prior simulation (blue) and the posterior simulations constrained by assimilating OMPS (red) and TROPOMI (green) formaldehyde observations. Panels show anthropogenic emissions (a), biogenic emissions (b), biomass burning emissions (c), and total emissions (d). The annual totals for each category are indicated in the legends.

period, which account for approximately 83% of the total annual difference between the two posterior datasets — with July and August alone contributing around 56%. This leads to different estimates of annual emissions: the annual total constrained by OMPS assimilation is estimated at 40.82 Tg, while that constrained by TROPOMI is 34.83 Tg, both differing from the prior estimate of 39.26 Tg. To more accurately assess the regional emission responses under different observational constraints, the consistent and inconsistent months are discussed separately in the ~~prior and the posterior NMVOC emissionsfrom anthropogenic activities, biogenic plus biomass burning are separately plotted in Figure 6.~~ following sections.

~~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~~In months with high consistency, January-February, during the transition from winter to spring, both assimilation results show a reduction in emissions, with anthropogenic emissions exhibiting a particularly significant decline. This may be because the emission inventory includes winter heating emissions, while the actual heating demand has been reduced due to global warming, resulting in an overestimation (Lu et al., 2025; Xu et al., 2025). In spring, March-May, emissions gradually increase, likely driven by enhanced biogenic emissions due to rising temperatures and vigorous vegetation growth (Guenther et al., 1995; Monson et al., 20 . After November, as the season shifts from autumn to winter, emissions decline again, with notable fluctuations in biogenic emissions in October, though anthropogenic emissions remain the primary contributor to the overall trend. In the inconsistent months (June-September, corresponding to summer and autumn), discrepancies arise between OMPS and TROPOMI results. These differences may arise from variations in emission characterization during summer, marked by strong convection, high humidity, and elevated cloud and aerosol content, which differentially impact the retrieval of optical depth and column concentrations by OMPS and TROPOMI. Additionally, differences in AMF and vertical profiles may further contribute to these discrepancies (Lorente et al., 2017; De Smedt et al., 2021).

~~As shown in Figure 6 (c.2, d.4~~

A clear regional divergence in NMVOC emission increments after assimilation is revealed in Figure 5, with further analysis highlighting the sources of these discrepancies. In the NCP, OMPS assimilation results for July indicate a significant emission increase of 62.04%, while TROPOMI shows a smaller increase. In August and September, the two datasets exhibit opposing trends. In contrast, increment differences are relatively minor in March-April, with OMPS assimilation results showing an increase of approximately 14%, while TROPOMI remains largely unchanged. During January-February and November-December, both datasets display minimal changes. Overall, the prior inventory for the NCP appears underestimated in May-June. In the YRD, except for May and July, other months (e.g., June, August, and September) show opposing trends. In the PRD, nearly all months from March to September are classified as inconsistent. However, both regions demonstrate consistent emission reductions during the cold season, suggesting an overestimation in the a priori emission inventory for winter. In the SCB, negative emission increments during the warm season are particularly pronounced, generally exceeding 20%, with

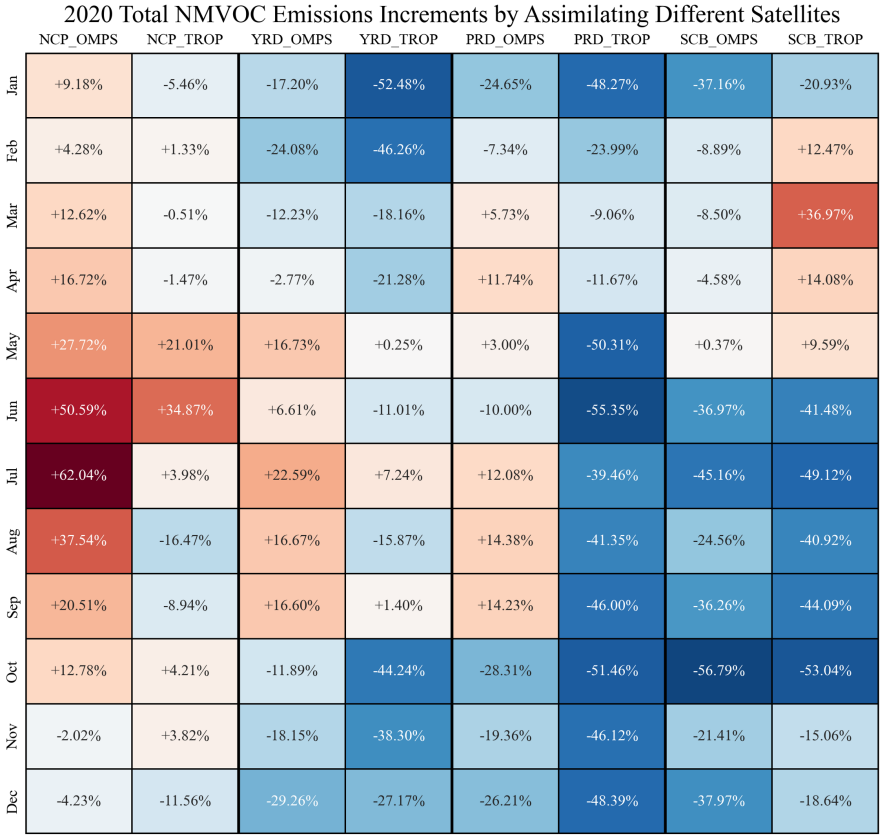

Figure 5. Monthly increments in total NMVOC emissions between the posterior and prior simulations derived from the assimilation of OMPS and TROPOMI formaldehyde observations over four key regions of China: the North China Plain, Yangtze River Delta, Pearl River Delta, and Sichuan Basin in 2020. Positive values indicate an increase in posterior emissions relative to the prior, while negative values indicate a decrease.

reductions in June-July surpassing 35%. During the cold season (January and October-December), ~~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~~ both datasets show consistent declines with comparable values. Months with lower consistency are primarily concentrated in February-May, indicating a likely overestimation in the a priori emission inventory for this region.

In 2020, anthropogenic emissions in China were influenced by the COVID-19 pandemic, leading to observable changes. To better evaluate the general applicability of the proposed method, it is also necessary to conduct a comparative analysis for the pre-pandemic year of 2019. 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,~~ S5 in the Supplement presents the total NMVOC emission increments for the four major regions in 2019, based on data assimilation of OMPS and TROPOMI observations. In the NCP region, strong consistency is again observed in June, with posterior emissions increasing by 57.71% and 30.09% from OMPS and TROPOMI assimilation, respectively, further confirming the underestimation of prior emissions in this period. In the ~~posterior estimates indicate that changesin 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.~~

c

3.3 Formaldehyde ~~total~~ columns evaluation

~~As~~

The spatial distributions of formaldehyde columns in February, May, August, and November 2020 are 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~~ 1. Panels (a.1-a.4) display the prior simulations of tropospheric columns, (b.1-b.4) present the posterior simulations of tropospheric columns assimilated by OMPS, (c.1-c.4) show the OMPS satellite observations of total columns, and (d.1-d.4) illustrate the TROPOMI satellite observations of tropospheric columns. In addition, 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~~ and posterior simulations of total columns for 2020 are also provided in the Supplementary Figure S7. As indicated by the vertical profiles in Figure 3, formaldehyde ~~levels~~ is mainly distributed below the tropopause. Comparisons between the prior and posterior results show that the differences between total and tropospheric columns are relatively small. Regarding the spatial patterns, high formaldehyde values in February are concentrated 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~~ , YRD, and PRD regions, with the posterior results showing an expanded 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,~~ area in the NCP but a reduced coverage in the YRD. In May, overall concentrations increase nationwide, with

particularly pronounced growth in the NCP and PRD. In August, concentrations increase in the NCP, YRD, and PRD, while they decrease in the SCB. In November, the changes are modest, but all four regions exhibit reduced concentrations.

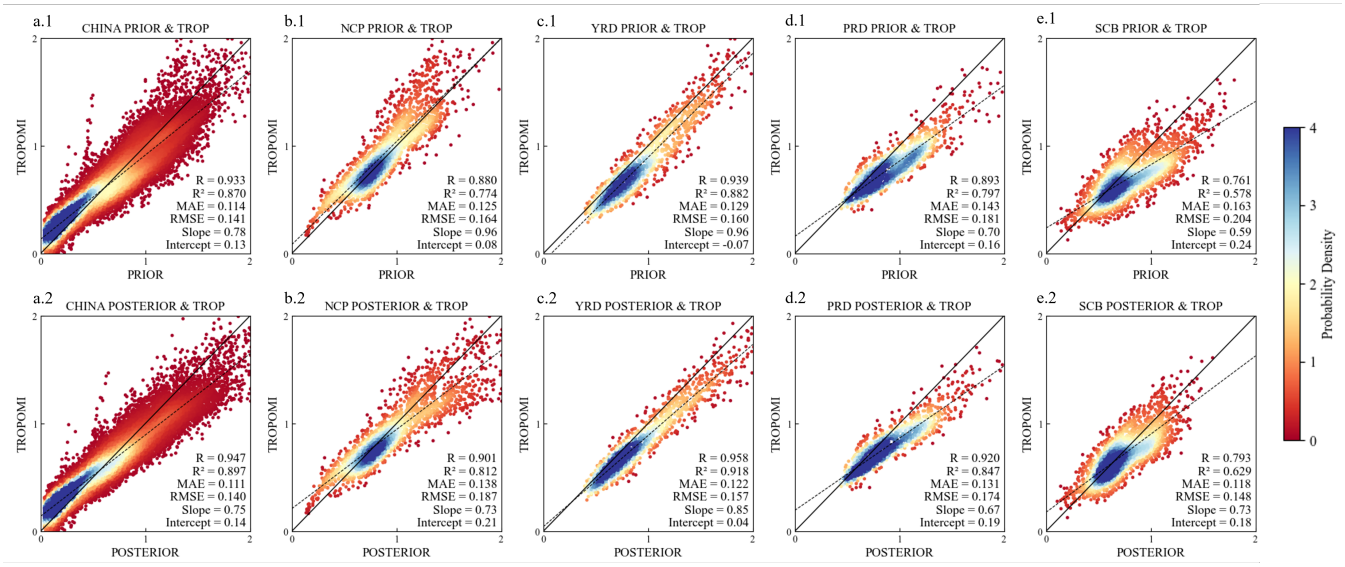

Figure 6. Scatter density plots comparing GEOS-Chem simulated formaldehyde columns with TROPOMI observations in 2020. Panels (a.1-e.1) show comparisons between prior simulations and TROPOMI, while panels (a.2-e.2) show comparisons between posterior simulations constrained by assimilating OMPS observations and TROPOMI. The regions considered are China (a), the North China Plain (b), the Yangtze River Delta (c), the Pearl River Delta (d), and the Sichuan Basin (e). The probability density of the data points is indicated by the color scale. The correlation coefficient (R), coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), regression slope, and intercept are reported in each panel.

The prior and OMPS-driven posterior simulations of formaldehyde tropospheric columns were compared with the TROPOMI formaldehyde tropospheric columns to evaluate the changes in ~~Inner Mongolia, Tibet, and the northwest were minimal. This is because the assigned background uncertainty, which is proportional to~~ formaldehyde. Scatter plots together with statistical metrics (R^2, R, MAE, and RMSE) for 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 below 5 $\times 10^{-4} kg/m^2$, and the actual values for some grid points in these regions are even less than~~ whole country and four subregions in 2020 are presented in Figure 6. The prior simulation already shows reasonably good performance (a.1-e.1), with most points distributed close to the 1 $\times 10^{-4} kg/m^2$. ~~Though assimilating OMPS observations~~;1 line and exhibiting strong correlations with observations. Nevertheless, further improvements are still possible. After assimilating OMPS data, the posterior ~~simulation in these regionsstill remains low, resulting in little change. As illustrated in Figure 7 (a.2, a.3)~~results compared with TROPOMI show higher R^2 values across all regions, ~~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~~ indicating strengthened correlations. For China and NCP, the improvements are comparable, with R^2 increasing by

about 0.027 (from 0.870 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~~ 0.897 for China, and from 0.774 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,~~ 0.812 for NCP). In the YRD, the improvement is more pronounced, with R^2 rising from 0.882 to 0.918, and the scatter around the regression line substantially reduced, with many outliers corrected. The most significant improvements occur in PRD and SCB, where R^2 increases by approximately 0.05. In these regions, the overestimations present in the prior simulations are effectively mitigated, particularly for high-value cases. In terms of RMSE and MAE, decreases are observed in all regions except NCP. A comparison between Figures (b.1) and (b.2) indicates improvements in the low- 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~~ mid-value ranges, whereas substantial overestimations appear in the high-value range. This issue is likely related to the instrumental errors of OMPS observations, as discussed in Sections 2.3.1 and 3.2, which introduce considerable uncertainties.

The monthly mean formaldehyde columns for 2020, derived from the prior simulation, posterior simulations constrained by the OMPS and TROPOMI observations, and satellite observations from OMPS and TROPOMI, are presented for China (a), NCP (b), YRD (c), PRD (d), and SCB (e) in Figure 7. At the national scale, the overall changes resulting from assimilation are relatively modest, with the main adjustments occurring in summer and early autumn. The OMPS-driven posterior results show ~~a significant improvement compared~~ increases relative 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~~in June-July, whereas the TROPOMI-assimilated results exhibit decreases in July-August compared to the prior. These discrepancies may be attributed to differences between the satellite products in their responses to biogenic emissions and photochemical processes under high-temperature and high-radiation conditions (De Smedt et al., 2018; Vigouroux et al., 2020)
.

Figure 8 (a) presents the increments between the posterior and prior simulations over the four regions when assimilating OMPS or TROPOMI observations, respectively. In the NCP, the posterior results constrained by both OMPS and TROPOMI show consistent increases in May-June, suggesting that the prior inventory may have underestimated the contributions from active photochemical production and anthropogenic emissions during summer (Wells et al., 2020). In the YRD and PRD, stronger consistency is observed in the cold season (January-March and October-December), with both posterior results showing decreases, which is consistent with reduced anthropogenic activity and lower formaldehyde production rates under wintertime conditions. The SCB exhibits more distinct characteristics, with OMPS and TROPOMI assimilation results consistently showing decreases in the second half of the year, particularly pronounced in June, July, and October. This pattern suggests that the prior emissions in this region were overestimated. Previous studies have highlighted that the large uncertainties in biogenic emissions in the SCB are critical factors influencing the accuracy of NMVOC emissions and simulations (Ma et al., 2019).

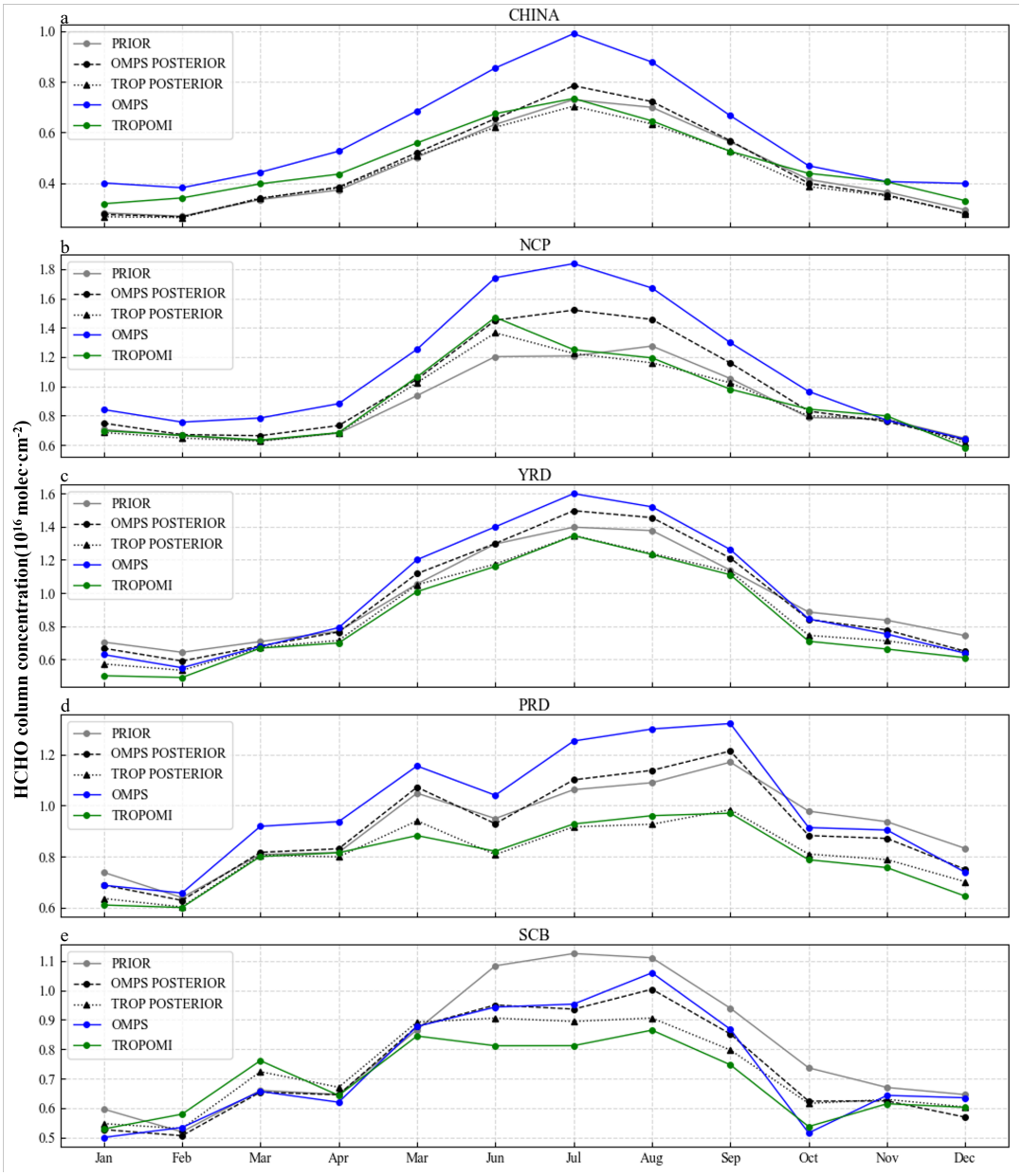

Figure 7. Monthly mean formaldehyde column concentrations in 2020 from the prior simulation (gray), posterior simulations constrained by assimilating OMPS (black dashed) and TROPOMI (black dotted) observations, and satellite observations from OMPS (blue) and TROPOMI (green). Panels show results over China (a), the North China Plain (b), the Yangtze River Delta (c), the Pearl River Delta (d), and the Sichuan Basin (e).

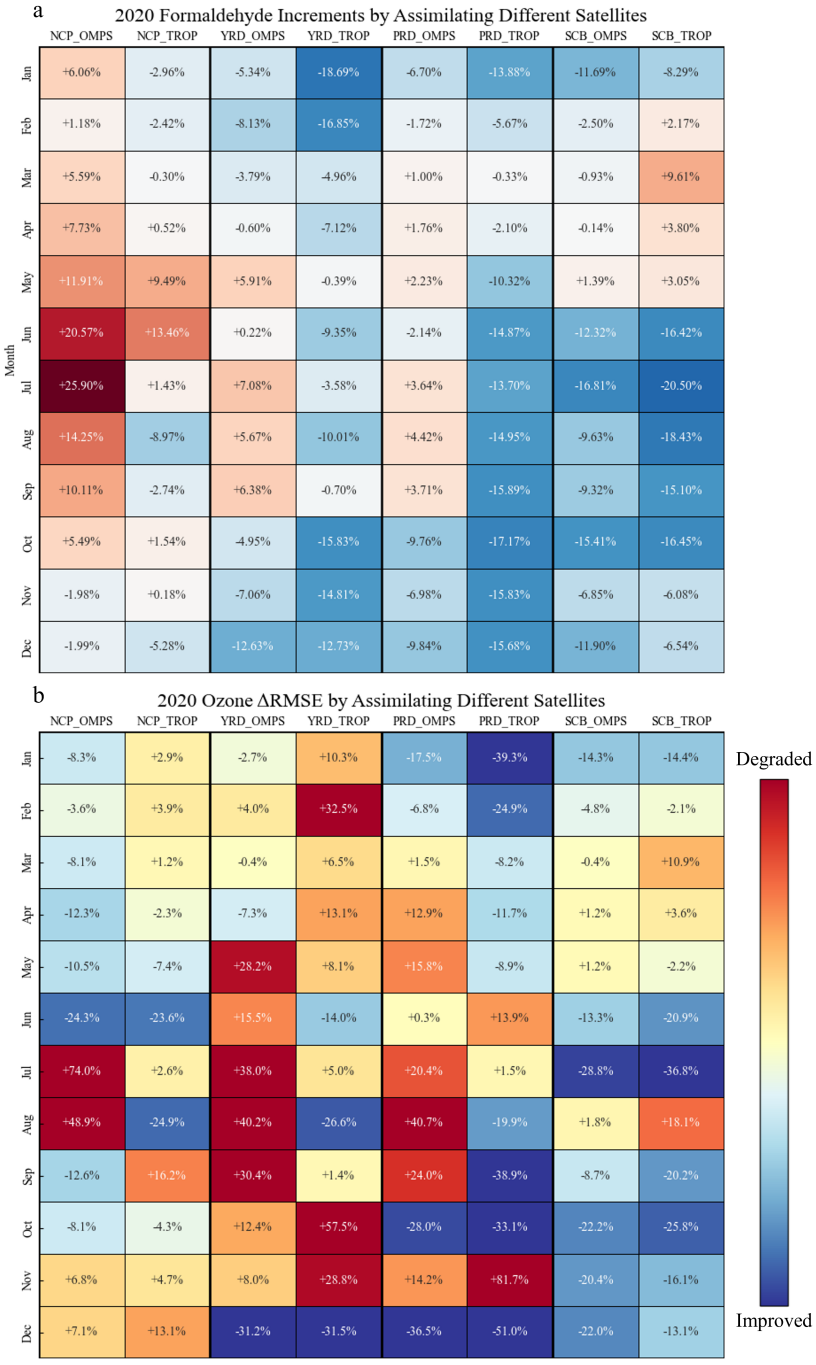

Figure 8. Monthly increments in (a) formaldehyde columns between posterior and prior simulations and (b) the relative changes in MDA8 ozone RMSE (ΔRMSE) after assimilating OMPS and TROPOMI observations in 2020. Results are shown for the North China Plain, Yangtze River Delta, Pearl River Delta, and Sichuan Basin. Positive values indicate an increase relative to the prior, while negative values indicate a decrease.

~~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.~~

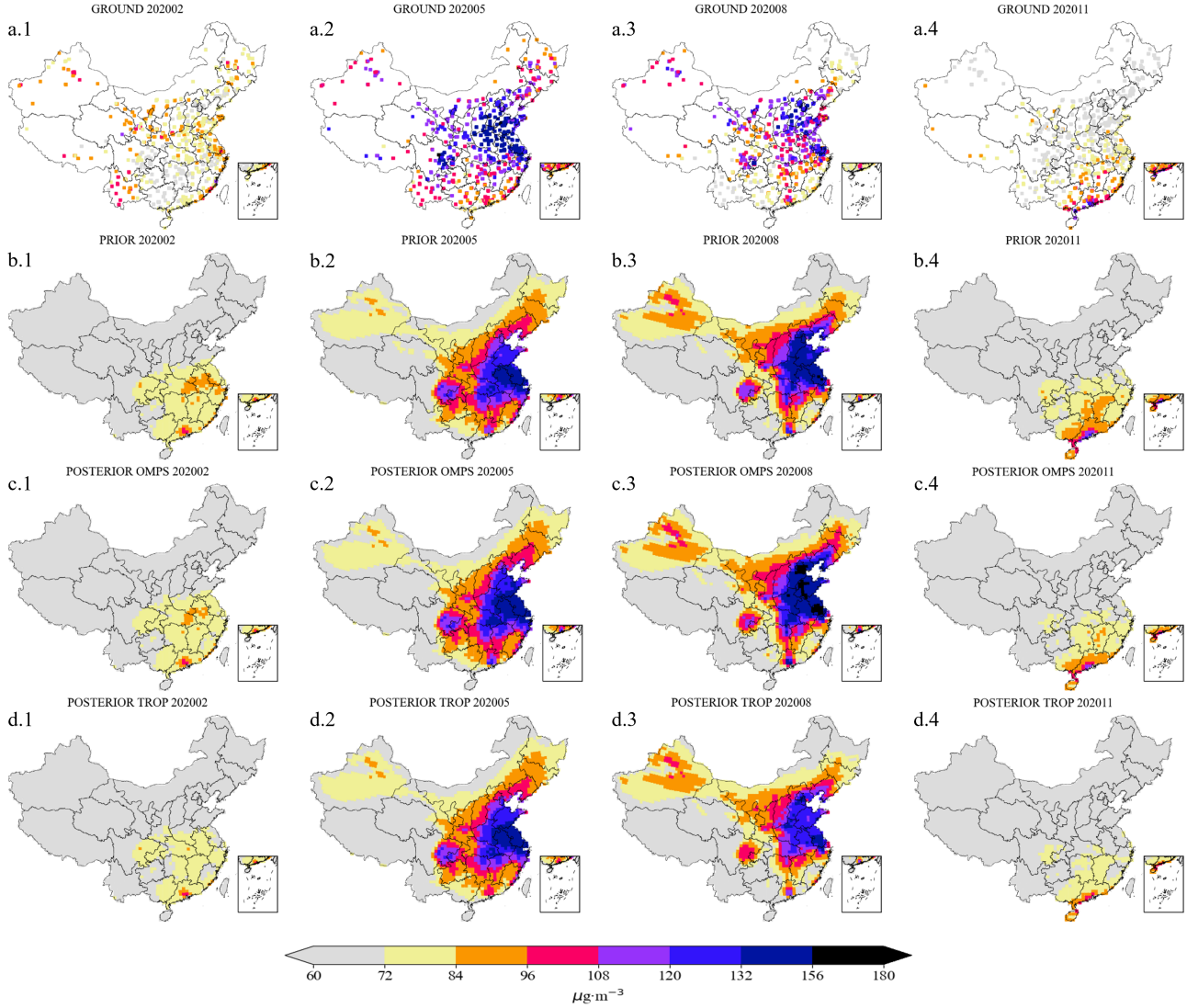

Figure 9. Spatial distributions of surface ozone concentrations in February, May, August, and November 2020. Panels (a.1-a.4) show ground-based observations, panels (b.1-b.4) show prior simulations, panels (c.1-c.4) show posterior simulations constrained by assimilating OMPS formaldehyde observations, and panels (d.1-d.4) show posterior simulations constrained by assimilating TROPOMI formaldehyde observations.

3.4 Impact of formaldehyde assimilation on ozone surface concentration

3.5 ~~Impact of Formaldehyde Assimilation on O_3 Surface Concentration~~

~~Formaldehyde is akey 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~~ The spatial distributions of observed 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)~~ozone at ground stations (a.1-a.4), together with the prior (b.1-b.4) and posterior simulations based on OMPS and TROPOMI assimilation (c.1-c.4, d.1-d.4), are shown in Figure 9. As shown in ~~Figure 9 (a), the frequency distribution histogram clearly demonstrates~~ panels (b.1-b.4), pronounced ozone hotspots are observed in NCP (February, May, and August), YRD (May and August), PRD (May, August, and November), and SCB (May and August). This is very similar to the observations shown in panels (a.1-a.4). It indicates 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~~ simulation captures the general patterns of ozone hotspots reasonably well, but notable biases remain. For example, ozone is clearly overestimated in PRD during February, May, and August, while underestimated in SCB during May and August. After assimilation with OMPS or TROPOMI, the ~~NCP region. As can be seen in Figure 9~~ posterior MDA8 ozone simulations retain the overall hotspot distribution, but the direction and magnitude of changes vary by region. For instance, in August, ozone concentrations increase in NCP and PRD with OMPS assimilation but decrease with TROPOMI assimilation. In February, both assimilation results decrease in YRD, although the decrease is more pronounced in the TROPOMI-based results. Moreover, many regional changes are difficult to discern visually from the spatial maps alone, highlighting the necessity of using statistical metrics to quantitatively assess ozone variations.

The RMSE values between the simulated MDA8 ozone and the ground-based observations are calculated. To better visualize the assimilation benefits, the RMSE variation either assimilating the TROPOMI or assimilating the OMPS in the four key regions are also shown in Figure 8 (b). Larger decreases in RMSE (darker blue) indicate more significant improvements, with the posterior ozone being closer to ground-based observations; conversely, larger increases in RMSE (darker red) indicate degraded performance, with the posterior ozone diverging further from the observations. In those inconsistent cases where the OMPS and TROPOMI posterior increments exhibit opposite signs (i.e., one increases while the other decreases), ozone simulation improvement is not guaranteed. For instance, in NCP during January-April and July, in YRD during June and

September, and in PRD during April, May, August, and September, one assimilation leads to improvement while the other indicates deterioration. Moreover, in several additional months both posteriors even show degradation, making it difficult to effectively evaluate the improvement in posterior ozone simulations. By contrast, ozone simulation improvements are clearly observed in consistent cases where the OMPS- and TROPOMI-constrained posteriors exhibit the same sign (i.e., both reductions in ΔRMSE). In NCP, ~~the RMSE (root mean square error) between the observed and simulated values decreasesfrom 24.44 $\mu g/m^3$ to 22.79 $\mu g/m^3$ for China and from 28.0 μg~~substantial improvements are observed in May and June, with the largest RMSE decrease in June, in agreement with the high-consistency pattern shown in Figure 8 (a). In YRD and PRD, RMSE decreases by more than 30% in December, representing the most significant improvement; in addition, PRD also shows clear improvements in January and October. These improvement months all correspond to periods of high consistency. In SCB, RMSE also decreases markedly during high-consistency months, including January, June, July, and September-December.

To further quantify ozone simulation improvements in consistent regions, statistics were performed for the months classified as consistent. Considering the similarity in monthly behavior between YRD and PRD, the two regions were combined in the analysis. The results indicate that the consistent regions include NCP in May-June, YRD/~~m^3 to 20.22 μg~~PRD in January-March and October-December, and SCB in January and June-December. Within these regions, except for March and November in YRD/~~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. 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). 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).~~ PRD and August in SCB, all other months show ozone simulation improvements. Overall, 13 out of the 16 consistent months exhibit improvements, accounting for 81.25%, with an average RMSE reduction of 24.7%. This result suggests that constraining NMVOC emissions through formaldehyde assimilation not only substantially improves formaldehyde simulations, but also exerts a positive impact on ozone simulations, with particularly significant improvements in regions and months characterized by high consistency.

To more robustly substantiate this conclusion, it is necessary to examine whether similar features can also be identified in 2019. In that year, OMPS and TROPOMI satellite observations were assimilated independently to constrain NMVOC emissions. The posterior-prior increments from the OMPS- and TROPOMI-driven assimilations, together with the changes in MDA8 ozone ΔRMSE, are presented in Figure S6 of the Supplement. In NCP, March, May, and June are identified as consistent months, during which the ozone RMSE values decrease, with the most pronounced improvement occurring in June. In YRD, the consistent months are February, October, and November, where the ozone improvements are relatively limited but nevertheless show better agreement with ground-based observations. In PRD, the consistent months include January, February, and June-December; with the exception of August, September, and November, the ozone RMSE decreases in the other months,

with notable improvements in June and July. In SCB, the two posterior datasets exhibit the highest level of consistency in 2019, with synchronous increases and decreases throughout the year. Ozone simulations in this region show better performance in all months except March and April, with particularly large improvements in June, July, and September-November, when the RMSE decreases by an average of 25.74%.

Across the four regions, 27 months are classified as consistent in 2019. Of these, 22 months exhibit improved ozone simulations, which corresponds to 81.48% of all consistent months, with both assimilations producing MDA8 ozone values closer to ground-based observations. This proportion differs from that of 2020 by only 0.23%, providing further evidence that ozone improvements are particularly significant in the months defined as consistent across the four regions.

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.~~ satellite-based formaldehyde retrievals from OMPS and TROPOMI were assimilated to constrain NMVOC emissions over China in 2020. The results demonstrate that assimilation corrects systematic biases in prior emission inventories and improves the simulation of formaldehyde columns and surface ozone in general. More importantly, by analyzing the consistency of posterior results via assimilating different formaldehyde products across regions and months, this work establishes a methodological framework to further assess the reliability of emission estimates with multiple satellite constraints.

~~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 observationsat grid resolutions of 0.5°×0.5°, 1°×1°, 2°×2°, and 4°×4° 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.~~ At the national scale, the OMPS- and TROPOMI-constrained posterior NMVOC emissions are broadly consistent across most months, with decreases in January-February and October-December and increases in March-May. The winter-spring decreases likely reflect overestimation of heating emissions in the prior inventory under reduced heating demand, whereas the spring increases are attributable to enhanced biogenic activity with rising temperatures. By contrast, notable discrepancies emerge in June-September—dominated by July-August—likely linked to strong convection,

high humidity, and elevated cloud/aerosol loading that differentially affect retrievals, compounded by AMF and vertical-profile differences. These discrepancies result in annual totals of 40.82 Tg (OMPS) and 34.83 Tg (TROPOMI), compared with 39.26 Tg in the prior. Regionally, NCP indicates prior underestimation in May-June consistently but pronounced divergences in July-September; YRD and PRD show warm-season inconsistencies but consistent cold-season decreases, suggesting wintertime overestimation in the prior inventory; SCB features substantial summer decreases (exceeding 20%, particularly in June-July) alongside consistent winter decreases, while several spring months also point to possible prior overestimation.

~~The monthly NMVOC emission inventories in China were then estimated by assimilating OMPS formaldehyde column observations. In general, the~~ Both the prior and 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~~ simulations capture the spatial distribution of the formaldehyde columns well. When comparing the prior simulation and the posterior simulation constrained by OMPS with TROPOMI satellite observations, ~~with the RMSE decreasing from 0.49 to 0.45 $\times 10^{16}$ 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 $\times 10^{16}$ 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~~ the prior already shows strong correlations, but further improvements are achieved after OMPS assimilation. The R^2 increases from 0.870 to 0.897 at the national scale and from 0.774 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~~ 0.812 in NCP; in YRD, the increase is larger, from 0.882 to 0.918; while the largest improvements are observed in PRD and SCB, with increases of about 0.05. Meanwhile, RMSE and MAE decrease in all regions except NCP. In NCP, the simulations improve in the low-to-middle value ranges, but overestimations remain in the high-value range, likely due to the large uncertainties introduced by OMPS instrumental errors.

~~Optimizing NMVOC emission inventories with formaldehyde observations resulted in~~ For ozone, comparison with surface MDA8 observations highlights 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 approachare 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.~~ high-consistency regions. In NCP, RMSE reductions are most pronounced in June, consistent with the strong emission and formaldehyde adjustments in this period. In YRD and PRD, December RMSE reductions exceed 30%, while additional improvements are found in PRD during January and October. In SCB, assimilation leads to persistent improvements from January through December, with notable reductions in June-July and the late autumn months. Overall, ozone improvements are observed in 13 of the 16 consistent months, which represent 81.25% of the total, with an average RMSE reduction of 24.7%.

To further test the robustness of our approach, OMPS and TROPOMI satellite observations were independently assimilated to constrain NMVOC emissions for 2019 (Figure S4). The spatial distribution of formaldehyde hotspots is similar to 2020 but with overall higher concentrations. At the regional scale, most consistent months between OMPS- and TROPOMI-constrained results indicate that the prior inventory underestimates emissions in NCP and overestimates them in YRD, PRD, and SCB. Importantly, 22 of the 27 consistent months (81.48%) show reduced ozone RMSE, with the largest improvements in SCB, confirming that consistent cases are strongly associated with enhanced ozone simulation performance. These findings also lend greater confidence to the optimized NMVOC emissions during the consistent months in these regions.

Future efforts should reassess assimilation performance with updated emission inventories and incorporate source-specific uncertainties, assigning different uncertainties to anthropogenic, biogenic, and biomass burning sectors, in order to better constrain their respective emissions. Moreover, because no independent validation data such as aircraft or FTIR measurements were available over China in 2020, future studies could further evaluate the assimilation results once such observational datasets become accessible.

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], and the Postgraduate Research & Practice Innovation Program of Jiangsu Province [grant number KYCX24_1528].

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.

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.

Boersma, K., Eskes, H., and Brinksma, E.: Error analysis for tropospheric NO2 retrieval from space, Journal of Geophysical Research: Atmospheres, 109, 2004.

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.

Cheng, J., Zhang, Y., Wang, T., Norris, P., Chen, W.-Y., and Pan, W.-P.: Thermogravimetric–Fourier transform infrared spectroscopy–gas chromatography/mass spectrometry study of volatile organic compounds from coal pyrolysis, Energy & Fuels, 31, 7042–7051, 2017.

Choi, J., Henze, D. K., Cao, H., Nowlan, C. R., González Abad, G., Kwon, H.-A., Lee, H.-M., Oak, Y. J., Park, R. J., Bates, K. H., et al.: An inversion framework for optimizing non-methane VOC emissions using remote sensing and airborne observations in Northeast Asia during the KORUS-AQ field campaign, Journal of Geophysical Research: Atmospheres, 127, e2021JD035 844, 2022.

Clarisse, L., Franco, B., Van Damme, M., Di Gioacchino, T., Hadji-Lazaro, J., Whitburn, S., Noppen, L., Hurtmans, D., Clerbaux, C., and Coheur, P.: The IASI NH 3 version 4 product: averaging kernels and improved consistency, Atmospheric Measurement Techniques Discussions, 2023, 1–31, 2023.

Cooper, M. J., Martin, R. V., Henze, D. K., and Jones, D.: Effects of a priori profile shape assumptions on comparisons between satellite NO 2 columns and model simulations, Atmospheric Chemistry and Physics, 20, 7231–7241, 2020.

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.

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.

Eskes, H. and Boersma, K.: Averaging kernels for DOAS total-column satellite retrievals, Atmospheric Chemistry and Physics, 3, 1285–1291, 2003.

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.

González Abad, G., Vasilkov, A., Seftor, C., Liu, X., and Chance, K.: Smithsonian astrophysical observatory ozone mapping and profiler suite (SAO OMPS) formaldehyde retrieval, Atmospheric Measurement Techniques, 9, 2797–2812, 2016.

Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., McKay, W., et al.: A global model of natural volatile organic compound emissions, Journal of Geophysical Research: Atmospheres, 100, 8873–8892, 1995.

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, M., Zheng, J., Yin, S., and Zhang, Y.: Trends, temporal and spatial characteristics, and uncertainties in biomass burning emissions in the Pearl River Delta, China, Atmospheric Environment, 45, 4051–4059, 2011.

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.

Hua, W., Lou, S., Huang, X., Xue, L., Ding, K., Wang, Z., and Ding, A.: Diagnosing uncertainties in global biomass burning emission inventories and their impact on modeled air pollutants, Atmospheric Chemistry and Physics, 24, 6787–6807, 2024.

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., Segers, A., Lin, H. X., Henzing, B., Wang, X., Heemink, A., and Liao, H.: Position correction in dust storm forecasting using LOTOS-EUROS v2.1: grid-distorted data assimilation v1.0, Geoscientific Model Development, 14, 5607–5622, https://doi.org/10.5194/gmd-14-5607-2021, https://gmd.copernicus.org/articles/14/5607/2021/, 2021.

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.

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.

Jung, J., Choi, Y., Mousavinezhad, S., Kang, D., Park, J., Pouyaei, A., Ghahremanloo, M., Momeni, M., and Kim, H.: Changes in the ozone chemical regime over the contiguous United States inferred by the inversion of NOx and VOC emissions using satellite observation, Atmospheric research, 270, 106 076, 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, L., Yang, W., Xie, S., and Wu, Y.: Estimations and uncertainty of biogenic volatile organic compound emission inventory in China for 2008–2018, Science of the Total Environment, 733, 139 301, 2020.

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., Xiao, Q., and Wang, B.: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test, Monthly Weather Review, 136, 3363–3373, 2008.

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.

Lorente, A., Folkert Boersma, K., Yu, H., Dörner, S., Hilboll, A., Richter, A., Liu, M., Lamsal, L. N., Barkley, M., De Smedt, I., et al.: Structural uncertainty in air mass factor calculation for NO 2 and HCHO satellite retrievals, Atmospheric Measurement Techniques, 10, 759–782, 2017.

Lu, X., Zhang, D., Wang, L., Wang, S., Zhang, X., Liu, Y., Chen, K., Song, X., Yin, S., Zhang, R., et al.: Establishment and verification of anthropogenic speciated VOCs emission inventory of Central China, Journal of Environmental Sciences, 149, 406–418, 2025.

Ma, M., Gao, Y., Wang, Y., Zhang, S., Leung, L. R., Liu, C., Wang, S., Zhao, B., Chang, X., Su, H., et al.: Substantial ozone enhancement over the North China Plain from increased biogenic emissions due to heat waves and land cover in summer 2017, Atmospheric Chemistry and Physics, 19, 12 195–12 207, 2019.

Miyazaki, K., Bowman, K. W., Yumimoto, K., Walker, T., and Sudo, K.: Evaluation of a multi-model, multi-constituent assimilation framework for tropospheric chemical reanalysis, Atmospheric Chemistry and Physics, 20, 931–967, 2020.

Monson, R. K., Grote, R., Niinemets, Ü., and Schnitzler, J.-P.: Modeling the isoprene emission rate from leaves, New Phytologist, 195, 541–559, 2012.

Nowlan, C. R., González Abad, G., Kwon, H.-A., Ayazpour, Z., Chan Miller, C., Chance, K., Chong, H., Liu, X., O'Sullivan, E., Wang, H., et al.: Global formaldehyde products from the Ozone Mapping and Profiler Suite (OMPS) nadir mappers on Suomi NPP and NOAA-20, Earth and Space Science, 10, e2022EA002 643, 2023.

Oomen, G.-M., Müller, J.-F., Stavrakou, T., De Smedt, I., Blumenstock, T., Kivi, R., Makarova, M., Palm, M., Röhling, A., Té, Y., et al.: Weekly derived top-down volatile-organic-compound fluxes over Europe from TROPOMI HCHO data from 2018 to 2021, Atmospheric Chemistry and Physics, 24, 449–474, 2024.

Palmer, P. I., Jacob, D. J., Chance, K., Martin, R. V., Spurr, R. J., Kurosu, T. P., Bey, I., Yantosca, R., Fiore, A., and Li, Q.: Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the Global Ozone Monitoring Experiment, Journal of Geophysical Research: Atmospheres, 106, 14 539–14 550, 2001.

Palmer, P. I., Jacob, D. J., Fiore, A. M., Martin, R. V., Chance, K., and Kurosu, T. P.: Mapping isoprene emissions over North America using formaldehyde column observations from space, Journal of Geophysical Research: Atmospheres, 108, 2003.

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.

Pei, D., Wang, A., Shen, L., and Wu, J.: Research on the Emission of Biogenic Volatile Organic Compounds from Terrestrial Vegetation, Atmosphere, 16, 885, 2025.

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.

Sakdapipanich, J. and Insom, K.: High-resolution gas chromatography-mass spectrometry, KGK Kautschuk Gummi Kunststoffe, 59, 382–387, 2006.

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., Choi, Y., Jeon, W., Woo, J.-H., Zhang, Q., and Kurokawa, J.-i.: Remote sensing evidence of decadal changes in major tropospheric ozone precursors over East Asia, Journal of Geophysical Research: Atmospheres, 122, 2474–2492, 2017.

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.

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, Atmospheric Chemistry and Physics, 21, 4825–4848, 2021.

Wells, K., Millet, D., Payne, V., Vigouroux, C., Aquino, C., De Mazière, M., de Gouw, J., Graus, M., Kurosu, T., Warneke, C., et al.: Next-generation isoprene measurements from space: Detecting daily variability at high resolution, Journal of Geophysical Research: Atmospheres, 127, e2021JD036 181, 2022.

Wells, K. C., Millet, D. B., Payne, V. H., Deventer, M. J., Bates, K. H., de Gouw, J. A., Graus, M., Warneke, C., Wisthaler, A., and Fuentes, J. D.: Satellite isoprene retrievals constrain emissions and atmospheric oxidation, Nature, 585, 225–233, 2020.

Wiedinmyer, C., Akagi, S., Yokelson, R. J., Emmons, L., Al-Saadi, J., Orlando, J., and Soja, A.: The Fire INventory from NCAR (FINN): A high resolution global model to estimate the emissions from open burning, Geoscientific Model Development, 4, 625–641, 2011.

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.

Wu, N., Geng, G., Xu, R., Liu, S., Liu, X., Shi, Q., Zhou, Y., Zhao, Y., Liu, H., Song, Y., et al.: Development of a high-resolution integrated emission inventory of air pollutants for China, Earth System Science Data, 16, 2893–2915, 2024.

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.

Xing, Y., Wang, G., Zhang, T., Shen, F., Meng, L., Wang, L., Li, F., Zhu, Y., Zheng, Y., He, N., et al.: VOC detections with optical spectroscopy, Prog. Electromagn. Res, 173, 71–92, 2022.

Xu, Y., Si, D., Cao, L., Yu, Y., Wu, X., and Chen, S.: The changes in heating and cooling energy demand in Beijing and Shanghai under global warming, Journal of Applied Meteorology and Climatology, 64, 413–424, 2025.

Xue, R., Wang, S., Li, D., Zou, Z., Chan, K. L., Valks, P., Saiz-Lopez, A., and Zhou, B.: Spatio-temporal variations in NO2 and SO2 over Shanghai and Chongming Eco-Island measured by Ozone Monitoring Instrument (OMI) during 2008–2017, Journal of Cleaner Production, 258, 120 563, 2020.

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.

Zhu, L., Jacob, D. J., Kim, P. S., Fisher, J. A., Yu, K., Travis, K. R., Mickley, L. J., Yantosca, R. M., Sulprizio, M. P., De Smedt, I., et al.: Observing atmospheric formaldehyde (HCHO) from space: validation and intercomparison of six retrievals from four satellites (OMI, GOME2A, GOME2B, OMPS) with SEAC 4 RS aircraft observations over the southeast US, Atmospheric chemistry and physics, 16, 13 477–13 490, 2016.

Zhu, L., Jacob, D. J., Keutsch, F. N., Mickley, L. J., Scheffe, R., Strum, M., González Abad, G., Chance, K., Yang, K., Rappenglück, B., et al.: Formaldehyde (HCHO) as a hazardous air pollutant: Mapping surface air concentrations from satellite and inferring cancer risks in the United States, Environmental Science & Technology, 51, 5650–5657, 2017.

Zhu, L., González Abad, G., Nowlan, C. R., Chan Miller, C., Chance, K., Apel, E. C., DiGangi, J. P., Fried, A., Hanisco, T. F., Hornbrook, R. S., et al.: Validation of satellite formaldehyde (HCHO) retrievals using observations from 12 aircraft campaigns, Atmospheric Chemistry and Physics, 20, 12 329–12 345, 2020.