

Point-by-point responses:

Specific comments

Lines 45–47: The documented SURF and GOS experiments compare alternative observing systems. Please revise the abstract so that joint surface–satellite assimilation is presented as a potential approach rather than a demonstrated result of MICA, unless a joint-assimilation experiment is documented.

In the paper we present only the SURF and GOS experiments. We have revised the text: “We find that inter-model spread is driven largely by horizontal resolution (Coefficient of Variation 10–20%), and that joint assimilation of satellite and surface observations will be a potential option for constraining emissions in regions with sparse monitoring or complex land–sea geometries.”

Lines 48–49: Please qualify “robust consistency” by specifying the regions, scales, and validation results that support this statement.

We have revised the sentence to read: “This dataset provides a foundation for regional and national methane budget assessments in Asia and for inverse model development, with cross-system consistency in regional emission estimates supported by evaluation against independent observations.” In the earlier part of Abstract already reports that validation against independent surface, aircraft, and balloon data shows substantial improvements in model performance after inversion, with significant reductions in model-observation root mean square error (RMSE).

Lines 79–80: Please clarify whether “divergences at finer scales” refers to differences among top-down estimates or between top-down and bottom-up estimates, and provide a concrete example.

The phrase “divergences at finer scales” refers to differences among top-down estimates, which is further highlights the value of multi-model intercomparisons. We have clarified this point in the revised text: “Despite providing critical insights into regional methane emissions, top-down estimates can differ substantially at finer spatial scales, particularly among individual countries, indicating persistent uncertainties (Janardanan et al., 2024; Miller et al., 2019; Saunois et al., 2016)”. A concrete example is provided in Table 4 of Saunois et al. (2016), which shows that the global total emissions estimated by an ensemble of 30 inversions have a relative range width less than 8%, whereas the relative ranges generally widen when the estimates are disaggregated to the regional scale.

Lines 136–139 and 298: The regional lists contain 23 entries, whereas line 298 states that the domain includes 28 countries. Please provide a complete list consistent with the country masks used in the analysis. In particular, check the inclusion of Baker Island and the omission of Sri Lanka, which is discussed in Section 4.3.

The study area comprises 24 Asian countries and island territories defined with an exclusive economic zone (EEZ) mask. East Asia (China, Mongolia, Japan, South Korea, and North Korea), South Asia (India, Pakistan, Nepal, Maldives, Bhutan, Bangladesh, Sri Lanka and Afghanistan), and Southeast Asia (Indonesia, Malaysia, Cambodia, Myanmar, Laos, Baker Island, the Philippines, Singapore, Thailand, Timor-Leste, and Vietnam). Sri Lanka and Baker Island are included in the study domain, although Baker Island, a U.S. territory in the central Pacific, is grouped with Southeast Asia based on its geographic proximity to this subregion within the mask. Given its negligible methane emissions, its inclusion has little effect on the total emissions estimated for Southeast Asia. The descriptions in Lines 136–139 and 298 have been harmonized for consistency.

Section 2.4, particularly lines 186–187: Please briefly summarize the treatment of methane losses, including OH fields and their temporal variability, and the initial and boundary/background concentrations, with references for details. Clarify the MIROC4-ACTM posterior loss correction and whether any loss terms are optimized. Briefly acknowledge the implications of these assumptions for emission and trend estimates.

OH fields are model-specific options in our study. We have revised section 2.4 to summarize the treatment of methane losses and of the atmospheric CH₄ initial and boundary/background conditions, with references to the relevant model descriptions.

MIROC4-ACTM applies a correction to posterior emissions to account for bias in the chemical loss calculated from the prior CH₄ distribution. Because chemical loss depends on atmospheric burden ($L = B/\tau$), the differences between modelled and observed background CH₄ at Syowa is used to estimate this correction. The correction changed global posterior emissions by -30 to $+20$ Tg CH₄ yr⁻¹ across the ensemble members during 2000–2020 (Belikov et al., 2024). It is a burden-based adjustment and not an independent optimization of OH or other chemical loss terms. The inferred emission magnitudes and trends therefore remain sensitive to the prescribed sinks and to the assumption that the model–observation difference at Syowa represents the global background CH₄ bias.

Line 193, Table 3, and Sections 3.4 and 4.3: Please clarify how sectoral emissions are derived across the inversion systems. List the six source categories optimized by NTFVAR and explain how they map onto the broader categories in Table 3. For each system, distinguish separately optimized sectors from those allocated using prior fractions or shared scaling factors when only total fluxes or broad source groups are optimized. In Section 4.3, specify whether the reported sectoral differences and trends refer to priors, posteriors, or both. Briefly explain what the existing NTFVAR/NTFVAR2 comparison shows about the dependence of posterior sectoral trends on prior assumptions, and qualify the conclusions accordingly.

Thank you for highlighting the need to clarify the derivation and interpretation of sectoral emissions. We have revised the model descriptions to state how each system obtains sectoral emissions. Sectoral emissions are derived differently among the participating inversion systems. NTFVAR optimizes time-varying flux corrections at a half-monthly time step for six source categories: five anthropogenic categories (agriculture, waste, coal, oil and gas, and biomass burning) and wetland. GEOS-Chem systems optimize total emissions and CTE-CH₄, LMDZ optimize broader source groups, and subsequently allocate the posterior fluxes to individual sectors using prior sectoral fractions. MIROC4-ACTM and NTLB do not optimize the total flux and do not provide sectoral products. The description has been added to the revision.

In Section 4.3, we now distinguish differences between prior and posterior sectoral emissions from trends calculated from the respective time series. We further discuss the NTFVAR/NTFVAR2 comparison to show that sectoral attribution and posterior trends can depend on the prior source partitioning and inversion configuration. Accordingly, we have qualified our conclusions and no longer interpret all posterior sectoral estimates as independently constrained by observations.

To quantify the dependence on the prior, in section 4.3, we further compared prior and posterior sectoral emissions of NTFVAR and NTFVAR2 for 2010–2021. Monthly emissions were summed to annual totals for 2010–2021 for 13 countries, their sum and three regional aggregates (South Asia, Southeast Asia, East Asia), in five sectors (agriculture, coal, gas and oil, waste and wetland), giving 85 target–sector pairs. Linear trends were fitted by ordinary least squares to the 12 annual values (nominal $p < 0.05$, no multiple-comparison correction), and were compared between two configurations for the priors, the posteriors, and the prior-to-posterior change in slope. In addition, we compared the trends unweighted ensemble mean in Figure 6 with and without NTFVAR2.

The two configurations agreed in slope direction for 62% of prior trends (53/85) and 82% of the posterior trends (70/85), and the sign of the prior-to-posterior change agreed in only 54% of pairs (46/85), so posterior agreement in direction does not imply agreement in magnitude or in how the inversion adjusted the prior. NTFVAR posterior trends largely retained their priors (no sign changes among 53 significant trends), whereas NTFVAR2 departed further from its priors (three sign changes among 49), and waste was the most consistent sector (17/17 significant and same direction in both). Excluding NTFVAR2 from the ensemble in section 3.4 left the posterior trend direction unchanged in 79/85 pairs. The six changes all occur for slopes that are indistinguishable from zero ($p \geq 0.46$ in both ensembles), and every trend that is significant in either ensemble keeps its direction. Significance changes in two pairs (Vietnam agriculture and South Asia oil and gas), and aggregate oil and gas emissions are about 5–8% higher without NTFVAR2. These results therefore show sensitivity to the inversion configuration, including the prior source partitioning, but do not isolate the effect of the prior. We have qualified the conclusions of Section 4.3 accordingly.

Line 198, Table 2 and Section 4.2: Please explain the NTLB resolutions reported as 27 km, 0.27°, and 0.5°. Indicate whether these refer to different transport, inversion, or output grids, and correct any inconsistent values or labels.

The NTLB inversion is performed on a $0.5^\circ \times 0.5^\circ$ grid. The footprints are driven directly by the NCEP Global Forecast System (GFS) data (0.25° resolution, 6-hourly time step). The total emission flux is optimized for the whole Asian region, without sector subdivision. We have corrected the values and labels in Table 2 and Section 4.2 accordingly.

Line 209, Tables 2–4 and Section 2.5: Please add the actual simulation periods for each configuration and experiment, preferably to Table 4, and consolidate the model-specific departures from the common protocol already mentioned in the text, including prior and observation products. Consistently distinguish the six inversion systems from the seven submitted configurations, identifying NTFVAR2 as an alternative configuration of NTFVAR.

In Table 4 we added a column listing the departures from the common protocol. To keep the table compact, we did not list the simulation periods, observation products and priors for every configuration; only departures from the common protocol are given. The text and tables now consistently distinguish the six inversion systems from the seven submitted configurations, and NTFVAR2 is identified as an alternative configuration of NTFVAR that uses the GCP prior.

Model	Experiment exp1(SURF) exp2(GOS)	Gridded Fluxes grid*.nc	Country Totals emis*.csv	CH ₄ mixing ratios conc*.csv	Validation val*.csv	Sectorer	Departures from common protocol
CTE-CH4	SURF, GOS	✓	✓	✓	✓	✓	
GEOS-Chem	SURF, GOS	✓	✓	✓	✓	✓	Short period (2018–2021)
LMDZ	SURF, GOS	✓	✓	✓	✓	✓	
MIROC4- ACTM	SURF	✓	✓	✓	✓		Observations from WDCGG only
NTFVAR	SURF, GOS	✓	✓	✓	✓	✓	
NTFVAR2 (alternative configuration of NTFVAR)	SURF	✓	✓	✓	✓	✓	GCP prior
NTLB	SURF	✓	✓	✓	✓		

Section 3.2: Please qualify the attribution to lower satellite sensitivity to local surface fluxes and briefly acknowledge other possible influences, such as observational sampling.

We agree that there might other factors effect. This larger uncertainty in GOS is likely partly attributable to the lower sensitivity of GOSAT retrievals to local surface fluxes, other factors such as observational sampling and spatial coverage might also contributed to the observed discrepancies.

Figures 5 and 6: Please specify the experiment(s) used. Clarify how NTFVAR2 is counted and whether membership changes across years or sectors, given the shorter GEOS-Chem record and the uneven availability of sectoral products in Table 4. For Figure 5, state whether anthropogenic and total emissions use the same members and whether each boxplot summarizes three-year means for individual members or pooled annual estimates. For Figure 6, specify whether the sectoral values are ensemble means or medians. Please explain how sources not shown and soil uptake are accounted for when relating the displayed sectoral values to the total fluxes reported elsewhere.

We have revised the titles and captions of Figures 5 and 6 to specify the experiments and inversion configurations used.

Figure 5 presents national-level boxplots of posterior methane emissions for 2019–2021, based on the SURF and GOS experiments. Anthropogenic emissions are calculated using the available anthropogenic estimates, excluding NTLB and MIROC4-ACTM, whereas total emissions include all inversion systems with available posterior total fluxes. For each configuration, the three-year mean is calculated first, and the boxplots summarize these member-specific means rather than pooled annual estimates. NTFVAR2 is included the same as other simulations. Figure 6 presents posterior country-level sectoral emissions for major emitting countries over 2010–2021, based on the SURF and GOS experiments from the four configurations providing sectoral estimates: CTE-CH₄, LMDZ, NTFVAR, and NTFVAR2. The plotted sectoral values are ensemble means.

We have also clarified the relationship between the displayed sectoral emissions and total fluxes reported elsewhere. The displayed sectors comprise agriculture, coal, oil and gas, waste, and wetlands. Total fluxes additionally account for soil uptake and other natural sources, including termites, geological sources, and freshwater emissions.

Please reconcile the apparent contradiction for Southeast Asia: lines 287–289 and Figure 4 indicate lower GOS emissions than SURF, whereas lines 540–541 state the reverse.

Thank you for noticing this inconsistency. We have corrected the text. The comparison of the SURF and GOS total estimates now reads consistently throughout: East Asia is significantly higher in SURF, Southeast Asia shows the reverse (significantly higher in GOS), and South Asia shows no significant difference.

Technical comments

Tables S1–S2: Please correct the latitude/longitude unit labels and check the ULD and DCD coordinates in Table S2. The ULD coordinate values appear interchanged, and the DCD entry appears inconsistent with Figure 1. Clarify whether these discrepancies are confined to the tables or also affect model sampling.

We have corrected the unit labels and the coordinates in Tables S1–S2: ULD is 37.48°N, 130.90°E and DCD is 37.32°N, 96.38°E. These were typographical errors confined to the tables; the coordinates in the observation files are correct, so model sampling is not affected.

Tables S1–S2: Please clarify whether the PDI entries represent distinct aircraft and surface datasets or whether the observation type in Table S1 requires correction. If the same observation records are available through both ObsPack and the additional datasets, briefly explain how duplicate records were handled.

The PDI was included In Tables S1 by mistake and has been removed.

Figure S3: Please check whether “SURF” in the caption should read “GOS”, and specify the averaging period for each panel.

Yes, the caption has been corrected to read “GOS” instead of “SURF”.

Please check the language and cross-references throughout, including “isd” at line 177, “expectations” at line 388, the duplicated “to” at line 545, and the section references for masks and aggregation.

Thank you. We have corrected these items and checked the language and cross-references throughout the manuscript, including the section references for masks and aggregation.