Signal-domain guided deep learning for gap-filling of XCO and XCH_4: a masked spatio-temporal fusion of TROPOMI and GEOS-Chem (2019–2023)

Chengkun An[1], Yuan Tian[1,2], Zhiwei Li[1], Qiaoyu Jiang[1], Peize Lin[4], Bowen Chang[1], Jingkai Xue[4], and Youwen Sun[2,3]

[1]Institutes of Physical Science and Information Technology, Anhui University, Hefei 230601, China
[2]National Key Laboratory of Opto-Electronic Information Acquisition and Protection Technology, Institutes of Physical Science and Information Technology, Anhui University, Hefei, Anhui, TS1China
[3]Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, HFIPS, Chinese Academy of Sciences, Hefei 230031, China
[4]School of Environmental Science and Optoelectronic Technology, University of Science and Technology of China, Hefei 230026, China

Correspondence: Yuan Tian (ytian@ahu.edu.cn) and Youwen Sun (ywsun@ahu.edu.cn)

Received: 23 December 2025 – Discussion started: 27 January 2026
Revised: 31 August 2026 – Accepted: 1 September 2026 – Published:

Abstract. Long-term, high-resolution monitoring of carbon monoxide (CO) and methane (CH_4) is essential for understanding their spatiotemporal variability and supporting climate mitigation strategies. However, satellite observations from instruments such as the TROPOspheric Monitoring Instrument (TROPOMI) are often spatially and temporally incomplete, while existing fusion methods still struggle to achieve both high accuracy and spatiotemporal continuity. Here, we propose a signal-domain fusion approach that combines three-dimensional discrete cosine transform (3D DCT) and singular value decomposition (SVD) to integrate TROPOMI retrievals with GEOS-Chem simulations. A lightweight residual U-Net is further employed to refine the initial reconstruction by learning residual fields from GEOS-Chem simulations and DCT/SVD reconstruction outputs, guided by a masked loss. The method generates global 0.25° and China-specific 0.05° daily gap-free XCO and XCH_4 datasets from 2019 to 2023. In the time-series comparison analysis at representative sites, the fused datasets generally follow the temporal variations observed by the Total Carbon Column Observing Network (TCCON) and TROPOMI. Missing-rate-threshold experiments further show that the fused products perform comparably to original TROPOMI retrievals under low and moderate missing-rate conditions and show improved performance under sparse TROPOMI coverage (MR $>$ 0.5), with R^2 values of 0.91 for XCO and 0.83 for XCH_4, along with reduced biases and standard deviations. The fused datasets also capture regional XCO increases in parts of North America, decreases over eastern China, and widespread XCH_4 growth, wildfire-related enhancements in Chongqing in 2022, and short-term variations over rice-growing regions. These results indicate that the proposed framework can reconstruct missing satellite observations with improved continuity and provide useful fused datasets for studying regional variability, event-related enhancements, and atmospheric composition changes. The generated datasets are publicly available at https://doi.org/10.5281/zenodo.22010891 (An et al., 2026).

1 Introduction

Carbon monoxide (CO) and methane (CH_4) are two critical atmospheric gases that play a key role in air quality monitoring and climate change research. CO is primarily produced through incomplete combustion of fossil fuels and biomass, as well as the oxidation of CH_4 and other hydrocarbons. Its main sink is the reaction with hydroxyl radicals (OH), which contributes to the formation of tropospheric ozone (O_3) and greenhouse gases such as carbon dioxide (CO_2) (Lelieveld et al., 2016; Spivakovsky et al., 2000). Due to its atmospheric lifetime of weeks to months, CO serves as an effective tracer for pollutant transport, facilitating the study of both horizontal and vertical atmospheric movements (Heald et al., 2003). CH_4 is a powerful greenhouse gas with a significantly longer atmospheric lifespan – around ten years – and a greater global warming potential compared to numerous other gases (Filonchyk et al., 2024; Heilig, 1994). It originates from both natural sources, such as wetlands, permafrost, and wildfires, and anthropogenic activities, including agriculture, livestock digestion, landfills, and fossil fuel extraction (Chai et al., 2016; Jackson et al., 2020). CH_4 plays a crucial role in atmospheric chemistry, influencing the oxidative capacity of the atmosphere and contributing to tropospheric ozone formation. Its rising concentration is a major driver of climate change (Lelieveld et al., 2016).

Although CO and CH_4 are often discussed separately in terms of their sources and impacts, they are chemically and dynamically coupled in the atmosphere. Both gases are predominantly removed by the hydroxyl radical (OH), meaning that changes in one can affect the atmospheric lifetime and concentration of the other through OH competition (Lelieveld et al., 2016). Moreover, they share major emission sources, particularly biomass burning and fossil-fuel-related activities such as incomplete combustion and flaring, where they can be emitted simultaneously (Andreae and Merlet, 2001; Akagi et al., 2011; Andreae, 2019). Their co-emission and co-transport imply that the spatial patterns of CO and CH_4 may be correlated near fire hotspots, industrial regions, and other combustion-influenced areas. Therefore, joint reconstruction of XCO and XCH_4 offers two key advantages: (1) shared spatial structure and common meteorological drivers can provide complementary constraints for deep learning models, improving generalization in data-sparse regions; (2) simultaneous monitoring can support more consistent estimation of emission ratios and source attribution, as demonstrated by studies using CH_4/CO or related trace-gas ratios to constrain biomass-burning and fossil-fuel methane emissions (Worden et al., 2013; Worden et al., 2017; Ramsden et al., 2022). A comprehensive understanding of these coupled gases is fundamental to improving air quality predictions and formulating climate mitigation policies.

Global satellite monitoring enables effective tracking and analysis of the sources, transport, and removal processes of atmospheric pollutants. At present, several satellite instruments provide observations of CO and/or CH_4, including the Measurements of Pollution in the Troposphere (MOPITT), which primarily focuses on CO retrievals (Deeter et al., 2003), as well as the Atmospheric Infrared Sounder (AIRS) (McMillan et al., 2005), Tropospheric Emission Spectrometer (TES) (Rinsland et al., 2006), and Interferometer for the Exploration of the Atmosphere in the Infrared (IASI) (Turquety et al., 2004), which are capable of retrieving multiple trace gases, including CO and CH_4, using different spectral channels and retrieval algorithms. These instruments offer critical data support for the global distribution of CO and CH_4 in the atmosphere.

The TROPOspheric Monitoring Instrument (TROPOMI) onboard the European Space Agency's Sentinel-5P mission provides higher daily global coverage and finer spatial resolution than previous observing missions. According to the official Sentinel-5P/TROPOMI Level-2 product table, the CO and CH_4 products are provided as total column products. In this study, XCO and XCH_4 refer to the column-averaged dry-air mole fractions of atmospheric CO and CH_4, respectively, derived from the corresponding TROPOMI total column products. TROPOMI has already been employed in numerous pertinent applications for XCO and XCH_4 investigations, including the calculation of emissions from biomass combustion (Goudar et al., 2023; Griffin et al., 2024). However, the continuity of TROPOMI atmospheric column retrievals is affected by clouds and aerosols, which can attenuate or scatter the reflected solar radiation measured by satellite sensors. As a result, observations may be incomplete, particularly in regions with high cloud cover and aerosol load – such as the humid tropics and high-latitude zones (Gao et al., 2023), Southeast Asia (Valerio et al., 2025), and eastern China during intense pollution episodes (Wang et al., 2021).

The generation of continuous TROPOMI XCO and XCH_4 products has been addressed through a range of gap-filling approaches. First, machine learning-based interpolation methods (Chen et al., 2022; Hu et al., 2022; Valerio et al., 2025; Wei et al., 2025) reconstruct missing observations by learning nonlinear spatiotemporal relationships. However, their performance strongly depends on the quality and representativeness of the training data, and biased or insufficient samples may lead to unreliable predictions. Second, enhanced spectral fitting algorithms (Borsdorff et al., 2019; Guanter et al., 2015; Schneising et al., 2023; Wang et al., 2020) improve retrieval continuity by refining the spectral inversion process, for example through higher-order polynomial fitting to reduce reflectance-induced biases, albeit at increased computational cost. Third, physics–data–driven coupling approaches, including model fusion, data assimilation, and hybrid machine learning frameworks, integrate satellite retrievals with chemical transport models or reanalysis datasets (Fritz et al., 2022; Schneising et al., 2023; Sicard et al., 2021). By combining physical constraints with data-driven learning or observational corrections, these methods improve spatial consistency and temporal continuity while

reducing retrieval bias. However, their performance remains dependent on uncertainties in prior model simulations and the representativeness of assimilated observations (Inness et al., 2022; Wang et al., 2023).

Despite these advances, current methods face challenges in balancing computational efficiency with physical consistency, particularly in regions with persistent cloud cover. To overcome these limitations, this study proposes a novel signal-domain guided spatio-temporal fusion framework. Our core objective is to generate daily global and regional continuous XCO and XCH_4 products (2019–2023) at high resolution (0.25° globally and 0.05° over China) by effectively leveraging complementary information from chemical transport modeling and frequency-domain representations.

The proposed method specifically addresses the shortcomings of existing approaches in three key aspects: (1) Addressing Data Sparsity and Non-linearity: Unlike pure machine learning interpolation that may struggle with unbalanced datasets, we employ a two-stage strategy. We first use 3D Discrete Cosine Transform (DCT) and Singular Value Decomposition (SVD) to perform low-rank signal-domain reconstruction (Rao and Yip, 2014; Wall et al., 2003). This step exploits the shared spatiotemporal structure between GEOS-Chem simulations and TROPOMI observations to approximate missing values in the frequency domain, providing a robust prior; (2) Enhancing Physical Consistency: To correct residual biases without overfitting, we implement a lightweight residual U-Net for pixel-level refinement (Ronneberger et al., 2015; Tang, 2022). Crucially, instead of using the TROPOMI data mask as a direct input feature, we incorporate it as a weighting mask in the loss function (Wei et al., 2022). This masked learning strategy ensures that the model is supervised exclusively by valid observations, thereby preserving the physical consistency of the DCT/SVD-derived signal-domain reconstructed fields; (3) Improving Efficiency and Accuracy: By fusing model-driven priors and observational constraints in both frequency and spatial domains, our method significantly reduces computational load compared to enhanced spectral fitting while avoiding the "black-box" limitation of pure data-driven methods.

Validation against independent TCCON observations demonstrates that the fused outputs outperform GEOS-Chem simulations alone and maintain accuracy comparable to or better than original TROPOMI retrievals in cloud-covered regions. This framework provides an efficient and interpretable solution for large-scale trace gas monitoring and offers new opportunities for atmospheric data assimilation and climate analysis.

2 Measurement and materials

2.1 Data description

2.1.1 TROPOMI XCO and XCH_4 products

This study employs TROPOMI Level-2 data products of column-averaged CO (XCO) and CH_4 (XCH_4). The TROPOMI instrument onboard the Sentinel-5 Precursor (S5P) satellite operates in a polar sun-synchronous orbit, and provides daily global XCO and XCH_4 observations at approximately 13:30 local solar time, thereby enabling near-global daily coverage. The pixel resolution of the TROPOMI XCO and XCH_4 products was improved from $7.0 \times 7.0\,km^2$ to $7.0 \times 5.5\,km^2$ in June 2019.

TROPOMI retrieves XCO and XCH_4 by measuring Earth-reflected radiation in the shortwave infrared (SWIR) spectral range (2305–2385 nm). For XCO, the retrieval shows high sensitivity to the tropospheric boundary layer under clear-sky conditions; however, this sensitivity is reduced under cloudy conditions due to changes in photon path length and scattering effects (Landgraf et al., 2016). For XCH_4, the operational retrieval algorithm simultaneously estimates column-averaged dry-air mole fractions of CH_4 together with atmospheric scattering parameters. High retrieval accuracy and precision are generally achieved under clear-sky conditions over land after strict filtering of scenes affected by clouds, aerosols, or high surface albedo variability (Hu et al., 2016).

The TROPOMI Level-2 CO and CH_4 products used in this study were obtained from the NASA GES DISC Sentinel-5P/TROPOMI archive for the period 2019–2023. We used the OFFL (Offline) datasets. Specifically, the CO product used is S5P_L2__CO_____HiR, Version 02 (DOI: https://doi.org/10.5270/S5P-bj3nry0TS3), while the CH_4 product used is S5P_L2__CH4____HiR, Version 02 (DOI: https://doi.org/10.5270/S5P-3lcdqivTS4). These Version 02 products include updated spectroscopic parameters and improved treatments for non-scattering clouds compared with earlier releases. The CO product was converted to XCO, and the CH_4 product is reported as XCH_4. Both XCO and XCH_4 are reported as column-averaged dry-air mole fractions in parts per billion (ppb). To ensure the quality and reproducibility of the satellite observations, we applied product-specific quality screening and retained only TROPOMI retrievals with a quality value (qa_value) greater than 0.5 for subsequent analysis, following the recommended quality-screening threshold for the Sentinel-5P CO and CH_4 Level-2 products.

2.1.2 GEOS-Chem chemical transport model

This study employs the GEOS-Chem model, a global three-dimensional atmospheric chemistry model driven by meteorological input from the Goddard Earth Observing System (GEOS) of the NASA Global Modeling and Assimilation Office (GMAO). We employed version 14.1.1 (http://acmg.

seas.harvard.edu/geos/, last access:TS5), which is driven by GEOS-FP meteorological data. The most recent GEOS-5 meteorological data product to be provided by NASA/GMAO is GEOS-FP ("forward processing") (http://gmao.com/gmao.html, last access:TS6). This product has a native horizontal resolution of 0.25° latitude × 0.3125° longitude and a temporal resolution of hourly data and 3-hourly data.

To balance computational cost and regional spatial detail, GEOS-Chem simulations were performed at a 2° × 2.5° horizontal resolution for global simulations, and at a 0.25° × 0.3125° nested resolution for regional simulations over China. The model time steps were set to 300 s for convective and advective transport, and 600 s for chemical processes.

The simulations incorporated detailed region-specific emission inventories. Anthropogenic emissions were sourced from the Air Pollutant Emission Inventory (APEI) v2016 for Canada, the National Emission Inventory (NEI) v2015-03 for North America, DICE-Africa for Africa (Marais and Wiedinmyer, 2016), and the MIX v1.1 inventory combined with HTAP for Asia (Li et al., 2017). Global aircraft and ship emissions were provided by the Community Emissions Data System (CEDS) (Hoesly et al., 2018), while biomass-burning CO emissions were obtained from the Quick Fire Emissions Dataset (QFED) versions 2.1, 2.2, and 2.4.

Based on these inputs and physical processes, GEOS-Chem simulates three-dimensional CO and CH_4 fields. However, XCO and XCH_4 are not direct native outputs; they were diagnosed during post-processing from the simulated vertical profiles. Specifically, the dry-air mole fraction profiles were vertically averaged using pressure-layer thickness as weighting, producing column-averaged dry-air mole fractions consistent with the definitions used by TROPOMI and TCCON. The resulting fields were converted to parts per billion (ppb) for use in the fusion framework.

In the proposed fusion framework, these GEOS-Chem simulations serve as large-scale physical priors. To match the spatial resolution of satellite observations, the global 2° × 2.5° GEOS-Chem fields were regridded to a 0.25° × 0.25° target grid, while the China 0.25° × 0.3125° fields were regridded to a 0.05° × 0.05° target grid. Regridding was performed using inverse distance weighting (IDW) to interpolate GEOS-Chem fields onto the target grids.

2.1.3 Total Carbon Column Observing Network (TCCON) measurements

TCCON employs a Fourier Transform Infrared Spectrometer (FTS) to measure direct solar light in order to determine the total column concentrations of greenhouse gases in the atmosphere, including carbon dioxide (CO_2), CH_4, CO, and others (Buschmann et al., 2016; Kiel et al., 2016; Sha et al., 2020; Yang et al., 2020). These data are extensively utilized to validate satellite remote sensing data (e.g., Chander et al., 2013; Imasu et al., 2023; Lin et al., 2024; Loew et al., 2017; Wu et al., 2019; etc.) and to evaluate the performance of climate models. Due to its high precision, standardized retrieval algorithm, and consistent calibration across sites, TCCON is widely regarded as a primary ground-based reference network for validating satellite greenhouse-gas products (Wunch et al., 2011; Herkommer et al., 2024). They are rigorously calibrated and validated with high accuracy and reliability. TCCON's observing stations are situated in numerous regions worldwide, providing a comprehensive understanding of the global greenhouse gas distribution. These stations are capable of observing a diverse array of climates and ecosystems. Harmonized processing is implemented for each site's data. In order to guarantee data consistency and comparability, the data from each station is processed in a consistent manner. TCCON's XCO and XCH_4 Dry Air Mole Fraction (Xgas) in ppb are employed in this study. Site-specific and time-scale data are accessible through the official website of TCCON (https://tccon.ornl.gov/, last access:TS7), and users may select the data that is most relevant to their research requirements. The validation of XCO and XCH_4 products from TROPOMI and GEOS-Chem is frequently conducted using TCCON (Borsdorff et al., 2019; Cogan et al., 2012; Inness et al., 2022; Schneising et al., 2019).

In this study, we employed the TCCON GGG2020 data release (Laughner et al., 2023), with site-specific datasets accessible through the official TCCON website (https://tccon.ornl.gov/, last access:TS8). The TCCON sites used in this study, shown in Fig. S1 and listed in Table S1 in the Supplement, were selected based on data availability during 2019–2023, temporal overlap with the TROPOMI and GEOS-Chem datasets, and geographical representativeness. The selected sites cover broad latitudinal ranges and different environmental conditions, including remote background regions, mid-latitude continental sites, high-latitude regions, island sites, and urban-influenced or emission-affected regions. All stations listed in Table S1 were included in the comprehensive TCCON validation whenever valid collocated samples were available after applying the temporal and spatial matching criteria described in Sect. 3.1. A subset of these stations was additionally selected for representative time-series and scatterplot examples based on data completeness and relatively continuous records; these illustrative examples were not used as the sole basis for the overall validation statistics.

Regarding data filtering, TCCON FTS instruments measure direct solar radiation, and the retrievals are therefore already subject to the network's standard clear-sky retrieval and quality-control procedures. In this study, we applied the standard quality control provided with the TCCON GGG2020 data release and did not impose additional cloud-cover filters on the TCCON measurements. Additional filtering related to satellite data availability, including the TROPOMI missing-rate criterion within the 2° collocation window, is described separately in Sect. 3.1.

2.2 Methodology

In this study, we propose a two-stage fusion framework that integrates physical modeling, signal-domain reconstruction, and deep learning-based residual correction to achieve continuous and accurate global mapping of atmospheric trace gases. The overall workflow is illustrated in Fig. 1, which presents the main components and their interconnections. In the first stage, a signal-domain spatio-temporal reconstruction is employed to exploit the low-frequency consistency and spatio-temporal correlations between the TROPOMI observations and GEOS-Chem simulations, effectively filling missing data regions caused by cloud cover or instrument limitations. In the second stage, a residual learning network based on a lightweight residual U-Net is introduced to refine the fused data by learning nonlinear and region-specific discrepancies between the preliminary reconstruction and the true satellite observations.

2.2.1 Data preprocessing

Data preprocessing is essential for ensuring the reliability of the fusion results. To reconcile the spatial resolution differences between satellite retrievals and model simulations, TROPOMI XCO and XCH_4 retrievals with qa_value below 0.5 were excluded, and the remaining observations were aggregated onto a global $0.25° \times 0.25°$ grid (720×1440) using area-weighted averaging (Wang et al., 2018). GEOS-Chem outputs were subsequently interpolated to the same grid using inverse distance weighting (Setianto and Triandini, 2013). The 0.25° spatial resolution was selected to balance the representation of mesoscale spatial variability with computational efficiency for multi-year global fusion analyses (Hu et al., 2024; Wang et al., 2023).

To evaluate the influence of the resolution mismatch between GEOS-Chem simulation results and the final fused product, we used the 2019 CO data to conduct a one-year resolution sensitivity experiment. In this experiment, we used GEOS-Chem simulation data at the native $2° \times 2.5°$ resolution and aggregated TROPOMI XCO to the $2° \times 2.5°$ resolution using the area-weighted averaging method. The related fusion calculation framework was directly implemented on the $2° \times 2.5°$ grid; subsequently, the obtained fused XCO data at the $2° \times 2.5°$ resolution were compared with the fused product obtained by aggregating the fused XCO data at the 0.25° resolution to the same $2° \times 2.5°$ grid using the area-weighted averaging method.

2.2.2 Averaging-kernel sensitivity test

We conducted an additional sensitivity test to examine the influence of TROPOMI averaging kernels (AKs) on the comparison between satellite retrievals and model simulations. For 2019, quality-screened TROPOMI CO Level-2 retrievals (qa_value > 0.5) were collocated with GEOS-Chem CO vertical profiles. Two GEOS-Chem XCO datasets were then generated for comparison. For the first dataset, GEOS-Chem XCO was diagnosed from the simulated CO profiles by vertically averaging the dry-air mole fractions using pressure-layer thickness as weighting, consistent with the standard post-processing procedure described above. For the second dataset, the GEOS-Chem CO profiles were first interpolated to the TROPOMI pressure layers and then smoothed using the TROPOMI CO column averaging kernels. Finally, both the original GEOS-Chem XCO and the AK-smoothed GEOS-Chem XCO were compared with TROPOMI XCO in ppb.

2.2.3 Spatio-temporal data fusion method based on signal domain reconstruction

Previous studies have shown that emission inventories typically underestimate GEOS-Chem simulation results (Hu et al., 2018; Liang et al., 2023). Despite these magnitude biases, GEOS-Chem simulations generally remain consistent with TROPOMI observations in terms of spatio-temporal variability, exhibiting similar temporal evolution patterns, including concurrent increases and decreases. This consistency suggests that the spatio-temporal correlation between GEOS-Chem simulations and TROPOMI observations can be effectively exploited. By leveraging this relationship, a data fusion framework based on signal-domain reconstruction can be used to generate datasets with improved spatial continuity and more complete coverage (Chen et al., 2022; He et al., 2022; Wang et al., 2021).

We assume that a spatio-temporal relationship exists between the GEOS-Chem simulations and TROPOMI observations for XCO and XCH_4, as expressed below:

$$\mathrm{XT} = f(\mathrm{XG}, \mathrm{Lat}, \mathrm{Lon}, \mathrm{Date}) \tag{1}$$

where Lat, Lon, and Date denote latitude, longitude, and time, respectively; XG represents the GEOS-Chem XCO and XCH_4 concentration field; and XT represents the corresponding TROPOMI observation field. In other words, the TROPOMI value at a given spatio-temporal coordinate can be approximated using the modeled concentration and its spatial and temporal information.

To simplify the representation of this relationship, Eq. (1) is reformulated as an element-wise scaling of the GEOS-Chem field by a spatio-temporal transformation matrix ρ:

$$\mathrm{XT} = \mathrm{XG} \cdot \rho \tag{2}$$

where $\cdot$ denotes element-wise multiplication, also known as the Hadamard product. The matrix ρ represents the local spatio-temporal relationship between GEOS-Chem and TROPOMI at each grid cell and time step. It has the same dimensions as XT and XG. Known values of ρ can be obtained at valid TROPOMI pixels by comparing XT and XG, while missing values of ρ need to be reconstructed.

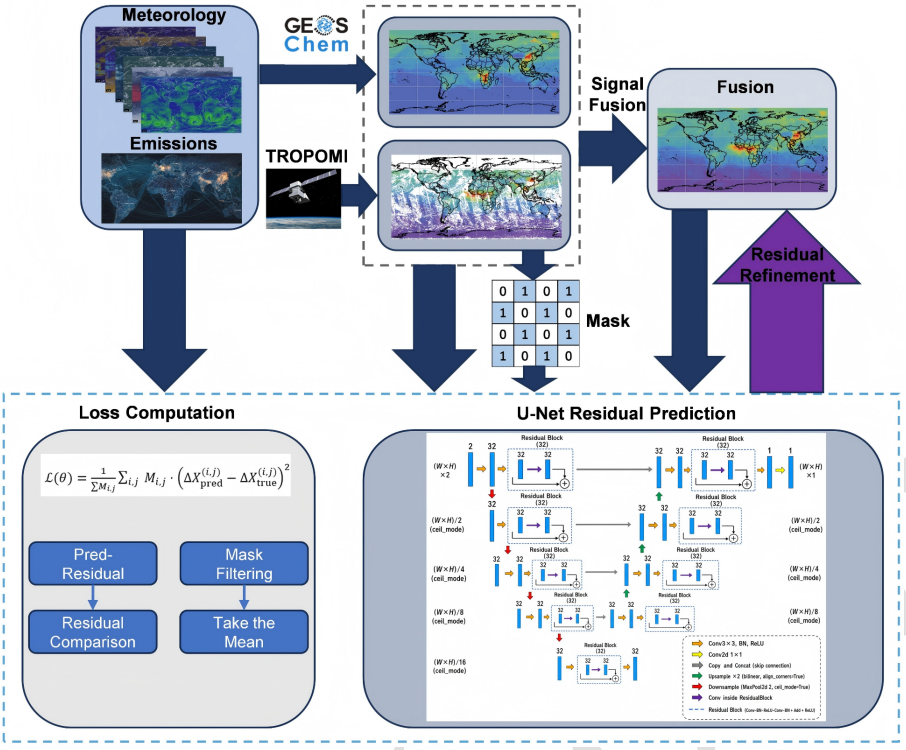

Figure 1. Overview of the proposed fusion framework combining physical modeling, signal-domain reconstruction, and deep learning-based residual correction. The process includes data preprocessing, DCT&SVD-based spatio-temporal fusion, and residual refinement via a lightweight residual U-Net architecture. Arrows indicate the information flow across different modules.

The parameter matrix ρ is assumed to be smooth in space and time, inspired by previous studies on filling missing values in spatio-temporal data and enhancing data smoothness using multidimensional discrete cosine transforms (Elharar et al., 2007; Garcia, 2010; Robinson and Kecman, 2003; Okolie and Smit, 2022; Peng et al., 2005; Rao and Yip, 2014; Wang et al., 2023). We therefore propose a spatio-temporal 3D matrix smoothing algorithm based on the discrete DCT and SVD to enhance data smoothness and fill missing values in spatio-temporal data. The method effectively handles spatio-temporal data with missing values while maintaining spatio-temporal correlation by combining spatio-temporal nearest-neighbor interpolation and regularized optimization techniques.

We find the spatio-temporal 3D matrix $\hat{\rho}$ that minimizes Eq. (3) by means of the 3D DCT. The objective function includes the residual term on the left-hand side and the smoothing term on the right-hand side[TS10]:

$$E\left(\hat{\rho}\right) = ||\omega^{\frac{1}{2}} * (\hat{\rho} - \rho)||^2 + \epsilon||\nabla^2\hat{\rho}||^2 \tag{3}$$

where $||\cdot||$ denotes the Euclidean paradigm, ω is a binary mask indicating the availability of a parameter corresponding to the spatio-temporal location of ρ, ϵ denotes the smoothing parameter, and ∇^2 denotes the Laplace operator. The satisfied condition $\hat{\rho}$ can be solved by iteration of Eq. (4):

$$\hat{\rho} = \alpha \mathrm{IDCT}_3\left(\Gamma^3 * \mathrm{DCT}_3\left(\omega * \left(\rho - \hat{\rho}\right) + \hat{\rho}\right)\right) + (1-\alpha)\,\hat{\rho} \tag{4}$$

where α is a parametric factor for accelerating convergence, Γ^3 denotes the 3D spatio-temporal filtering matrix associated with the smoothing term, which can be obtained through Eq. (5), and DCT_3 and IDCT_3 denote the 3D discrete cosine signal transform and its inverse transform, respectively.

$$\Gamma^3_{i_1,i_2,i_3} = \frac{1}{1 + \epsilon\sum_{k=1}^{3} 2\left[1 - \cos\frac{(i_k-1)\pi}{n_k}\right]} \tag{5}$$

Here, i_k denotes the i-th value along the k-th dimension and n_k denotes the size of ρ along the k-th dimension. This means that the value at each position of this three-dimensional spatio-temporal filtering matrix is completely determined by its position. The closer its position is to the element at position (1, 1, 1), the larger (the closer it is to 1) the value is, and vice versa. The value at position (1, 1, 1) is 1. Since the low-frequency components of the discrete cosine transformed signal matrix are mainly located close to position (1, 1, 1), this filtering matrix allows us to search for a smoothed ρ in the frequency domain. In this study, the total number of iterations is empirically set to 100, α is set to 0.75, and ϵ is selected through an automated search within 10^{-3} to

10^{-1}, using the value that minimizes the reconstruction error on valid observations. This optimization is performed independently for XCO and XCH_4 and for the global and China domains.

Equations (6) and (7) define the forward and inverse three-dimensional discrete cosine transforms used in the iterative update in Eq. (4). Specifically, DCT_3 transforms the spatio-temporal parameter matrix ρ from the latitude–longitude–time domain into the frequency domain, where the dominant large-scale and slowly varying structures are represented by low-frequency coefficients. After the smoothing filter Γ^3 in Eq. (5) is applied to these coefficients, $IDCT_3$ transforms the filtered signal back to the original spatio-temporal domain. Therefore, Eqs. (6) and (7) provide the mathematical basis for the frequency-domain smoothing and reconstruction of ρ.

The 3D DCT, denoted as DCT_3, is defined as:

$$F(u,v,w) = \frac{2}{\sqrt{NMP}} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \sum_{z=0}^{P-1} f(x,y,z) \cos\left(\frac{\pi(2x+1)u}{2N}\right) \cos\left(\frac{\pi(2y+1)v}{2M}\right) \cos \tag{6}$$

where $u = 0, 1, 2, \ldots, N-1$, $v = 0, 1, 2, \ldots, M-1$, and $w = 0, 1, 2, \ldots, P-1$. N, M, and P represent the size of the signal in each of the three dimensions.

The inverse 3D DCT, denoted as $IDCT_3$, is defined as:

$$f(x,y,z) = \frac{2}{\sqrt{NMP}} \sum_{u=0}^{N-1} \sum_{v=0}^{M-1} \sum_{w=0}^{P-1} F(u,v,w) \cos\left(\frac{\pi(2x+1)u}{2N}\right) \cos\left(\frac{\pi(2y+1)v}{2M}\right) \cos \tag{7}$$

where $x = 0, 1, 2, \ldots, N-1$, $y = 0, 1, 2, \ldots, M-1$, and $z = 0, 1, 2, \ldots, P-1$. N, M, and P represent the size of the signal in each of the three dimensions. Through the repeated application of DCT_3 frequency-domain filtering, and $IDCT_3$, missing values in ρ can be reconstructed while preserving the dominant low-frequency spatio-temporal structure.

Furthermore, the 3D discrete cosine transform requires a complete 3D matrix for signal conversion. Consequently, we first interpolate ρ with missing values using the spatio-temporal autocorrelation property of ρ. Subsequently, SVD is applied to retain the dominant components of the reconstructed parameter matrix and reduce weak noisy components. In this study, 80 % of the cumulative singular-value energy is retained as a practical and computationally efficient truncation threshold for the preliminary reconstruction. The rule of SVD is illustrated in Eq. (8), which is used to preserve the primary components of ρ for the iteration in Eq. (4):

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{8}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{U} \in \mathbb{R}^{m\times m}$ are orthogonal TS11matrices whose column vectors are called left singular vectors; $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ is a diagonal matrix whose diagonal elements are the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$, and r is the rank of the matrix $\mathbf{A}$. $\mathbf{V} \in \mathbb{R}^{n\times n}$ is an orthogonal matrix whose column vectors are called right singular vectors.

To evaluate whether this empirical threshold materially affects the reconstruction, we conducted an additional threshold-sensitivity experiment for the 2019 daily global XCO reconstruction. Daily GEOS-Chem XCO simulations and collocated TROPOMI CO retrievals on the $0.25° \times 0.25°$ grid were used to construct the TROPOMI/GEOS-Chem ratio fields. The same DCT-SVD reconstruction procedure was repeated using cumulative singular-value energy thresholds of 70 %, 75 %, 80 %, 85 %, 90 %, and 95 %, with the 80 % case used as the reference. The results are summarized in Table S8 and Fig. S2.

Calculate the total energy of the singular values as $E_{\text{total}} = \sum_{i=1}^{r} \sigma_i^2$, and find the smallest k such that the first k singular values account for at least 80 % of the total energy, i.e., $(\sum_{i=1}^{k} \sigma_i^2)/E_{\text{total}} \geq 0.8$. Then, setting the last r–k smaller singular values in $\mathbf{\Sigma}$ to zero yields the truncated diagonal matrix $\mathbf{\Sigma}_k = \text{diag}(\sigma_1, \sigma_2 \ldots, \sigma_k, 0, \ldots, 0)$. Finally, reconstruct the approximation matrix using the truncated singular value matrix $\mathbf{A}_k = U\,\Sigma_k V^T$.

2.2.4 Deep residual refinement via learning-based mask reconstruction

To further enhance the reconstruction quality, we introduced a residual learning module based on a deep neural network to refine the preliminary DCT/SVD reconstruction. Although the signal-domain reconstruction captures the dominant spatiotemporal structures shared by GEOS-Chem and TROPOMI, local nonlinear discrepancies may still remain because of retrieval noise, regional emission variability, transport errors, and meteorological influences. Therefore, a residual U-Net was employed to learn pixel-level corrections between the preliminary fused field and valid TROPOMI observations.

Because the target grids are finer than the native GEOS-Chem grids, the residual-learning step should be interpreted as observation-constrained reconstruction rather than pure model downscaling. GEOS-Chem provides the large-scale physical background, while the residual network learns corrections between the preliminary DCT/SVD reconstruction and valid TROPOMI observations on the target grid. Therefore, fine-scale variations in the fused datasets are constrained primarily by satellite-observed spatial patterns and auxiliary meteorological information, rather than being generated solely from the coarse GEOS-Chem fields.

Residual learning objective

Let X_{DCT} denote the preliminary fused XCO/XCH_4 field obtained from the DCT/SVD-based reconstruction, and X_{TROPOMI} be the valid observational values from the satellite. The residual between the fusion estimate and the true

value (only available at observed locations) is defined as:

$$\Delta X_{\text{true}} = X_{\text{TROPOMI}} - X_{\text{DCT}}, \text{where } M = 1 \quad (9)$$

Here, $M \in \{0,1\}^{H\times W}$ is a binary mask indicating the presence (1) or absence (0) of valid satellite data. The goal is to train a model $\mathcal{F}_\theta(\cdot)$ parameterized by θ to predict the residual ΔX_{pred} across the entire domain:

$$\Delta X_{\text{pred}} = \mathcal{F}_\theta(X_{\text{G}}, X_{\text{DCT}}, \mathbf{A}) \quad (10)$$

where X_{G} denotes the GEOS-Chem full-coverage simulation data; X_{DCT} denotes the DCT/SVD-reconstructed preliminary fusion; $\mathbf{A}$ denotes the auxiliary information such as meteorological fields and emission inventories. Precursor meteorological data simulated by GEOS-Chem and emission inventory were used in this study. The final fused data is obtained by correcting the DCT/SVD estimate with the predicted residual:

$$X_{\text{fused}} = X_{\text{DCT}} + \Delta X_{\text{pred}} \quad (11)$$

Loss function design

The model is trained using only the valid observations, i.e., locations where $M = 1$. The masked loss was used because reliable TROPOMI targets are available only at valid retrieval pixels; pixels affected by clouds, aerosols, or retrieval failures do not provide trustworthy reference values. The mask therefore avoids penalizing the network at invalid or missing satellite pixels. Although the loss is computed only at valid pixels, the trained residual model is applied to the full spatial domain because its inputs, including GEOS-Chem simulations and the preliminary DCT/SVD reconstruction, are spatially complete or reconstructed before residual learning. Thus, the model transfers residual relationships learned from valid observations to missing regions with similar spatiotemporal patterns represented by the two input fields, following the logic of recent masked-learning atmospheric gap-filling studies (Wei et al., 2023; Sun et al., 2026). The loss function is designed to minimize the residual error at these observed locations:

$$\mathcal{L}(\theta) = \frac{1}{\sum M_{i,j}} \sum\nolimits_{i,j} M_{i,j} \cdot \left(\Delta X_{\text{pred}}^{(i,j)} - \Delta X_{\text{true}}^{(i,j)}\right)^2 \quad (12)$$

This masked mean squared error ensures that the learning focuses on valid TROPOMI retrievals. However, because valid retrievals may be spatially or temporally clustered in cloud-free regions, the masked loss may introduce sampling bias and affect extrapolation in persistently cloudy or aerosol-affected areas where valid observations are sparse. This risk is partly mitigated by the use of spatially complete predictors, including GEOS-Chem simulations and the preliminary DCT/SVD reconstruction, which provide physical and spatiotemporal constraints across the full domain. Nevertheless, fused data in regions with persistent data gaps should be interpreted with caution.

Model architecture and training

In this study, we employed a lightweight residual U-Net architecture to predict the full-domain residual field ΔX_{pred}, which represents the correction from the DCT/SVD reconstruction to the expected TROPOMI observation. The network inputs include two channels: the GEOS-Chem simulation and the DCT/SVD reconstruction output (Wang et al., 2025). The TROPOMI mask is not used as an input but is instead applied during the loss computation to focus learning only on valid pixels.

The residual U-Net consists of an encoder–decoder structure with skip connections and lightweight residual blocks. The network contains 10 convolutional layers in total, including the convolutional layers inside the residual blocks. All hidden convolutional layers use 32 filters, with two input channels and one output channel for the predicted residual field. The input and output spatial dimensions remain unchanged (i.e., 720×1440 at 0.25° resolution), enabling pixel-wise learning of spatial residuals. The model was trained using daily global maps from 2019 to 2023. Training was performed for 5 epochs with a batch size of 4. We used the Adam optimizer with a fixed learning rate of 1×10^{-4}, and no learning-rate scheduler was applied. No dropout layer or data augmentation was used in the final configuration. To improve numerical stability, all input variables were normalized before training. The final output is truncated to avoid physically implausible corrections, by enforcing:

$$\left|\Delta X_{\text{pred}}\right| < \gamma \cdot \sigma_{\text{TROPOMI}}, \gamma = 3 \quad (13)$$

where σ_{TROPOMI} denotes the standard deviation of observed valid values, and γ is a hyperparameter controlling confidence bounds. This truncation is applied only to the neural-network-predicted residual correction, not directly to the final XCO or XCH_4 concentration fields. Its purpose is to prevent unphysical residual corrections arising from unstable extrapolation in regions with sparse valid TROPOMI observations. We acknowledge that such truncation may partly dampen extremely large residual corrections and could therefore reduce the magnitude of rare local extremes. However, because the truncation is imposed on the residual term rather than on the full concentration fields, the large-scale and physically consistent extreme signals already present in the GEOS-Chem simulations, TROPOMI observations, or DCT/SVD reconstruction are retained. Thus, the constraint mainly suppresses anomalous correction spikes while preserving the dominant extreme-event signals.

2.2.5 Ablation study design

To explicitly isolate the contribution of each component in the proposed two-stage framework, we conducted an ablation study using site-based validation across all TCCON stations for the year 2021. Five configurations were compared: (1) only DCT fusion (Garcia, 2010), (2) DCT and SVD mixed-

signal reconstruction (Bengherabi et al., 2008; Majhi and Pal, 2021), (3) residual CNN (Jiang et al., 2024), (4) residual U-Net (Ronneberger et al., 2015; Yan et al., 2022), and (5) residual XGBoost (Naseem et al., 2024). The performance was assessed using the coefficient of determination (R^2), root mean square error (RMSE), and mean bias (μ). This design allows the contribution of the DCT/SVD signal-domain reconstruction stage and the residual-correction stage to be evaluated separately. The site-based validation results are summarized in Tables S2 and S3.

2.2.6 Evaluation scheme

Our evaluation methodology includes TCCON site validation, spatial-distribution assessment, and application-oriented analyses. For TCCON site validation, GEOS-Chem simulations, TROPOMI observations, and fused XCO and XCH_4 data are compared with TCCON measurements at both individual-site and all-site scales. The evaluation metrics include the coefficient of determination (R^2), root mean square error (RMSE), mean bias (μ), and standard deviation of the bias (σ) (Karunasingha, 2022; Kobayashi and Salam, 2000; Wang et al., 2023). These metrics are used as descriptive validation statistics, and no formal hypothesis tests or p-value thresholds are applied. The spatial-distribution assessment includes comparisons among GEOS-Chem simulations, TROPOMI observations, and fused XCO and XCH_4 data over multiple temporal scales, including multi-year averages, seasonal variations, and annual metrics. Application-oriented analyses include the evaluation of TROPOMI data availability, the responses of XCO and XCH_4 to extreme events, and the comparison of multi-year column concentration trends across different regions.

3 Results and discussion

3.1 TCCON site validation

TCCON provides a high-precision and widely used ground-based reference for validating satellite greenhouse-gas products. Therefore, TCCON-based validation was used as the primary benchmark for evaluating the fused XCO and XCH_4 datasets. The TCCON sites are sparsely and unevenly distributed across the globe, with higher concentrations in North America, Europe, East Asia, and a limited number of remote oceanic and high-altitude regions. Because TROPOMI overpasses each region at approximately 13:30 local time, TCCON observations within $13{:}30 \pm 1$ h local solar time were averaged to obtain daily ground-based validation data. For spatial consistency, GEOS-Chem simulations, TROPOMI observations, and the fused datasets were extracted within a 2° radius around each TCCON site. We acknowledge that TCCON provides near-point-scale column measurements, whereas the TROPOMI and fused datasets represent spatially averaged grid-cell values at 0.25° resolution. Therefore, a spatial representativeness mismatch is unavoidable, especially in regions with strong horizontal gradients, complex terrain, coastal influences, or localized emission sources. The 2° radius was used as a compromise to include a sufficient number of satellite and gridded data samples around each station while limiting the influence of distant air masses.

To evaluate the performance of the fused data under sparse TROPOMI coverage, the TROPOMI missing rate was used as an additional selection criterion. For each site-day collocation, the missing rate was defined as the fraction of grid cells within the 2° radius that had no valid TROPOMI retrieval after quality filtering:

$$\mathrm{MR} = 1 - \frac{N_{\mathrm{valid}}}{N_{\mathrm{total}}} \tag{14}$$

where N_{total} is the total number of 0.25° grid cells within the 2° radius and N_{valid} is the number of grid cells with valid TROPOMI retrievals after applying the qa_value > 0.5 criterion. Grid cells without satellite retrievals or with qa_value ≤ 0.5 were treated as missing. We selected site-day observation cases with MR > 0.5, corresponding to less than 50 % valid TROPOMI coverage within the collocation window, to evaluate the performance of the fused XCO and XCH_4 datasets under sparse TROPOMI coverage conditions.

3.1.1 TCCON-based validation setup and stage-wise ablation

Based on the same TCCON collocation strategy, we first performed an ablation analysis to clarify the contribution of each component in the proposed two-stage framework before evaluating the final fused datasets. As summarized in Tables S2 and S3, the comparison between "Only DCT" and "DCT/SVD mixed-signal reconstruction" isolates the contribution of the SVD-based low-rank constraint, while the comparison between "DCT/SVD mixed-signal reconstruction" and the residual-learning configurations isolates the contribution of the residual-correction stage. For XCO, adding SVD to DCT increased R^2 from 0.8227 to 0.8335 and reduced RMSE from 10.439 to 10.1166 ppb. The residual U-Net further improved R^2 to 0.8450 and reduced RMSE to 9.7616 ppb. For XCH_4, adding SVD increased R^2 from 0.7056 to 0.7320 and reduced RMSE from 17.5156 to 16.7111 ppb, while the residual U-Net further improved R^2 to 0.7598 and reduced RMSE to 15.8212 ppb. These results indicate that both the DCT/SVD signal-domain reconstruction stage and the residual U-Net correction stage make measurable contributions to the final fused datasets.

After confirming the contribution of each component, we further evaluated the final fused datasets against TCCON observations at representative stations and across all qualified station-day collocations.

3.1.2 Sensitivity to the SVD energy threshold

The sensitivity experiment showed that the reconstructed XCO fields were highly stable across the tested SVD energy thresholds. Relative to the 80 % reference case, all threshold experiments yielded spatial correlation coefficients higher than 0.9998, RMSE differences below 0.62 ppb, spatial standard deviation ratios close to unity, and hotspot overlap ratios greater than 0.97 (Table S8). The cumulative singular-value energy analysis further showed that the first, first two, and first three singular values explained approximately 92.0 %, 96.6 %, and 97.5 % of the total energy, respectively (Fig. S2). These results indicate that the XCO reconstruction is not sensitive to the exact SVD truncation threshold within the tested range of 70 %–95 %, supporting the use of 80 % as a stable and computationally efficient threshold.

3.1.3 Validation of the final fused datasets

Figures 2 and 3 illustrate representative time series of daily GEOS-Chem simulations, TROPOMI observations, fused data, and TCCON XCO and XCH_4 measurements at selected ground stations. The ground stations were selected according to the following criteria. For each gas, we first screened TCCON stations that retained a relatively large number of valid spatially collocated data pairs after applying the $13{:}30 \pm 1$ h temporal window, the 2° spatial matching radius, and the TROPOMI quality filter (qa_value > 0.5). Among these qualified stations, we further selected those with relatively continuous multi-year records and limited long data gaps to illustrate the temporal consistency among GEOS-Chem, TROPOMI, the fused data, and TCCON observations. Therefore, easttroutlake01, lamont01, and parkfalls01 were selected as representative XCO time-series examples, while edwards01, nicosia01, and pasadena01 were selected as representative XCH_4 examples. These selected stations were used only for illustrative time-series comparison, whereas the quantitative site-level and overall validation statistics were calculated separately using all qualified TCCON station-day collocations under the corresponding validation criteria. Compared with TCCON observations, both GEOS-Chem XCO and XCH_4 simulations show systematic underestimation; nevertheless, they capture similar temporal variations, i.e., comparable increase–decrease patterns over time. This indicates that GEOS-Chem can provide useful spatiotemporal reference information for the fusion process. The temporal variations of both TROPOMI and fused XCO and XCH_4 are generally consistent with those observed by TCCON. Compared with TROPOMI, the fused data show reduced biases μ (-0.79, -3.58, and -1.08 ppb for XCO; -5.43, -2.49, and -4.34 ppb for XCH_4) and lower standard deviations σ (9.93, 4.26, and 5.73 ppb for XCO; 5.99, 6.87, and 6.77 ppb for XCH_4) at the selected representative sites.

GEOS-Chem simulations show a systematic low bias relative to TCCON observations. The raw, uncorrected GEOS-Chem simulations are used as the physical prior in the fusion and reconstruction framework. The mean-bias adjustment is applied only for the diagnostic GEOS-Chem–TCCON comparison and is not used in generating the fused datasets. For each TCCON station and each gas, we calculate the mean difference between GEOS-Chem simulations and TCCON observations over all valid collocated days, and then subtract this mean bias from the GEOS-Chem time series at that station. This procedure removes the station-mean offset while preserving the model's day-to-day variability, allowing a clearer assessment of whether GEOS-Chem captures the temporal variations observed by TCCON. All figures and tables involving this bias-adjusted GEOS-Chem diagnostic comparison are explicitly labeled to indicate that the GEOS-Chem results were bias-corrected before validation.

Following the representative time-series comparison, we further evaluated site-level statistical agreement using scatterplots at selected representative stations under sparse TROPOMI coverage conditions in Figs. 4 and 5. This selection was based on the number of valid collocated samples after applying the missing-rate criterion, rather than on validation performance. Specifically, for each TCCON station, we first applied the qa_value > 0.5 quality filter, the $13{:}30 \pm 1$ h temporal window, the 2° spatial collocation radius, and the site-day missing-rate criterion (MR > 0.5). Subsequently, stations with a relatively large number of valid collocated samples after applying the missing-rate screening criterion were selected for visualization, whereas the overall validation statistics were calculated using all qualified high-missing-rate site-day collocations rather than only these representative stations. For XCO validation, burgos01, izana01, and nicosia01 were selected as representative sites (Fig. 4). At burgos01 and izana01, the fused data show improved agreement with TCCON compared with GEOS-Chem simulations and the original TROPOMI retrievals, as indicated by higher coefficients of determination (R^2), lower root-mean-square errors (RMSE), smaller biases (μ), and lower standard deviations (σ). At nicosia01, the fused data do not show a clear improvement over TROPOMI but remain comparable, indicating that the fusion method does not degrade the validation performance at this site. For XCH_4 validation, burgos01, rikubetsu01, and xianghe01 were selected as representative sites (Fig. 5). At burgos01 and xianghe01, the fused data show improved statistical consistency with TCCON relative to GEOS-Chem and modest improvements relative to TROPOMI across several evaluation metrics, including R^2, RMSE, μ, and σ. At rikubetsu01, the fused data remain comparable to TROPOMI, further indicating the robustness of the fusion method across different sites. Tables S4 and S5 present the validation results for all qualified individual sites for XCO and XCH_4, respectively. Overall, as shown in Tables S4 and S5, the fused data improve at least one of the three reported site-level metrics relative to the original TROPOMI retrievals at 14 of 18 XCO sites (77.8 %) and 10 of 17 XCH_4

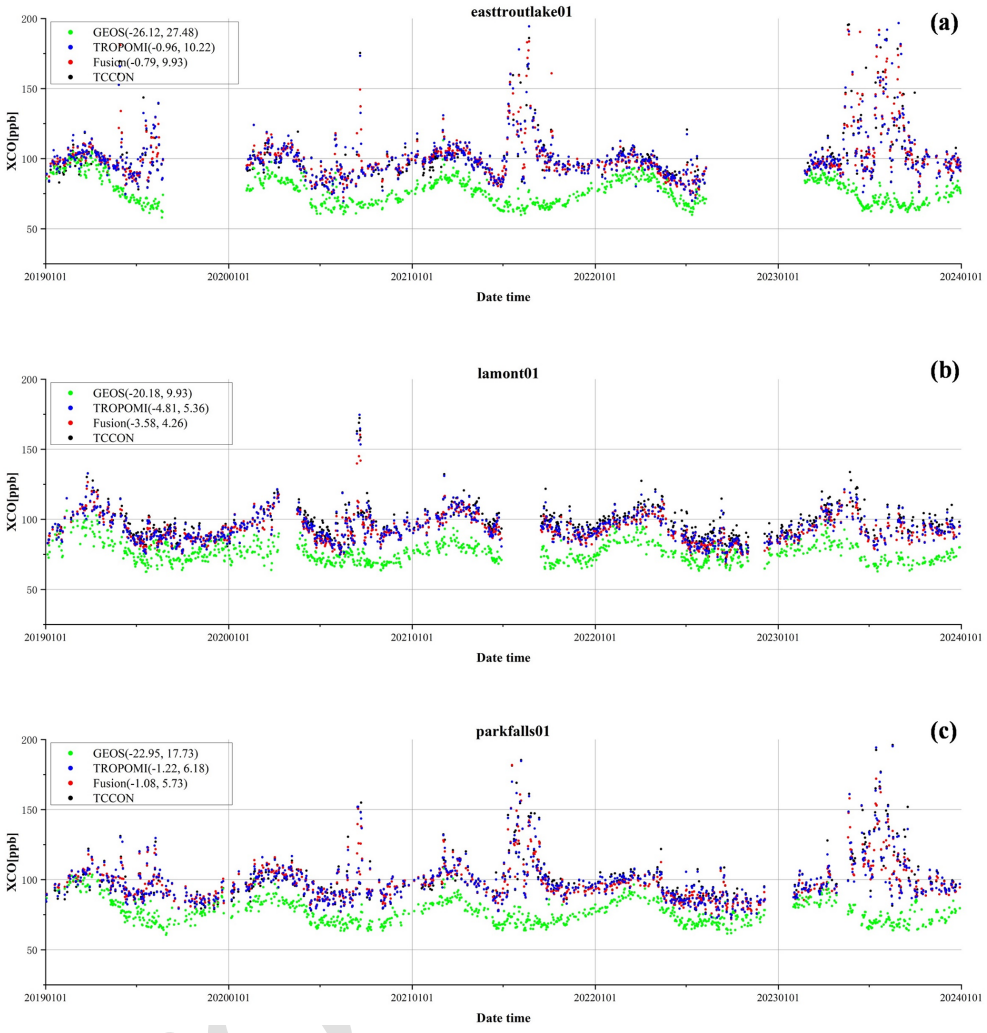

Figure 2. Scatterplots of the time series for daily GEOS-Chem, TROPOMI, fused, and TCCON XCO data at **(a)** easttroutlake01, **(b)** lamont01, and **(c)** parkfalls01. The first and second numbers in parentheses represent μ and σ, respectively. Unit: parts per billion for μ and σ.

sites (58.8 %). Here, improvement refers to higher R^2, lower RMSE, or smaller $|\mu|$.

Figure 6 presents the overall validation results for XCO and XCH_4 across all qualified TCCON site-day collocations under sparse TROPOMI coverage conditions (MR > 0.5). For XCO, the fused dataset substantially reduces the systematic low bias of GEOS-Chem and shows improved agreement with TCCON compared with the original TROPOMI retrievals across several validation metrics. The fused XCO dataset achieves an R^2 of 0.91, which is higher than that of TROPOMI, indicating an improved ability to capture the variability observed by TCCON. Its RMSE and standard deviation of the bias are 5.65 and 5.48 ppb, respectively, both lower than those of TROPOMI. The mean bias (μ) is also comparable to that of TROPOMI, suggesting that the fusion process preserves the overall magnitude while reducing random deviations. For XCH_4, the fused dataset also shows improved statistical consistency with TCCON. The fused XCH_4 dataset achieves an R^2 of 0.83 and a lower standard deviation of the bias than TROPOMI under the same validation protocol. Although its RMSE and mean bias are comparable to those of TROPOMI, the reduced dispersion indicates that the fused dataset provides more stable estimates under sparse-observation conditions. Overall, these results demonstrate that the proposed fusion approach improves the consistency and robustness of XCO and XCH_4 estimates relative to GEOS-Chem and maintains comparable or improved performance relative to TROPOMI, particularly under high-missing-rate conditions.

To demonstrate the performance of our fused datasets under different TROPOMI coverage, we conducted experiments using different MR thresholds. Tables S6 and S7 further quantify how the performance of the fused datasets varies with TROPOMI coverage. For XCO, when the miss-

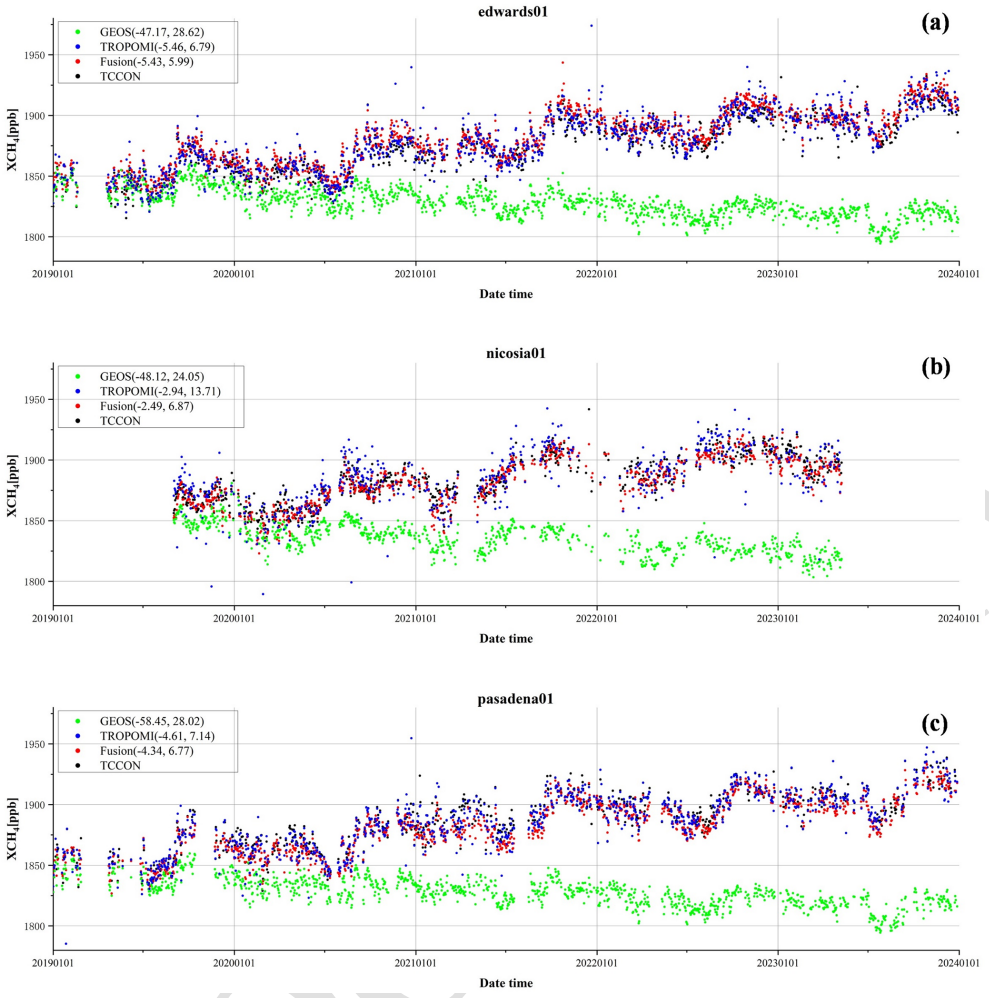

Figure 3. Scatterplots of the time series for daily GEOS-Chem, TROPOMI, fused, and TCCON XCH_4 data at **(a)** edwards01, **(b)** nicosia01, and **(c)** pasadena01. The first and second numbers in parentheses represent μ and σ, respectively. Unit: parts per billion for μ and σ.

ing rate is low or moderate (MR < 0.3 and $0.3 \leq$ MR ≤ 0.5), TROPOMI shows slightly higher R^2 values and lower RMSE values than the fused dataset, indicating the high reliability of TROPOMI retrievals under relatively good coverage conditions. However, under sparse TROPOMI coverage conditions (MR > 0.5), the performance of TROPOMI decreases, with its R^2 decreasing to 0.89 and RMSE increasing to 6.16 ppb. In contrast, the fused XCO data maintain a higher R^2 value of 0.91 and a lower RMSE value of 5.65 ppb, indicating greater robustness under high-missing-rate conditions. For XCH_4, a similar pattern is observed. TROPOMI performs better than the fused product under low and moderate missing-rate conditions, with higher R^2 and lower RMSE. However, when MR > 0.5, TROPOMI performance decreases markedly, with R^2 declining to 0.81 and RMSE increasing to 15.84 ppb. The fused XCH_4 data show a slightly higher R^2 of 0.83 and a lower RMSE of 15.40 ppb under the same high-missing-rate condition. Although the fused XCH_4 bias is larger than that of TROPOMI, the higher R^2 and lower RMSE suggest that the fused dataset reduces overall errors under sparse-observation conditions. Overall, under conditions of good TROPOMI coverage, the fused XCO and XCH_4 datasets perform comparably to the TROPOMI datasets. When TROPOMI retrievals are sparse, the fused datasets show relatively high reliability. These results indicate that our fusion framework is effective for gap-filling applications.

To further position the proposed fused datasets within the current landscape of atmospheric composition products, we compared our TCCON-based validation results with reported accuracies of representative existing products in the literature, rather than conducting a direct pixel-level comparison. Direct product-to-product comparison is not straightforward because existing products differ in target species, input observations, assimilation or fusion methods, spatial and temporal resolutions, and validation protocols. For example, CAMS-EGG4 provides global greenhouse-gas reanalysis fields for CO_2 and CH_4 by combining model and satellite

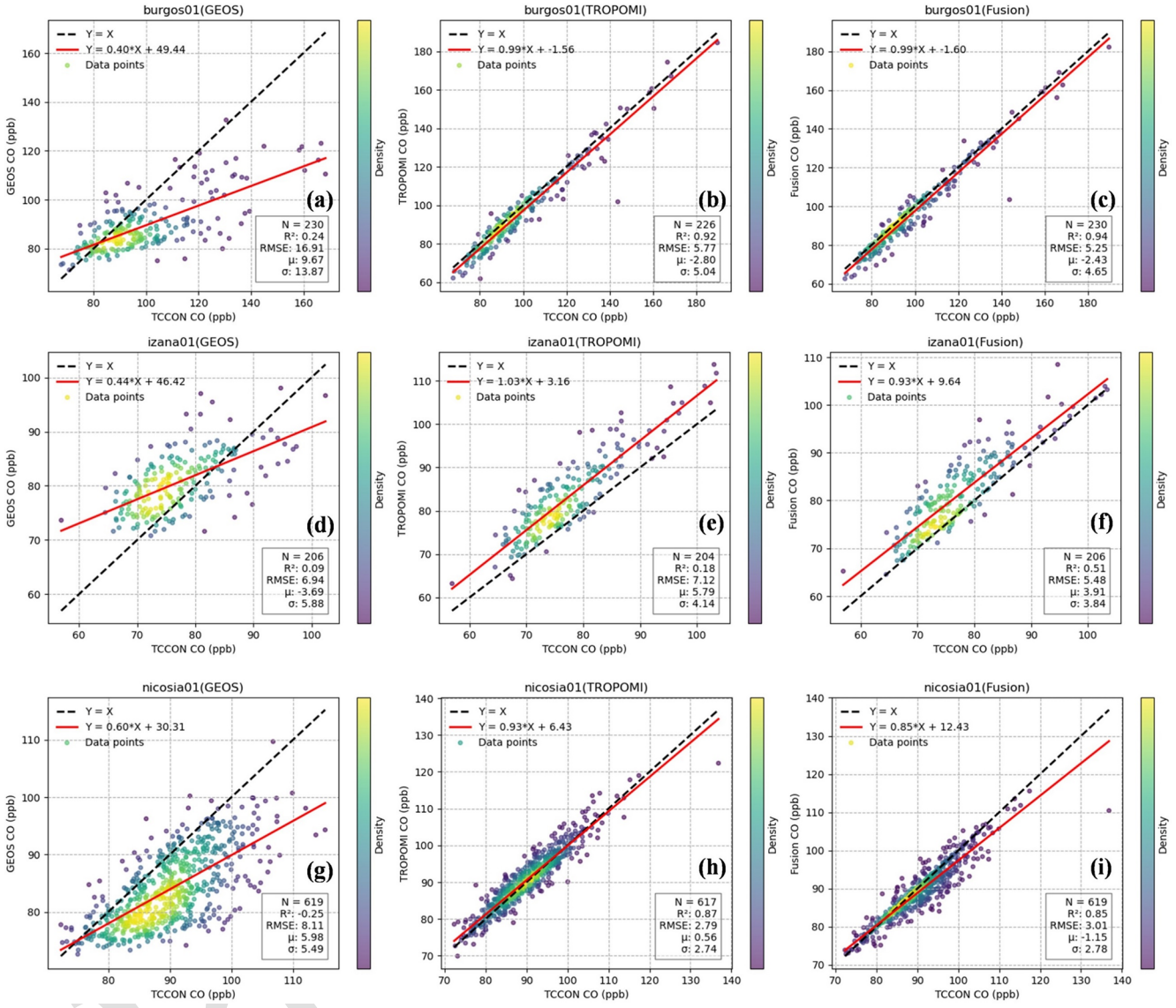

Figure 4. Density scatterplots of the independent validation results for GEOS-Chem, TROPOMI, and fused XCO data at burgos01, izana01, and nicosia01 under sparse TROPOMI coverage conditions (MR > 0.5). Panels **(a)**, **(d)**, **(g)**; **(b)**, **(e)**, **(h)**; and **(c)**, **(f)**, **(i)** correspond to GEOS-Chem, TROPOMI, and fused XCO data, respectively. Black dotted and red solid lines indicate the 1 : 1 line and fitted regression line, respectively. X: TCCON data; Y: GEOS-Chem, TROPOMI, and fused data. Units: ppb for RMSE, μ, and σ.

information within the ECMWF Integrated Forecasting System, but it does not provide a TROPOMI-based XCO product directly comparable to the fused XCO dataset developed here (Agustí-Panareda et al., 2023). Agustí-Panareda et al. (2023) reported that CAMS-EGG4 generally shows monthly systematic and random errors for CO_2 and CH_4 within about 1 % during 2003–2020, while also noting caveats related to changes in assimilated observations and fixed underlying emissions. Wang et al. (2023) developed a daily seamless 0.25° global XCO_2 and XCH_4 fusion product from GOSAT, OCO-2, and CAMS-EGG4, and reported TCCON validation statistics of $\sigma \approx 1.18$ ppm for XCO_2 and $\sigma \approx 11.3$ ppb for XCH_4, with R^2 values of 0.91–0.95 for XCO_2 and 0.90 for XCH_4. In comparison, our fused datasets achieve R^2 values of 0.91 for XCO and 0.83 for XCH_4 against TCCON. Although these values are not strictly comparable because of differences in target gases, input data, and validation settings, they indicate that the proposed TROPOMI–GEOS-Chem fusion framework provides competitive TCCON-based validation performance while extending daily gap-free mapping to both XCO and XCH_4 at global 0.25° resolution and China-specific 0.05° resolution.

Note that averaging kernels (AKs) were not explicitly applied in the TCCON comparison. This choice was made

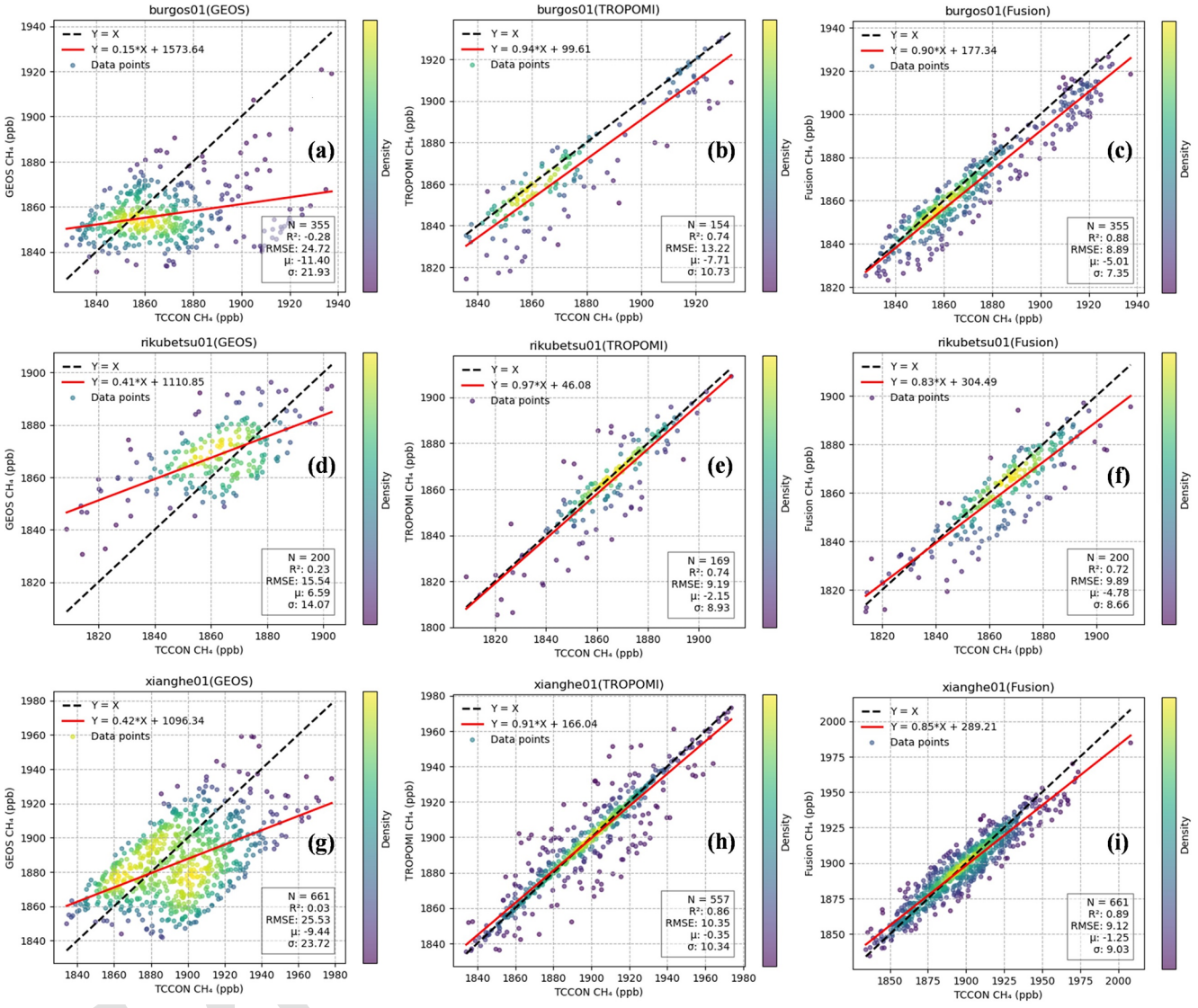

Figure 5. Density scatterplots of the independent validation results for GEOS-Chem, TROPOMI, and fused XCH_4 data at burgos01, rikubetsu01, and xianghe01 under sparse TROPOMI coverage conditions (MR > 0.5). Panels **(a)**, **(d)**, **(g)**; **(b)**, **(e)**, **(h)**; and **(c)**, **(f)**, **(i)** correspond to GEOS-Chem, TROPOMI, and fused XCH_4 data, respectively. Black dotted and red solid lines indicate the 1 : 1 line and fitted regression line, respectively. X: TCCON data; Y: GEOS-Chem, TROPOMI, and fused data. Units: ppb for RMSE, μ, and σ.

because the validation focuses on column-averaged dry-air mole fractions after strict quality filtering and spatial–temporal aggregation, for which AK-related smoothing differences are expected to be smaller than retrieval noise, sampling gaps, and spatial representativeness errors. Previous TROPOMI XCH_4/XCO validation studies have shown that AK-related adjustments can be small when column averaging kernels are close to unity in the lower atmosphere (Schneising et al., 2026), although rigorous validation studies still account for smoothing uncertainty and a priori profile effects (Sha et al., 2021). Nevertheless, we acknowledge that ignoring AKs may introduce residual smoothing uncertainty, especially under cloudy conditions or in regions with strong vertical gradients.

3.2 Multi-scale spatiotemporal analysis

Figure S3 illustrates the comparison of worldwide yearly GEOS-Chem, TROPOMI, fused XCO, and XCH_4 for the years 2020 and 2022. The fused data exhibit a similar geographical distribution with TROPOMI. Despite a substantial underestimation of GEOS-Chem, it retains a robust geographic distribution alignment with TROPOMI, providing a critical reference for data fusion, and this underestimation is markedly altered post-fusion.

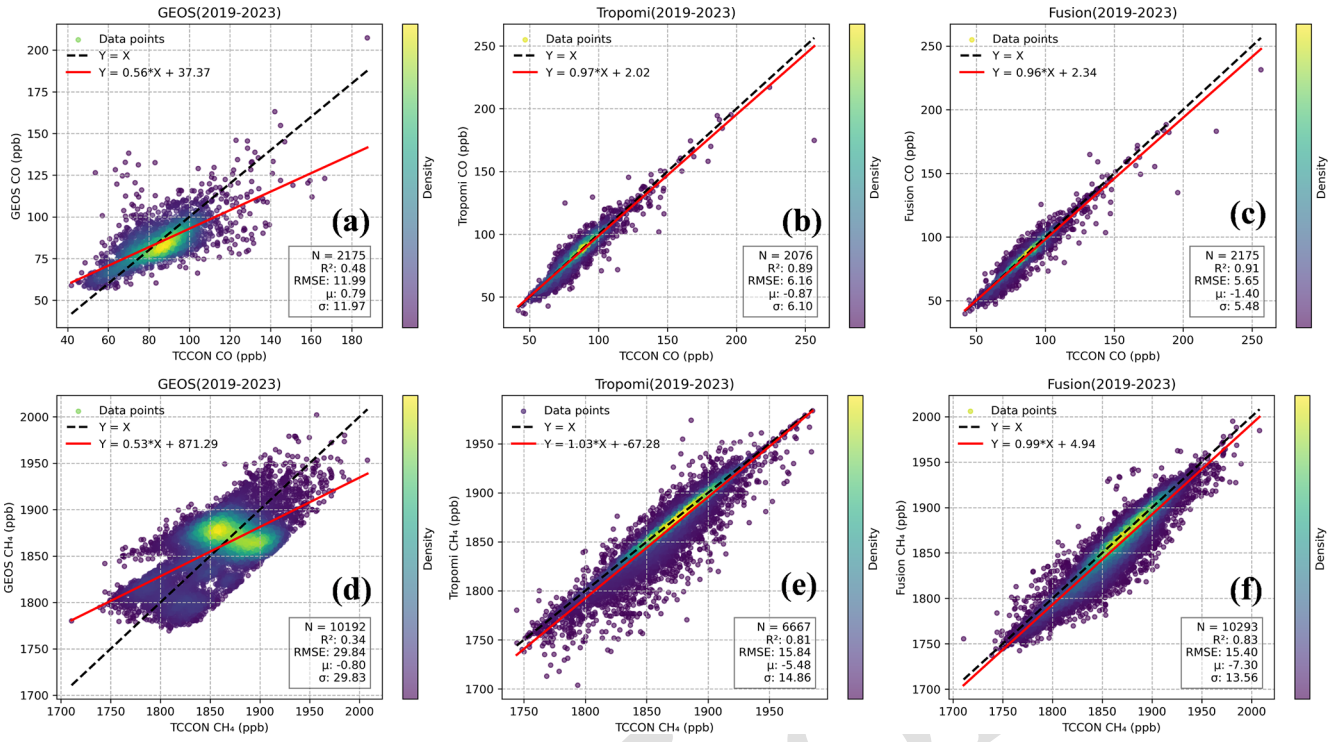

Figure 6. Density scatterplots of the overall validation results for GEOS-Chem, TROPOMI, and fused XCO and XCH_4 data at all corresponding validation sites under sparse TROPOMI coverage conditions (MR > 0.5). GEOS-Chem results were bias-corrected before validation. Panels **(a)**, **(d)**; **(b)**, **(e)**; and **(c)**, **(f)** correspond to GEOS-Chem, TROPOMI, and fused data, respectively. Black dotted and red solid lines indicate the 1 : 1 line and fitted regression line, respectively. X: TCCON data; Y: GEOS-Chem, TROPOMI, and fused data. Units: ppb for RMSE, μ, and σ.

Figure S4 illustrates the global distribution of fused XCO and XCH_4 for three representative days in 2020 and 2022, respectively. The fusion findings, as seen in the image, offer comprehensive information on atmospheric CO and CH_4, distinctly revealing their worldwide geographical distribution. For comparison, Fig. S5 illustrates the global distribution of XCO and XCH_4 observed by TROPOMI on these corresponding days. Meteorological factors have resulted in several gaps in the satellite observations, particularly evident in the XCH_4 data, when compared to Fig. S3. Figure S3 illustrates that the fused data addresses the deficiencies in geographical and temporal information, so improving data continuity while preserving the integrity of the satellite observations.

Figure 7 illustrates the global multi-year average distributions of fused XCO and XCH_4 for the period 2019–2023. Elevated concentrations of both gases are predominantly observed across Asia, particularly over China and India. For XCO, distinct high-value regions are also evident in Central Africa and northern South America. Figures S6 and S7 illustrate the seasonal averages of the fused global XCO and XCH_4 data from 2019 to 2023. The seasonal variations in the geographical distribution are distinctly captured by the fusion results. Notably, XCO exhibits more pronounced spatiotemporal variability compared to XCH_4. Specifically, CO tends to be spatially concentrated in certain regions, whereas CH_4 displays a relatively more uniform global distribution.

Figures S8 and S9 illustrate the spatial distribution of the annual rates of change in global XCO and XCH_4 from 2019 to 2023. For each grid cell, the rate of change is estimated by applying ordinary least-squares linear regression to the five annual mean values. Based on this analysis, XCO shows an increasing tendency over North America, whereas Central Africa and Eastern China show decreasing tendencies during 2019–2023 (Fig. S8). For XCH_4, the global annual mean increases over the same period, with a larger positive rate of change observed over Central Africa (Fig. S9).

3.3 Local high-resolution data analysis

To enhance the refinement capability of our fused data, we employed GEOS-Chem to simulate 0.25×0.3125 nested gridded data for the Chinese region. Utilizing the same methodology, we refined the fused data for this region, achieving a grid accuracy of 0.05°, which served as the basis for our analysis of the local area in China.

Figures S10 and S11 illustrate the variations in XCO and XCH_4 from the TROPOMI and fusion datasets during the late August 2022 Chongqing hill fire. To quantify the con-

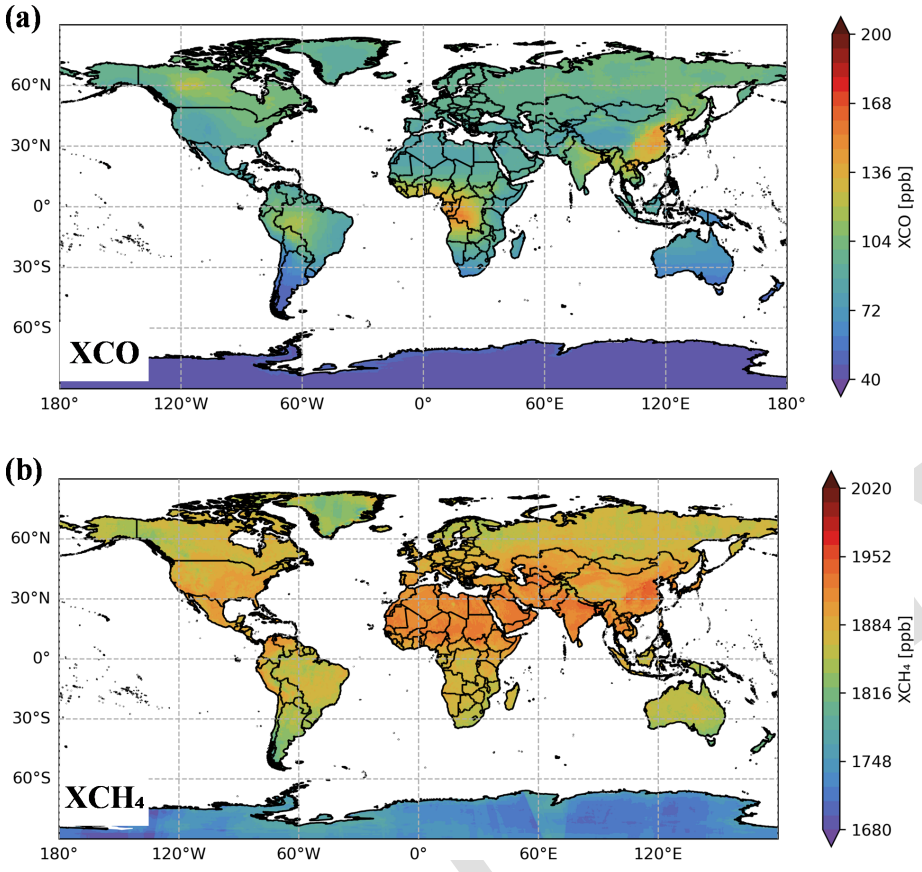

Figure 7. Multi-year mean global concentration distributions of fused **(a)** XCO and **(b)** XCH_4 data from 2019 to 2023. Color bars indicate XCO and XCH_4 concentrations in parts per billion.

centration enhancements, the fire-affected area is defined as a localized domain of 29°–30.5° N and 105.5°–107.5° E, based on the spatial extent of thermal anomalies and smoke plumes observed during the peak of the event. The active fire period is defined as 21–31 August 2022, while 10–20 August 2022 is used as the reference period. The reference period is selected as an adjacent pre-fire baseline, rather than a climatological baseline, to reduce the influence of broader seasonal variability and to better isolate short-term concentration changes associated with the fire episode. Mean XCO and XCH_4 values are calculated by averaging all grid cells within the defined fire-affected domain for each day. The reported enhancements are then calculated as the difference between the mean concentrations during the active fire period and those during the reference period. Based on this definition, the fused data show average increases of 17.1 ppb for XCO and 24.5 ppb for XCH_4 during the fire period. The comparison also indicates that the fused datasets preserve local spatial features observed by TROPOMI while providing more complete spatial coverage over the fire-affected region.

Rice paddies are an important source of CH_4 emissions, while associated agricultural activities, such as crop-residue burning, may also affect regional CO variability. Therefore, we further analyzed the short-term temporal changes in XCO and XCH_4 over the major rice-growing regions of Northeast and Southeast China to evaluate the applicability of the high-resolution fused datasets in agricultural emission-related regions. Figures 8 and 9 show the temporal variations in XCO and XCH_4 derived from TROPOMI and the fused data over these regions, respectively. To reduce the influence of the seasonal cycle, the short-term rates of change were derived using harmonic regression with annual and semiannual terms, and the corresponding 95 % confidence intervals are provided in Tables S9 and S10.

The results show that XCO exhibits negative short-term rates of change in both the satellite and fused datasets across the rice cultivation areas of Northeast and Southeast China. In Northeast China, the XCO rate is $-2.36\,\mathrm{ppb\,yr^{-1}}$ TS12 for TROPOMI and $-2.28\,\mathrm{ppb\,yr^{-1}}$ for the fused data. In Southeast China, the XCO rate is $-1.29\,\mathrm{ppb\,yr^{-1}}$ for TROPOMI and $-1.76\,\mathrm{ppb\,yr^{-1}}$ for the fused data. The consistent negative rates indicate that the fused dataset preserves the main decreasing XCO signal observed by TROPOMI. Differences between the two datasets may be partly related to the uneven temporal sampling of TROPOMI caused by clouds, aerosols,

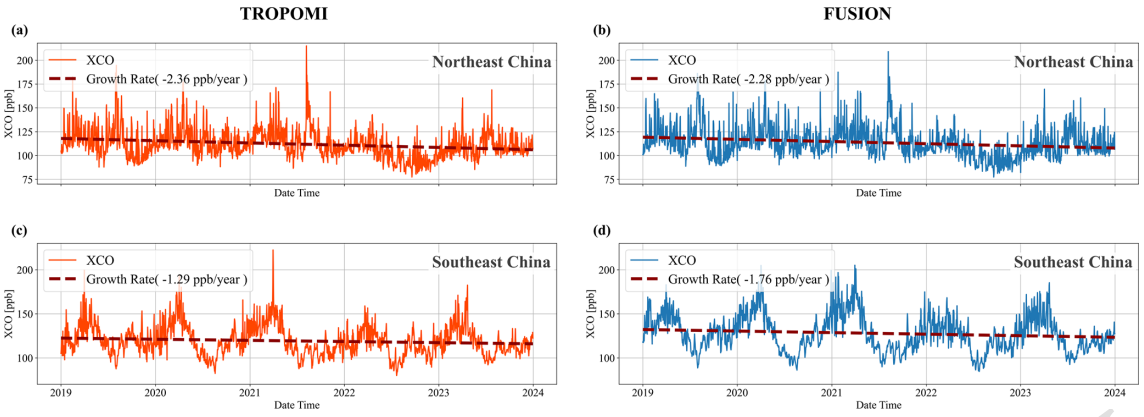

Figure 8. Daily time series of TROPOMI and fused XCO data over rice-growing regions in **(a, b)** Northeast China and **(c, d)** Southeast China. Red dashed lines indicate short-term rates of change derived from harmonic regression after accounting for the seasonal cycle, with values shown in parentheses. Unit: parts per billion per year.

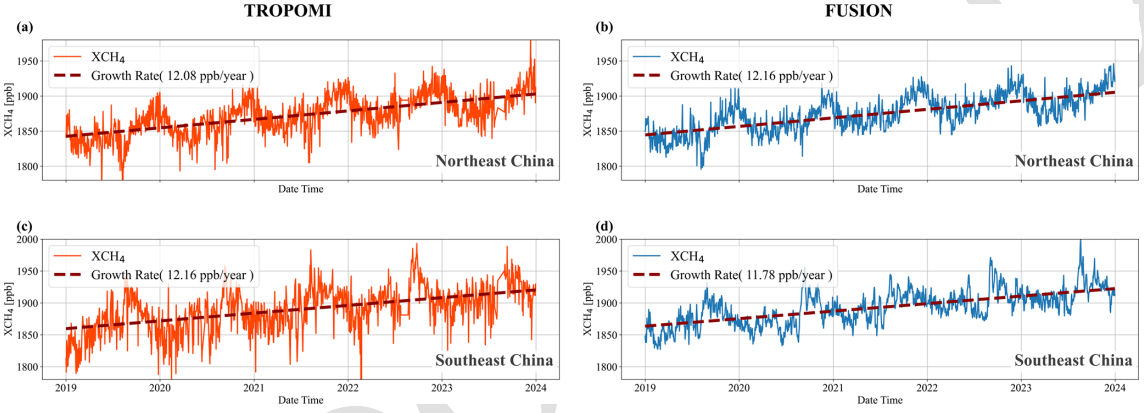

Figure 9. Daily time series of TROPOMI and fused XCH_4 data over rice-growing regions in **(a, b)** Northeast China and **(c, d)** Southeast China. Red dashed lines indicate short-term rates of change derived from harmonic regression after accounting for the seasonal cycle, with values shown in parentheses. Unit: parts per billion per year.

and retrieval filtering, whereas the fused dataset provides spatially and temporally continuous fields.

In contrast, XCH_4 exhibits positive short-term rates of change in both the satellite and fused datasets. In Northeast China, the XCH_4 rate is 12.08 ppb yr^{-1} for TROPOMI and 12.16 ppb yr^{-1} for the fused data. In Southeast China, the XCH_4 rate is 12.16 ppb yr^{-1} for TROPOMI and 11.78 ppb yr^{-1} for the fused data. The close agreement between the two datasets indicates that the fused dataset preserves the main increasing XCH_4 signal observed by TROPOMI while providing improved temporal continuity.

3.4 Discussion of uncertainties and limitations

Although the proposed fusion framework substantially improves the spatial continuity of TROPOMI XCO and XCH_4 products and provides daily gap-free fields, several limitations should be considered when using the fused datasets.

Figures 10–11 show the temporal and latitudinal distributions of daily biases between GEOS, Fused, TROPOMI, and TCCON. Overall, GEOS exhibits a clear systematic underestimation of both XCO and XCH_4 relative to TCCON, especially after 2020, when strong negative biases appear at many sites. After fused, the biases are substantially reduced, with colors generally shifting from dark blue to light blue or near white. This indicates that the fusion method effectively corrects the systematic underestimation in GEOS and improves its consistency with TCCON. Compared with GEOS, TROPOMI shows smaller overall biases, but it has more missing data and still exhibits localized positive or negative biases at some sites and periods. Overall, the fusion product performs better than GEOS in reducing biases and improving reliability, although residual uncertainties remain at individual sites.

AKs are an important source of uncertainty in comparisons between satellite retrievals and model simulations because they describe the vertical sensitivity of TROPOMI re-

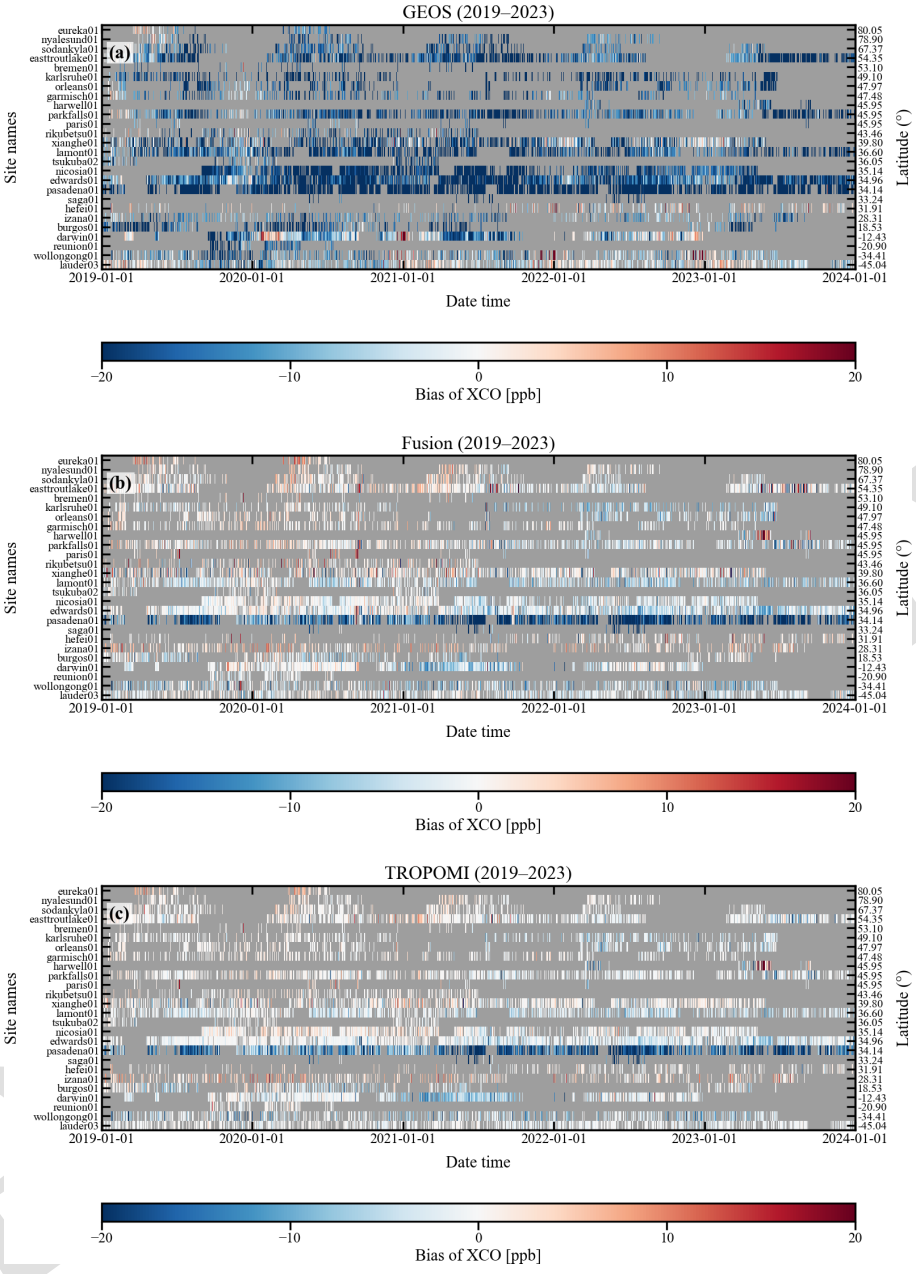

Figure 10. Heat maps of the biases between daily **(a)** GEOS, **(b)** fused, and **(c)** TROPOMI XCO and TCCON XCO over time and latitude. Color ramps stand for the biases of XCO. Background colors (grey) indicate the missing data.

trievals. To evaluate this effect, we conducted an AK sensitivity test using the 2019 CO data by comparing GEOS-Chem XCO with TROPOMI XCO before and after applying the TROPOMI CO column averaging kernels. As shown in Table S11, applying the TROPOMI CO column averaging kernels slightly increased the R^2 from 0.543 to 0.565, while the RMSE increased from 28.17 to 28.66 ppb and the mean bias changed from −17.09 to −18.34 ppb. These results indicate that AK treatment has a limited influence on the column-concentration-based fusion results in this experiment.

Another concern relates to the dependence of the fused datasets on the accuracy of the GEOS-Chem prior and the representativeness of valid TROPOMI observations. GEOS-Chem provides physically consistent large-scale background information, but it carries uncertainties from emission inventories, transport, chemistry, boundary conditions, vertical mixing, and other model parameterizations. In regions or periods where valid TROPOMI retrievals are persistently sparse – such as the Intertropical Convergence Zone, high-latitude zones, or heavily polluted areas with dense cloud or

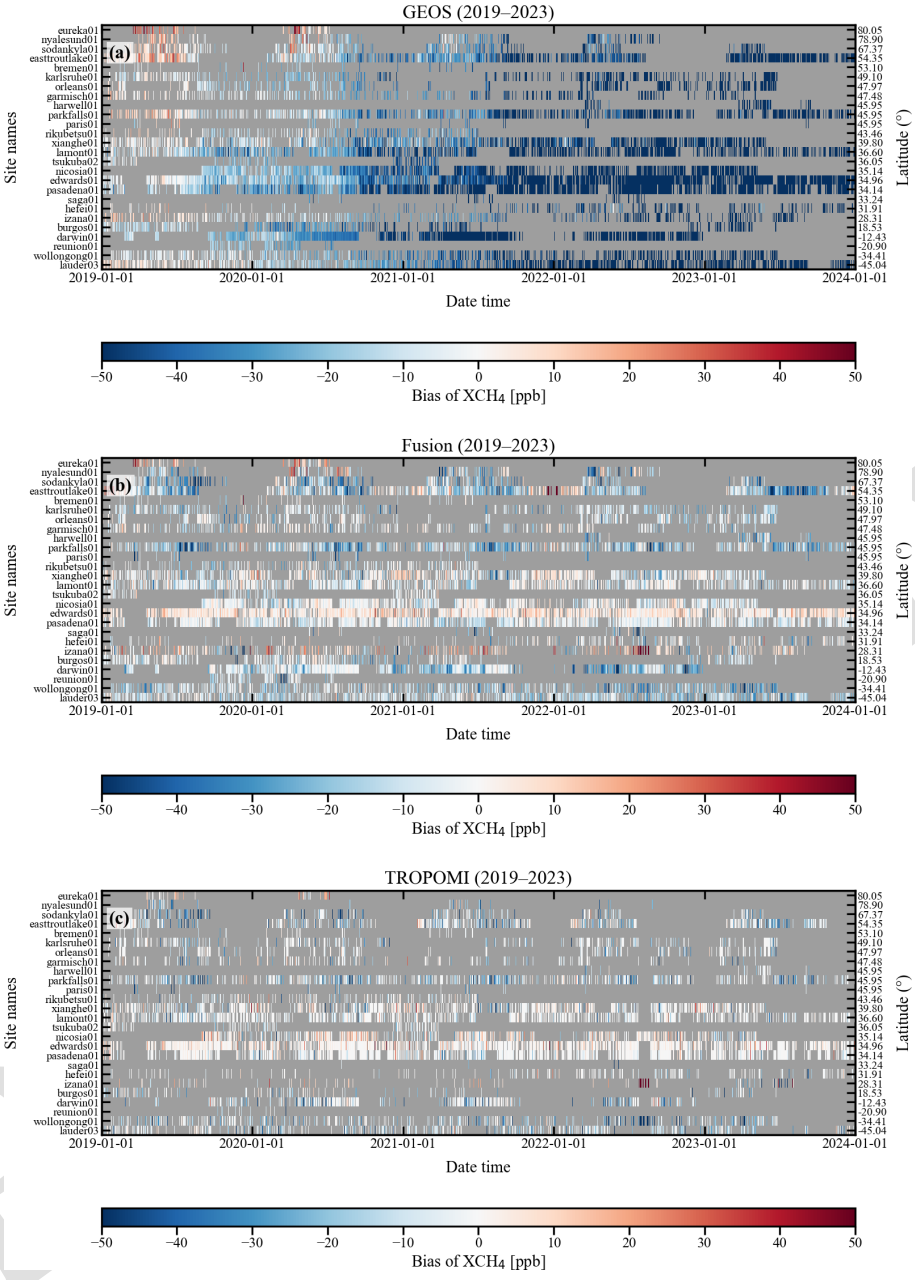

Figure 11. Heat maps of the biases between daily **(a)** GEOS, **(b)** fused, and **(c)** TROPOMI XCH_4 and TCCON XCH_4 over time and latitude. Color ramps stand for the biases of XCH_4. Background colors (grey) indicate the missing data.

aerosol cover – the reconstruction relies more strongly on the GEOS-Chem prior and the DCT/SVD spatiotemporal structure. Consequently, systematic biases in the prior model may propagate into the fused fields. Users are therefore advised to consult the obs_mask variable provided in the dataset to assess observational support and to interpret the reconstructed values cautiously in persistently missing regions.

Figure S12 shows that the fused XCO data generated directly at the native 2° × 2.5° resolution (Fig. S12a) and the 0.25° fused XCO data aggregated to the same 2° × 2.5° grid (Fig. S12b) present relatively similar global spatial distributions, with relatively high XCO over East Asia, central Africa, and parts of South America. The quantitative comparison results in Fig. S13 also show high consistency, with $R^2 = 0.926$, RMSE = 6.93 ppb, and mean bias = 0.15 ppb. These results indicate that the 0.25° fused XCO data preserve the coarse-scale spatial structure, and that the resolution mismatch does not have a substantial influence on the coarse-scale reconstruction fidelity of XCO.

The masked residual-learning strategy itself introduces uncertainty in persistently missing regions. Because the residual U-Net is trained only at valid TROPOMI pixels, the model learns the statistical relationship between the preliminary reconstruction error and the input predictors under observed conditions. When applied to unobserved regions – especially those with persistent cloud cover or aerosol contamination – the model infers the most likely residual pattern given similar meteorological and spatiotemporal conditions, rather than the exact unobserved atmospheric state on a specific day. This does not invalidate the gap-filling framework, but it implies that reconstructed values in long-term data gaps should be understood as observation-constrained estimates rather than direct measurements. Users should be aware of this extrapolation uncertainty when analyzing local short-lived anomalies in persistently cloudy regions.

Finally, possible discontinuities associated with changes in the input satellite product should be considered. The TROPOMI pixel size changed around June 2019 from 7.0 km × 7.0 km to 7.0 km × 5.5 km, which may affect sampling density and local spatial representativeness. Although the fusion framework uses gridding, quality filtering, and residual correction to mitigate such effects, subtle artifacts or temporal discontinuities may remain. Caution is advised when interpreting local-scale or short-term variations that span this instrumental transition.

In summary, the fused datasets are most suitable for studying daily to seasonal spatiotemporal variability, improving data continuity and spatial coverage, examining regional-scale events (e.g., biomass burning emissions), and supporting comparative analyses during 2019–2023. For fine-scale local anomaly detection in persistently missing regions, for applications in areas lacking independent ground-based validation, or for studies requiring strict profile-sensitive validation or long-term climate trend attribution, users should consult the observation-support variables, including obs_mask, MR, and observation_fraction, and carefully consider all the limitations discussed above.

4 Fused dataset summary

The final data products generated in this study consist of daily gap-free fused XCO and XCH_4 datasets for the period 2019–2023. Two spatial scales are provided. The global product contains daily XCO and XCH_4 fields at 0.25° × 0.25° resolution, covering the global domain. The regional product provides higher-resolution daily XCO and XCH_4 fields over China at 0.05° × 0.05° resolution. Both products are reported as column-averaged dry-air mole fractions in parts per billion (ppb).

The datasets are distributed in NetCDF4 (.nc/.nc4) format. Each monthly file contains gridded daily fields with the dimensions "time", "lat", and "lon". The "time" coordinate is defined using CF-style physical time units, for example, "days since 2019-01-01 00:00:00", with "time = 0" corresponding to the first day of the month. Users can therefore retrieve a specific daily field by selecting the corresponding "time" index in the monthly NetCDF file. The coordinate variables "lat" and "lon" define the spatial grid of each product.

The main data variables are "XCO" and "XCH_4", representing the fused column-averaged dry-air mole fractions of CO and CH_4, respectively. In addition, an obs_mask variable is provided to indicate whether the value at each grid cell is a TROPOMI observation or reconstructed by the fusion framework. The obs_mask variable is a dimensionless binary mask, where obs_mask = 1 indicates that the value at that grid cell and day is a TROPOMI observation, while obs_mask = 0 indicates that the value was generated by the proposed reconstruction method.

An "observation_fraction" variable is also included in each monthly file. This variable is calculated as the mean of "obs_mask" over the "time" dimension and represents the fraction of days in that month for which valid TROPOMI observations are available at each grid cell. For example, "observation_fraction = 1" indicates that valid satellite observations are available for all days in the month at that grid cell, whereas "observation_fraction = 0" indicates that all daily values at that grid cell are reconstructed by the fusion framework. This information is provided to prevent users from interpreting the entire gap-free product as purely observation-based.

Because of file-size and repository upload limitations, the daily data are organized into monthly NetCDF files rather than being provided as a single file for the entire study period. This monthly organization does not indicate temporal averaging. Each monthly file contains all daily fields for that calendar month.

The revised NetCDF files include CF-style metadata, including coordinate attributes, variable long names, units, missing-value definitions, temporal coverage, spatial resolution, product region, data source, author, institution, and contact information. The main variables included in the released NetCDF files are summarized as follows: "XCO" (ppb), fused column-averaged dry-air mole fraction of carbon monoxide; "XCH_4" (ppb), fused column-averaged dry-air mole fraction of methane; "obs_mask" (unitless binary mask), indicating whether the grid-cell value is derived from a valid satellite observation ("1") or reconstructed by the proposed fusion method ("0"); "observation_fraction" (unitless fraction from 0 to 1), indicating the monthly fraction of days with valid satellite observations at each grid cell; "time", the daily time coordinate within each monthly file, defined using CF-style units; and "lat" and "lon", the latitude and longitude coordinates of the gridded product.

5 Data availability

The global (0.25°) and China-specific (0.05°) daily gap-free XCO and XCH_4 datasets (2019–2023) generated in this study are openly available in the Zenodo repository at https://doi.org/10.5281/zenodo.22010891 (An et al., 2026).

6 Conclusions

This study developed a two-stage signal-domain guided spatio-temporal fusion framework for generating continuous XCO and XCH_4 products by integrating GEOS-Chem simulations with TROPOMI satellite observations. The framework first applies 3D DCT and SVD to reconstruct low-rank signal-domain structures, thereby providing a robust prior for missing-value approximation. A lightweight residual convolutional neural network is then used to refine pixel-level residuals under observational constraints. Using this framework, we generated daily continuous XCO and XCH_4 products for 2019–2023 at 0.25° global resolution and 0.05° resolution over China.

Validation against independent TCCON observations shows that the fused datasets improve spatiotemporal continuity while still maintaining good consistency with TCCON data. In the time-series comparison analysis at representative sites, the fused XCO and XCH_4 products generally conform to the temporal variation patterns observed by TCCON and TROPOMI, while reducing the systematic underestimation phenomenon associated with GEOS-Chem simulations. To further evaluate the performance under different TROPOMI data coverage rates, we conducted experiments based on missing-rate thresholds, with three MR intervals: $MR < 0.3$, $0.3 \leq MR \leq 0.5$, and $MR > 0.5$. The results show that when $MR < 0.3$ and $0.3 \leq MR \leq 0.5$, the fused products perform comparably to the original TROPOMI retrievals; whereas under sparse TROPOMI data coverage conditions ($MR > 0.5$), the fused XCO product obtained a higher R^2 value and a lower RMSE value than the TROPOMI product, with $R^2 = 0.91$ and $RMSE = 5.65$ ppb. The fused XCH_4 product obtained a higher R^2 value and a lower RMSE value, with $R^2 = 0.83$ and $RMSE = 15.40$ ppb. Compared with GEOS-Chem simulations, the fused products significantly reduced systematic bias.

The fused datasets also captured meaningful spatial and temporal patterns of atmospheric CO and CH_4. During 2019–2023, the reconstructed products revealed regional XCO increases in parts of North America, decreases over eastern China, and widespread XCH_4 growth. The high-resolution China product further resolved regional-scale features, including enhanced XCO and XCH_4 during the 2022 Chongqing wildfire event. Over rice-growing regions, the fused datasets preserved the main short-term temporal signals observed by TROPOMI while reducing the influence of uneven sampling caused by satellite data gaps.

By combining the spatiotemporal continuity of GEOS-Chem, the observational constraints of TROPOMI, and signal-domain reconstruction, the fused datasets provide more complete and statistically consistent representations of global and regional XCO and XCH_4 variations. The fusion process also mitigates the systematic underestimation observed in GEOS-Chem simulations. Such model biases may arise from multiple sources, including uncertainties in emission inventories, chemical mechanisms, atmospheric transport, boundary conditions, vertical mixing, and other model parameterizations. Therefore, the fused products can serve as an improved basis for analyzing long-term spatial patterns, seasonal variability, and regional changes in atmospheric CO and CH_4.

Despite these advances, several limitations remain. The fusion framework depends partly on the quality of GEOS-Chem simulations, and residual uncertainties may persist due to imperfect emission inventories, transport representation, chemical mechanisms, and parameterizations. Future work should further improve the prior model fields, refine the fusion strategy, and incorporate additional independent observations, such as GOSAT/GOSAT-2 for CH_4, MOPITT or IASI for CO, and expanded ground-based measurements, to further enhance the robustness and applicability of the dataset. Overall, this study provides continuous, high-resolution, and validated datasets of XCO and XCH_4 that can support future research on atmospheric composition, air quality, emission monitoring, and climate-related applications.

Supplement. The supplement related to this article is available online at [the link will be implemented upon publication].

Author contributions. CA supplemented the related experiments, revised and wrote the final manuscript, and reconstructed the dataset. ZL contributed to the study design, generated the preliminary dataset, and wrote the initial manuscript draft. QJ, PL, BC, and JX contributed to the study design. YT and YS provided constructive comments on the manuscript. All authors contributed to the study.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. This research was jointly supported by the National Key Research and Development Program of China (grant no. 2022YFC3700103), the National Natural Science Foundation of China (grant no. 42205151), the National Science Fund for Excellent Young Scholars (grant no. 62322514), and the Anhui Provincial Academician Workstation for Innovation in Optoelectronic Information Acquisition and Application.

Financial support. This research has been supported by the NAME OF FUNDER (grant no. GRANT AGREEMENT NO).

TS13

Review statement. This paper was edited by Alexander Kokhanovsky and reviewed by four anonymous referees.

References

Agustí-Panareda, A., Barré, J., Massart, S., Inness, A., Aben, I., Ades, M., Baier, B. C., Balsamo, G., Borsdorff, T., Bousserez, N., Boussetta, S., Buchwitz, M., Cantarello, L., Crevoisier, C., Engelen, R., Eskes, H., Flemming, J., Garrigues, S., Hasekamp, O., Huijnen, V., Jones, L., Kipling, Z., Langerock, B., McNorton, J., Meilhac, N., Noël, S., Parrington, M., Peuch, V.-H., Ramonet, M., Razinger, M., Reuter, M., Ribas, R., Suttie, M., Sweeney, C., Tarniewicz, J., and Wu, L.: Technical note: The CAMS greenhouse gas reanalysis from 2003 to 2020, Atmos. Chem. Phys., 23, 3829–3859, https://doi.org/10.5194/acp-23-3829-2023, 2023.

Akagi, S. K., Yokelson, R. J., Wiedinmyer, C., Alvarado, M. J., Reid, J. S., Karl, T., Crounse, J. D., and Wennberg, P. O.: Emission factors for open and domestic biomass burning for use in atmospheric models, Atmos. Chem. Phys., 11, 4039–4072, https://doi.org/10.5194/acp-11-4039-2011, 2011.

An, C., Tian, Y., Li, Z., Jiang, Q., Lin, P., Chang, B., Xue, J., and Sun, Y.: Signal-Domain Guided Deep Learning for Gap-Filling of XCO and XCH_4: A Masked Spatio-Temporal Fusion of TROPOMI and GEOS-Chem (2019–2023), Zenodo [data set], https://doi.org/10.5281/zenodo.22010891, 2026.

Andreae, M. O.: Emission of trace gases and aerosols from biomass burning – an updated assessment, Atmos. Chem. Phys., 19, 8523–8546, https://doi.org/10.5194/acp-19-8523-2019, 2019.

Andreae, M. O. and Merlet, P.: Emission of trace gases and aerosols from biomass burning, Global Biogeochem. Cycles, 15, 955–966, https://doi.org/10.1029/2000GB001382, 2001.

Bengherabi, M., Mezai, L., Harizi, F., Guessoum, A., and Cheriet, M.: Score Fusion of SVD and DCT-RLDA for Face Recognition, in: 2008 First Workshops on Image Processing Theory, Tools and Applications, Presented at the 2008 First Workshops on Image Processing Theory, Tools and Applications, pp. 1–8, https://doi.org/10.1109/IPTA.2008.4743776, 2008.

Borsdorff, T., aan de Brugh, J., Schneider, A., Lorente, A., Birk, M., Wagner, G., Kivi, R., Hase, F., Feist, D. G., Sussmann, R., Rettinger, M., Wunch, D., Warneke, T., and Landgraf, J.: Improving the TROPOMI CO data product: update of the spectroscopic database and destriping of single orbits, Atmos. Meas. Tech., 12, 5443–5455, https://doi.org/10.5194/amt-12-5443-2019, 2019.

Buschmann, M., Deutscher, N. M., Sherlock, V., Palm, M., Warneke, T., and Notholt, J.: Retrieval of xCO_2 from ground-based mid-infrared (NDACC) solar absorption spectra and comparison to TCCON, Atmos. Meas. Tech., 9, 577–585, https://doi.org/10.5194/amt-9-577-2016, 2016.

Chai, X., Tonjes, D. J., and Mahajan, D.: Methane emissions as energy reservoir: Context, scope, causes and mitigation strategies, Prog. Energ. Combust., 56, 33–70, https://doi.org/10.1016/j.pecs.2016.05.001, 2016.

Chander, G., Hewison, T. J., Fox, N., Wu, X., Xiong, X., and Blackwell, W. J.: Overview of Intercalibration of Satellite Instruments, IEEE T. Geosci. Remote, 51, 1056–1080, https://doi.org/10.1109/TGRS.2012.2228654, 2013.

Chen, J., Shen, H., Li, X., Li, T., and Wei, Y.: Ground-level ozone estimation based on geo-intelligent machine learning by fusing in-situ observations, remote sensing data, and model simulation data, Int. J. Appl. Earth Obs., 112, 102955, https://doi.org/10.1016/j.jag.2022.102955, 2022.

Chen, X., Huang, Y., Nie, C., Zhang, S., Wang, G., Chen, S., and Chen, Z.: A long-term reconstructed TROPOMI solar-induced fluorescence dataset using machine learning algorithms, Sci. Data, 9, 427, https://doi.org/10.1038/s41597-022-01520-1, 2022.

Cogan, A. J., Boesch, H., Parker, R. J., Feng, L., Palmer, P. I., Blavier, J.-F. L., Deutscher, N. M., Macatangay, R., Notholt, J., Roehl, C., Warneke, T., and Wunch, D.: Atmospheric carbon dioxide retrieved from the Greenhouse gases Observing SATellite (GOSAT): Comparison with ground-based TCCON observations and GEOS-Chem model calculations, J. Geophys. Res.-Atmos., 117, D21301, https://doi.org/10.1029/2012JD018087, 2012.

Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J.-F., Ho, S.-P., Yudin, V., Attié, J.-L., Packman, D., Chen, J., Mao, D., and Drummond, J. R.: Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, J. Geophys. Res., 108, 4399, https://doi.org/10.1029/2002JD003186, 2003.

Elharar, E., Stern, A., Hadar, O., and Javidi, B.: A Hybrid Compression Method for Integral Images Using Discrete Wavelet Transform and Discrete Cosine Transform, J. Display Technol., 3, 321–325, https://doi.org/10.1109/JDT.2007.900915, 2007.

Filonchyk, M., Peterson, M. P., Zhang, L., Hurynovich, V., and He, Y.: Greenhouse gases emissions and global climate change: Examining the influence of CO_2, CH_4, and N_2O, Sci. Total Environ., 935, 173359, https://doi.org/10.1016/j.scitotenv.2024.173359, 2024.

Fritz, T. M., Eastham, S. D., Emmons, L. K., Lin, H., Lundgren, E. W., Goldhaber, S., Barrett, S. R. H., and Jacob, D. J.: Implementation and evaluation of the GEOS-Chem chemistry module version 13.1.2 within the Community Earth System Model v2.1, Geosci. Model Dev., 15, 8669–8704, https://doi.org/10.5194/gmd-15-8669-2022, 2022.

Gao, M., Xing, Z., Vollrath, C., Hugenholtz, C. H., and Barchyn, T. E.: Global observational coverage of onshore oil and gas methane sources with TROPOMI, Sci. Rep., 13, 16759, https://doi.org/10.1038/s41598-023-41914-8, 2023.

Garcia, D.: Robust smoothing of gridded data in one and higher dimensions with missing values, Comput. Stat. Data An., 54, 1167–1178, https://doi.org/10.1016/j.csda.2009.09.020, 2010.

Goudar, M., Anema, J. C. S., Kumar, R., Borsdorff, T., and Landgraf, J.: Plume detection and emission estimate for biomass burning plumes from TROPOMI carbon monoxide observations using APE v1.1, Geosci. Model Dev., 16, 4835–4852, https://doi.org/10.5194/gmd-16-4835-2023, 2023.

Griffin, D., Chen, J., Anderson, K., Makar, P., McLinden, C. A., Dammers, E., and Fogal, A.: Biomass burning CO emissions: exploring insights through TROPOMI-derived emissions and emission coefficients, Atmos. Chem. Phys., 24, 10159–10186, https://doi.org/10.5194/acp-24-10159-2024, 2024.

Guanter, L., Aben, I., Tol, P., Krijger, J. M., Hollstein, A., Köhler, P., Damm, A., Joiner, J., Frankenberg, C., and Landgraf, J.: Potential of the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence, Atmos. Meas. Tech., 8, 1337–1352, https://doi.org/10.5194/amt-8-1337-2015, 2015.

He, C., Ji, M., Grieneisen, M. L., and Zhan, Y.: A review of datasets and methods for deriving spatiotemporal distributions of atmospheric CO_2, J. Environ. Manage., 322, 116101, https://doi.org/10.1016/j.jenvman.2022.116101, 2022.

Heald, C. L., Jacob, D. J., Fiore, A. M., Emmons, L. K., Gille, J. C., Deeter, M. N., Warner, J., Edwards, D. P., Crawford, J. H., Hamlin, A. J., Sachse, G. W., Browell, E. V., Avery, M. A., Vay, S. A., Westberg, D. J., Blake, D. R., Singh, H. B., Sandholm, S. T., Talbot, R. W., and Fuelberg, H. E.: Asian outflow and trans-Pacific transport of carbon monoxide and ozone pollution: An integrated satellite, aircraft, and model perspective, J. Geophys. Res., 108, 2003JD003507, https://doi.org/10.1029/2003JD003507, 2003.

Heilig, G. K.: The greenhouse gas methane (CH_4): Sources and sinks, the impact of population growth, possible interventions, Popul. Environ., 16, 109–137, https://doi.org/10.1007/BF02208779, 1994.

Herkommer, B., Alberti, C., Castracane, P., Chen, J., Dehn, A., Dietrich, F., Deutscher, N. M., Frey, M. M., Groß, J., Gillespie, L., Hase, F., Morino, I., Pak, N. M., Walker, B., and Wunch, D.: Using a portable FTIR spectrometer to evaluate the consistency of Total Carbon Column Observing Network (TCCON) measurements on a global scale: the Collaborative Carbon Column Observing Network (COCCON) travel standard, Atmos. Meas. Tech., 17, 3467–3494, https://doi.org/10.5194/amt-17-3467-2024, 2024.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Hu, H., Hasekamp, O., Butz, A., Galli, A., Landgraf, J., Aan de Brugh, J., Borsdorff, T., Scheepmaker, R., and Aben, I.: The operational methane retrieval algorithm for TROPOMI, Atmos. Meas. Tech., 9, 5423–5440, https://doi.org/10.5194/amt-9-5423-2016, 2016.

Hu, J., Jia, J., Ma, Y., Liu, L., and Yu, H.: A Reconstructed Global Daily Seamless SIF Product at 0.05 Degree Resolution Based on TROPOMI, MODIS and ERA5 Data, Remote Sens., 14, 1504, https://doi.org/10.3390/rs14061504, 2022.

Hu, K., Liu, Z., Shao, P., Ma, K., Xu, Y., Wang, S., Wang, Y., Wang, H., Di, L., Xia, M., and Zhang, Y.: A review of satellite-based CO_2 data reconstruction studies: Methodologies, challenges, and advances, Remote Sens., 16, 3818, https://doi.org/10.3390/rs16203818, 2024.

Hu, L., Keller, C. A., Long, M. S., Sherwen, T., Auer, B., Da Silva, A., Nielsen, J. E., Pawson, S., Thompson, M. A., Trayanov, A. L., Travis, K. R., Grange, S. K., Evans, M. J., and Jacob, D. J.: Global simulation of tropospheric chemistry at 12.5 km resolution: performance and evaluation of the GEOS-Chem chemical module (v10-1) within the NASA GEOS Earth system model (GEOS-5 ESM), Geosci. Model Dev., 11, 4603–4620, https://doi.org/10.5194/gmd-11-4603-2018, 2018.

Imasu, R., Matsunaga, T., Nakajima, M., Yoshida, Y., Shiomi, K., Morino, I., Saitoh, N., Niwa, Y., Someya, Y., Oishi, Y., Hashimoto, M., Noda, H., Hikosaka, K., Uchino, O., Maksyutov, S., Takagi, H., Ishida, H., Nakajima, T. Y., Nakajima, T., and Shi, C.: Greenhouse gases Observing SATellite 2 (GOSAT-2): mission overview, Prog. Earth Planet. Sci., 10, 33, https://doi.org/10.1186/s40645-023-00562-2, 2023.

Inness, A., Aben, I., Ades, M., Borsdorff, T., Flemming, J., Jones, L., Landgraf, J., Langerock, B., Nedelec, P., Parrington, M., and Ribas, R.: Assimilation of S5P/TROPOMI carbon monoxide data with the global CAMS near-real-time system, Atmos. Chem. Phys., 22, 14355–14376, https://doi.org/10.5194/acp-22-14355-2022, 2022.

Jackson, R. B., Saunois, M., Bousquet, P., Canadell, J. G., Poulter, B., Stavert, A. R., Bergamaschi, P., Niwa, Y., Segers, A., and Tsuruta, A.: Increasing anthropogenic methane emissions arise equally from agricultural and fossil fuel sources, Environ. Res. Lett., 15, 071002, https://doi.org/10.1088/1748-9326/ab9ed2, 2020.

Jiang, Y., Si, C., and Yang, L.: Improvement Strategies for Mask R-CNN in Satellite Image Analysis, in: 2024 3rd International Conference on Electronics and Information Technology (EIT), Presented at the 2024 3rd International Conference on Electronics and Information Technology (EIT), pp. 739–744, https://doi.org/10.1109/EIT63098.2024.10762551, 2024.

Karunasingha, D. S. K.: Root mean square error or mean absolute error? Use their ratio as well, Inform. Sciences, 585, 609–629, https://doi.org/10.1016/j.ins.2021.11.036, 2022.

Kiel, M., Wunch, D., Wennberg, P. O., Toon, G. C., Hase, F., and Blumenstock, T.: Improved retrieval of gas abundances from near-infrared solar FTIR spectra measured at the Karlsruhe TCCON station, Atmos. Meas. Tech., 9, 669–682, https://doi.org/10.5194/amt-9-669-2016, 2016.

Kobayashi, K. and Salam, M. U.: Comparing Simulated and Measured Values Using Mean Squared Deviation and its Components, Agron. J., 92, 345–352, https://doi.org/10.2134/agronj2000.922345x, 2000.

Landgraf, J., aan de Brugh, J., Scheepmaker, R., Borsdorff, T., Hu, H., Houweling, S., Butz, A., Aben, I., and Hasekamp, O.: Carbon monoxide total column retrievals from TROPOMI shortwave infrared measurements, Atmos. Meas. Tech., 9, 4955–4975, https://doi.org/10.5194/amt-9-4955-2016, 2016.

Laughner, J. L., Roche, S., Kiel, M., Toon, G. C., Wunch, D., Baier, B. C., Biraud, S., Chen, H., Kivi, R., Laemmel, T., McKain, K., Quéhé, P.-Y., Rousogenous, C., Stephens, B. B., Walker, K., and Wennberg, P. O.: A new algorithm to generate a pri-

ori trace gas profiles for the GGG2020 retrieval algorithm, Atmos. Meas. Tech., 16, 1121–1146, https://doi.org/10.5194/amt-16-1121-2023, 2023.

Lelieveld, J., Gromov, S., Pozzer, A., and Taraborrelli, D.: Global tropospheric hydroxyl distribution, budget and reactivity, Atmos. Chem. Phys., 16, 12477–12493, https://doi.org/10.5194/acp-16-12477-2016, 2016.

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

Liang, R., Zhang, Y., Chen, W., Zhang, P., Liu, J., Chen, C., Mao, H., Shen, G., Qu, Z., Chen, Z., Zhou, M., Wang, P., Parker, R. J., Boesch, H., Lorente, A., Maasakkers, J. D., and Aben, I.: East Asian methane emissions inferred from high-resolution inversions of GOSAT and TROPOMI observations: a comparative and evaluative analysis, Atmos. Chem. Phys., 23, 8039–8057, https://doi.org/10.5194/acp-23-8039-2023, 2023.

Lin, P., Tian, Y., Borsdorff, T., Li, Z., Landgraf, J., Wu, H., Xue, J., Ding, D., Ye, H., Zhu, Y., and Liu, C.: TROPOMI unravels transboundary transport pathways of atmospheric carbon monoxide in Tibetan Plateau, Sci. Total Environ., 952, 175942, https://doi.org/10.1016/j.scitotenv.2024.175942, 2024.

Loew, A., Bell, W., Brocca, L., Bulgin, C. E., Burdanowitz, J., Calbet, X., Donner, R. V., Ghent, D., Gruber, A., Kaminski, T., Kinzel, J., Klepp, C., Lambert, J., Schaepman-Strub, G., Schröder, M., and Verhoelst, T.: Validation practices for satellite-based Earth observation data across communities, Rev. Geophys., 55, 779–817, https://doi.org/10.1002/2017rg000562, 2017.

Majhi, M. and Pal, A. K.: An image retrieval scheme based on block level hybrid dct-svd fused features, Multimed. Tools Appl., 80, 7271–7312, https://doi.org/10.1007/s11042-020-10005-5, 2021.

Marais, E. A. and Wiedinmyer, C.: Air Quality Impact of Diffuse and Inefficient Combustion Emissions in Africa (DICE-Africa), Environ. Sci. Technol., 50, 10739–10745, https://doi.org/10.1021/acs.est.6b02602, 2016.

McMillan, W. W., Barnet, C., Strow, L., Chahine, M. T., McCourt, M. L., Warner, J. X., Novelli, P. C., Korontzi, S., Maddy, E. S., and Datta, S.: Daily global maps of carbon monoxide from NASA's Atmospheric Infrared Sounder, Geophys. Res. Lett., 32, 2004GL021821, https://doi.org/10.1029/2004GL021821, 2005.

Naseem, S., Mahmood, T., Khan, A. R., Farooq, U., Nawazish, S., Alamri, F. S., and Saba, T.: Image Fusion Using Wavelet Transformation and XGboost Algorithm, Comput. Mater. Contin., 79, 801–817, https://doi.org/10.32604/cmc.2024.047623, 2024.

Okolie, C. J. and Smit, J. L.: A systematic review and meta-analysis of Digital elevation model (DEM) fusion: pre-processing, methods and applications, ISPRS J. Photogramm., 188, 1–29, https://doi.org/10.1016/j.isprsjprs.2022.03.016, 2022.

Peng, J., Kim, C.-S., and Kuo, C.-C. J.: Technologies for 3D mesh compression: A survey, J. Vis. Commun. Image Represent., 16, 688–733, https://doi.org/10.1016/j.jvcir.2005.03.001, 2005.

Ramsden, A. E., Ganesan, A. L., Western, L. M., Rigby, M., Manning, A. J., Foulds, A., France, J. L., Barker, P., Levy, P., Say, D., Wisher, A., Arnold, T., Rennick, C., Stanley, K. M., Young, D., and O'Doherty, S.: Quantifying fossil fuel methane emissions using observations of atmospheric ethane and an uncertain emission ratio, Atmos. Chem. Phys., 22, 3911–3929, https://doi.org/10.5194/acp-22-3911-2022, 2022.

Rao, K. R. and Yip, P.: Discrete cosine transform: algorithms, advantages, applications, Academic Press, San Diego, ISBN 9780125802031, 2014.

Rinsland, C. P., Luo, M., Logan, J. A., Beer, R., Worden, H., Kulawik, S. S., Rider, D., Osterman, G., Gunson, M., Eldering, A., Goldman, A., Shephard, M., Clough, S. A., Rodgers, C., Lampel, M., and Chiou, L.: Nadir measurements of carbon monoxide distributions by the Tropospheric Emission Spectrometer instrument onboard the Aura Spacecraft: Overview of analysis approach and examples of initial results, Geophys. Res. Lett., 33, 2006GL027000, https://doi.org/10.1029/2006GL027000, 2006.

Robinson, J. and Kecman, V.: Combining support vector machine learning with the discrete cosine transform in image compression, IEEE T. Neural Networ., 14, 950–958, https://doi.org/10.1109/TNN.2003.813842, 2003.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, edited by: Navab, N., Hornegger, J., Wells, W. M., Frangi, A. F., Springer International Publishing, Cham, pp. 234–241, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Schneising, O., Buchwitz, M., Reuter, M., Bovensmann, H., Burrows, J. P., Borsdorff, T., Deutscher, N. M., Feist, D. G., Griffith, D. W. T., Hase, F., Hermans, C., Iraci, L. T., Kivi, R., Landgraf, J., Morino, I., Notholt, J., Petri, C., Pollard, D. F., Roche, S., Shiomi, K., Strong, K., Sussmann, R., Velazco, V. A., Warneke, T., and Wunch, D.: A scientific algorithm to simultaneously retrieve carbon monoxide and methane from TROPOMI onboard Sentinel-5 Precursor, Atmos. Meas. Tech., 12, 6771–6802, https://doi.org/10.5194/amt-12-6771-2019, 2019.

Schneising, O., Buchwitz, M., Hachmeister, J., Vanselow, S., Reuter, M., Buschmann, M., Bovensmann, H., and Burrows, J. P.: Advances in retrieving XCH_4 and XCO from Sentinel-5 Precursor: improvements in the scientific TROPOMI/WFMD algorithm, Atmos. Meas. Tech., 16, 669–694, https://doi.org/10.5194/amt-16-669-2023, 2023.

Schneising, O., Bovensmann, H., Buchwitz, M., Buschmann, M., Deutscher, N. M., Griffith, D. W. T., Hachmeister, J., Hase, F., Iraci, L. T., Kivi, R., Morino, I., Ohyama, H., Petri, C., Reuter, M., Robinson, J., Roehl, C., Sha, M. K., Shiomi, K., Strong, K., Sussmann, R., Té, Y., Velazco, V. A., Vrekoussis, M., Wang, W., Warneke, T., Weidmann, D., Wunch, D., Zhou, M., and Bösch, H.: TROPOMI/WFMD v2.0: Improved retrievals of XCH_4 and XCO with XGBoost-based quality filtering, Atmos. Meas. Tech., 19, 2407–2435, https://doi.org/10.5194/amt-19-2407-2026, 2026.

Setianto, A. and Triandini, T.: Comparison of kriging and inverse distance weighted (IDW) interpolation methods in lineament extraction and analysis, J. Appl. Geol., 5, 21–29, https://doi.org/10.22146/jag.7204, 2013.

Sha, M. K., De Mazière, M., Notholt, J., Blumenstock, T., Chen, H., Dehn, A., Griffith, D. W. T., Hase, F., Heikkinen, P., Hermans, C., Hoffmann, A., Huebner, M., Jones, N., Kivi, R., Langerock, B., Petri, C., Scolas, F., Tu, Q., and Weidmann, D.: Intercomparison of low- and high-resolution infrared spectrometers for ground-

based solar remote sensing measurements of total column concentrations of CO_2, CH_4, and CO, Atmos. Meas. Tech., 13, 4791–4839, https://doi.org/10.5194/amt-13-4791-2020, 2020.

Sha, M. K., Langerock, B., Blavier, J.-F. L., Blumenstock, T., Borsdorff, T., Buschmann, M., Dehn, A., De Mazière, M., Deutscher, N. M., Feist, D. G., García, O. E., Griffith, D. W. T., Grutter, M., Hannigan, J. W., Hase, F., Heikkinen, P., Hermans, C., Iraci, L. T., Jeseck, P., Jones, N., Kivi, R., Kumps, N., Landgraf, J., Lorente, A., Mahieu, E., Makarova, M. V., Mellqvist, J., Metzger, J.-M., Morino, I., Nagahama, T., Notholt, J., Ohyama, H., Ortega, I., Palm, M., Petri, C., Pollard, D. F., Rettinger, M., Robinson, J., Roche, S., Roehl, C. M., Röhling, A. N., Rousogenous, C., Schneider, M., Shiomi, K., Smale, D., Stremme, W., Strong, K., Sussmann, R., Té, Y., Uchino, O., Velazco, V. A., Vigouroux, C., Vrekoussis, M., Wang, P., Warneke, T., Wizenberg, T., Wunch, D., Yamanouchi, S., Yang, Y., and Zhou, M.: Validation of methane and carbon monoxide from Sentinel-5 Precursor using TCCON and NDACC-IRWG stations, Atmos. Meas. Tech., 14, 6249–6304, https://doi.org/10.5194/amt-14-6249-2021, 2021.

Sicard, P., Crippa, P., De Marco, A., Castruccio, S., Giani, P., Cuesta, J., Paoletti, E., Feng, Z., and Anav, A.: High spatial resolution WRF-Chem model over Asia: Physics and chemistry evaluation, Atmos. Environ., 244, 118004, https://doi.org/10.1016/j.atmosenv.2020.118004, 2021.

Spivakovsky, C. M., Logan, J. A., Montzka, S. A., Balkanski, Y. J., Foreman-Fowler, M., Jones, D. B. A., Horowitz, L. W., Fusco, A. C., Brenninkmeijer, C. A. M., Prather, M. J., Wofsy, S. C., and McElroy, M. B.: Three-dimensional climatological distribution of tropospheric OH: Update and evaluation, J. Geophys. Res., 105, 8931–8980, https://doi.org/10.1029/1999JD901006, 2000.

Sun, L., Zhang, X., Fan, Y., Wang, Z., and Sun, X.: Downscaling aerosol optical depth by fusing satellite retrieval and model simulation using artificial intelligence technology, Atmos. Res., 328, 108411, https://doi.org/10.1016/j.atmosres.2025.108411, 2026.

Tang, Z.: Gap-filling using machine learning: implementations and applications in remote sensing, PhD thesis, University of Helsinki, Helsingin yliopisto, http://hdl.handle.net/10138/343816 (last access: TS14), ISBN: 978-951-51-6600-5, 2022.

Turquety, S., Hadji-Lazaro, J., Clerbaux, C., Hauglustaine, D. A., Clough, S. A., Cassé, V., Schlüssel, P., and Mégie, G.: Operational trace gas retrieval algorithm for the Infrared Atmospheric Sounding Interferometer, J. Geophys. Res., 109, 2004JD004821, https://doi.org/10.1029/2004JD004821, 2004.

Valerio, A., Chen, Y., Liu, C., Chen, Y., and Lin, C.: A Hybrid Regression–Kriging–Machine Learning Framework for Imputing Missing TROPOMI NO_2 Data over Taiwan, Remote Sens., 17, 2084, https://doi.org/10.3390/rs17122084, 2025.

Wall, M. E., Rechtsteiner, A., and Rocha, L. M.: Singular Value Decomposition and Principal Component Analysis, in: A Practical Approach to Microarray Data Analysis, edited by: Berrar, D. P., Dubitzky, W., and Granzow, M., Springer US, Boston, MA, pp. 91–109, https://doi.org/10.1007/0-306-47815-3_5, 2003.

Wang, M., Jiang, Z., Chen, X., Han, W., Zhu, L., He, T.-L., and Shen, Y.: Daily seamless dataset of HCHO concentrations: Vertical relationship between surface and column HCHO in China in 2019–2022, Atmos. Environ., 362, 121546, https://doi.org/10.1016/j.atmosenv.2025.121546, 2025.

Wang, Y., Qi, Q., and Liu, Y.: Unsupervised segmentation evaluation using area-weighted variance and Jeffries-Matusita distance for remote sensing images, Remote Sens., 10, 1193, https://doi.org/10.3390/rs10081193, 2018.

Wang, Y., Yuan, Q., Xiao, R., Li, T., and Zhang, L.: Recovery of the Carbon Monoxide Product from S5P-TROPOMI by Fusing Multiple Datasets: A Case Study in Hubei Province, China, in: IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 5529–5532, https://doi.org/10.1109/IGARSS39084.2020.9323092, 2020.

Wang, Y., Yuan, Q., Li, T., Tan, S., and Zhang, L.: Full-coverage spatiotemporal mapping of ambient $PM_{2.5}$ and PM_{10} over China from Sentinel-5P and assimilated datasets: Considering the precursors and chemical compositions, Sci. Total Environ., 793, 148535, https://doi.org/10.1016/j.scitotenv.2021.148535, 2021.

Wang, Y., Yuan, Q., Li, T., Yang, Y., Zhou, S., and Zhang, L.: Seamless mapping of long-term (2010–2020) daily global XCO_2 and XCH_4 from the Greenhouse Gases Observing Satellite (GOSAT), Orbiting Carbon Observatory 2 (OCO-2), and CAMS global greenhouse gas reanalysis (CAMS-EGG4) with a spatiotemporally self-supervised fusion method, Earth Syst. Sci. Data, 15, 3597–3622, https://doi.org/10.5194/essd-15-3597-2023, 2023.

Wei, C., Fan, H., Xie, S., Wu, C.-Y., Yuille, A., and Feichtenhofer, C.: Masked Feature Prediction for Self-Supervised Visual Pre-Training, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, New Orleans, LA, USA, pp. 14648–14658, https://doi.org/10.1109/CVPR52688.2022.01426, 2022.

Wei, J., Li, Z., Lyapustin, A., Wang, J., Dubovik, O., Schwartz, J., Sun, L., Li, C., Liu, S., and Zhu, T.: First close insight into global daily gapless 1 km $PM_{2.5}$ pollution, variability, and health impact, Nat. Commun., 14, 8349, https://doi.org/10.1038/s41467-023-43862-3, 2023.

Wei, Q., Song, W., Dai, B., Wu, H., Zuo, X., Wang, J., Chen, J., Li, J., Li, S., and Chen, Z.: Spatiotemporal estimation of surface NO2 concentrations in the Pearl River Delta region based on TROPOMI data and machine learning, Atmos. Pollut. Res., 16, 102353, https://doi.org/10.1016/j.apr.2024.102353, 2025.

Worden, J., Wecht, K., Frankenberg, C., Alvarado, M., Bowman, K., Kort, E., Kulawik, S., Lee, M., Payne, V., and Worden, H.: CH_4 and CO distributions over tropical fires during October 2006 as observed by the Aura TES satellite instrument and modeled by GEOS-Chem, Atmos. Chem. Phys., 13, 3679–3692, https://doi.org/10.5194/acp-13-3679-2013, 2013.

Worden, J. R., Bloom, A. A., Pandey, S., Jiang, Z., Worden, H. M., Walker, T. W., Houweling, S., and Röckmann, T.: Reduced biomass burning emissions reconcile conflicting estimates of the post-2006 atmospheric methane budget, Nat. Commun., 8, 2227, https://doi.org/10.1038/s41467-017-02246-0, 2017.

Wu, X., Xiao, Q., Wen, J., You, D., and Hueni, A.: Advances in quantitative remote sensing product validation: Overview and current status, Earth-Sci. Rev., 196, 102875, https://doi.org/10.1016/j.earscirev.2019.102875, 2019.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network, Philos. Trans. R. Soc. A, 369, 2087–2112, https://doi.org/10.1098/rsta.2010.0240, 2011.

Yan, C., Fan, X., Fan, J., and Wang, N.: Improved U-Net Remote Sensing Classification Algorithm Based on Multi-Feature Fusion Perception, Remote Sens., 14, https://doi.org/10.3390/rs14051118, 2022.

Yang, Y., Zhou, M., Langerock, B., Sha, M. K., Hermans, C., Wang, T., Ji, D., Vigouroux, C., Kumps, N., Wang, G., De Mazière, M., and Wang, P.: New ground-based Fourier-transform near-infrared solar absorption measurements of XCO_2, XCH_4 and XCO at Xianghe, China, Earth Syst. Sci. Data, 12, 1679–1696, https://doi.org/10.5194/essd-12-1679-2020, 2020.

Remarks from the typesetter

TS1 Please confirm the removal of P.R. for consistency.
TS2 Please confirm the running title or provide an alternative.
TS3 Please provide a citation and corresponding reference list entry.
TS4 Please provide a citation and corresponding reference list entry.
TS5 Please provide a working link and a date of last access.
TS6 Please provide a working link and a date of last access.
TS7 Please provide a working link and a date of last access.
TS8 Please provide a working link and a date of last access.
TS9 Please confirm the change here and above.
TS10 Please note that we do not use asterisks to indicate multiplication. Please check all asterisks used throughout the paper and let us know whether they can be replaced with $\times$ or $\cdot$.
TS11 Please check throughout the text that all vectors are denoted by bold italics and matrices by bold roman.
TS12 Please note that units have been changed to exponential format throughout the text. Please check all instances.
TS13 Please note that there is funding information given in the acknowledgements, but you did not indicate any funding upon manuscript registration. Therefore, we were not able to complete the financial support statement. Please provide the missing information and double-check your acknowledgements to see whether repeated information can be removed from the acknowledgements. Thanks.
TS14 Please provide date of last access.